Molecular Dissection of Microsatellite Instable Colorectal Cancer

Eduardo Vilar¹ and Josep Tabernero²

ABSTRACT
Colorectal cancer was one of the first solid tumors to be classified on the basis of molecular profiling. Microsatellite instability has allowed researchers to distinguish a specific subtype of colorectal cancer that has a clearly identified molecular origin (mismatch repair deficiency), arises on a hereditary and sporadic basis, is linked to a clear clinicopathologic profile, and has prognostic implications. Inconclusive predictive data along with a paucity of targeted drug development have prevented this molecular classification system from being implemented in the clinical setting. New high-throughput genomic data have validated it, thus stressing the fact that it is ready to be applied clinically.

Significance: Application of a molecular classification of colorectal cancer in the clinical arena is an unmet promise. Recent results of large-scale genomic analyses have provided confirmation and further insights into the molecular biology of already known colorectal cancer subgroups. The quintessential example is the microsatellite instability subgroup, which has been well characterized during the past 2 decades. Future drug development and clinical research initiatives in colorectal oncology should consider these and other known cancer subgroups and start targeting these selected patient populations.

Cancer Discov; 3(4); 1–10. © 2012 AACR.

INTRODUCTION
Colorectal cancer is the third most common cancer in both men and women and constitutes a major health problem in Western countries. Approximately 51,690 people will have died of this disease in 2012 in the United States (1). Nonetheless, the incidence and mortality rates have been declining during the last 2 decades, mainly owing to improvements in screening, early diagnosis, and the introduction of new therapeutic agents in the adjuvant and metastatic settings.

The study of the molecular biology of colorectal cancer has reached important landmarks. Routine use of colonoscopy enabled the characterization of the steps leading to progression of a polyoid lesion into a carcinoma at the molecular level, the so-called adenoma-to-carcinoma sequence. This model provided insights into the role of alterations in the main oncogenes (e.g., KRAS and BRAF) and tumor suppressor genes (e.g., APC and TP53) in the biology of colorectal cancer (2). Studies of the contribution of other molecular mechanisms, such as epigenetics (e.g., methylation profiling) and microRNAs (miRNA), have enriched our understanding of the oncogenic process of the colonic epithelium (3–5). This accumulation of knowledge has translated into drug development and biomarker discovery that has already had an impact on the care of patients with colorectal cancer. The best example of this translational knowledge is the implementation of monoclonal antibodies against the EGF receptor (EGFR) and the use of the mutation status of KRAS in selecting patients to be treated with these agents in the clinic (6).

Technical improvements in the chemistry of the PCR, routine use of sequencing techniques, and implementation of other high-throughput platforms have enabled the profiling of the most common mutations in colorectal cancer and have contributed to better delineation of molecular subgroups. Colorectal cancer has been classified by the presence of insertion-deletions in loci containing microsatellite repeats into 3 groups: microsatellite instability high (MSI-H), microsatellite instability low (MSI-L), and microsatellite stable (MSS; ref. 7). Other classification systems consider the presence of chromosomal instability, dividing colorectal cancer into the categories of chromosomal instability positive (CIN+) or chromosomal instability negative (CIN-; ref. 3) or the level of methylation in different markers, identifying the cancers as CpG island methylation phenotype (CIMP) high (CIMP-H), low, or negative (5).

Determination of the mutation status of major oncogenes
(e.g., *KRAS*, *BRAF*, *NRAS*, and *PIK3CA*) and tumor suppressor genes (e.g., *TP53* and *PTEN*) has added more information to these classifications (3). These 3 molecular subgroups have clinical and pathologic correlates, thus establishing patterns for distinguishing phenotypes in the clinical setting. These classification systems complement each other and represent different approaches to the same problem (Fig. 1).

Despite this accumulation of molecular knowledge, the objective of tailoring colorectal cancer treatments based on molecular profiling has yet to be attained. Molecular subgroups are not routinely evaluated and considered in the daily clinical arena for making therapeutic decisions. Here, we review selected research findings characterizing the prominent and classically described molecular subgroup of colorectal cancer based on MSI. We review the basis for distinguishing this subgroup at the molecular level and describe the clinical and pathologic correlates underlying this phenotype. We then dissect the 3 existing molecular classifications of colorectal cancer and their relation to MSI. Finally, we update the prognostic and predictive implications of the MSI classification system and establish connections between available cytotoxic and targeted therapies and the MSI subgroup.

**Dissection of the Genomic Landscape of Colorectal Cancer**

A comprehensive assessment of the genomic landscape of colorectal cancer was undertaken for the first time by Sjoblom and colleagues (8) and Wood and colleagues (9) from the Vogelstein group. The results of this work were generated using first-generation sequencing (also known as Sanger sequencing) adapted for a high-throughput scale. The initial genomic reference used was the consensus coding sequences database (8). This dataset was later expanded by including additional annotated sequences from the Reference Sequence database (9). A total of 11 tumors were studied in the discovery phase, and 24 additional samples provided grounds for validation of the results. The large number of collected data made it necessary to implement sophisticated bioinformatic approaches to determine the somatic status of the alterations, the likelihood that each alteration contributed to tumorigenesis, and the functional consequences. The genomic reference to establish the somatic status of mutations was the DNA extracted from 2 matched normal tissue samples. The cancer mutation prevalence (CaMP) score reflected the probability that a particular gene had a prominent role in the tumor biology (a cancer driver gene) by having higher mutational rates than what was expected randomly based on the background of a particular tumor. Finally, mutations were analyzed with several algorithms to predict the functional consequences for the final protein product.

A total of 519 genes and 751 mutations were identified. The list of candidate cancer driver genes totaled 140 and included the usual suspects previously observed by other researchers (e.g., *APC*, *KRAS*, *PIK3CA*, *TP53*, *SMAD4*, *PTEN*, and *RB1*) but also new actors (e.g., *RET*, *ALK*, *SMAD3*, and *NF1*; Table 1). The median number of nonsilent mutations observed per colorectal cancer case was 76, a striking number because 15 was the median number of cancer driver genes mutated per tumor. This fact highlights that a large number of alterations found in tumors are just “passengers” and therefore have no biologic or therapeutic relevance.

