

# Defective Mismatch Repair, Microsatellite Mutation Bias, and Variability in Clinical Cancer Phenotypes

Sandeep N. Shah<sup>1,2</sup>, Suzanne E. Hile<sup>1</sup>, and Kristin A. Eckert<sup>1,2</sup>

## Abstract

Microsatellite instability is associated with 10% to 15% of colorectal, endometrial, ovarian, and gastric cancers, and has long been used as a diagnostic tool for hereditary nonpolyposis colorectal carcinoma-related cancers. Tumor-specific length alterations within microsatellites are generally accepted to be a consequence of strand slippage events during DNA replication, which are uncorrected due to a defective postreplication mismatch repair (MMR) system. Mutations arising within microsatellites associated with critical target genes are believed to play a causative role in the evolution of MMR-defective tumors. In this review, we summarize current evidence of mutational biases within microsatellites arising as a consequence of intrinsic DNA sequence effects as well as variation in MMR efficiency. Microsatellite mutational biases are generally not considered during clinical testing; however, we suggest that such biases may be clinically significant as a factor contributing to phenotypic variation among microsatellite instability-positive tumors. *Cancer Res*; 70(2); 431–5. ©2010 AACR.

## Introduction

Microsatellites are short, repetitive elements in genomic DNA consisting of one to six bases per repeat unit, and are nonrandomly distributed throughout the human genome (for reviews, see refs. 1, 2). Microsatellites located in intragenic regions (promoters, 3'-untranslated regions, and introns) can be important regulators of gene expression by influencing transcription rate, RNA stability, splicing efficiency, and RNA-protein interactions. Because microsatellite sequences are highly polymorphic in human populations, genetic changes in some microsatellites can be important evolutionarily by affecting phenotype. Microsatellites located within intergenic regions might also have functional roles in chromatin organization and recombination.

Postreplication mismatch repair (MMR) is a critical mechanism for maintaining microsatellite stability through the correction of base substitution mismatches and insertion/deletion events (for review, see ref. 3). Premutational intermediates are detected and processed by heterodimers of the MutS and MutL family of proteins. The MutS $\alpha$  complex (MSH2 and MSH6) can recognize base mispairs as well as single-base insertion/deletion loops (IDL) in either the parental or nascent strand. MutS $\beta$  (MSH2 and MSH3) can recognize IDLs of two to four bases in addition to single-base

IDLs. MutL heterodimers are recruited by MutS complexes to the mismatch or IDL, and consist of MLH1 dimerized with either PMS2 (MutL $\alpha$ ), PMS1 (MutL $\beta$ ), or MLH3 (MutL $\gamma$ ). MutL $\alpha$  is regarded as the major MutL complex for the repair of both large and small IDLs, whereas MutL $\gamma$  and MutL $\beta$  have little or no known role in MMR, respectively.

## Clinical Phenotype of Microsatellite Mutation

Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), is characterized by early onset tumors in the proximal colon as well as extracolonic regions (for review, see ref. 4). Lynch syndrome results from the germ line inheritance of mutations in MMR genes and subsequent MMR deficiency arising during neoplastic cell evolution. Such MMR deficiency can be readily diagnosed as tumor-specific microsatellite instability (MSI), using a panel of defined microsatellite markers. Patients that are homozygotes or compound heterozygotes for mutations in *MSH2*, *MSH6*, *MLH1*, or *PMS2* genes could present with childhood hematologic and brain malignancies as well as features of neurofibromatosis (e.g., café au lait spots), in addition to early onset gastrointestinal neoplasms.

