Tetranucleotide Microsatellite Mutational Behavior Assessed in Real Time: Implications for Future Microsatellite Panels

Maide Ö. Raeker,*‡ Jovan Pierre-Charles,*‡ and John M. Carethers§

*Division of Gastroenterology and Hepatology; ‡Department of Internal Medicine; and §Department of Human Genetics and Rogel Cancer Center, University of Michigan, Ann Arbor, Michigan

SUMMARY

Frameshift mutation behavior of human tetranucleotide microsatellites is not known, yet is the key marker to identify elevated microsatellite alterations at selected tetranucleotide repeats. In plasmid-transfected, DNA mismatch repair–deficient cell models, we discovered a deletional frameshift bias for tetranucleotide sequences with deletion of initial insertions through intermediate heteroduplexes.

BACKGROUND & AIMS: Fifty percent of colorectal cancers show elevated microsatellite alterations at selected tetranucleotide repeats (EMAST) and are associated with inflammation, metastasis, and poor patient outcome. EMAST results from interleukin 6–induced nuclear-to-cytosolic displacement of the DNA mismatch repair protein Mutated S Homolog 3, allowing frameshifts of dinucleotide and tetranucleotide but not mononucleotide microsatellites. Unlike mononucleotide frameshifts that universally shorten in length, we previously observed expansion and contraction frameshifts at tetranucleotide sequences. Here, we developed cell models to assess tetranucleotide frameshifts in real time.

METHODS: We constructed plasmids containing native (AAAG)18 and altered-length ([AAAG]15 and [AAAG]12) human D9S242 locus that placed enhanced green fluorescent protein +1 bp/-1 bp out-of-frame for protein translation and stably transfected into DNA mismatch repair–deficient cells for clonal selection. We used flow cytometry to detect enhanced green fluorescent protein–positive cells to measure mutational behavior.

RESULTS: Frameshift mutation rates were 31.6 to 71.1 × 10^−4 mutations/cell/generation and correlated with microsatellite length (r^2 = 0.986, P = .0375). Longer repeats showed modestly higher deletion over insertion rates, with both equivalent for shorter repeats. Accumulation of more deletion frameshifts contributed to a distinct mutational bias for each length (overall: 77.8% deletions vs 22.2% insertions), likely owing to continual deletional mutation of insertions. Approximately 78.9% of observed frameshifts were 1 AAAG repeat, 16.1% were 2 repeats, and 5.1% were 3 or more repeats, consistent with a slipped strand mispairing mutation model.

CONCLUSIONS: Tetranucleotide frameshifts show a deletion bias and undergo more than 1 deletion event via intermediates, with insertions converted into deletions. Tetranucleotide markers added to traditional microsatellite instability panels will be able to determine both EMAST and classic microsatellite instability, but needs to be assessed by multiple markers to account for mutational behavior and intermediates. (Cell Mol Gastroenterol Hepatol 2020;9:689–704; https://doi.org/10.1016/j.jcmgh.2020.01.006)

Keywords: DNA Mismatch Repair; MSH3; Tetranucleotide Microsatellites; Frameshift Mutation; EMAST.
As many as 50% of all colorectal cancers (CRCs) show some form of DNA mismatch repair (MMR) defect that drives its pathogenesis and/or progression. For instance, germline mutations in the MMR genes MSH2, MLH1, PMS2, and MSH6 cause Lynch syndrome, a heritable condition in which patients may develop CRC and other gastrointestinal and female reproductive tract cancers. Lynch syndrome CRCs represent approximately 3% of all CRCs. Another inherited human condition is constitutional MMR deficiency, in which patients received 2 mutated MMR genes in their germline, 1 from each Lynch syndrome parent. Constitutional MMR deficiency is extremely rare, and constitutes only a tiny fraction of 1% of all CRCs. Somatic inactivation of MLH1 via hypermethylation of its promoter region is a common cause for sporadic microsatellite instability (MSI) of CRCs, and constitutes approximately 15% of all CRCs. Lynch-like syndrome, a condition in which CRCs show 2 somatic MMR gene mutations, is observed in 1%–2% of all CRCs. In all of the 4 conditions described earlier, tumors show the biomarker MSI-high (MSI-H), and, in general, the outcome of patients with an MSI-H tumor are more favorable compared with patients with a microsatellite stable (MSS) tumor. Another form of MSI is termed elevated microsatellite alterations at selected tetranucleotide repeats (EMAST). EMAST is observed in 50% of all sporadic CRCs and its detection requires the use of tetranucleotide microsatellite markers that are not currently present in MSI panels. The outcome of patients with EMAST tumors contrasts sharply with patients with MSI-H tumors; patients with EMAST tumors show localized inflammation in the tumor, often show metastasis, and have poor prognosis compared with patients without EMAST tumors. Because EMAST is a somatically acquired biomarker associated with inflammation, it also can be termed inflammation-associated microsatellite alterations.

Microsatellites are tandem repeats of short DNA sequences that are 1–6 base pairs (bp) long, with each unit repeated 6 to as many as 30 times, and constitutes approximately 3% of the human genome. These short-sequence DNA repeats are distributed throughout the genome within introns, promoters, and other cis-regulatory regions, and in some exons, and are prone to high mutation rates because of their motif reiterations. It is thought that DNA polymerase pauses at repetitive sequences and transiently dissociates from the template strand during replication. Subsequent incorrect reannealing of the template and newly synthesized strands generate unpaired nucleotide loops on either the newly synthesized strand or the template strand, which results in addition or deletion of 1 or more repeat units of the microsatellites during the next round of DNA replication, and manifests as frameshifts in mononucleotide, dinucleotide, and tetranucleotide microsatellites in the absence of repair. DNA MMR function is essential for recognizing and correcting these intermediate insertion/deletion loops to maintain genome stability. Defective MMR results in accumulation of frameshift mutations at microsatellites that manifests as MSI, and MSI-H is a pathogenic genetic pathway that directs the development and progression of several cancers, including CRC. Clinically, MSI-H CRCs are more resistant to 5-fluorouracil chemotherapy because of a lack of MMR, and more sensitive to immune checkpoint blockade owing to hypermutated genomes and subsequent generation of immune-responsive neoantigens, and increased cell expression of programmed cell death receptor/programmed cell death ligand-1.