This landmark work of Vogelstein and colleagues has been essential to understanding the biology of colorectal cancer. In addition, the total number of mutated cancer driver genes that were detected was amenable to high-throughput synthetic lethal and chemical screens, which prompted the exploration of new treatment avenues for colorectal cancer. However, the number of tumor samples used for both the discovery analysis and the validation cohort was relatively limited because of the technical complexity of large-scale sample sequencing at that time. Finally, this research did not take into account the molecular subtypes already recognized on the basis of MSI status or CIN.

The most comprehensive molecular analysis of colorectal cancer was conducted and published recently by The
Radic hypermethylation ofing MSI-H and CIMP-H clustered together, principally owing to MSI-H and CIMP-H. It is not surprising that tumors display hypermutated tumors was composed of those displaying a rate into hypermutated and nonhypermutated. The group divided tumors using an arbitrary cutoff of the mutational rate of mutations displayed. Therefore, subsequent analysis revealed 2 clearly differentiated groups of tumors by the sequencanalysis was completed in 224 tumors and notCancer Genome Atlas (TCGA) Network (10). Targeted exome paired samples. Assessment of the mutational rate sequencing analysis was completed in 224 tumors and normal paramed sequencing analysis was completed in 224 tumors and normal paired samples. Assessment of the mutational rate revealed 2 clearly differentiated groups of tumors by the rate of mutations displayed. Therefore, subsequent analysis divided tumors using an arbitrary cutoff of the mutational rate into hypermutated and nonhypermutated. The group of hypermutated tumors was composed of those displaying MSI-H and CIMP-H. It is not surprising that tumors displaying MSI-H and CIMP-H clustered together, principally owing to the fact that the majority of MSI-H tumors arise from sporadic hypermethylation of MLH1, one of the markers defining the CIMP status. In contrast, the group of nonhypermutated tumors mapped very well with CIN+ tumors.

The TCGA census of genes significantly mutated in hypermutated and nonhypermutated tumors was fundamentally different (Table 1), emphasizing the different biology of these subtypes. In this regard, the TCGA analysis that stratifies the mutational census by tumor subtype addressed a weakness of the work by Vogelstein and colleagues (8, 9). Hypermutated tumors were found to be mutated in ACVR2A (63%), APC (51%), TGFB2 (51%), BRAF (46%), MSH3 (40%), and FZD3 (29%), whereas nonhypermutated tumors harbored mutations in APC (81%), TP53 (60%), KRAS (43%), TTN (31%), PIK3CA (18%), SMAD4 (9%), and CTNNB1 (5%). Not surprisingly, the list of genes among the hypermutated tumors included several containing coding microsatellite tracts, such as TGFB2 (used as a marker in MSI analysis and a key component of the TGF-β pathway) and ACVR2A and MSH3 (also prototypic markers of this subtype; ref. 11). Interestingly, those genes containing microsatellite tracts and harboring secondary mutations owing to mismatch deficiency did not have high MutSig scores. The fact that this score has the same goal as the CaMP reiterates the concern that secondary mutations are just passengers that reflect high levels of genomic instability (12). It is worth noting that although the mutated genes differed substantially between the subtypes, several common pathways (such as Wnt) were observed, as reflected by the high mutational rate in APC in both subtypes and frequent mutations in FZD3 and CTNNB1 in the hypermutated and nonhypermutated subtypes, respectively. In addition, the phosphoinositide 3-kinase (PI3K) and RAS/extracellular signal–regulated kinase (ERK) pathways exhibited mutated KRAS and PIK3CA in the nonhypermutated subtype and mutated BRAF in the hypermutated subtype.

An important contribution of the TCGA study has been the genomic analysis dissecting 2 distinct subgroups of colorectal cancers. This stratified analysis should be considered a stepping stone for future colorectal cancer research, and the scientific community should embrace this classification not only in the research arena but also in the clinical setting.

### Microsatellite Instability in Colorectal Cancer

The first molecular subtype of colorectal cancers to be described was the MSI subgroup of tumors. Several research groups have contributed to the characterization of the MSI phenotype at the clinical, pathologic, and molecular levels. The molecular fingerprints of these tumors are the presence of insertions or deletions of mono-, di-, tri-, or tetranucleotides in microsatellite regions widespread in the entire genome (13–15). The DNA repair mechanism in charge of correcting these errors is the MMR system, which is deficient in this tumor subtype, thus introducing a high level of genomic instability in noncoding and coding regions (16). The pathogenesis of this deficiency is due to a germline mutation in one of the genes of the MMR system (MLH1, MSH2, MSH6, PMS2, or TACSTD1/EPICAMY) or the presence of hypermethylation in the promoter of MLH1. Therefore, MSI tumors can have a genetic origin (Lynch syndrome) or a sporadic origin (hypermethylated MSI tumors; ref. 7).

Molecular recognition of MSI occurs by using 2 approaches: MSI analysis and immunohistochemical analysis of MMR proteins (17). The former assesses the length of a panel of microsatellite markers in the tumor and a normal reference (either normal mucosa or the germline) by using fragment analysis of PCR products labeled with fluorescent dyes, and the latter determines the expression level of the proteins in charge of maintaining the integrity of microsatellite tracts. Differences in the length of 2 or more markers (the standard

---

**Table 1. Census of the 20 most significantly mutated genes in colorectal cancer, identified using large-scale genomic analysis**