Germ line mutations in patients with a history of HNPCC are most frequently seen in the *MSH2* or *MLH1* genes. Due to partial redundancy and differences in substrate specificity, *MSH6* (MutS $\alpha$  specific) alterations are less common in HNPCC, whereas genetic predisposition to HNPCC has not been shown for *MSH3* (MutS $\beta$  specific) mutations (5). Similarly, *PMS2* (MutL $\alpha$  specific) mutations are rarely observed in patients with HNPCC, as compared with *MLH1* mutations, even though *PMS2* is the primary binding partner of *MLH1*. This discrepancy in the frequency of *MLH1* versus *PMS2* mutations might be due, in part, to the redundancy between *PMS2* and *MLH3* activities because *MLH3* has similar biochemical properties

**Authors' Affiliations:** <sup>1</sup>Department of Pathology, Gittlen Cancer Research Foundation and <sup>2</sup>The Intercollege Graduate Degree Program in Genetics, Pennsylvania State University College of Medicine, Hershey, Pennsylvania

**Corresponding Author:** Kristin A. Eckert, Department of Pathology, The Pennsylvania State College of Medicine, 500 University Drive, Hershey, PA 17033. Phone: 717-531-4065; Fax: 717-531-5634; E-mail: kae4@psu.edu.

doi: 10.1158/0008-5472.CAN-09-3049

©2010 American Association for Cancer Research.

to PMS2 (6). However, investigations of MLH3 (MutL $\gamma$  specific) and PMS1 (MutL $\beta$  specific) have shown only a minor role in microsatellite stability (3), and it is unclear whether mutations in these genes are involved in HNPCC-like phenotypes (5).

Although MMR proteins act in a single biochemical pathway, loss of individual components results in both similar and distinct clinical phenotypes. Patients carrying heterozygous mutations in either the *MLH1* or *MSH2* genes display classic HNPCC, whereas patients carrying heterozygous *MSH6* and *PMS2* mutations usually display a decreased penetrance and later onset of HNPCC (5). Genotypic-phenotypic correlations for HNPCC are complicated, as patients carrying different variant alleles within the same gene could have distinct clinical phenotypes. For example, a recent study described a patient with an *MLH1-G67E* mutation and family history of HNPCC that displayed both classic and atypical tumors, distinct from families with *MLH1-G67R* mutations that only show classic tumors (7). Clinical phenotypes could also vary widely among patients having the same allelic mutation. For example, many carriers with an *MSH2-A636P* allele displayed only colorectal carcinomas, whereas other patients also developed endometrial or ovarian carcinomas (8). Variable presentations of tumor types among patients with HNPCC might reflect the stochastic nature of genetic alterations (mutation timing, frequency, and specificity) that result from the mutator phenotype initiated by defective MMR.

### Microsatellite Mutational Biases Due To Intrinsic DNA Features

Mutation rates of common microsatellites found in the human genome are extremely variable. The combined mutation rate estimates from analyses of parent-child allele transmissions and experimental determinations in somatic, nontumorigenic human cells range from a low of  $\sim 10^{-6}$  to a high of  $\sim 10^{-2}$  mutations per locus per generation (for reviews, see refs. 1, 9). Multiple extrinsic factors influence microsatellite mutation rates, including flanking sequence, recombination rate, sex, and age. Undoubtedly, however, the vast majority of the observed variation in mutagenesis is due to features intrinsic to the repeated DNA sequence itself, including motif unit size (mononucleotide, dinucleotide, trinucleotide, etc.), length (number of units), and sequence composition (1, 9, 10). In general, motif size is inversely related to microsatellite allele mutation frequency, whereas motif length is directly related to microsatellite mutation frequency. Motif sequence can induce mutational variation in several ways. First, the precise sequence composition of an allele affects thermostability and DNA secondary structure potential, both of which may be factors contributing to the significant differences in microsatellite allele mutation rates observed in human cells (11). Second, the purity of a microsatellite directly affects mutability, as interrupted alleles have a significantly lower mutation rate. Third, the complexity of a microsatellite locus (simple, compound, or complex) has been implicated as a factor in mutability (1). In addition to differences in absolute

mutation frequency, the types of mutational events arising within microsatellite alleles are dependent on intrinsic DNA features. Mathematical modeling using human genome data has shown that expansions are most prevalent in short microsatellites, whereas long microsatellites tend to contract (1). Nontumorigenic human cells also display distinct mutational biases for contractions versus expansion mutations within dinucleotide and tetranucleotide microsatellites due to differences in motif sequence composition (9, 11).