The type of MSI observed in tumors is dependent on the defect in MMR. Using a panel of at least 5 mononucleotide (eg, [A]n) and dinucleotide (eg, [CA]n) microsatellite markers, the National Institutes of Health Consensus conference defined MSI-H cancers as tumors with 2 or more of 5 markers frameshifted when compared with nontumor tissue, MSI-Low (MSI-L) as having only 1 frameshifted marker, and MSS as having no markers frameshifted. This definition more accurately identified sporadic MSI-H CRCs, which have somatic hypermethylation of MLH1 driving its pathogenesis, and identified most Lynch syndrome patients, in which a germline MMR mutation drives its pathogenesis and multiple organ risk for cancer. Inflammation-associated microsatellite alterations or EMAST is observed in a variety of cancers including CRCs. EMAST is not identified with the National Institutes of Health Consensus marker panel because that panel only contains mononucleotide and dinucleotide microsatellite markers. EMAST (in the absence of MSI-H) is an acquired defect that is triggered by an interleukin 6-induced nuclear-to-cytosol shift of the MMR protein MSH3, causing subsequent dinucleotide and tetranucleotide (eg, [AAAG]n) longer frameshifts of genomic microsatellites. Overall, tumors defective for MLH1, MSH2, or PMS2 would show mononucleotide, dinucleotide, and tetranucleotide frameshifts, whereas tumors defective for MSH6 would manifest mostly mononucleotide and some dinucleotide frameshifts, and tumors with isolated MSH3 dysfunction (EMAST) would show dinucleotide and tetranucleotide instability and no mononucleotide frameshifts. The observation that MSI-L tumors nearly always show dinucleotide without mononucleotide instability indicate that MSI-L is driven

**Abbreviations used in this paper:** bp, base pair; CRC, colorectal cancer; DIG, Digoxigenin; EGFP, enhanced green fluorescent protein; EMAST, elevated microsatellite alterations at selected tetranucleotide repeats; FALCOR, Fluctuation Analysis CalculatOR; MMR, DNA mismatch repair; MR, mutation resistant; MSI, microsatellite instability; MSI-L, microsatellite instability low; MSI-H, microsatellite instability high; MSS, microsatellite stable; PCR, polymerase chain reaction.
by MSH3 dysfunction and thus is a component of EMAST.5

We and others previously have shown that mono-nucleotide and dinucleotide microsatellites in the absence of MMR consistently and uniformly frameshift to shorter microsatellite lengths.35–40 For this to occur, the insertion/deletion loop needs to occur on the template DNA strand, allowing the newly synthesized DNA strand to anneal a shortened complementary microsatellite sequence.35–39 In the absence of MLH1 or with knock-down of MSH3, we initially observed both deletions and insertion length changes at tetranucleotide sequences, suggesting that insertion/deletion loops form at these longer microsatellites on both the template and newly synthesized DNA strand.32 Here, we created cell models capable of detecting both insertion and deletion frameshifts to examine the mutational dynamics of human tetranucleotide microsatellites in the absence of MMR. Understanding the dynamics helps to understand the function of MMR specifically at longer repeat units, and particularly because tetranucleotide sequences are being considered for future microsatellite panels to identify EMAST.

Results

Creation of +1 and -1 bp Frameshift Reporting Model Systems to Measure Tetranucleotide Microsatellite Length Insertion and Deletion Frameshifts

We used the human D9S242 tetranucleotide microsatellite locus to create our model systems because it has been found highly mutable within CRCs and is a commonly used marker locus to determine EMAST.5,6,28 D9S242 contains 18 repeats of AAAG in its native state; we aimed to investigate insertions and deletions from its native length and at modified lengths of (AAAG)15 and (AAAG)12 to assess the effects of microsatellite length on mutational behavior.
within the same locus. The studied tetranucleotide microsatellite lengths cover the similar range of lengths used for EMAST marker panels.\textsuperscript{6,26,30,31}

We built on our prior experience in measuring mononucleotide microsatellite deletion mutation\textsuperscript{37-40} by creating pIREShygB–enhanced green fluorescent protein (EGFP) plasmids capable of measuring both deletional and insertional events for tetranucleotide sequences. As outlined in Figure 1, we constructed an EGFP reporter system containing the 14 bp flanking both the 5’ and 3’ ends of the various-length \textit{D9S242} tetranucleotide microsatellite sequences immediately after the start codon for expression of EGFP. The flanking sequences are important in maintaining the native mutability of the microsatellite.\textsuperscript{39,40} To accurately measure deletion and insertion mutations, we then placed the EGFP protein reading frame +1 bp and -1 bp out-of-frame at the end of the inserted tetranucleotide sequences to prevent EGFP translation (Figure 1A). In this manner, a deletion frameshift of 1 AAAG repeat in the +1 bp out-of-frame system would restore the protein reading frame to express EGFP; similarly, an insertion frameshift of 1 AAAG repeat in the -1 bp out-of-frame system would restore the protein reading frame to express EGFP. For negative controls, to determine any background mutation rates, we interrupted each tetranucleotide length microsatellite by mutating 2 nucleotides at every third repeat to prevent frameshift mutation (mutation resistant [MR]) and placed the sequence -1 bp and +1 bp out-of-frame (Figure 1B). For positive controls, we used the same interrupted sequences as used for the negative controls but placed the sequences in-frame with the EGFP protein reading frame (Figure 1C). All constructs and inserted sequences used for models are listed in Table 1.

We used the MMR-deficient colon cancer cell line HCT116 for our model system because it is devoid of mismatch repair function (MLH1- and MSH3-deficient) (Figure 2A), thus allowing accumulation of microsatellite frameshifts for which we can measure.\textsuperscript{41} We assessed EGFP expression via fluorescent microscopy after transfection of the tetranucleotide microsatellite constructs into HCT116 cells and clonal selection (Figure 2B–D). By 48 hours after transfection, in-frame MR constructs readily expressed EGFP, whereas out-of-frame MR constructs did not, confirming their designed roles as positive and negative controls, respectively (Figure 2B and C). Cells containing the inserted \textit{(AAAG)}\textsubscript{18} in-frame with the EGFP reading frame expressed EGFP by 48 hours after transfection, verifying that a functional EGFP protein was present if in-frame with the tetranucleotide microsatellite (Figure 2D).

We used flow cytometry to quantify the accumulation of EGFP-expressing cells, gated by EGFP expression in the positive controls and nontransfected cells (Figure 3). We assessed EGFP expression in the monoclonal cells weekly for 5 weeks. EGFP expression emerged approximately 2 weeks after stable transfection, with all tetranucleotide length variants accumulating EGFP-positive cells in a linear fashion and at similar rates between the -1 bp

---

**Table 1. Microsatellite Plasmid Constructs With Inserted DNA Microsatellite Sequences**

| Microsatellite plasmid constructs | DNA sequences of microsatellite region in plasmids |
|---------------------------------|-----------------------------------------------|
| \textbf{D9S242-MR-18-inframe} | AAGAAAGAAAGAAGACTGAAAGAAAGAGCGAAAGAAAGATCGAAAGAAAGACGGAAAGAAAGACTGAAAGAAAGATGGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-MR-18-(+1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-MR-18-(1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-MR-12-inframe} | AAGAAAGAAAGAAGACTGAAAGAAAGAGCGAAAGAAAGATCGAAAGAAAGACGGAAAGAAAGATGGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-MR-12-(+1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-MR-12-(1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-18-inframe} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-18-(+1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-18-(1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-12-inframe} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-12-(+1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{D9S242-AAAG-12-(1 bp) OF} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
| \textbf{OF, out-of-frame.} | AAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAGAGAGAAAGG |
and +1 bp out-of-frame constructs (Figure 4A). MR-negative controls for both the +1 bp and -1 bp reporter systems showed very low levels of background spontaneous EGFP expression, with 15–300 EGFP-positive cells 5 weeks after transfection (Figure 4B) compared with accumulated EGFP expression from frameshifts measured in experimental tetranucleotide microsatellite constructs with 3000–15,000 EGFP-positive cells 5 weeks after transfection (Figure 4A).

