Published in final edited form as:

J Pathol. 2013 January ; 229(1): 99–110. doi:10.1002/path.4087.

MLH1-Silenced and Non-Silenced Subgroups of Hypermutated Colorectal Carcinomas Have Distinct Mutational Landscapes

Lawrence A. Donehower^{1,2,*}, Chad J. Creighton^{1,3,*}, Nikolaus Schultz⁴, Eve Shinbrot¹, Kyle Chang¹, Preethi H. Gunaratne^{1,5}, Donna Muzny¹, Chris Sander⁴, Stanley R. Hamilton⁶, Richard A. Gibbs¹, and David Wheeler¹

¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030

²Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030

³Dan L. Duncan Cancer Center Division of Biostatistics, Department of Medicine, Baylor College of Medicine, Houston, TX 77030

⁴Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065

⁵Department of Biology and Biochemistry, University of Houston, Houston, TX 77204

⁶Division of Pathology and Laboratory Medicine, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030

Abstract

Approximately 15% of colorectal carcinomas (CRC) exhibit a hypermutated genotype accompanied by high levels of microsatellite instability (MSI-H) and defects in DNA mismatch repair. These tumors, unlike the majority of colorectal carcinomas, are often diploid, exhibit frequent epigenetic silencing of the *MLH1* DNA mismatch repair gene, and have a better clinical prognosis. As an adjunct study to The Cancer Genome Atlas consortium that recently analyzed 224 colorectal cancers by whole exome sequencing, we compared the 35 CRC (15.6%) with a hypermutated genotype to those with a non-hypermutated genotype. We found that 22 (63%) of hypermutated CRC exhibited transcriptional silencing of the *MLH1* gene, a high frequency of *BRAF* V600E gene mutations and infrequent *APC* and *KRAS* mutations, a mutational pattern

Correspondence to: Lawrence A. Donehower, Department of Molecular Virology and Microbiology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. larryd@bcm.edu.

*Co-first authors

Conflict of Interest Statement: None of the authors have any conflicts of interest as defined by the guidelines for *Journal of Pathology*.

Statement of Author Contributions:

L.A. Donehower initiated and carried out the majority of the analyses of the paper and did much of the writing of the paper.

C. Creighton did a significant amount of bioinformatic analyses, generated all of the figures for this paper, contributed to the writing, and performed virtually all of the statistics.

N. Schultz provided the initial data and analyses of the hypermutated CRC that formed the foundation for this paper.

E. Shinbrot generated the final CRC mutation data used for the CRC validation set.

K. Chang generated all of the homopolymer repeat mutation data used for Figure 6.

P. Gunaratne performed some of the analyses for the microRNA data presented in Figure 3.

D. Muzny supervised the generation of much of the sequencing data used in this paper. C. Sander supervised the initial data analyses of the TCGA CRC data that was the foundation for this paper.

S. Hamilton is a physician/scientist with expertise in clinical aspects of CRC who assisted in the writing and editing of this paper.

R. Gibbs is the director of the Baylor College of Medicine Human Genome Sequencing Center who provided the resources and overall guidance for this study.

D. Wheeler is the director of this research and supervised its progress at all steps and performed writing and editing of the paper.

List of Online Supporting Information:

Supplementary Table 1: Mutations in Key Genes of the 250 CRC Validation Set

significantly different from their non-hypermuted counterparts. However, the remaining 13 (37%) hypermutated CRC lacked *MLH1* silencing, contained tumors with the highest mutation rates (“ultramutated” CRC), and exhibited higher incidences of *APC* and *KRAS* mutations, but infrequent *BRAF* mutations. These patterns were confirmed in an independent validation set of 250 exome-sequenced CRC. Analysis of mRNA and microRNA expression signatures revealed that hypermutated CRC with *MLH1* silencing had greatly reduced levels of WNT signaling and increased *BRAF* signaling relative non-hypermuted CRC. Our findings suggest that hypermutated CRC include one subgroup with fundamentally different pathways to malignancy than the majority of CRC. Examination of *MLH1* expression status and frequencies of *APC*, *KRAS*, and *BRAF* mutation in CRC may provide a useful diagnostic tool that could supplement the standard microsatellite instability assays and influence therapeutic decisions.

Keywords

colorectal cancer; microsatellite instability; *MLH1*; *APC*; *KRAS*; *BRAF*; WNT signaling; mutation rate

Introduction

Colon and rectal cancers (CRCs) are the third most frequently arising cancer in both men and women in the United States [1]. While progress has been made in CRC treatment, mortality rates remain high [2]. Some inherited and somatically arising CRC exhibit genetic instability in simple tandem repeat sequences called microsatellite DNA, often associated with defective DNA mismatch repair [3,4]. About 15% of CRC exhibit microsatellite instability (MSI) [5]. Pathologically, these tumors tend to arise in the proximal colon, display lymphocytic infiltration, are often mucinous, and respond to chemotherapeutics differently than microsatellite stable tumors [5]. About 75% of CRCs with microsatellite instability display hypermethylation and transcriptional silencing of the *MLH1* gene [6,7]. A frequent genetic alteration in CRC with MSI is the oncogenic *BRAF* V600E mutation. *BRAF* mutations in sporadic CRC have been associated with increased mortality [8].

Truncating mutations in one or both adenomatous polyposis coli (*APC*) alleles occur early during CRC evolution, occurring at roughly 80% frequency [9]. *APC* truncating mutations result in constitutive activation of WNT-mediated growth signaling [9]. Later genetic events in CRC include oncogenic *KRAS* mutations and inactivating p53 tumor suppressor mutations [10]. *KRAS* mutation status has been a useful marker in CRC therapeutic interventions with epidermal growth factor receptor-specific antibodies. CRC with wildtype *KRAS* were more likely to show therapeutic benefit when treated with these therapeutic antibodies [11]. Sporadic CRC with and without MSI display distinctive patterns of mutations in cancer-associated genes. Unlike CRC with microsatellite stability, CRC with MSI exhibit reduced frequencies of *APC*, p53 and *KRAS* mutations, as well as increases in *TGFBR2* mutations and the aforementioned *BRAF* mutations [12–15].