Mechanistically, microsatellite alleles expand or contract through strand slippage events during DNA synthesis. One potential source of the above mutational biases may be differential stabilization of slipped strand intermediates through formation of stable secondary DNA structures within microsatellite alleles. In this mechanism, both the DNA sequence content and the DNA strand location (leading versus lagging strand) of a microsatellite contribute to mutational bias. DNA repair has also been shown to impart mutation biases.

### Microsatellite Mutational Biases Due To Defective MMR

The repair specificity of individual MMR proteins is an especially important consideration for disease phenotype. DNA repair by each MutS heterodimer is substrate-specific within mononucleotide, dinucleotide, or tetranucleotide repeats: MutS $\alpha$  is expected to repair premutational intermediates only within mononucleotide repeats, whereas MutS $\beta$  potentially repairs premutational intermediates within mononucleotide, dinucleotide, trinucleotide, and tetranucleotide repeats (12). In yeast and *Escherichia coli* models, the efficiency of MMR within microsatellites is dependent on motif size, length, and sequence composition of an allele (9). Mutation rate variations among GT/CA loci within the yeast genome have been attributed to differential MMR efficiencies (13). We have recently shown that differential MMR efficiency by PMS2 contributes to mutation rate biases at mononucleotide, dinucleotide, and tetranucleotide alleles in nontumorigenic human cells (14). Thus, available experimental data show that MMR contributes to mutation frequency biases within microsatellites.

We and others also have observed differences in the mutational specificity of microsatellites due to defects in specific MMR genes (Table 1). In mice, a deficiency of *MSH2* gives rise to 78% 1-bp deletions in a mononucleotide allele; in contrast, loss of *MSH3* or *MSH6* results in only 61% or 47% deletion mutations within mononucleotides, respectively (15). For dinucleotide GT/CA alleles, loss of yeast *MSH3* results in a bias favoring deletions, whereas loss of yeast *MSH6* results in the opposite bias, favoring insertions. In *MSH2*-deficient yeast, the majority of mutations at [AG/TC]<sub>5</sub> within the *APC* gene are insertions, whereas deletions are more frequent in an [A/T]<sub>6</sub> tract of the same gene (16). Human *APC* in MMR-defective colorectal tumors displays a reverse trend in which deletions are readily observed within the [AG/TC]<sub>5</sub> tract, whereas insertions are common at [A/T]<sub>6</sub> (16). The human *APC* data was obtained from a mutation database

**Table 1.** Microsatellite mutational biases due to loss of specific MMR gene functions

MMR gene deficiency	Species	STR	Del/Ins	Reference
<i>MSH2</i>	Mouse	(G/C) <sub>7-8</sub>	4:1	(15)
	Yeast	(GT/CA) <sub>16.5</sub>	1:1	(12)
<i>MSH6</i>	Mouse	(G/C) <sub>7-8</sub>	1:1	(15)
	Yeast	(G/C) <sub>18</sub>	3:1	(12)
		(GT/CA) <sub>16.5</sub>	1:3.5	(12)
<i>MSH3</i>	Mouse	(G/C) <sub>7-8</sub>	2:1	(15)
	Yeast	(GT/CA) <sub>16.5</sub>	4:1	(12)
<i>MLH1</i>	Mouse	(G/C) <sub>7-8</sub>	4:1	(15)
		(A/T) <sub>23-27</sub>	19:1	(17)
<i>PMS2</i>	Mouse	(AC/TG) <sub>22-33</sub>	2:1	(17)
		(G/C) <sub>7-8</sub>	2:1	(15)
		(A/T) <sub>23-27</sub>	6:1	(17)
	Human	(AC/TG) <sub>22-33</sub>	1:2	(17)
		(GT/CA) <sub>10</sub>	1:11	Unpublished observations
		(TTCC/AAGG) <sub>9</sub>	1:21	(14)
	(TTTC/AAAG) <sub>9</sub>	1:4	(14)	

representing patients of various MMR backgrounds, and as such, mutation spectra are a combination of various MMR defects. The discrepancy between yeast and human data could also indicate selection for the tumor phenotype.