Tetranucleotide Microsatellite Frameshifts Increase in Mutational Frequency Over Time, With Longer Lengths More Prone to Frameshifts

We determined the mutational frequency of MR (negative control) and tetranucleotide microsatellite (experimental) sequences (proportion of frameshifted mutant cells within the analyzed cells) by using the web-based Fluctuation Analysis CalculatOR (FALCOR) rate calculator that is based on Luria-Delbruck fluctuation analysis (see the Materials and Methods section for more detail). The FALCOR rate calculator is accessible at http://www.mitochondria.org/protocols/FALCOR.html. The average mutant frequency for duplicate cell cultures at each time point showed a near-linear increase with both the -1 bp (insertion) and +1 bp (deletion) out-of-frame constructs (Figure 5). The mutational frequency was higher for longer-length tetranucleotide microsatellites over shorter ones, with (AAAG)$_{18}$ > (AAAG)$_{15}$ > (AAAG)$_{12}$ (Figure 5A). The background mutational frequency observed in MR negative control cells were at least 1 log lower than experimental cells (Figure 5B), which was used to subtract from experimental cell data to accurately determine subsequent mutational rates. MR -1 bp (insertion) construct cells showed a small linear increase over time that was completely absent in the MR +1 bp (deletion)
construct cells (Figure 5B). However, mutational frequencies of either MR -1 bp or +1 bp out-of-frame negative control cells did not increase with increased tetranucleotide repeat length, indicating that the occurrence of background frameshifts is independent of the number of repeats in the construct.

We calculated the insertion and deletion frameshift mutation rate (mutations per cell per generation) for each tetranucleotide construct using the Ma-Sandri-Sarkar Maximum Likelihood Estimator method, corrected for their matched negative controls (Table 2). Longer tetranucleotide sequences showed higher mutation rates over shorter-length sequences with (AAAG)_{18} > (AAAG)_{15} > (AAAG)_{12} (Table 2), and the increase in mutation rates was significant for longer-length sequences for both -1 bp (insertion) or +1 bp (deletion) systems except for detection of deletion frameshifts between (AAAG)_{18} and (AAAG)_{15} (Table 3). In contrast for the (AAAG)_{12} microsatellite, the +1 bp constructs measuring deletion frameshifts showed higher mutation rates than the -1 bp constructs measuring insertion frameshifts (Table 2). To ensure that the observed mutation rate differences were not the result of differences in proliferation of the clonal cells (thus affecting DNA replication rates), we determined the cell growth rates and doubling times of both EGFP-positive and EGFP-negative cells for the (AAAG)_{18} -1 bp and +1 bp systems (Figure 6). We observed identical growth rates and doubling times for all versions of cells containing the various constructs, indicating that the measured frameshift mutation rates were not influenced by differences in DNA replication rates (Figure 6). Overall, longer-length tetranucleotide microsatellites are more prone to frameshifts than shorter ones, and deletion frameshift mutation rates are higher than that of insertion frameshift mutation rates for (AAAG)_{18} and (AAAG)_{15}, but equivalent for (AAAG)_{12} microsatellites. The combined -1 bp and +1 bp mutation rates were linear as a function of tetranucleotide microsatellite length (Figure 7).

**Tetranucleotide Frameshifts Show a Distinct Deletion Bias in its Mutational Spectra**

Six weeks after exponential growth, we sorted EGFP-positive and EGFP-negative cells to determine the degree of accumulation of frameshift alterations within the (AAAG)_{18}, (AAAG)_{15}, and (AAAG)_{12} microsatellites via colony polymerase chain reaction (PCR) amplification and DNA sequencing (Figures 8 and 9 and Table 4). Some colonies from both +1 bp and -1 bp systems contained overlapping DNA sequences indicative of intermediate mutant cells containing heteroduplex DNA, with the -1 bp system measuring insertions showing overall 28.5% of cells as heteroduplex compared with 8.5% of the +1 bp system measuring deletions (Figures 8 and 9). EGFP-negative cells from the +1 bp system for each tetranucleotide length was more than 90% mutation free (Figure 8A–C) whereas EGFP-negative cells from the -1 bp system were 79%–89% mutation free (Figure 9A–C) owing to the higher percentage of heteroduplex DNA. In examining the 709 EGFP-positive,
single-cell colonies from both the -1 bp and +1 bp systems, we noted that some of the (AAAG)\textsubscript{18}, (AAAG)\textsubscript{15}, and (AAAG)\textsubscript{12} microsatellites were at native length (Figures 8D–F and 9D–F). There are 3 potential explanations: (1) daughter cells from DNA heteroduplexes contain the native length microsatellite and retain EGFP positivity, (2) there may be more than 1 copy of the plasmid in cell clones, and/or (3) there are continual mutations after insertion frameshifts that revert back to the microsatellite’s native length. To examine plasmid copy number, we

![Figure 4](image_url)

**Figure 4.** Similar rates of increase in EGFP-positive cells transfected with -1 bp and +1 bp tetranucleotide constructs for each length. (A) (AAAG)\textsubscript{18} (left) (AAAG)\textsubscript{15} (middle), and (AAAG)\textsubscript{12} (right) -1 bp and +1 bp constructs in EGFP-positive cells up to 6 weeks after transfection. Note the Y-axis scale is higher for longer-length constructs. (B) EGFP expression in cells transfected with negative controls did not show an increase over time for +1 bp out-of-frame constructs (left), but showed some length-independent increase with -1 bp out-of-frame negative control constructs (right). Note that the Y-axis scale for negative control transfected cells are almost negligible compared with their uninterrupted repeat counterparts in panel A. OF, out-of-frame.

![Figure 5](image_url)

**Figure 5.** (A) Mutant frequencies (proportion of mutated cells within analyzed cells) increased over time during an exponential cell growth period for uninterrupted tetranucleotide microsatellite repeats. The average mutant frequency of (AAAG)\textsubscript{18} was 1.6-fold higher than that of (AAAG)\textsubscript{15}, and 4.2-fold higher than that of (AAAG)\textsubscript{12} within +1 bp OF system (detecting deletions), and were 2.1- and 3.5-fold higher, respectively, within -1 bp OF system (detecting insertions). The average mutant frequencies of (AAAG)\textsubscript{15} were 2.6- and 1.7-fold higher than those of (AAAG)\textsubscript{12} tracks in +1 bp out-of-frame and -1 bp out-of-frame systems, respectively. (B) Mutant frequencies of interrupted repeats (negative controls) in +1 bp out-of-frame system were approximately 11-fold smaller than those in the -1 bp out-of-frame system. Note that the Y-axis scale is markedly smaller in panel B compared with panel A.
performed Southern blot analysis, which showed that cells transfected with (AAAG)\textsubscript{18} in the -1 bp system and (AAAG)\textsubscript{12} in +1 bp system possessed a single plasmid, whereas other cells harbored 2 or 3 copies of its plasmid (Figure 10). Thus, it remains possible that not all copies of the tetranucleotide microsatellite sequence in EGFP-positive cells frameshifted and could be responsible for the detection of native length sequences. However, this fact would not alter the calculation of the mutation rates significantly because the cell had at least 1 copy frameshifted to become EGFP positive. The fact that cells transfected with (AAAG)\textsubscript{18} in the -1 bp system and (AAAG)\textsubscript{12} in the +1 bp system had only 1 plasmid copy suggests that the third possibility of continual deletion mutations after insertion frameshifts likely occurs, allowing reversion back to the microsatellite’s native length sequence.