The Velculescu, Kinzler, and Vogelstein laboratories have pioneered genomic surveys of colorectal cancers, identifying novel altered genes and chromosomal aberrations [16–18]. Recently, The Cancer Genome Atlas (TCGA) research consortium analyzed 224 tumor/normal CRC pairs by integrating data from whole exome sequencing, DNA copy number variation, promoter methylation, and global mRNA and microRNA expression [19]. In the TCGA study CRC were subdivided into 15% hypermutated (>12 non-silent exonic mutations/megabase) and 85% non-hypermuted (<12 non-silent exonic mutations/megabase) populations. Hypermuted CRC were distinguished by high rates of *MLH1* gene silencing and high MSI. Here, in an adjunct paper to the TCGA study, we analyze key

features of the hypermutated CRC. We show that the hypermutated CRC can be usefully subcategorized into *MLH1*-silenced and non-*MLH1*-silenced groups. The *MLH1*-silenced hypermutated CRC exhibit low numbers of *APC* mutations and high rates of *BRAF* V600E mutations. Analysis of mRNA and miRNA expression patterns indicates that this group of tumors may not rely on activated WNT signaling for tumor initiation and progression. Thus, we propose that classification of CRC by *MLH1* silencing status in addition to MSI status may provide an alternative and useful way to predict CRC pathobiology and therapy responses.

Materials and Methods

Sample and Data Acquisition

The analyses described here were from 474 tumor/normal CRC on which whole exome sequencing was performed under the auspices of The Cancer Genome Atlas (TCGA) Consortium. Details of sample acquisition, DNA sequencing, DNA copy number studies, RNA expression, miRNA expression, microsatellite instability analyses, and epigenomic analyses were described in the parent TCGA CRC publication [19] (which featured 224 of the 474 tumors). The 224 samples in the discovery cohort were sequenced by the genome centers at Washington University School of Medicine and Baylor College of Medicine [19]. The 250 samples in the validation cohort were sequenced at the Baylor College of Medicine Human Genome Sequencing Center. All data collected by the TCGA Consortium provided for full patient privacy protection and were in accordance with established ethical standards.

Comparison of gene mutational patterns

The 224 CRC discovery set (featured in the initial publication [19]) that were whole exome sequenced were divided into two categories based on mutation rates: hypermutated (>12 non-silent mutations per 10⁶ exonic nucleotides) and non-hypermutated (<12 non-silent mutations per 10⁶ exonic nucleotides). The 35 hypermutated CRC were then subdivided into 22 *MLH1*-silenced CRC and 13 non-*MLH1*-silenced CRC, and were analyzed separately from non-hypermutated CRC. While all 22 *MLH1*-silenced hypermutated CRC appeared to have normal germline mismatch repair genes (*MSH2*, *MSH6*, *MLH1*, and *PMS2*), 2 of the 13 non-*MLH1*-silenced hypermutated tumors may have been Lynch syndrome CRC, as patient AA-A00A exhibited a germline frameshift mutation in *PMS2* and patient AA-A01R displayed a frameshift mutation in *MSH6*. Of the 189 non-hypermutated CRC, 163 had complete methylation data including methylation status of the *MLH1* gene. Validated mutations in *APC*, *KRAS*, *NRAS*, *BRAF*, *TP53*, *SMAD2*, *SMAD3*, *SMAD4*, *ACVR2A*, *ACVR1B*, and *TGFBR2* were determined for all tumors and mutation frequencies of individual genes or gene groups (e.g. TGF- β pathway members) in each of the three CRC categories were compared by Fisher's exact test or chi-square test as indicated.

A second validation set of 250 CRC was analyzed by whole exome sequencing and global methylation status (including *MLH1*). Of this second set, 29 (11.6%) were categorized as *MLH1*-methylated (top 20% of methylation beta values, based on Illumina array probe cg02279071 and the initial analysis of the discovery set) and hypermutated (>12 mutations per exome Mb), 16 (6.4%) were hypermutated and did not exhibit *MLH1* methylation, while the remaining 205 (82%) CRC were classified as non-hypermutated (<12 mutations/Mb).

Comparison of mRNA and microRNA expression

To compare mRNA expression patterns, we acquired the RNAseq data for those tumors among the 224 discovery CRC with this data available. The tumors were subdivided into the three categories, *MLH1*-silenced hypermutated (N = 20), non-*MLH1*-silenced hypermutated (N = 9) and non-hypermutated (N = 140). Mean expression levels for each of 20,533 genes

in the three CRC categories were calculated and ratios of mean expression between *MLH1*-silenced CRC and non-hypermuted CRC was determined. Student's t test was used to identify those genes whose expression were significantly different between these two categories. Those genes with the lowest P values by t test were considered to be most significantly different in their expression.

Gene expression signature analysis

Gene transcription signature of pathways were defined as follows: p53, canonical bound and up-regulated p53 gene targets, as catalogued in the p53 IARC database (<http://www-p53.iarc.fr/TargetGenes.html>); *KRAS*, genes differentially expressed between *KRAS*-dependent and *KRAS*-independent cell lines, as previously defined by Singh *et al.* [20]; *BRAF*, genes induced in melanocytes after activation of BRAF ($P < 0.05$, t-test, fold > 1.4, using GSE13827 dataset); TGF- β , genes differentially expressed in TGF-beta-treated epithelial cell lines, as previously defined by Padua *et al.* [21]; Using the RNA-seq expression profiles, the TCGA colorectal tumors were scored for relative activity of a given pathway signature; for signatures consisting of up-regulated genes the signature score within each tumor was defined as the Spearman's rank sum statistic of the RPKM values of the genes; for signature consisting of both up-regulated and down-regulated genes, the signature score was defined as the Spearman's rank sum statistic of the up-regulated genes, minus the rank sum statistic of the down-regulated genes. Java Treeview [22] represented mutation/LOH/gene signature patterns as heat maps.