Loss of specific members of the MutL family also produces distinct mutational biases. For mononucleotide [G/C]<sub>7-8</sub> alleles in mice, 80% of mutations in the absence of MLH1 were deletions, compared with only 64% with PMS2 loss (Table 1; ref. 15). The difference between MutL members (*MLH1* and *PMS2*) was even more dramatic for an AC/TG dinucleotide locus, in that MLH1-deficient mice displayed a bias in favor of deletions whereas PMS2-deficient mice displayed a bias in favor of expansions (17). In our recent publication, we showed that a bias in favor of expansions within microsatellites was also observed in PMS2-deficient human cells (14). As shown in Table 1, deficiency of PMS2 increases expansion mutations as compared with deletions within the [TTCC/AAGG]<sub>9</sub> and [TTTC/AAAG]<sub>9</sub> tetranucleotide repeats, as well as within a [GT/CA]<sub>10</sub> dinucleotide repeat. This bias may be significant *in vivo*, as loss of PMS2 allows nuclear localization of MLH3 (6), thereby potentially changing the mutation profile. We proposed that repair proteins such as MLH1-MLH3 might be able to compensate for PMS2 loss, as MMR components have very distinct effects during DNA repair. Consistent with this interpretation, knockout mice with PMS2/MLH3 deficiency display much greater MSI than PMS2-deficient mice (18). Taken together, these experimental data show that loss of PMS2 generates microsatellite mutational biases that are quite distinct from those generated by loss of MLH1.

### Target Genes of MSI

Three mechanisms have been described by which alterations in microsatellite allele length directly affect neoplastic

progression: (a) gene inactivation due to frameshift mutations within exonic microsatellites; (b) gene activation due to enhanced transcription factor binding of microsatellites within promoter regions; and (c) microsatellite allele length differences that affect gene expression levels.

Mutation of mononucleotide microsatellites within exonic regions has been extensively analyzed in MSI colorectal tumors (reviewed in ref. 19). Classes of genes known to be altered in MSI tumors include DNA repair, cell growth and signaling, apoptosis, and transcription factors. For example, the *TGFBR2* gene contains an (A)<sub>10</sub> repeat which is mutated in 86% of MSI-positive tumors, compared with just 0.6% of MSI-negative tumors (20). It is also interesting to note that several of the MMR genes, i.e., *MSH6*, *MSH3*, *PMS2*, and *MLH3*, also contain mononucleotide microsatellites. This property has been proposed to allow fine-tuning of mutation rates for adaptation during evolution (21) and may also play a role in the immune response (22). MMR is highly involved in class switch recombination and somatic hypermutation, and thus, control of immunoglobulin diversity. Switch regions contain highly repetitive G/C-rich sequences, and loss of individual MMR proteins has been shown to differentially regulate the frequency of recombination within these regions.

Intragenic microsatellites located within introns or 5'- and 3'-untranslated regions should also be considered as potential target genes for MSI. For example, experiments to elucidate the mechanism of gene regulation by the EWS/FLI oncogenic transcription factor found that a minimum of four consecutive GGAA sequences need to be present in the promoter of target genes for transcriptional activity. Importantly, transactivation increased with each additional unit of GGAA, and differences in microsatellite size and sequence have been suggested to be a factor determining susceptibility to Ewing's sarcoma (23). Microsatellite allele length polymorphisms are known to be correlated with differential gene

expression. A relevant example for neoplasia is the epidermal growth factor receptor gene, which contains a polymorphic (CA)<sub>14–21</sub> unit repeat within intron 1. Variation in the number of (CA) units causes differing levels of epidermal growth factor receptor gene expression and concomitant protein production (shorter alleles produce greater protein), which may be a predictor of clinical outcome in breast cancer (24).

The mutator phenotype model of cancer, which is appropriate for the description of MMR-deficient tumors, postulates that tumors evolve as a heterogeneous collection of genetically distinct cells, each having different, but frequently overlapping, patterns of mutation (25). Direct analyses of mononucleotide alleles in MSI colorectal carcinomas have shown that strand slippage mutations are as abundant in short, intergenic microsatellites as within intragenic, coding repeats (26). Thus, in addition to microsatellite sequence changes that are phenotypically selected during tumor cell evolution, a high frequency of random (“passenger”) microsatellite changes is expected to arise throughout the genome of individual MMR-deficient tumor cells.