<table>
<thead>
<tr>
<th>Rank</th>
<th>CaMP score</th>
<th>MutSig score—hypermutated</th>
<th>MutSig score—nonhypermutated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>APC</td>
<td>BRAF</td>
<td>APC</td>
</tr>
<tr>
<td>2</td>
<td>KRA5</td>
<td>APC</td>
<td>TP53</td>
</tr>
<tr>
<td>3</td>
<td>TP53</td>
<td>DMD</td>
<td>PK3CA</td>
</tr>
<tr>
<td>4</td>
<td>PIK3CA</td>
<td>CASPB</td>
<td>KRAS</td>
</tr>
<tr>
<td>5</td>
<td>FBXW7</td>
<td>MIER3</td>
<td>FBXW7</td>
</tr>
<tr>
<td>6</td>
<td>CSM3D</td>
<td>ATP6V0D2</td>
<td>SMAD4</td>
</tr>
<tr>
<td>7</td>
<td>TNN</td>
<td>SLC9A9</td>
<td>TCF7L2</td>
</tr>
<tr>
<td>8</td>
<td>NAV3</td>
<td>PTPN12</td>
<td>TTN</td>
</tr>
<tr>
<td>9</td>
<td>SMAD4</td>
<td>CDC27</td>
<td>NRAS</td>
</tr>
<tr>
<td>10</td>
<td>EPHA4</td>
<td>MAP7</td>
<td>CTNNB1</td>
</tr>
<tr>
<td>11</td>
<td>MAP2K7</td>
<td>FZD3</td>
<td>GRIK3</td>
</tr>
<tr>
<td>12</td>
<td>EPBH6</td>
<td>MYO1B</td>
<td>SMAD2</td>
</tr>
<tr>
<td>13</td>
<td>PTEN</td>
<td>TCFER1</td>
<td>KIAA1B04</td>
</tr>
<tr>
<td>14</td>
<td>ADAMTS5L3</td>
<td>SLTRK6</td>
<td>ACVR1B</td>
</tr>
<tr>
<td>15</td>
<td>GUCY1A2</td>
<td>ACDT12</td>
<td>GPC6</td>
</tr>
<tr>
<td>16</td>
<td>SMAD2</td>
<td>TPTE</td>
<td>EDNRB</td>
</tr>
<tr>
<td>17</td>
<td>ORS1E1</td>
<td>RARB</td>
<td>SOX9</td>
</tr>
<tr>
<td>18</td>
<td>LAMA1</td>
<td>PTEN</td>
<td>WBCR17</td>
</tr>
<tr>
<td>19</td>
<td>C10orf137</td>
<td>LEPREL1</td>
<td>FAM123B</td>
</tr>
<tr>
<td>20</td>
<td>TCF7L2</td>
<td>SMAD4</td>
<td>ATM</td>
</tr>
</tbody>
</table>

NOTE: CaMP score–based ranking was retrieved from supplementary materials provided by Wood et al. (9). Mutation Significance (MutSig) score–based ranking was retrieved from the TCGA Network. Genes commonly mutated across all 3 analyses are shown in dark gray, those common to CaMP score analysis and MutSig score for hypermutated tumors in light gray, and those common to CaMP score analysis and MutSig score for nonhypermutated tumors in blue.

Cancer Genome Atlas (TCGA) Network (10). Targeted exome sequencing analysis was completed in 224 tumors and normal paired samples. Assessment of the mutational rate revealed 2 clearly differentiated groups of tumors by the rate of mutations displayed. Therefore, subsequent analysis divided tumors using an arbitrary cutoff of the mutational rate into hypermutated and nonhypermutated. The group of hypermutated tumors was composed of those displaying MSI-H and CIMP-H. It is not surprising that tumors displaying MSI-H and CIMP-H clustered together, principally owing to the fact that the majority of MSI-H tumors arise from sporadic hypermethylation of MLH1, one of the markers defining the CIMP status. In contrast, the group of nonhypermutated tumors mapped very well with CIN+ tumors.

The TCGA census of genes significantly mutated in hypermutated and nonhypermutated tumors was fundamentally different (Table 1), emphasizing the different biology of these subtypes. In this regard, the TCGA analysis that stratifies the mutational census by tumor subtype addressed a weakness of the work by Vogelstein and colleagues (8, 9). Hypermutated tumors were found to be mutated in ACVR2A (63%), APC (51%), TGFB2 (51%), BRAF (46%), MSH3 (40%), and FZD3 (29%), whereas nonhypermutated tumors harbored mutations in APC (81%), TP53 (60%), KRAS (43%), TTN (31%), PIK3CA (18%), SMAD4 (9%), and CTNNB1 (5%). Not surprisingly, the list of genes among the hypermutated tumors included several containing coding microsatellite tracts, such as TGFB2 (used as a marker in MSI analysis and a key component of the TGF-β pathway) and ACVR2A and MSH3 (also prototypic markers of this subtype; ref. 11). Interestingly, those genes containing microsatellite tracts and harboring secondary mutations owing to mismatch deficiency did not have high MutSig scores. The fact that this score has the same goal as the CaMP reiterates the concern that secondary mutations are just passengers that reflect high levels of genomic instability (12). It is worth noting that although the mutated genes differed substantially between the subtypes, several common pathways (such as Wnt) were observed, as reflected by the high mutational rate in APC in both subtypes and frequent mutations in FZD3 and CTNNB1 in the hypermutated and nonhypermutated subtypes, respectively. In addition, the phosphoinositide 3-kinase (PI3K) and RAS/extracellular signal–regulated kinase (ERK) pathways exhibited mutated KRAS and PIK3CA in the nonhypermutated subtype and mutated BRAF in the hypermutated subtype.

An important contribution of the TCGA study has been the genomic analysis dissecting 2 distinct subgroups of colorectal cancers. This stratified analysis should be considered a stepping stone for future colorectal cancer research, and the scientific community should embrace this classification not only in the research arena but also in the clinical setting.

**Microsatellite Instability in Colorectal Cancer**

The first molecular subtype of colorectal cancers to be described was the MSI subgroup of tumors. Several research groups have contributed to the characterization of the MSI phenotype at the clinical, pathologic, and molecular levels. The molecular fingerprints of these tumors are the presence of insertions or deletions of mono-, di-, tri-, or tetranucleotides in microsatellite regions widespread in the entire genome (13–15). The DNA repair mechanism in charge of correcting these errors is the MMR system, which is deficient in this tumor subtype, thus introducing a high level of genomic instability in noncoding and coding regions (16). The pathogenesis of this deficiency is due to a germline mutation in one of the genes of the MMR system (MLH1, MSH2, MSH6, PMS2, or TACSTD1/EPICAMY) or the presence of hypermethylation in the promoter of MLH1. Therefore, MSI tumors can have a genetic origin (Lynch syndrome) or a sporadic origin (hypermethylated MSI tumors; ref. 7).