Results

Hypermuted CRC with *MLH1* gene silencing have infrequent *APC* mutations

In the recently reported TCGA study, 224 CRC normal/tumor sample pairs were analyzed by whole exome sequencing, global DNA copy number, promoter methylation, mRNA and miRNA expression [19]. Of 224 CRC, 35 (15.6%) exhibited "hypermethylation" with at least 12.0 non-silent mutations per exome megabase. All but one of the remaining tumors had a mutation rate below 7.0 non-silent mutations per megabase and were classified "non-hypermuted" [19]. For our discovery studies we compared the 35 hypermutated and 189 non-hypermuted tumors for which there was complete exome sequencing data. As a validation step we analyzed an additional 250 CRC (not represented in the original TCGA study) by exome sequencing and assessed their global methylation status. Among the validation set, 45 of 250 CRC were categorized as hypermutated and the remaining 205 CRC were non-hypermuted (Figure 1). Four CRC had extremely high mutation rates and were classified "ultramutated".

In the TCGA study the majority of hypermutated tumors exhibited high microsatellite instability (MSI-H) and silenced *MLH1* expression [19]. Overall, hypermutated CRC had fewer *APC*, *KRAS*, and *TP53* mutations compared to the non-hypermuted CRC [19]. In contrast, mutations in TGF- β signaling genes and *BRAF* were dramatically elevated in the hypermutated tumors. To determine whether these mutational patterns were more specific to a subset of the hypermutated CRC, we stratified the 224 CRC discovery set by *MLH1* silencing or MSI status. When the hypermutated tumors were subcategorized according to *MLH1* silencing status (*MLH1* promoter methylation and low *MLH1* expression), such tumors had a very low *APC* mutation rate (Figure 2A). In the 224 tumor discovery set, only 3 of 22 *MLH1*-silenced hypermutated CRC displayed a truncating *APC* mutation, while 12 of 13 non-*MLH1*-silenced hypermutated CRC exhibited at least one truncating *APC* mutation ($P = 3 \times 10^{-12}$) (Figure 2A,B). Consistent with previous reports, 152 of 189 (80.4%) non-hypermuted CRC contained at least one truncating *APC* mutation. In

addition, hypermutated CRC with *MLH1* silencing had the highest frequency of MSI-H (Figure 2A).

Stratification of the 45 hypermutated CRC in the 250 tumor validation set by *MLH1* promoter methylation status indicated 29 *MLH1* methylated CRC and 16 non-*MLH1* methylated CRC. Only 6 of 29 (20.7%) *MLH1* methylated CRC contained *APC* truncating mutations, whereas 8 of 16 (50%) non-*MLH1* methylated CRC contained truncating *APC* mutations (Figure 2C; Table S1). In contrast, 176 of 205 (85.9%) non-hypermutated CRC exhibited at least one truncating *APC* mutation (Figure 2C, Figure S1). Thus, the significantly reduced *APC* mutation rate in the *MLH1* methylated hypermutated CRC of the validation set corroborated the low *APC* mutation numbers observed in the original discovery set.

For the discovery set, the non-*MLH1*-silenced hypermutated CRC exhibited more variation in MSI status. While 21 of 22 *MLH1*-silenced hypermutated tumors displayed MSI-H (high microsatellite instability), only 6 of 13 non-*MLH1*-silenced CRC were in this category (and non-hypermutated CRC rarely had MSI-H). Despite having lower frequencies of MSI-H, the non-*MLH1*-silenced hypermutated CRC had higher mean mutation rates (100 non-silent mutations per megabase) than their *MLH1*-silenced counterparts (27.8 non-silent mutations per megabase) ($P = 0.032$). The hypermutated CRC with the seven highest mutation rates were considered “ultramutated” and were all non-*MLH1*-silenced tumors (Figure 2A).

***MLH1*-silenced hypermutated CRC display evidence of reduced WNT signaling**

Truncating mutations in *APC* in colorectal cancer result in stabilization and nuclear localization of beta-catenin and transcriptional activation of WNT signaling targets [9]. About 80% of CRC have *APC* truncating mutations that activate WNT signaling, but *MLH1*-silenced hypermutated CRC have low frequencies of *APC* truncating mutations (Figure 2). This suggests that most *MLH1*-silenced CRC may not have activated WNT signaling. To test this, we compared RNA expression profiles as determined by RNAseq. Using the 224 tumor discovery set, we compared RNA expression levels of 20,533 genes in 20 *MLH1*-silenced hypermutated CRC and 9 non-*MLH1*-silenced hypermutated CRC to that in 140 non-hypermutated CRC for which complete RNA expression data was available. The genes were then sorted by those most differentially regulated between each of the two hypermutated CRC categories and the non-hypermutated CRC. As expected, when comparing the *MLH1*-silenced CRC to the non-hypermutated CRC, the most differentially regulated gene was *MLH1*, with 4-fold higher expression in the non-hypermutated CRC compared to the hypermutated *MLH1*-silenced CRC ($P = 2.29 \times 10^{-41}$) (Figure 3A). The second most differentially regulated gene was *AXIN2*, a WNT signaling regulator and target, which was increased 7-fold in the non-hypermutated CRC compared to the *MLH1*-silenced hypermutated CRC ($P = 5.61 \times 10^{-27}$) (Figure 3A).