EGFP-positive cells from the +1 bp system designed to measure deletion frameshift mutations showed remarkable consistency for losing 1 repeat unit of the tetranucleotide microsatellite (~70% of colonies for each of [AAAG]\textsubscript{18}, [AAAG]\textsubscript{15}, and [AAAG]\textsubscript{12}) (Figure 8D–F and Table 4). Deletion of 2 repeat units of the tetranucleotide microsatellite occurred in 4.3% to 6.7% of colonies among the 3 repeat lengths (Table 4). Insertions were observed for only 3.6% of colonies from (AAAG)\textsubscript{18}.

EGFP-positive cells from the -1 bp system designed to measure insertion frameshift mutations showed evidence for a broad and diverse mutational spectrum compared with the +1 bp system, showing both insertions and deletions at DNA sequencing for all 3 tetranucleotide microsatellite lengths (Figure 9D–F). Although designed to measure 1 repeat unit insertions, we observed that only 23.5% of (AAAG)\textsubscript{18}, 3.6% of (AAAG)\textsubscript{15}, and 16.3% of (AAAG)\textsubscript{12} constructs in cells had tetranucleotide repeat insertions (Table 4). With (AAAG)\textsubscript{18}, some insertions consisted of more than 2 tetranucleotide repeat units (Table 4). Among the (AAAG)\textsubscript{16} and (AAAG)\textsubscript{15} constructs, deletions made up the majority of clones observed (outside of heteroduplexed DNA), which suggests continual deletion frameshift mutation of insertions (and formation of heteroduplexes) (Table 4). Deletions consisted of up to 2 repeat units for the (AAAG)\textsubscript{12} and (AAAG)\textsubscript{15} microsatellites, and as many as 8 repeat units for the (AAAG)\textsubscript{18} microsatellite, indicating that longer tetranucleotide sequences have more opportunity for larger deletion slippage. In combination between the +1 bp and -1 bp systems, there were 77.8% deletions vs 22.2% insertions, with 78.9% of all observed frameshifts being 1 AAAG repeat unit, 16.1% being 2 repeat units, and 5.1% being 3 or more repeat units. Overall, our results suggest that tetranucleotide repeat expansions (insertions) are not favored and are more prone to be converted to shorter lengths in the absence of MMR activity.

**Discussion**

In this report, we created cell reporting systems containing tetranucleotide microsatellite sequences derived from the human locus D9S242 to measure both deletion and insertion frameshift mutation behavior in real time and accumulation over time in the absence of competent DNA MMR. We observed the following using these created MMR-deficient cell reporting systems: (1) tetranucleotide microsatellite mutation frequency increases with the length of the microsatellite, (2) tetranucleotide microsatellite mutation rates are linear as a function of microsatellite length, (3) deletion mutation rates for longer tetranucleotide microsatellites are higher than insertion rates but equivalent between deletions and insertions for shorter tetranucleotide microsatellites, and (4) there is accumulation of excessive deletion mutation events compared with insertion events by more than a 3-to-1 ratio over time for tetranucleotide microsatellites. We show that tetranucleotide frameshift mutation has a distinct deletional mutational bias over time. This study measured and compared insertion and deletion

**Table 2.** Mutation Rates of Microsatellites in MMR-Deficient Cells

| Microsatellites | -1 bp OF system, means ± SEM | +1 bp OF system, means ± SEM | P value\textsuperscript{a} |
|-----------------|------------------------------|------------------------------|------------------|
| D9S242 (AAAG)\textsubscript{18} | 34.5 x 10\textsuperscript{-4} ± 0.06 x 10\textsuperscript{-4} | 36.6 x 10\textsuperscript{-4} ± 0.04 x 10\textsuperscript{-4} | .0012 |
| D9S242 (AAAG)\textsubscript{15} | 23.4 x 10\textsuperscript{-4} ± 0.83 x 10\textsuperscript{-4} | 32.0 x 10\textsuperscript{-4} ± 0.90 x 10\textsuperscript{-4} | .0198 |
| D9S242 (AAAG)\textsubscript{12} | 16.4 x 10\textsuperscript{-4} ± 1.51 x 10\textsuperscript{-4} | 15.2 x 10\textsuperscript{-4} ± 1.08 x 10\textsuperscript{-4} | .5983 |

NOTE. Mutation rates were calculated by the Ma-Sandri-Sarkar Maximum Likelihood Estimator method and corrected for their negative controls. Rates are expressed as mutations at microsatellite sequence per cell per generation. OF, out-of-frame.

\textsuperscript{a}Between mutation rates of -1 bp OF and +1 bp OF systems.

**Table 3.** Statistical Significance of Mutation Rate Increase

| Microsatellite lengths | P value\textsuperscript{a} | Microsatellite lengths | P value\textsuperscript{a} |
|------------------------|-----------------|------------------------|-----------------|
| -1 bp OF system        |                 | +1 bp OF system        |                 |
| 18 vs 15               | .0232           | 18 vs 15               | .0600           |
| 18 vs 12               | .0263           | 18 vs 12               | .0159           |
| 15 vs 12               | .0413           | 15 vs 12               | .0039           |

OF, out-of-frame. \textsuperscript{a}The t test with 1-tailed distribution and unequal variance was used.
mutation rates and the resultant mutational spectra for a human tetranucleotide microsatellite sequence.

In comparing tetranucleotide microsatellite frameshifts with mononucleotide microsatellite frameshifts, it is not surprising that longer repeat lengths showed higher mutation rates, and our findings for tetranucleotide microsatellite frameshifts mirror those observed for mononucleotide microsatellite frameshifts.38 Because sequences flanking the microsatellite can modify the mutation rate of a microsatellite sequence,39,40 we used the same flanking sequence in all of our experiments to remove this potential variable from mutation rate measurement. Heteroduplex DNA readily was observed in our constructs for both the -1 bp and +1 bp systems and also have been reported previously.35–38 Heteroduplexes carry unequal lengths of the microsatellite on each DNA strand and indicate transition via intermediate alleles. Each heteroduplex may undergo more than 1 mutation transition event (eg, insertion followed by deletion back to the native original number of repeats). Our observation of native original tetranucleotide microsatellite lengths, particularly in the -1 bp system, indicate reversion of insertion to deletion to return to the original number of repeat units.