Molecular recognition of MSI occurs by using 2 approaches: MSI analysis and immunohistochemical analysis of MMR proteins (17). The former assesses the length of a panel of microsatellite markers in the tumor and a normal reference (either normal mucosa or the germline) by using fragment analysis of PCR products labeled with fluorescent dyes, and the latter determines the expression level of the proteins in charge of maintaining the integrity of microsatellite tracts. Differences in the length of 2 or more markers (the standard...
Bethesda panel uses 5 markers, but some panels use more markers, and therefore the cutoff is 30% of the markers for MSI-H or the absence of expression of one of the MMR proteins is indicative of MSI-H (17–19). Instability noted in one marker (or in less than 30% if more than 5 markers are assessed) is indicative of MSI-L. The absence of instability in any of the markers tested corresponds to MSS status. It is worth noting that tumors displaying MSI-L have been traditionally grouped with MSS tumors.

The genetic instability in MSI primarily reflects the variation in microsatellite tracts, which introduces secondary mutations in coding regions of cancer driver genes and also instability in passenger genes. The census of genes reported to harbor mutations in microsatellite repeats in MSI tumors contains more than 30 genes that are involved in very diverse cellular functions and pathways (11). This “instabilome” has been systematically assessed in the past using PCR techniques and more recently by exome sequencing (10, 20). In fact, the inclusion of MSI-H among the group of hypermutated tumors in the TCGA study stresses the point that this subgroup accumulates a high proportion of insertion-deletions in secondary target genes compared with the proportion in tumors displaying MSS (10). In addition, as pointed out in the previous section, the mutation profile outside microsatellite regions is fundamentally different in MSI tumors than in other colorectal cancers and may reflect a particular pattern for certain mutational hotspots that may be dependent on, or independent of, MMR deficiency.

Tumors displaying MSI are strongly associated with mutations in BRAF, particularly among sporadic MSI tumors caused by hypermethylation of MLH1 (21, 22). In the clinical setting, the absence of BRAF mutations in Lynch syndrome cases has been used to rule out a sporadic origin of MSI tumors displaying a genetic background (23, 24). At the same time, MSI tumors exhibit a relative lack of KRAS mutations. In addition, several studies have suggested a higher mutational rate in different actors of the PI3K/AKT/mTOR pathway among MSI tumors, such as PIK3CA and PTEN (25–28), although these results need to be validated in additional cohorts of patients. The Wnt pathway is frequently deregulated in both MSI and MSS tumors. The Wnt pathway has been found indeed to be the primary driving force controlling the cell fate of stem cells and their progeny (also known as transit-amplifying cells); both types of cells are localized in the intestinal crypts. It has been shown that activation of the Wnt pathway at physiologic levels is required to maintain the crypt and phenotype (29). Furthermore, activation of the Wnt pathway either by inducing the expression of β-catenin–dependent transcription factors (such as TCF4) or deleting the tumor suppressor gene APC leads to a rapid proliferation of the stem cell compartment in the crypts and the generation of adenomas. Therefore, these findings provide evidence that adenoma cells represent the transformed counterparts of the proliferative crypt progenitors and indicate the essential role of the Wnt pathway in this process (30). The Wnt-related genes harboring alterations are different in MSI and MSS tumors. Compared with MSS tumors, MSI tumors exhibit more frequent mutations in TGFBR2, TCF7L2, and FZD3 and less frequent mutations in APC (10). This mutational profile translates into differences at the expression level as well. Gene expression profiles have been used successfully to profile MSI tumors (31–35) and cell lines (36) by using gene expression microarrays. However, the level of overlap between the gene lists generated by these studies has been partial and limited.

This molecular background on MSI leads to a recognizable clinicopathologic phenotype. Several studies have shown that MSI tumors tend to be right sided and diagnosed at earlier stages than MSS tumors. In fact, several studies have reported that the prevalence of MSI in colorectal cancer ranges from 8% to 20% (37), although the actual values might differ according to stage distribution (7). MSI is more frequently observed among tumors diagnosed as stage II colorectal cancer (20%) than as stage III colorectal cancer (12%); ref. 38 and even less often among stage IV tumors (~4%; ref. 39). In addition, sporadic MSI cases are generally diagnosed in older patients (at least 70 years of age) and patients with familial cases are younger (less than 50 years of age), both showing a U-shaped age distribution (40). From a pathologic point of view, MSI tumors have high histologic grades, a mucinous phenotype with prominent numbers of tumor-infiltrating lymphocytes, a lack of dirty necrosis, and a Crohn-like host response (41, 42).

**Chromosomal Instability in Colorectal Cancer: The Flip Side of the Coin**

The vast majority of colorectal cancers (85%) present genomic instability in the form of aneuploidy, thus reflecting a wide range of chromosomal gains and losses (Fig. 1) (3, 43). This phenomenon is a classic hallmark event of cancer. Colorectal tumors tend to present and accumulate specific chromosomal abnormalities, a fact that points toward the essential role of chromosomal aberrations in tumorigenesis and argues against a random effect linked to the carcinogenesis process itself. Classic karyotyping analysis, allelotyping efforts, and lately high-throughput array techniques have consistently shown recurrent losses in chromosomes 1p, 8p, 15q, 17p, and 18q and gains in chromosomes 7p, 7q, 8q, 13q, and 20q (10, 44, 45).

Despite research efforts, the origin of CIN in colorectal cancer remains unknown. Several genes involved in the cell cycle, particularly at checkpoints of cell cycle phases, have been suggested to be responsible for CIN (e.g., BUB1, ATM, ATR, BRCA1, BRCA2, STK15, PLK1, MRE11 and FBXW7 ref. 45). CIN can introduce another way to subclassify colorectal tumors. However, it does not allow further distinguishing between MSI and MSS tumors. MSI tumors tend to be diploid and lack chromosomal abnormalities (CIN-), whereas MSS is associated with an abundance of chromosomal gains and losses (instable karyotype, CIN+). The technical determination of the status of CIN is not as well established and standardized as the determination of MSI, and standard criteria to define CIN status as opposed to MSI are lacking (19).