The reduced *AXIN2* in the hypermutated *MLH1*-silenced CRC indicated that WNT signaling is downregulated in these tumors. We compared RNA expression datasets for over-represented gene ontology categories using the software provided by the Database for Annotation, Visualization and Integrated Discovery (DAVID) and the Broad Institute Gene Set Enrichment Analysis (GSEA) Websites. Both software packages showed that WNT signaling pathway genes were significantly over-represented among the most differentially regulated genes between *MLH1*-silenced CRC and non-hypermutated CRC. Invariably the WNT signaling genes were expressed at higher levels in the non-hypermutated CRC compared to the *MLH1*-silenced hypermutated CRC (Figure 3A). We also compared gene expression of the non-*MLH1*-silenced hypermutated tumors to the non-hypermutated CRC and these tumors generally displayed higher expression of WNT genes than their *MLH1*-

silenced counterparts but less expression than the non-hypermuted CRC (Figure 3A). Heat maps were generated from the expression data for 41 WNT antagonists, 41 WNT agonists, and 17 WNT targets and graphically show differences in expression among the three categories of CRC (Figure 3B). Non-hypermuted tumors show the highest levels of WNT gene expression, while the *MLH1*-silenced hypermutated CRC show the lowest overall levels of WNT gene expression (Figure 3B). High expression of WNT antagonist genes in the non-hypermuted CRC may seem inconsistent with elevated WNT signaling, but deregulated WNT signaling resulting from *APC* loss triggers an intense but ineffective negative feedback loop accompanied by high expression of WNT antagonists [23,24]. Thus, based on expression data and *APC* mutation patterns, there appears to be a marked reduction in WNT signaling in the *MLH1*-silenced hypermutated CRC compared to the non-hypermuted CRC.

Global microRNA expression patterns were consistent with reduced WNT signaling in *MLH1*-silenced CRC. In the 224 tumor discovery set, we compared the three categories of CRC for differences in expression of each miRNA species and found several dozen that were highly up- or downregulated in the *MLH1*-silenced CRC compared to the non-hypermuted CRC (Figure 3C,D). For the 17 most significantly downregulated miRNAs in the *MLH1*-silenced CRC, we found that 9 of these miRNAs were reported in the literature to be upregulated by WNT and/or TGF- β signaling (Figure 3C). Thus, the non-hypermuted CRC showed evidence of increased WNT and TGF- β signaling based on miRNA signatures. The non-*MLH1*-silenced hypermutated tumors showed intermediate levels of downregulation for these 17 miRNAs. Thus, microRNA expression data support the mRNA expression analyses indicating that the *MLH1*-silenced CRC display decreased WNT signaling activity.

High rates of *BRAF* mutation and low rates of *KRAS* mutation in *MLH1*-silenced CRC

BRAF mutations are increased and *KRAS* mutations are reduced in MSI-H CRC [5]. When we stratified the discovery set hypermutated CRC into *MLH1*-silenced and non-*MLH1*-silenced groups, only 1 of 22 (4%) *MLH1*-silenced CRC had a canonical oncogenic *KRAS* mutation (mutation of codons 12, 13, or 61), while 5 of 13 (38%) *MLH1*-non-silenced hypermutated CRC had oncogenic *KRAS* or *NRAS* mutations (Figure 4A,B) ($P = 0.02$). The 38% *KRAS* mutation incidence in the non-*MLH1*-silenced hypermutated CRC was not significantly different from the 48% RAS mutation frequency observed in the non-hypermuted tumors (Figure 4B). In contrast, 15 of 22 (68%) *MLH1*-silenced CRC had the oncogenic *BRAF* V600E mutation, whereas none of the 13 *MLH1*-non-silenced CRC and only 5 of 189 (2.6%) non-hypermuted CRC exhibited this mutation (Figure 4A,C). Analysis of *BRAF* AND RAS gene expression signatures (defined previously in cell line studies) showed that the *MLH1*-silenced CRC displayed evidence of enhanced *BRAF* signaling and reduced RAS signaling compared to their non-*MLH1*-silenced hypermutated and non-hypermuted counterparts (Figure 4F). The presence of oncogenic *KRAS* mutations correlated with higher levels of a “*KRAS* dependency” gene expression signature [16] ($P < 0.01$, t-test), while the presence of *BRAF* V600E mutations correlated with a higher *BRAF*-activation signature ($P < 0.0001$, t-test).

Analysis of the validation set for *KRAS*, *NRAS*, and *BRAF* mutations showed similar patterns to those observed in the discovery set. Only 10% of *MLH1* methylated and hypermutated CRC exhibited canonical *KRAS* or *NRAS* mutations, while 25% of non-*MLH1* methylated and hypermutated CRC and 47% of non-hypermuted CRC contained RAS mutations (Figure 4D; Table S1). Moreover, *BRAF* mutations were very frequent in the *MLH1* methylated hypermutated CRC relative to the other CRC categories. Over 59% of *MLH1* methylated and hypermutated CRC exhibited *BRAF* V600E mutations, while 13% of non-*MLH1*

methyated hypermutated CRC and 3% of non-hypermutated CRC contained *BRAF* mutations (Figure 4E; Table S1).

***MLH1*-silenced and non-silenced hypermutated CRC exhibit reduced p53 mutations and increased TGF- β pathway mutations**

We examined mutation frequencies in tumor suppressors, including *TP53* and TGF- β pathway genes, as these genes are frequently altered in CRC [17,25,26]. In the 224 tumor discovery set, *TP53* mutations were significantly less frequent in the hypermutated CRC compared to the non-hypermutated CRC. Almost 60% of non-hypermutated CRC displayed *TP53* mutations while only 18% of *MLH1*-silenced and 23% of non-*MLH1*-silenced hypermutated CRC exhibited *TP53* mutations (Figure 5A,B). Analysis of *TP53* status in the 250 tumor validation set showed a similar pattern as the discovery set, with 24%, 37%, and 67% p53 mutations in the *MLH1*-silenced, non-*MLH1*-silenced, and non-hypermutated CRC, respectively (Figure 5C). Analysis of a p53 expression signature in the discovery set confirmed that p53 target genes were more upregulated ($P < 0.0001$, t-test) in the hypermutated CRC compared to the non-hypermutated CRC (Figure 5G), and that p53 signature correlated with both p53 somatic mutation and copy loss ($P < 10^{-10}$ for each).

The CRC hypermutated tumors in the discovery and validation sets also exhibited high rates of mutations in the TGF- β pathway member genes such as *SMAD2/3/4*, *ACVR1B*, *ACVR2A*, and *TGFBR2*, while such mutations in the non-hypermutated CRC were much lower in frequency (Figure 5A,D,E). There were no significant differences in TGF- β pathway member mutations between the *MLH1*-silenced and non-*MLH1*-silenced hypermutated CRC. In many cases, the TGF- β pathway alterations were the result of frameshift mutations observed in long mononucleotide repeats that are present in the coding sequences of *ACVR1B*, *ACVR2A*, and *TGFBR2*. When the TGF- β RNA expression signatures were compared among the tumors, TGF- β signaling was decreased in a large fraction of the *MLH1*-silenced CRC relative to the majority of non-HM CRC (Figure 5F).