Our observed deletional mutational bias for tetranucleotide microsatellite frameshifts indicates that there is a preference for competent MMR to correct unpaired loops on the template DNA strand over unpaired loops on the newly synthesized strand during DNA replication. Insertions do occur for tetranucleotide microsatellite slippage mistakes, but because of heteroduplex intermediates, insertions often are converted into deletions, contracting the number of the microsatellite repeat units. We note the mutational spectra of colonies that possessed 1 sole genotype (combining both +1 bp and -1 bp systems) showed that (AAAG)12 had 62.3% deletions (86.4% had 1 repeat unit) and 37.7% insertions (all 1 repeat unit), (AAAG)15 had 97.1% deletions (75% had 1 repeat unit) and 2.9% insertions (all 1 repeat unit) and (AAAG)18 had 79.1% deletions (68.6% had 1 repeat unit) and 20.9% insertions (93.9% had 1 repeat unit and 6.1% had ≥4 repeat units). Thus, the shorter the tetranucleotide microsatellite, the more likely that a deletion is only 1 repeat unit, which should be taken into account on fragment analysis if used in MSI panels.

Tetranucleotide microsatellite sequences are not currently a part of the classic MSI panel to determine MSI-H or MSS in cancers because the panel solely contains mononucleotide and dinucleotide microsatellite markers.27 This panel misses determination of inflammation-associated microsatellite alterations or EMAST, which is the most common MMR defect in CRC as well as being observed in multiple other cancers.5,28 EMAST also is observed in non-neoplastic tissues such as ulcerative colitis and hamartomatous polyps.43,44 Because EMAST prognosticates patients with several cancers, including those with CRC, determining its presence has value.5–7 EMAST also may have value in determining the risk of neoplasia from ulcerative colitis.43 EMAST is caused by nuclear-to-cytosol displacement of MSH3 as a result of inflammation and/or oxidative stress, and can be determined only if the MSI panel has mononucleotide, dinucleotide, and tetranucleotide microsatellite markers present to examine the biochemical signature of isolated MSH3 deficiency and exclude MLH1, PMS2, and MSH2 function loss.5,7,15,31,43,44 Thus, tetranucleotide microsatellite markers may be added to traditional MSI panels as the importance of EMAST as a prognostic marker becomes more realized.

Here, we examined the dynamics of tetranucleotide frameshift mutation in the absence of DNA MMR because of an initial observation that appeared to be different than what we observed with mononucleotide frameshift mutation.32 There were some limitations to our study. Because some cells contained more than 1 copy of some constructs, there were more targets in those cells for
frameshifts to potentially occur and theoretically could modify the calculation of the mutation rate slightly. This is despite that restoration of the EGFP reading frame in just 1 single construct would be enough to detect the cell as EGFP positive by flow cytometry. On the other hand, with some cells having more than 1 copy of a construct, the detection of some native-length tetranucleotide sequences could mean that there was EGFP positivity with some constructs without frameshifts. Thus, our detection of native-length tetranucleotide sequences could mean there was a construct that did not frameshift mutate in a cell, or that heteroduplexes that expanded subsequently contracted down to the native length. Based on cells in which there was clearly only 1 copy of the construct, heteroduplexes with continual deletion mutation after insertion frameshifts occur, with reversion back to the microsatellite’s native length.

In summary, tetranucleotide microsatellites that undergo frameshift mutation when MMR is defective can expand and contract in length owing to insertion and deletion of repeat units but shows a distinct deletional mutation bias suggesting that MMR preferentially repairs insertion/deletion loops on the template DNA strand. Longer tetranucleotide microsatellites have an increased propensity for larger DNA slippage events in the absence of MMR.

Materials and Methods

Generation of EGFP Frameshift Reporting Plasmids

We developed microsatellite frameshift detection systems for the D9S242 human tetranucleotide locus using the previously established pIREShyg2-EGFP vector.\textsuperscript{15–40} The
pIREShyg2-EGFP plasmid vector contains pmel and AscI sites just after the start codon of EGFP and was linearized by using these restriction enzymes (New England Biolabs, Ipswitch, MA) to generate a 3’ blunt end (pmel site) and a 5’-GGCG overhang (AscI site) for directional cloning of microsatellite inserts. Compatible forward and reverse AAAG repeat oligos at different lengths ([AAAG]_{18}, [AAAG]_{15}, and [AAAG]_{12}) surrounded by the D9S242 locus flanking nucleotide sequences (14 bp from both 5’ and 3’ directions) with a 5’-CGCG overhang and 3’ blunt end were synthesized (Integrated DNA Technologies, Inc, San Diego, CA), annealed and then directionally cloned within the EGFP gene between the pmel and Ascl recognition site right after the ATG start codon to generate repeat plasmid-containing constructs (pIREShyg2-EGFP[AAAG]_{18}, pIREShyg2-EGFP[AAAG]_{15}, and pIREShyg2-EGFP[AAAG]_{12}). Thereafter, the protein translation reading frame of downstream EGFP within each construct was shifted by +1 and -1 bp using the Quick-change II site-directed mutagenesis kit (Stratagene, La Jolla, CA) so that plasmids do not express EGFP protein. Deletion and insertion of 1 or more AAAG repeat units resulting from slippage mutations restores the reading frame of EGFP, and EGFP-expressing cells were quantified by flow cytometry to determine mutation rates and spectra (Figure 1A). Negative control plasmids were constructed by changing 2 nucleotides in every third AAAG repeat for each length to prevent frameshifts and placed out-of-frame with EGFP (Figure 1B). A mutation-resistant, microsatellite repeat–interrupted, positive control plasmid in frame with the EGFP sequence that expressed EGFP protein was prepared to use for calibration of the flow cytometry (Figure 1C).

Plasmids were transformed into Stbl2 competent bacteria (Life Technologies, Inc, Rockville, MD); ampicillin-resistant single colonies were amplified in overnight cultures
and DNA was isolated from bacterial cells using the mini prep kit (Qiagen, Valencia, CA) as per the manufacturer’s instructions. Isolated DNAs were subjected to PCR amplifications using primers flanking the inserted microsatellite sequences (forward: 5’-ATTGACGCAAATGGGCGGTA-3'; reverse: 5’-CTTCATGTGGTCGGGGTAGC-3') followed by DNA sequencing using sequencing primers (forward: 5’-GAACCCACTGCTTACTGGCT-3'; reverse: 5’-GCTGAACCTTGGCGGCTTTA-3') to confirm the accuracy of the inserted microsatellites.