**The Intersection between MSI and CIMP**

The discovery of the presence of concordant methylation events in several genes (e.g., CDKN2A, MGMT, and MLH1) has fostered the study of epigenetics in colorectal cancer and added a new dimension to the molecular classification of colorectal cancer. Comprehensive studies of more than 30 loci have confirmed the identification of a cluster of methylation events in a specific subgroup of genes, thus leading to the identification of CIMP subgroups. Because methylation is a physiologic
process related to aging, initially controversy arose regarding the correct interpretation of this phenomenon and its specific relationship with cancer. In addition, standardization of the technique used to quantify the level of gene methylation and the panel of markers (MINT1, MINT2, MINT31, CDKN2A, and MLH1) is the most commonly used) has allowed for homogenization of the definition of the CIMP phenotype (3, 5).

The fact that the majority of MSI tumors arise owing to the epigenetic inactivation of MLH1 has linked the CIMP phenotype to MSI and vice versa since the initial description of CIMP and has also generated controversy regarding the independent existence of these subgroups (Fig. 1). Correlative clinical studies have identified a clinicopathologic profile of CIMP-H tumors that coincides with most characteristics of the MSI phenotype: right-sided location, older age at diagnosis, higher frequency in female patients, poor differentiation, and association with the presence of BRAF mutations. In addition, the preneoplastic lesion observed in CIMP cases, called sessile serrated adenoma, displays a characteristic pattern resembling the features of hyperplastic polyposis (46). However, the CIMP subgroup includes a broader spectrum of tumors and has a poorer outcome than MSI tumors. Therefore, the MSI subgroup is more homogeneous in terms of its natural history, genotypic, and phenotypic characteristics, so both the CIMP status and the MSI status need to be taken into consideration to refine a prognostic assessment.

Furthermore, it is clear that a new wave of molecular studies conducting a multidimensional assessment of the molecular profile of colorectal cancer, using sequencing and expression data, analysis of copy number variation, methylation, and miRNA profiling, need to take into account the 3 main classifications: MSI, CIN, and CIMP. The intersection of all of these annotations with the known tumor subtypes will allow fulfillment of the as yet unmet promise of a molecular subclassification. Recently, an unsupervised cluster analysis of gene expression data has shown 6 different subgroups among a total of 750 patients. These subgroups gained biologic and clinical sense upon annotation of the genomic aberrations present in tumors (CIN status by CGH arrays), as well as the anatomic location (right versus left), the status of MSI and CIMP, and the presence of the principal mutations (KRAS, BRAF, and TP53 status). Not surprisingly, the most segregated subtypes constituted tumors displaying MSI-H and CIMP-H, harboring BRAF mutations, and located in the right side of the colon. This finding provides further support for the value of MSI as a main classifier of colorectal cancer (47). The contribution of tumor location in colorectal cancer biology needs to be studied in more detail. Molecular studies conducted in a subset of stage II/III colorectal tumor samples collected in the context of the large multinational randomized controlled trial Pan-European Trial in Adjuvant Colon Cancer 3 (PETACC-3) showed an important contribution of tumor location in terms of prognostic assessment. However, the final publication of those results reported only a modest role of location as a determining factor for prognostic classification (HR for overall survival, right vs. left 1.29, \( P < 0.04 \); but not statistically significant for relapse-free survival). Therefore, it has yet to be determined whether location has a prominent role in the subclassification of colorectal cancer or is just a surrogate factor of other molecular events such as MSI and CIN (48).

Role of MSI as Prognostic and Predictive Factor of Classic Chemotherapeutic Drugs

The role of MSI as a prognostic marker has long been controversial. However, it is well accepted that colorectal tumors diagnosed as stage I, II, or III have a better prognosis if they exhibit MSI rather than MSS. This survival advantage is concordant with the fact that MSI tumors experience lower rates of tumor recurrence, especially at distant sites, than do MSS tumors (49), although this effect is most likely restricted to stage II rather than to stage III tumors (50). This effect has already been extensively reviewed by others and analyzed in a meta-analysis (7, 37, 38). Emerging data show the poor prognosis of MSI colorectal cancer that progresses to stage IV (51). The predominant role of BRAF mutations in stage IV MSI colorectal cancer is likely the main contributor to this worse outcome. In fact, tumors harboring BRAF mutations have shown worse outcomes in terms of overall and progression-free survival than have wild-type tumors (52–54). However, results are conflicting among studies that point toward a modulatory effect of MSI over BRAF mutations in survival (51, 55, 56). The PETACC-3 survival analysis of patients with stage II or III disease stratified by MSI and BRAF status did not confirm the prognostic effect of BRAF on relapse of the disease, but those patients with tumors harboring BRAF mutations had a poorer prognosis once they had relapses (50). Although the number of patients in that trial did not allow enough power to detect statistically significant differences, this result suggested that BRAF affects outcome among patients with MSI advanced colorectal cancer (50).

The value of MSI as a predictive marker of response to 5-fluorouracil (5-FU), irinotecan, and other chemotherapeutic agents remains controversial. Conflicting results on 5-FU have been published during the past decade (Table 2; refs. 7, 57), possibly owing to the retrospective and single-institution nature of most of the studies, the use of different methods and criteria to evaluate the presence of MSI, and inadequate interpretation of data. More recently, 2 large retrospective analyses from several randomized trials confirmed that the detrimental effect on survival of the use of adjuvant 5-FU-based chemotherapy is restricted to stage II cases of colorectal cancer (58) and is not applicable to stage III cases (49). An interesting finding recently published by Sinicropi and colleagues (49) was that the benefit of 5-FU for patients with stage III colorectal cancer displaying MSI may be restricted to those harboring germline mutations in MMR genes (i.e., Lynch syndrome cases). This finding points to molecular differences within the MSI subgroup that would condition different responses to 5-FU.