***MLH1*-silenced CRC show higher mononucleotide frameshifts but lower overall mutation rates compared to non-*MLH1*-silenced hypermutated CRC**

TGF- β pathway member genes such as *ACVR1B*, *ACVR2A*, and *TGFBR2* show increased frameshift mutations within mononucleotide repeats within their coding sequences [5,19]. We expanded this frameshift analysis to 20 genes that had mononucleotide repeats of six nucleotides or longer. In the 224 tumor discovery set, the hypermutated tumors displayed dramatically higher frameshift mutation rates in those 20 genes compared to the non-hypermutated tumors (Figure 6A,B). The *MLH1*-silenced CRC showed significantly higher overall mononucleotide frameshift rates compared to their *MLH1*-non-silenced hypermutated counterparts (Figure 6B). For the 20 examined genes, the average rates were 5.14 mononucleotide repeat frameshift mutations per *MLH1*-silenced CRC, 2.08 per non-*MLH1*-silenced CRC, and 0.05 per non-hypermutated CRC. The difference between the *MLH1*-silenced and non-silenced hypermutated CRC was significant ($P = 0.0044$). This pattern was repeated in the 250 tumor validation set (Figure 6D). When a set of genes that included the 20 genes with mononucleotide repeats and 6 genes important for DNA mismatch repair were analyzed in the discovery set tumors, missense and nonsense mutation rates were higher in the non-*MLH1*-silenced CRC (8.08 mutations/tumor) than the *MLH1*-silenced CRC (1.86 mutations/tumor) and the non-hypermutated CRC (0.21 mutations/tumor) (Figure 6C). The non-*MLH1*-silenced CRC had significantly more missense and nonsense mutations in the various mismatch repair genes and homopolymer containing genes than did the *MLH1*-silenced CRC ($P = 0.028$) (Figure 6C). However, in the validation tumor set this difference in the two hypermutated tumor types was less pronounced (Figure 6E).

Discussion

Sporadic colorectal cancers (CRC) are usually subclassified according to whether they display microsatellite instability (MSI) or not [5]. The 15% of CRC with MSI are frequently associated with distinctive molecular features, including defects in mismatch repair, high genome mutation rates, and epigenetic silencing of the *MLH1* mismatch repair gene. Another useful way to subcategorize CRC is by mutation rate. The TCGA study showed that over 15% of all CRC had distinctly elevated gene mutation rates (>12 non-silent mutations/10⁶ exonic nucleotides) compared to the remainder of the CRC. The hypermutated CRC had a high fraction of tumors with MSI-H (27/35) and *MLH1* silencing (22/35). The TCGA CRC study also revealed that the hypermutated CRC had fewer *APC*, *KRAS*, and *TP53* mutations, while these tumors also displayed higher numbers of *BRAF* and TGF- β pathway-related mutations. These mutation patterns were observed in both the discovery set of 224 CRC and the validation set of 250 CRC (Table 1, Table S1).

The gene mutation differences between hypermutated and non-hypermutated CRC suggested that these tumors might undergo distinct pathways to tumorigenesis. To examine this possibility further, we subdivided the 35 hypermutated CRC by MSI and *MLH1* silencing status and found that *MLH1* silencing best defined a subcategory in which *APC* and *KRAS* mutations were infrequent and *BRAF* V600E mutations were frequent. The *MLH1*-silenced hypermutated CRC were almost invariably MSI-H, while hypermutated CRC without *MLH1* silencing were highly variable in MSI status. This latter subcategory of hypermutated CRC also displayed high *APC* and *KRAS* and low *BRAF* mutation rates similar to that of the non-hypermutated tumors. However, both categories of hypermutated CRC do exhibit high rates of TGF- β pathway mutations and low frequencies of p53 mutations relative to non-hypermutated CRC. These studies suggest that *MLH1* silencing defines a subset of about 10% of sporadic CRC that are distinct from both non-hypermutated CRC and other hypermutated CRC that do not exhibit *MLH1* silencing.

Truncating mutations in the *APC* tumor suppressor gene are the signature mutations of colorectal cancer, occurring in roughly 80% of all tumors. The low *APC* mutation rates in *MLH1*-silenced hypermutated CRC were not accompanied by increased mutation rates in other WNT pathway regulators, suggesting that these tumors do not initiate or progress by augmented WNT signaling. This was supported by analysis of CRC mRNA expression levels showing that expression of WNT regulatory genes and target genes were significantly reduced in the *MLH1*-silenced CRC relative to non-hypermutated CRC. In addition, the CRC microRNA analyses showed that upregulated miRNAs associated with augmented WNT signaling are overrepresented in the non-hypermutated CRC compared to the *MLH1*-silenced CRC. The combined mutational and expression data argue that enhanced WNT signaling does not play a major role in initiation or progression of *MLH1*-silenced hypermutated CRC, in contrast to the non-hypermutated CRC. The non-*MLH1*-silenced hypermutated CRC generally displayed intermediate levels of WNT signaling (Figure 3, Table 1).

If enhanced WNT signaling is not important in the *MLH1*-silenced CRC, what is the positive growth stimulus that drives these particular tumors? The frequent presence of oncogenic *BRAF* V600E mutations the *MLH1*-silenced CRC suggest that this event be a key driver of tumor cell proliferation. Since *BRAF* is a major downstream target of RAS signaling, it is not surprising that *KRAS* and *NRAS* oncogenic mutations are rare in the *MLH1*-silenced CRC while occurring much more frequently in the non-*MLH1*-silenced CRC that do not have *BRAF* mutations. We hypothesize that positive growth signaling in the *MLH1*-silenced CRC is driven in part by *BRAF* activation, whereas proliferation in the non-silenced CRC is driven largely by enhanced WNT and RAS signaling.