### Stable Transfection and Generation of Monoclonal Cell Lines

The human colon adenocarcinoma cancer cell line HCT116 (MLH1\(^{-}\) and MSH3\(^{-}\)) was obtained from American Type Culture Collection (Manassas, VA) and cultured in Iscove’s Modified Dulbecco’s Medium + GlutaMax (cat. 31980030; Life Technologies, Inc) supplemented with 10% fetal bovine serum (cat. 16000-044; Life Technologies, Inc), 1% penicillin-streptomycin (50 mg/mL solution, cat. 15140122; Life Technologies, Inc), and 0.2% Normocin (50 mg/mL solution, cat ant-nr-2; InvivoGen, San Diego, CA). Cells were seeded in 6-well plates 2 days before transfection and transfected with 2 μg of one of the prepared plasmid constructs (Table 1) using the Amaxa Cell Line Nucleofector Kit V (cat. VCA-1003) (optimized for the HCT116 cell line) and the Amaxa Nucleofector II device by following the manufacturer’s instructions on nucleofection of adherent cell lines (Lonza, Cologne, Germany).

Selection by hygromycin B (50 mg/mL solution, cat. 10687-010; Life Technologies) started 1 day after transfection at the predetermined required dose for HCT116 cells (400 μg/mL) for 14 days. After 2 weeks of selection, polyclonal cells were pooled and cultured in media containing 200 μg/mL hygromycin B and subsequently subjected to single-cell sorting using MoFlo Astrios (Beckman Coulter, Inc, Brea, CA) to establish monoclonal cell lines. Genomic DNA was isolated from monoclonal cell lines using the QIAamp DNA mini kit (cat. 51304; Qiagen) and analyzed by PCR and DNA sequencing for microsatellite inserts and their accuracy as described earlier.

### Sorting and Analysis of Cells by Flow Cytometry

MoFlo Astrios with Summit V6.2.3.15613 software (Beckman Coulter, Brea, CA) was used for cell sorting and analysis of mutant cells. Approximately 10,000 nonfluorescent cells from each monoclonal line were sorted onto 12-well plates with media containing 200 μg/mL hygromycin B. Sorted cells were expanded to keep them in exponential growth in 100-mm plates for 5 weeks and duplicate cultures were analyzed each week to quantify the EGFP-expressing mutant population. At approximately 80% confluency, 100-mm cell plates were trypsinized (0.05% trypsin–EDTA, cat. 25300-054; Life Technologies), pelleted, and resuspended in 2 mL Dulbecco’s Phosphate-Buffered Saline containing 1% bovine serum albumin with the addition of 2 drops of ReadiDrop Propidium Iodide (cat. 135-1101; BioRad, Hercules, CA) and then 200,000 alive, single cells were analyzed to determine EGFP-expressing frameshift mutant population and mutation rates. Propidium iodide–stained and unstained HCT116 cells and HCT116 cells transfected with positive control plasmid expressing EGFP were used as controls to set up gates in a flow cytometer.

### Estimation of Mutation Rates, Frequencies, and Spectra

#### Mutation rates

We used the program FALCOR, a web tool designed for use with Luria-Delbruck fluctuation analysis to calculate the mutation rate and frequency.\(^{42}\) Mutation rates were calculated by the Ma-Sandri-Sarkar Maximum Likelihood Estimator method and expressed as mutations at microsatellites per cell per generation.

#### Mutation spectra

At the end of the sixth week, cells from EGFP-negative (nonmutated) and EGFP-positive (mutated) populations were sorted on MoFlo Astrios for genomic DNA isolation. Forward (5’-ATTGACGCAAATGGGCGGTA-3’) and reverse (5’-CTTCATGTGGCGGCTTTA-3’) primers were used for PCR amplification of a 550-bp DNA fragment.

### Table 4. Mutational Spectra of +1 bp and -1 bp Systems by DNA Sequencing

| Microsatellites | -1 bp OF system | +1 bp OF system |
|-----------------|-----------------|-----------------|
|                 | Deletions       | Insertions      |
| D9S242\(\text{AAAG}\)\(_{12}\) | 4.1% deletions | 76.7% insertions |
|                 | 40.0% one-repeat | 94.6% one-repeat |
|                 | 60.0% two-repeat | 5.4% two-repeat  |
|                 | 38.8% heteroduplexes | 4.1% heteroduplexes |
|                 | 40.8% wild type | 19.2% wild type  |
|                 | 16.3% one-repeat insertions | |
| D9S242\(\text{AAAG}\)\(_{15}\) | 37.5% deletions | 70.2% deletions  |
|                 | 28.6% one-repeat | 95.7% one-repeat |
|                 | 71.4% two-repeat | 4.3% two-repeat  |
|                 | 26.8% heteroduplexes | 11.9% heteroduplexes |
|                 | 32.1% wild type | 17.9% wild type  |
|                 | 3.6% one-repeat insertions | |
| D9S242\(\text{AAAG}\)\(_{18}\) | 36.0% deletions | 71.7% deletions  |
|                 | 24.2% one-repeat | 93.2% one-repeat |
|                 | 48.5% two-repeat | 6.7% two-repeat  |
|                 | 13.6% three-repeat | 9.0% heteroduplexes |
|                 | 13.6% five or more repeat | 15.7% wild type  |
|                 | 15.3% heteroduplexes | 3.6% one-repeat insertions |

NOTE. Wild-type refers to the native sequence in the construct. Each repeat unit is 1 AAAG. OF, out-of-frame.
containing the inserted microsatellite region from genomic DNAs using Q5 Hot Start High-Fidelity 2× Master Mix (cat. M0494S; NEB, Ipswich, MA). PCR amplicons were run on 1.7% agarose gel and approximately 550-bp DNA fragments were cut out, DNAs were isolated, and then cloned into pMiniT vector using the NEB PCR cloning kit (cat. E1202). Individual colonies from transformation plates were screened for their microsatellite inserts by colony PCR. Q5 Hot Start High-Fidelity 2× Master Mix and cloning analysis forward (5'-ACCTGCCAACCAAAGCGAGAAC-3') and reverse (5'-TCAGGGTTATTGTCTCATGAGCG-3') primers were used for colony PCR. Q5 Hot Start High-Fidelity 2× Master Mix and cloning analysis forward (5'-ACCTGCCAACCAAAGCGAGAAC-3') and reverse (5'-TCAGGGTTATTGTCTCATGAGCG-3') primers were used for colony PCR. Q5 Hot Start High-Fidelity 2× Master Mix and cloning analysis forward (5'-ACCTGCCAACCAAAGCGAGAAC-3') and reverse (5'-TCAGGGTTATTGTCTCATGAGCG-3') primers were used for colony PCR. After the PCR, amplicons were cleaned up with ExoSAP-IT (Affymetrix, Santa Clara, CA) and then analyzed for repeat insertions and deletions by DNA sequencing to determine mutational spectra.