Some preclinical (59–62) and clinical (39, 50, 63–65) data have suggested a selective sensitivity of MSI tumors to irinotecan (Table 3). Although the molecular basis of this increased sensitivity remains partially elusive, different research groups have linked it to a deficiency in the DNA repair mechanism involved in the correction of double-strand breaks induced by irinotecan (61, 62). Clinical data derived from retrospective reviews of patients enrolled in clinical trials have generated conflicting results. The post hoc analysis of the Cancer and Leukemia Group B (CALGB) 89803 trial, which was originally designed to evaluate the efficacy of irinotecan, bolus 5-FU, and folinic acid compared with weekly bolus of
5-FU as adjuvant therapy in stage III cases, showed a trend toward greater 5-year disease-free survival for patients with MSI-H tumors treated with the combined regimen, although it did not reach statistical significance (64). The retrospective analysis of 1,254 patients included in the PETACC3 trial, which studied the effect of irinotecan, infusional 5-FU, and folinic acid as adjuvant therapy compared with infusional 5-FU and folinic acid alone in stage II and III cases, failed to show improved disease-free survival for MSI stage II patients (50). Therefore, the role of MSI as a predictive factor for chemotherapy is restricted to avoiding the administration of adjuvant 5-FU to patients with MSI stage II colorectal cancer. At present, MSI should not be considered a validated marker for establishing treatment decisions regarding irinotecan-based regimens.

**Targeted Drug Therapies for MSI Tumors**

The unmet promise in colorectal cancer oncology is the development of targeted therapies for molecularly defined subgroups. In this regard, the MSI subgroup is one of the best candidates because its molecular etiology is already well understood, the mutational profile is well defined, and a clear genotype-phenotype relationship is established. Targeting the MSI subgroup can be achieved by studying the sensitivity of drugs that exploit the abnormal functioning of the canonical MMR pathway. Alternative functions of the MMR system have been recently described, thus making them potential targets as well. Furthermore, the presence of BRAF mutations in sporadic MSI tumors makes this subgroup ideal for testing BRAF inhibitors (Table 4).

**Systems biology tools** have emerged as an agnostic way to approach drug development and identify new drugs and targets for specific populations of patients. The Connectivity Map is a tool developed by the Broad Institute (Cambridge, MA) that relates gene expression changes induced by a wide panel of small molecules (∼1,500 drugs) in a discrete number of cell line models of different histologic origins with a reference gene expression profile (66). This reference profile constitutes a query profile defining a particular biologic question (e.g., a tumor subgroup or a pattern of resistance to a specific agent). We have queried the Connectivity Map using gene signatures characterizing MSI colorectal cancers. The lists of drugs obtained were combined using sophisticated bioinformatics criteria, and a final short list with candidate compounds was obtained for further testing. This approach suggested that inhibitors of PI3K and mTOR could be specific compounds useful for treating MSI tumors.

**Table 2. Clinical studies analyzing the effect of 5-FU in cohorts of MSI colorectal cancer**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study type</th>
<th>Patients, n</th>
<th>Tumor stage</th>
<th>MSI-H, %</th>
<th>Follow-up duration, mo</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elsaleh et al. (78)</td>
<td>R</td>
<td>656</td>
<td>III</td>
<td>8.5</td>
<td>54</td>
<td>Benefit</td>
</tr>
<tr>
<td>Hemminki et al. (79)</td>
<td>P NR</td>
<td>95</td>
<td>III</td>
<td>12</td>
<td>31</td>
<td>Benefit</td>
</tr>
<tr>
<td>Liang et al. (80)</td>
<td>P NR</td>
<td>244</td>
<td>IV</td>
<td>21.3</td>
<td>NA</td>
<td>Benefit</td>
</tr>
<tr>
<td>Ribic et al. (42)</td>
<td>R from RCT</td>
<td>570</td>
<td>II/III</td>
<td>16.7</td>
<td>88.8</td>
<td>Detriment</td>
</tr>
<tr>
<td>Carethers et al. (81)</td>
<td>R</td>
<td>204</td>
<td>II/III</td>
<td>17.6</td>
<td>43.7</td>
<td>None</td>
</tr>
<tr>
<td>Benatti et al. (82)</td>
<td>R</td>
<td>1,263</td>
<td>All</td>
<td>20.3</td>
<td>64</td>
<td>None</td>
</tr>
<tr>
<td>Jover et al. (83)</td>
<td>P NR</td>
<td>754</td>
<td>All</td>
<td>8.8</td>
<td>24.3</td>
<td>None</td>
</tr>
<tr>
<td>Lamberti et al. (84)</td>
<td>P NR</td>
<td>416</td>
<td>All</td>
<td>12.5</td>
<td>32.9</td>
<td>None</td>
</tr>
<tr>
<td>Kim et al. (85)</td>
<td>R from RCT</td>
<td>542</td>
<td>II/III</td>
<td>18.1</td>
<td>60</td>
<td>None</td>
</tr>
<tr>
<td>Sargent et al. (58)</td>
<td>R from RCT</td>
<td>1,027</td>
<td>II/III</td>
<td>16</td>
<td>60</td>
<td>Detriment</td>
</tr>
<tr>
<td>Des Guetz et al. (57, 65)</td>
<td>MA</td>
<td>3,690</td>
<td>II/III</td>
<td>14</td>
<td>NA</td>
<td>None</td>
</tr>
<tr>
<td>Sinicrope et al. (49)</td>
<td>R from RCT</td>
<td>2,141</td>
<td>II/III</td>
<td>16.1</td>
<td>96</td>
<td>Benefit for Lynch syndrome None for sporadic MSI</td>
</tr>
</tbody>
</table>

*NOTE: The vast majority of these studies were observational studies, and some had retrospectively reviewed data collected in the context of randomized controlled trials.*