The *MLH1*-silenced CRC exhibited significantly higher rates of frameshifts in mononucleotide repeats, including key genes of the TGF- β signaling pathway. In particular, the *MLH1*-silenced CRC exhibited frequent frameshifts in mononucleotide repeats present in mismatch repair genes *MSH3* and *MSH6*. Because *MLH1*, *MSH3*, and *MSH6* play an important role in resolving small insertions/deletions in newly replicated DNA that arise from DNA polymerase slippage on long mononucleotide repeats [27], absence of their function in CRC would be expected to greatly increase the frequency of frameshifts in genes containing such repeats. Another interesting observation was that the non-*MLH1*-silenced hypermutated CRC displayed higher mutation rates than their *MLH1*-silenced counterparts. The mechanisms for this extreme mutability in the non-silenced cohort remain unclear, but this group of CRC had frequent mutations in the exonuclease domain of the POLE DNA repair polymerase [19].

Finally, does the reclassification of CRC by *MLH1* silencing status instead of by MSI status provide any clinical benefit? We found that *APC*, *BRAF*, and *KRAS* mutation status was more definitive in *MLH1*-silenced CRC compared to MSI-H CRC. The 68% frequency of *BRAF* mutations in the *MLH1*-silenced CRC make this subgroup particularly attractive candidates for treatment with BRAF AND MEK inhibitors [28,29] and unattractive candidates for treatment with anti-EGFR molecules such as cetuximab [30]. Thus, we propose that a simple test to identify *MLH1* expression levels in CRC could potentially be as useful as the microsatellite instability test for clinical diagnoses and therapeutic decisions.

Acknowledgments

This study was supported by a grant to the Baylor College of Medicine Human Genome Sequencing Center U24 CA143843.

List of Abbreviations

CRC	colorectal cancers
DAVID	Database for Annotation, Visualization and Integrated Discovery
GSEA	Gene Set Enrichment Analysis
HM	hypermutated
Mb	megabase(s)
LOH	loss of heterozygosity
MSI	microsatellite instability
MSS	microsatellite stable
RPKM	Reads Per Kilobase of exon model per Million mapped reads
TCGA	The Cancer Genome Atlas (multi-institutional research consortium)

References

1. Registries CfDCaPNPoC. 2007 Top Ten Cancers. U.S. Department of Health and Human Services; 2011. United States Cancer Statistics.
2. Wolpin BM, Mayer RJ. Systemic treatment of colorectal cancer. *Gastroenterology*. 2008; 134:1296–1310. [PubMed: 18471507]
3. Ionov Y, Peinado MA, Malkhosyan S, et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*. 1993; 363:558–561. [PubMed: 8505985]

4. Parsons R, Li GM, Longley MJ, et al. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell*. 1993; 75:1227–1236. [PubMed: 8261516]
5. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010; 138:2073–2087. [PubMed: 20420947]
6. Kane MF, Loda M, Gaida GM, et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res*. 1997; 57:808–811. [PubMed: 9041175]
7. Toyota M, Ahuja N, Ohe-Toyota M, et al. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA*. 1999; 96:8681–8686. [PubMed: 10411935]
8. Ogino S, Noshio K, Kirkner GJ, et al. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut*. 2009; 58:90–96. [PubMed: 18832519]
9. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer*. 2001; 1:55–67. [PubMed: 11900252]
10. Walther A, Johnstone E, Swanton C, et al. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer*. 2009; 9:489–499. [PubMed: 19536109]
11. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med*. 2008; 359:1757–1765. [PubMed: 18946061]
12. Salahshor S, Kressner U, Pahlman L, et al. Colorectal cancer with and without microsatellite instability involves different genes. *Genes Chromosomes Cancer*. 1999; 26:247–252. [PubMed: 10502323]
13. Frattini M, Balestra D, Suardi S, et al. Different genetic features associated with colon and rectal carcinogenesis. *Clin Cancer Res*. 2004; 10:4015–4021. [PubMed: 15217933]
14. Samowitz WS, Slattery ML, Sweeney C, et al. APC mutations and other genetic and epigenetic changes in colon cancer. *Mol Cancer Res*. 2007; 5:165–170. [PubMed: 17293392]
15. Albuquerque C, Baltazar C, Filipe B, et al. Colorectal cancers show distinct mutation spectra in members of the canonical WNT signaling pathway according to their anatomical location and type of genetic instability. *Genes Chromosomes Cancer*. 2010; 49:746–759. [PubMed: 20544848]
16. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006; 314:268–274. [PubMed: 16959974]
17. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science*. 2007; 318:1108–1113. [PubMed: 17932254]
18. Leary RJ, Lin JC, Cummins J, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc Natl Acad Sci USA*. 2008; 105:16224–16229. [PubMed: 18852474]
19. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–337. [PubMed: 22810696]
20. Singh A, Greninger P, Rhodes D, et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell*. 2009; 15:489–500. [PubMed: 19477428]
21. Padua D, Zhang XH, Wang Q, et al. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell*. 2008; 133:66–77. [PubMed: 18394990]
22. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*. 2004; 20:3246–3248. [PubMed: 15180930]
23. Lustig B, Jerchow B, Sachs M, et al. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol Cell Biol*. 2002; 22:1184–1193. [PubMed: 11809809]
24. Boerboom D, White LD, Dalle S, et al. Dominant-stable beta-catenin expression causes cell fate alterations and Wnt signaling antagonist expression in a murine granulosa cell tumor model. *Cancer Res*. 2006; 66:1964–1973. [PubMed: 16488995]
25. Markowitz S, Wang J, Myeroff L, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science*. 1995; 268:1336–1338. [PubMed: 7761852]