Protein Isolation and Western Blot Analysis

**Protein isolation.** Cells were cultured in 100-mm cell plates, at the 70%–90% confluence levels, washed with ice-cold Dulbecco’s Phosphate-Buffered Saline twice, and then lysed in the plate with 300 μL radioimmunoprecipitation assay buffer (cat. 89900; Thermo Scientific–Pierce), supplemented with 2 mmol/L phenylmethylsulfonyl fluoride (cat. P7626; Sigma) and protease inhibitor cocktail (cat. P-2714; Sigma). Cell lysates were transferred into prechilled 1.5-mL microcentrifuge tubes, vortexed, and then incubated on ice for 15–30 minutes, with vortexing at 10-minute intervals. Lysates were clarified by centrifugation at 14,000 rpm (13,000 × g) at 4°C for 15 minutes and stored at −80°C until ready for analysis. Protein concentrations were determined using the Thermo Scientific Pierce BCA Protein Assay Kit (product 23225) according to the manufacturer’s instructions.

**Protein gel electrophoresis, Western blot, and antibody labeling.** Proteins were prepared and separated on 4%–12% NuPAGE Bis-Tris precast mini gel using the XCell SureLock Mini-Cell electrophoresis unit (Invitrogen) according to the manufacturer’s instructions. Approximately 35 μg protein was loaded into each well and run along with Novex Sharp prestained protein standards from Invitrogen (cat. LC5800) and all blue precision plus protein standards.
from Bio-Rad (cat. 161-0393). Protein samples were transferred onto a polyvinylidene difluoride membrane by Western blot using the XCell II Blot Module (Invitrogen) according to the manufacturer’s instructions. Membranes were blocked with SuperBlock T20 (tris-buffered saline) blocking buffer (cat. 37536; Thermo Scientific) and then incubated with the following primary antibodies: hMLH1 1:1000 (cat. 554073; BD Biosciences, San Jose, CA), hMSH2 1:1000 (cat. 556349; BD Biosciences), hMSH6 1:500 (cat. 610919 BD Biosciences), hMSH3 1:500 (cat. 611390; BD Biosciences), and α-tubulin 1:7500 (cat. 62204; Thermo Scientific). Primary antibodies were detected using the Pierce goat anti-mouse IgG horseradish peroxidase secondary antibody (Thermo Scientific) 1:5000. Restore Western Blot Stripping Buffer (cat. 21059; Thermo Scientific) was used to remove the antibody from the blot for reprobing with a different antibody. The blots were developed using Immobilon Western Chemiluminescent horseradish-peroxidase substrate (Millipore, Burlington, MA) and visualized by ImageQuant LAS4000 (GE Healthcare Bio-Sciences Corp, Piscataway, NJ).

Southern Blot Analysis

DNA probe. A Digoxigenin-labeled DNA probe was synthesized using the DIG High Prime DNA Labeling and Detection Starter Kit II according to the manufacturer’s instructions (cat. 11585614910; Roche, Basel, Switzerland). The pIREShygB-EGFP plasmid was digested with kpnI and run on 1% agarose gel and the 4.0-kb band DNA fragment was cut out and purified to use as the DNA template (Figure 10).

Southern blot. Genomic DNAs from transfected and nontransfected HCT116 cells were digested with BamHI, which cuts plasmid DNA once outside of the probe region, and purified with phenol-chloroform extraction. The DNA fragments were separated by electrophoresis on 0.8% agarose gel overnight at 20 volts. DNA gels were depurinated in 250 mmol/L HCl for 10 minutes, denatured in 0.5 mol/L NaOH, 1.5 mol/L NaCl for 30 minutes, neutralized in 0.5 mol/L Tris-HCl (pH 7.5), 1.5 mol/L NaCl for 30 minutes, equilibrated in 10× standard saline citrate and Southern blotted onto positively charged nylon membranes (11209299001; Roche). The blots were UV cross-linked, hybridized with the DIG-labeled 4.0-kb DNA probe at 50°C, and detected according to the manufacturer’s instructions. The blots were visualized by ImageQuant LAS4000 (GE Healthcare Bio-Sciences Corp). Copy number standards were prepared by spiking BamHI digests of nontransfected HCT116 genomic DNA with 0, 1, 3, 5, 10, 25, and 100 copies of the pIREShygB-EGFP plasmid, electrophoresing on 0.8% gel, blotting onto membrane, and hybridizing with the DIG-labeled 4.0-kb DNA probe at 50°C as described earlier.

References

1. Carethers JM, Stoffel EM. Lynch syndrome and Lynch syndrome mimics: the growing complex landscape of hereditary colon cancer. World J Gastroenterol 2015; 21:9253–9261.

2. Carethers JM, Jung BH. Genetics and genetic biomarkers in sporadic colorectal cancer. Gastroenterology 2015;149:1177–1190.

3. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135:1079–1099.

4. Goel A, Boland CR. Microsatellite instability in colorectal cancer. Gastroenterology 2010;138:2073–2087.

5. Carethers JM, Koi M, Tseng-Rogenski S. EMAST is a form of microsatellite instability that is initiated by inflammation and modulates colorectal cancer progression. Genes 2015;6:185–205.

6. Deveraj B, Lee A, Cabrera BL, Miyai K, Luo L, Ramamoorthy S, Keku T, Sandler RS, McGuire KL, Carethers JM. Relationship of EMAST and microsatellite instability among patients with rectal cancer. J Gastrointest Surg 2010;14:1521–1528.

7. Garcia M, Choi C, Kim HR, Daoud Y, Toyiama Y, Takahashi M, Goel A, Boland CR, Koi M. Association between recurrent metastasis from stage II and III primary colorectal and moderate microsatellite instability. Gastroenterology 2012;143:48–50.

8. Koi M, Garcia M, Choi C, Kim H-R, Koike J, Hemmi H, Nagasaka T, Okugawa Y, Toyiama Y, Kitajima T, Imaoka H, Kusunoki M, Chen YH, Mukherjee B, Boland CR, Carethers JM. Microsatellite alterations with allelic loss on 9p24.2 signify less aggressive colorectal cancer metastasis. Gastroenterology 2016;150:944–955.

9. Koi M, Tseng-Rogenski SS, Carethers JM. Inflammation-associated microsatellite alterations: mechanisms and significance in the prognosis of patients with colorectal cancer. World J Gastrointest Oncol 2018;10:1–14.

10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Graftham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Showkineen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Steidley K, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W.
Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platerz M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor ML, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Homischer K, Nordsieck G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLyasgaht A, Miklukos M, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stapula E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Geyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patroinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Wetterstrand KA, Patroinos A, Shizuya H, Choi S, Chen YJ, Szustakowski J. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.

11. Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Gen Biol 2003;4:F13.

12. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet 2004;5:435–445.

13. Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. Frameshift mutations and the genetic code. Cold Spring Harbor Symp Quant Biol 1967;31:77–84.

14. Carethers JM, Hawn MT, Chauhan DP, Luce MC, Marra G, Koli M, Boland CR. Competency in mismatch repair prohibits clonal expansion of cancer cells treated with N-methyl-N′-nitro-N-nitrosoguanidine. J Clin Invest 1996;98:199–206.

15. Carethers JM. Hereditary, sporadic and metastatic colorectal cancers are commonly driven by specific spectrums of defective DNA mismatch repair components. Trans Am Clin Climatol Assoc 2016;127:81–97.