**Table 3. Clinical studies analyzing the effect of irinotecan in cohorts of MSI colorectal cancers**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study type</th>
<th>Patients, n</th>
<th>Tumor stage</th>
<th>MSI-H, %</th>
<th>Follow-up duration, mo</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koopman et al. (39)</td>
<td>R from RCT</td>
<td>515</td>
<td>IV</td>
<td>3.5</td>
<td>43</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>Fallik et al. (63)</td>
<td>R</td>
<td>72</td>
<td>IV</td>
<td>9.7</td>
<td>NA</td>
<td>Benefit</td>
</tr>
<tr>
<td>Bertagnolli et al. (64)</td>
<td>R from RCT</td>
<td>723</td>
<td>III</td>
<td>13.3</td>
<td>79.8</td>
<td>Benefit</td>
</tr>
<tr>
<td>Tejpar et al. (50)</td>
<td>R from RCT</td>
<td>1,254</td>
<td>II/III</td>
<td>22/12</td>
<td>68</td>
<td>No benefit</td>
</tr>
</tbody>
</table>

5-FU as adjuvant therapy in stage III cases, showed a trend toward greater 5-year disease-free survival for patients with MSI-H tumors treated with the combined regimen, although it did not reach statistical significance (64). The retrospective analysis of 1,254 patients included in the PETACC3 trial, which studied the effect of irinotecan, infusional 5-FU, and folinic acid as adjuvant therapy compared with infusional 5-FU and folinic acid alone in stage II and III cases, failed to show improved disease-free survival for MSI stage II patients (50). Therefore, the role of MSI as a predictive factor for chemotherapy is restricted to avoiding the administration of adjuvant 5-FU to patients with MSI stage II colorectal cancer. At present, MSI should not be considered a validated marker for establishing treatment decisions regarding irinotecan-based regimens.

**Targeted Drug Therapies for MSI Tumors**

The unmet promise in colorectal cancer oncology is the development of targeted therapies for molecularly defined subgroups. In this regard, the MSI subgroup is one of the best candidates because its molecular etiology is already well understood, the mutational profile is well defined, and a clear genotype-phenotype relationship is established. Targeting the MSI subgroup can be achieved by studying the sensitivity of drugs that exploit the abnormal functioning of the canonical MMR pathway. Alternative functions of the MMR system have been recently described, thus making them potential targets as well. Furthermore, the presence of BRAF mutations in sporadic MSI tumors makes this subgroup ideal for testing BRAF inhibitors (Table 4).

**Systems biology tools** have emerged as an agnostic way to approach drug development and identify new drugs and targets for specific populations of patients. The Connectivity Map is a tool developed by the Broad Institute (Cambridge, MA) that relates gene expression changes induced by a wide panel of small molecules (∼1,500 drugs) in a discrete number of cell line models of different histologic origins with a reference gene expression profile (66). This reference profile constitutes a query profile defining a particular biologic question (e.g., a tumor subgroup or a pattern of resistance to a specific agent). We have queried the Connectivity Map using gene signatures characterizing MSI colorectal cancers. The lists of drugs obtained were combined using sophisticated bioinformatics criteria, and a final short list with candidate compounds was obtained for further testing. This approach suggested that inhibitors of PI3K and mTOR could be specific compounds useful for treating MSI tumors.

**Table 2. Clinical studies analyzing the effect of 5-FU in cohorts of MSI colorectal cancer**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study type</th>
<th>Patients, n</th>
<th>Tumor stage</th>
<th>MSI-H, %</th>
<th>Follow-up duration, mo</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elsaleh et al. (78)</td>
<td>R</td>
<td>656</td>
<td>III</td>
<td>8.5</td>
<td>54</td>
<td>Benefit</td>
</tr>
<tr>
<td>Hemminki et al. (79)</td>
<td>P NR</td>
<td>95</td>
<td>III</td>
<td>12</td>
<td>31</td>
<td>Benefit</td>
</tr>
<tr>
<td>Liang et al. (80)</td>
<td>P NR</td>
<td>244</td>
<td>IV</td>
<td>21.3</td>
<td>NA</td>
<td>Benefit</td>
</tr>
<tr>
<td>Ribic et al. (42)</td>
<td>R from RCT</td>
<td>570</td>
<td>II/III</td>
<td>16.7</td>
<td>88.8</td>
<td>Detriment</td>
</tr>
<tr>
<td>Carethers et al. (81)</td>
<td>R</td>
<td>204</td>
<td>II/III</td>
<td>17.6</td>
<td>43.7</td>
<td>None</td>
</tr>
<tr>
<td>Benatti et al. (82)</td>
<td>R</td>
<td>1,263</td>
<td>All</td>
<td>20.3</td>
<td>64</td>
<td>None</td>
</tr>
<tr>
<td>Jover et al. (83)</td>
<td>P NR</td>
<td>754</td>
<td>All</td>
<td>8.8</td>
<td>24.3</td>
<td>None</td>
</tr>
<tr>
<td>Lamberti et al. (84)</td>
<td>P NR</td>
<td>416</td>
<td>All</td>
<td>12.5</td>
<td>32.9</td>
<td>None</td>
</tr>
<tr>
<td>Kim et al. (85)</td>
<td>R from RCT</td>
<td>542</td>
<td>II/III</td>
<td>18.1</td>
<td>60</td>
<td>None</td>
</tr>
<tr>
<td>Sargent et al. (58)</td>
<td>R from RCT</td>
<td>1,027</td>
<td>II/III</td>
<td>16</td>
<td>60</td>
<td>Detriment</td>
</tr>
<tr>
<td>Des Guetz et al. (57, 65)</td>
<td>MA</td>
<td>3,690</td>
<td>II/III</td>
<td>14</td>
<td>NA</td>
<td>None</td>
</tr>
<tr>
<td>Sinicrope et al. (49)</td>
<td>R from RCT</td>
<td>2,141</td>
<td>II/III</td>
<td>16.1</td>
<td>96</td>
<td>Benefit for Lynch syndrome None for sporadic MSI</td>
</tr>
</tbody>
</table>

*NOTE: The vast majority of these studies were observational studies, and some had retrospectively reviewed data collected in the context of randomized controlled trials.*

MA, meta-analysis; NA, not assessed; NR, nonrandomized; P, prospective; R, retrospective; RCT, randomized clinical trial; mo, months.