26. Hempen PM, Zhang L, Bansal RK, et al. Evidence of selection for clones having genetic inactivation of the activin A type II receptor (ACVR2) gene in gastrointestinal cancers. *Cancer Res.* 2003; 63:994–999. [PubMed: 12615714]
27. Kolodner RD, Marsischky GT. Eukaryotic DNA mismatch repair. *Curr Opin Genet Dev.* 1999; 9:89–96. [PubMed: 10072354]
28. Shepherd C, Puzanov I, Sosman JA. B-RAF inhibitors: an evolving role in the therapy of malignant melanoma. *Curr Oncol Rep.* 2010; 12:146–152. [PubMed: 20425073]
29. Nissan MH, Solit DB. The "SWOT" of BRAF inhibition in melanoma: RAF inhibitors, MEK inhibitors or both? *Curr Oncol Rep.* 2011; 13:479–487. [PubMed: 21997758]
30. Lin JS, Webber EM, Senger CA, et al. Systematic review of pharmacogenetic testing for predicting clinical benefit to anti-EGFR therapy in metastatic colorectal cancer. *Am J Cancer Res.* 2011; 1:650–662. [PubMed: 21779535]

Figure 1.

Mutation frequencies in human colorectal cancer delineate hypermutated and non-hypermutated patient groups. Points to the left of the vertical dashed line represent hypermutated tumors. The four tumors at the far left with the highest mutation rates are considered “ultramutated”. Data are from 250 colorectal tumors in the “validation” patient cohort (frequencies for the 224 tumor “discovery” cohort featured in [19]). For each tumor, rectal or colon cancer status, MSI status, *MLH1* methylation status, hypermutation status, as well as *APC*, *KRAS*, and *BRAF* mutational status are indicated along the bottom. Y-axis scale is log base 10.

Figure 2.

Subclassification of CRC into three groups based on mutation frequency and *MLH1* silencing status. (A) The 35 hypermutated CRC and 10 representative non-hypermutated CRC from the 224 tumor discovery set are listed. The hypermutated tumor identification (ID) numbers and mutation rates (number of non-silent mutations per megabase of exonic sequence) are shaded pink, except for the seven most hypermutated tumors which are shaded blue. The ten representative non-hypermutated tumors are shaded in orange. The 35 hypermutated CRC are further subdivided into 22 tumors with *MLH1* silencing (light blue shading) and 13 tumors without such silencing (yellow shading). Each tumor is further characterized for microsatellite instability (MSI), CpG Island Methylator Phenotype (CIMP) status [19], and *APC* gene mutation status. Those tumors with truncating *APC* mutations are shaded in green with the specific amino acid mutation indicated. Tumors with missense mutations are indicated in light green boxes and are less likely to be functionally important. MSI-H: high levels of microsatellite instability; MSI-L: low levels of microsatellite instability; MSI-S: stable, no significant microsatellite instability. CIMP-H: high CIMP; CIMP-L: low CIMP; Cluster 3/4: lowest CIMP levels [19]. (B) Graph showing relative frequency of *APC* truncating mutations in *MLH1*-silenced hypermutated CRC (N = 22), non-*MLH1*-silenced hypermutated CRC (N = 13), and non-hypermutated CRC (N = 189) of the discovery set. (C) Graph showing relative frequency of *APC* truncating mutations in *MLH1* methylated hypermutated CRC (N = 29), non-*MLH1* methylated hypermutated CRC (N = 16), and non-hypermutated CRC (N = 205) of the 250 tumor validation set. Relative statistical significance for differences between each category are indicated by P-values (one-sided Fisher's exact test).

Figure 3.

WNT regulatory genes and target genes are downregulated in *MLH1*-silenced CRC. (A) Representative WNT signaling regulatory genes and target genes are significantly downregulated in the *MLH1*-silenced hypermutated CRC. Table includes gene expression relative to non-hypermutated CRC ("non-HM"), both for *MLH1*-silenced CRC and non-*MLH1*-silenced hypermutated CRC, with corresponding t-test comparisons. WNT regulatory genes and target genes showing the most differential regulation by P values are shown. (B) Heat map showing that canonical WNT regulatory genes and target genes as a group appear downregulated (blue) in *MLH1*-silenced CRC, but upregulated (yellow) in non-hypermutated CRC. (C) Downregulated miRNAs in *MLH1*-silenced CRC relative to non-hypermutated CRC are disproportionately associated with enhanced WNT and TGF-beta signaling. Results are based on expression profiles for 706 miRNAs in 20 *MLH1*-silenced CRC, 7 non-*MLH1*-silenced hypermutated CRC, and 94 non-hypermutated CRC. Those miRNAs showing the most differential expression between *MLH1*-silenced hypermutated CRC and non-hypermutated CRC are shown, and miRNAs associated in the literature with either activation of WNT or TGF- β pathway signaling are indicated. (D) Expression heat map showing top miRNAs differentially regulated ($P < 0.001$) in *MLH1*-silenced CRC (indicated by red part of bar just above the heat map) compared to non-hypermutated CRC. Non-hypermutated CRC and *MLH1*-silenced hypermutated CRC exhibit very different patterns of miRNA expression. Yellow, high expression; blue, low expression. MSI/CIMP status: red, high; blue, low.

Figure 4.

MLH1-silenced CRC exhibit high incidence of *BRAF* V600E mutations and low incidence of oncogenic RAS mutations. (A) Table showing infrequent oncogenic *KRAS* mutations (green boxes) and frequent *BRAF* V600E mutations (green boxes) in *MLH1*-silenced CRC and frequent *KRAS* and no *BRAF* mutations in non-*MLH1*-silenced hypermutated CRC and representative non-hypermutated CRC of the 224 tumor discovery set. *KRAS* and *NRAS* mutations that are not the canonical codon 12, 13, and 61 oncogenic mutations are shown in light green boxes. The *BRAF* mutation that is not a V600E mutation is also indicated in light green. (B,C) Graph showing relative frequency of oncogenic RAS (*NRAS* or *KRAS*) (B) or *BRAF* V600E (C) mutations in *MLH1*-silenced hypermutated CRC (N = 22), non-*MLH1*-silenced hypermutated CRC (N = 13), and non-hypermutated CRC (all non-*MLH1*-silenced) (N = 189) of the discovery set. Statistical significance for differences between each category are indicated (RAS, chi-square test; *BRAF*, one-sided Fisher's exact test). (D,E) Graph showing relative frequency of oncogenic RAS (*NRAS* or *KRAS*) (D) or *BRAF* V600E (E) mutations in *MLH1*-methylated hypermutated CRC (N = 29), non-*MLH1*-methylated hypermutated CRC (N = 16), and non-hypermutated CRC (N = 205) of the 250 tumor validation set. (F) *BRAF* and RAS gene expression signatures correlate with *BRAF* and RAS mutational status in *MLH1*-silenced and non-silenced CRC in the discovery set. Somatic mutation events for *KRAS*, *NRAS*, and *BRAF* are shown across the 224 TCGA colorectal