16. Modrich P, Lahue R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 1996;65:101–133.

17. Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. Chem Rev 2006;106:302–323.

18. Carethers JM. Microsatellite instability pathway and EMAST in colorectal cancer. Curr Colorectal Cancer Rep 2017;13:73–80.

19. Carethers JM, Chauhan DP, Fink D, Nebel S, Bresalier RS, Howell SB, Boland CR. Mismatch repair proficiency and in vitro response to 5-fluorouracil. Gastroenterology 1999;117:123–131.

20. Carethers JM, Smith EJ, Behling CA, Nguyen L, Tajima A, Doctolero RT, Cabrera BL, Goel A, Arnold CA, Miyai K, Boland CR. Use of 5-fluorouracil and survival in patients with microsatellite unstable colorectal cancer. Gastroenterology 2004;126:394–401.

21. Tajima A, Hess MT, Cabrera BL, Kolodner RD, Carethers JM. The mismatch repair complex hMutSα recognizes 5-fluorouracil-modified DNA: implications for chemosensitivity and resistance. Gastroenterology 2004;127:1678–1684.

22. Iwaizumi M, Tseng-Rogenski S, Carethers JM. DNA mismatch repair proficiency executing 5-fluorouracil cytotoxicity in colorectal cancer cells. Cancer Biol Ther 2011;12:756–764.

23. Tajima A, Iwaizumi M, Tseng-Rogenski S, Cabrera BL, Carethers JM. Both hMutSα and hMutSβ complexes participate in 5-fluorouracil cytotoxicity. PLoS One 2011;6:e28117.

24. Hamaya Y, Guarinos C, Tseng-Rogenski SS, Iwaizumi M, Das R, Joer R, Castells A, Lior X, Andreu M, Carethers JM. Efficacy of 5-fluorouracil adjuvant therapy for patients with EMAST-positive stage II/III colorectal cancers. PLoS One 2015;10:e0127591.

25. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Lubes BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzti TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhaijee F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardoll DM, Papadopoulos N, Kinzler KW, Zhou S, Comish TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B, Diaz LA Jr. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372:2509–2520.

26. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Lubes BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzti TS, Lee JJ, Greten TF, Duffy AG, Giombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bekar KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA, Diaz LA Jr. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357:409–413.

27. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248–5257.

28. Watson MM, Berg M, Soreide K. Prevalence and implications of elevated microsatellite alterations at selected tetranucleotides in cancer. Br J Cancer 2014;111:823–827.
29. Haugen AC, Goel A, Yamada K, Marra G, Nguyen T-P, Nagasaka T, Kanazawa S, Koike J, Kikuchi Y, Zhong X, Arita M, Shibuya K, Oshimura M, Hemmi H, Boland CR, Koi M. Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. Cancer Res 2008;68:8465–8472.

30. Lee S-Y, Chung H, Deveraj B, Iwaizumi M, Han HS, Hwang DY, Seong MK, Jung BH, Carethers JM. Microsatellite alterations at selected tetranucleotide repeats are associated with morphologies of colorectal neoplasias. Gastroenterology 2010;139:1519–1525.

31. Venderbosch S, van Lent-van Vliet S, deHaan AF, Nagtegaal ID. EMAST is associated with a poor prognosis in microsatellite instable metastatic colorectal cancer. PLoS One 2015;10:e0124538.

32. Tseng-Rogenski SS, Chung H, Wilk MB, Zhang S, Iwaizumi M, Carethers JM. Oxidative stress induces nuclear-to-cytosol shift of hMSH3, a potential mechanism for EMAST in colorectal cancer cells. PLoS One 2012;7:e50541.

33. Campregher C, Schmid G, Ferk F, Knasmuller S, Khare V, Kortum B, Dammann K, Lang M, Scharl T, Spittler A, Roig AI, Shay JW, Genzer C, Gasche C. MSH3-deficiency initiates EMAST without oncogenic transformation of human colon epithelial cells. PLoS One 2012;7:e50541.

34. Tseng-Rogenski S, Hamaya Y, Choi D, Carethers JM. Interleukin 6 alters localization of hMSH3, leading to DNA mismatch repair defects in colorectal cancer cells. Gastroenterology 2015;148:579–589.

35. Gasche C, Chang CL, Natarajan L, Goel A, Rhee S, Young DJ, Arnold CN, Boland CR. Identification of frame-shift intermediate mutant cells. Proc Natl Acad Sci U S A 2003;100:1914–1919.

36. Campregher C, Scharl T, Nemeth M, Honeder C, Jascur T, Boland CR, Gasche C. The nucleotide composition of microsatellites impacts both replication fidelity and mismatch repair in human colorectal cells. Hum Mol Genet 2010;19:2648–2657.

37. Chung H, Young DJ, Lopez CG, Le T-A, Lee JK, Ream-Robinson D, Huang SC, Carethers JM. Mutation rates of TGFBR2 and ACVR2 coding microsatellites in human cells with defective DNA mismatch repair. PLoS One 2008;3:e3463.

38. Chung H, Lopez CG, Holmstrom J, Young DJ, Lai JF, Ream-Robinson D, Carethers JM. Both microsatellite length and sequence context determine frameshift mutation rates in defective DNA mismatch repair. Hum Mol Genet 2010;19:2638–2647.

39. Chung H, Lopez CG, Young DJ, Lai JF, Holmstrom J, Ream-Robinson D, Cabrera BL, Carethers JM. Flanking sequence specificity determines coding microsatellite heteroduplex and mutation rates with defective DNA mismatch repair (MMR). Oncogene 2010;29:2172–2180.

40. Chung H, Chaudhry J, Lai JF, Young DJ, Carethers JM. Flanking nucleotide specificity for DNA mismatch repair deficient frameshifts within activin receptor 2 (ACVR2). Mutat Res 2012;729:73–80.

41. Koi M, Umar A, Chauhan DP, Cherian SP, Carethers JM, Kunkel TA, Boland CR. Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N’-nitro-N-nitrosoguanidine-tolerance in colon tumor cells with homozygous MSH1 mutation. Cancer Res 1994;54:4308–4312.

42. Hall BM, Ma C-X, Liang P, Singh KV. Fluxation Analysis CalculOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics 2009;25:1564–1565.

43. Munakata K, Koi M, Kitajima T, Tseng-Rogenski SS, Uemura M, Matsuno H, Kawai K, Sekido Y, Mizushima T, Toiyama Y, Yamada T, Mano M, Mita E, Kusunoki M, Mori M, Carethers JM. Inflammation-associated microsatellite alterations caused by MSH3 dysfunction are prevalent in ulcerative colitis and increase with neoplastic advancement. Clin Transl Gastroenterol 2019;10:e00105.

44. Huang SC, Lee JK, Smith EJ, Doctolero R, Tajima A, Beck SE, Weidner N, Carethers JM. Evidence for an hMSH3 defect in familial hamartomatous polyps. Cancer 2011;117:492–500.