**Table 3. Clinical studies analyzing the effect of irinotecan in cohorts of MSI colorectal cancers**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study type</th>
<th>Patients, n</th>
<th>Tumor stage</th>
<th>MSI-H, %</th>
<th>Follow-up duration, mo</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koopman et al. (39)</td>
<td>R from RCT</td>
<td>515</td>
<td>IV</td>
<td>3.5</td>
<td>43</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>Fallik et al. (63)</td>
<td>R</td>
<td>72</td>
<td>IV</td>
<td>9.7</td>
<td>NA</td>
<td>Benefit</td>
</tr>
<tr>
<td>Bertagnolli et al. (64)</td>
<td>R from RCT</td>
<td>723</td>
<td>III</td>
<td>13.3</td>
<td>79.8</td>
<td>Benefit</td>
</tr>
<tr>
<td>Tejpar et al. (50)</td>
<td>R from RCT</td>
<td>1,254</td>
<td>II/III</td>
<td>22/12</td>
<td>68</td>
<td>No benefit</td>
</tr>
</tbody>
</table>
tumors. Preclinical testing revealed that colorectal cancer cell lines displaying MSI were more sensitive than MSS to single-agent first-generation PI3K and mTOR inhibitors (35). In addition, this in silico assessment revealed a preferential activity of COX-2 inhibitors and nonsteroidal anti-inflammatory drugs as potential drug candidates in MSI tumors. This preclinical observation was confirmed later by the Colorectal Adenoma/Carcinoma Prevention Program 2 (CAPP2) study of the effect of aspirin as a chemopreventive agent for Lynch syndrome patients. An initial report of the CAPP2 study did not show benefits in terms of colorectal cancer prevention in this population (67); however, an update published after longer follow-up revealed that aspirin prevented colorectal and Lynch syndrome–related tumors (68).

High-throughput methods have been used to reveal connections between MSI and potential targeted therapies based on the concept of synthetic lethality. Screening with siRNA using conditional cell line models for MLH1 and MSH2 deficiency revealed a synthetic lethal interaction between MMR deficiency and silencing or inhibition of the PTEN-induced putative kinase 1 gene (PINK1; ref. 69). A screen restricted to DNA polymerases allowed identification of synthetic lethal interactions between MSH2 and the polymerase B gene (POLB) and between MLH1 and the polymerase G gene (POLG) (70). A large-scale chemotherapeutic drug screen found that an MSH2-deficient cell line model was specifically more sensitive for the classic chemotherapeutic drug methotrexate (71). The synthetic lethal interactions of PINK1, POLB, and POLG and the preferential effect of methotrexate in MMR-deficient models have been linked to the accumulation of reactive oxygen species (72) and point to the involvement of the MMR system in cell functions other than its canonical pathway (i.e., recognition of base-to-base mismatches). Nonetheless, the MMR pathway may also be involved directly or indirectly in the repair of double-strand breaks through the homologous recombination pathway. This suggested a new role of the MMR pathway to include the direct participation of the MMR proteins in the repair of double-strand breaks or the introduction of secondary mutations in genes involved in the homologous recombination pathway by the MMR deficiency. In particular, the activity of the PARP inhibitor ABT-888 in colorectal cancer cell lines displaying MSI has been linked to a mutation in a microsatellite tract.

**CONCLUSIONS**

Colorectal cancer can be subclassified and dissected variously by using genomic, proteomic, methylation, and mitochondrial RNA analysis. Molecular subgroups have been identified, described at the molecular level, and correlated with clinicopathologic profiles. Studies have shown the value of these
molecular subgroups for prognostic and predictive purposes. We have focused on colorectal cancer displaying MSI. Despite the wealth of information that has been reviewed here, molecular annotations and classifications are still far from being used in daily clinical practice. The past decade has witnessed spectacular advances in the development of molecular tumor subclassification and its implementation in the clinical arena for not only histologic malignancies but also solid tumors. An example illustrating this approach is breast cancer. The scientific and clinical community involved in the treatment of breast cancer has embraced molecular classification of this disease and has incorporated it into clinical research, implementing clinical trials that are molecularly based for each subtype. Moreover, the care of patients who have a diagnosis of breast cancer is more heavily based on the specific subtype of tumor than it is for patients with any other kind of cancer. Unfortunately, this kind of advancement has not occurred with colorectal cancer. In reviewing the data collected for the MSI subgroup, we are making a call for action to the colorectal cancer community to make advances toward implementing these molecular classification systems that are now even more evident after the publication of massive high-throughput analysis by the TCGA Network. The first step is thinking about the molecular subtype of each patient seen in the clinic and then going back to the drawing board to design studies that take these classifications into account. MSI could be the first, but other subtypes, such as MSS tumors dissected by mutational analysis or CIMP status, could be immediately pursued as well.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: E. Vilar, J. Tabernero
Writing, review, and/or revision of the manuscript: E. Vilar, J. Tabernero
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): E. Vilar

Acknowledgments
The authors thank Elizabeth Hess for editing this manuscript.

Grant Support
This work was supported in part by the Conquer Cancer Foundation of the American Society of Clinical Oncology, Young Investigator Award (to E. Vilar), and by The University of Texas MD Anderson Cancer Center Core Support Grant (P30 CA016672).

Received October 12, 2012; revised December 13, 2012; accepted December 18, 2012; published OnlineFirst March 01, 2013.

REFERENCES
Molecular Subclassification of Colorectal Cancer


50. Des Guetz G, Uzzan B, Nicolas P, Schischmanoff O, Perret GY, Morere JF. Microsatellite instability does not predict the efficacy of...
Molecular Dissection of Microsatellite Instable Colorectal Cancer

Eduardo Vilar and Josep Tabernero

Cancer Discovery  Published OnlineFirst March 1, 2013.

Updated version  Access the most recent version of this article at: doi:10.1158/2159-8290.CD-12-0471

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.