tumors. In addition, gene expression signatures for KRAS and BRAF activity *in vitro* (defined using public expression databases, see Methods) were applied to the TCGA colorectal expression data, in order to score each tumor for relative signature activity (yellow: more active, blue: less active).

Figure 5.

Hypermuted CRC exhibit high incidence of TGF- β pathway mutations and low incidence of p53 mutations. (A) Table from the 224 tumor discovery set showing that both categories of hypermutated CRC exhibit frequent mutations in genes that regulate TGF- β signaling and infrequent mutations in p53, relative to their non-hypermutated counterparts. Mutations in six frequently mutated TGF- β pathway genes (*SMAD2*, *SMAD3*, *SMAD4*, *ACVR1B*, *ACVR2A*, *TGFB2*) and *TP53* are indicated by green boxes. (B,C) Graphs showing relative frequency of TGF- β pathway (B) or *TP53* (C) mutations in *MLH1*-silenced hypermutated CRC (N = 22), non-*MLH1*-silenced hypermutated CRC (N = 13), and non-hypermutated CRC (all non-*MLH1*-silenced) (N = 189) from the discovery set. (D,E) Graphs showing relative frequency of TGF- β pathway (D) or *TP53* (E) mutations in *MLH1*-silenced hypermutated CRC (N=29), non-*MLH1*-silenced hypermutated CRC (N=16) and non-hypermutated CRC (N=205) of the validation set. Statistical significance for differences between each category is indicated (TGF- β , chi-squared test; p53, one-sided Fisher's exact test). (F) TGF- β pathway mutations in the discovery set correlate with a reduced TGF- β expression signature as well as *MLH1* silencing. Somatic mutation and copy loss-of-heterozygosity (LOH) patterns are shown for TGF- β regulator genes, with corresponding TGF- β activity levels as inferred from gene target expression (yellow: high activity). (G) Increased p53 expression signature correlates with relative absence of p53 mutations in CRC with *MLH1* silencing in the discovery set. *TP53* mutation and LOH are shown, in addition to average expression of p53-upregulated target genes.

Figure 6. Mutational patterns differ between *MLH1*-silenced and non-silenced hypermutated CRC. (A) *MLH1*-silenced CRC display increased frameshift mutations in mononucleotide repeats compared to their non-*MLH1*-silenced hypermutated counterparts. Mutation patterns in 23 genes of 22 *MLH1*-silenced hypermutated CRC (N = 22), 13 non-*MLH1*-silenced hypermutated CRC (N=13), and 10 representative non-hypermutated CRC of the discovery set are shown. Of the 23 genes shown, 20 (indicated by orange shading) have long mononucleotide repeats that are frequently subject to frameshift mutations in hypermutated CRC. The six genes at the left are associated with DNA mismatch repair. Red boxes indicate frameshift mutations in long mononucleotide repeats and green boxes indicate all other types of non-silent mutations. Total numbers of each mutation for each tumor are shown in the right two columns. Because some genes have more than one mutation in a tumor the mutation totals do not simply correlate with the number of red or green boxes for a particular tumor. (B) Graph shows mean number of frameshift mutations per discovery set tumor in mononucleotide repeat sequences contained within 20 genes known to be affected by such mutations in colorectal cancers with microsatellite instability. Each bar represents the mean number of such mutations for each tumor among *MLH1*-silenced hypermutated CRC (N = 22), non-*MLH1*-silenced hypermutated CRC (N=13), or non-hypermutated CRC (N = 189). (C) Non-*MLH1*-silenced hypermutated CRC display higher rates of non-homopolymer frameshift mutations than their *MLH1*-silenced counterparts. Graph shows the mean number of non-homopolymer mutations in 23 genes (including six DNA mismatch repair-associated genes) in each tumor for the three separate CRC categories of the discovery set. (D) Mean numbers of mononucleotide frameshift mutations for each tumor category in the 250 tumor validation set CRC. (E) Mean numbers of non-homopolymer mutations for each tumor category in the validation set CRC. Statistical significance for differences in each type of mutation between CRC categories are indicated by P values (Fisher's exact test).

Table 1

Summary of Key Genetic Attributes of Colorectal Cancers (Combined Discovery and Validation Sets)

	Hypermethylated CRC*		Non-Hypermethylated CRC*
	<i>MLH1</i> -silenced*	Non- <i>MLH1</i> -silenced*	
Fraction of CRC	51/474 (10.8%)	29/474 (6.1%)	394/474 (83.1%)
Fraction MSI-H	48/51 (94.1%)	18/29 (62.1%)	2/394 (0.5%)
Mean Mutation Rate	32.26/Mb	59.82/Mb	4.09/Mb
“Ultramutated” CRC	1/51 (2.0%)	10/29 (34.5%)	0/394 (0%)
<i>APC</i> Mutations/Tumor	12/51 (23.5%)	31/29 (107%)	440/394 (117%)
WNT Gene Expression	+	++	++++
<i>KRAS</i> Mutations/Tumor	4/51 (7.8%)	9/29 (31.0%)	167/394 (42.4%)
<i>BRAF</i> Mutations/Tumor	32/51 (62.7%)	2/29 (6.9%)	12/394 (3.0%)

* Includes CRC from 224 tumor discovery cohort and *MLH1* methylated CRC from 250 tumor validation cohort