## Perspective

Long regarded as “junk DNA,” microsatellites are emerging as important regulators of genome function and gene expression. Historically, the primary clinical use of microsatellites has been as markers of genome instability. However, a more direct function of microsatellite instability is to create genetic diversity for tumor cell evolution or immune system adaptability. Mutational biases (e.g., high versus low mutability;

insertion versus deletion events) within microsatellites are caused by intrinsic DNA features and by loss of a specific MMR protein. Such biases will directly determine the complement of microsatellite-associated target genes that are affected in MSI tumors. In addition, individual germ line polymorphisms in microsatellite sequence purity and length will affect the inherent mutability of a particular microsatellite allele. We postulate that somatic mutational biases, together with germ line microsatellite polymorphisms, may underlie the tissue-specific and individual-specific pattern of tumor formation observed among patients with HNPCC and other MSI. The unique pattern of mutational events—error type, affected unit size, and length of the microsatellite—arising within an MMR-defective tumor cell population will thus shape an individual’s tumor genome landscape, leading to a distinct clinical phenotype. Understanding the genotypic-phenotypic relationships in MMR-deficient tumors is necessary to develop better predictive models of disease progression and novel therapeutic approaches.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Grant Support

NIH grant no. RO1 CA100060 (K. Eckert) and the Jake Gittlen Cancer Research Foundation.

Received 8/14/09; revised 9/22/09; accepted 9/24/09; published OnlineFirst 1/12/10.

## References

- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 2004;5:435–45.
- Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 2004; 21:991–1007.
- Hsieh P, Yamane K. DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech Ageing Dev* 2008;129:391–407.
- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 2009;76:1–18.
- Peltomäki P. Lynch syndrome genes. *Fam Cancer* 2005;4:227–32.
- Korhonen MK, Raevaara TE, Lohi H, Nystrom M. Conditional nuclear localization of hMLH3 suggests a minor activity in mismatch repair and supports its role as a low-risk gene in HNPCC. *Oncol Rep* 2007; 17:351–4.
- Clyne M, Offman J, Shanley S, et al. The G67E mutation in hMLH1 is associated with an unusual presentation of Lynch syndrome. *Br J Cancer* 2009;100:376–80.
- Foulkes WD, Thiffault I, Gruber SB, et al. The founder mutation MSH2\*1906GÄC is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am J Hum Genet* 2002;71:1395–412.
- Eckert KA, Hile SE. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog* 2009;48:379–88.
- Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res* 2008;18:30–8.
- Eckert KA, Yan G, Hile SE. Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells. *Mol Carcinog* 2002;34:140–50.
- Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol* 1997;17:2851–8.
- Hawk JD, Stefanovic L, Boyer JC, Petes TD, Farber RA. Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. *Proc Natl Acad Sci U S A* 2005;102:8639–43.
- Shah SN, Eckert KA. Human postmeiotic segregation 2 exhibits biased repair at tetranucleotide microsatellite sequences. *Cancer Res* 2009;69:1143–9.
- Hegan DC, Narayanan L, Jirik FR, Edelmann W, Liskay RM, Glazer PM. Differing patterns of genetic instability in mice deficient in the mismatch repair genes Pms2, Mlh1, Msh2, Msh3 and Msh6. *Carcinogenesis* 2006;27:2402–8.
- Kazunori Otsuka TS, Shibata H, Kato S, Sakayori M, Shimodaira H, Kanamaru R, Ishioka C. Analysis of the human APC mutation spectrum in a *Saccharomyces cerevisiae* strain with a mismatch repair defect. *Int J Cancer* 2003;103:624–30.
- Yao X, Buermeyer AB, Narayanan L, et al. Different mutator phenotypes in Mlh1- versus Pms2-deficient mice. *Proc Natl Acad Sci U S A* 1999;96:6850–5.

18. Chen PC, Dudley S, Hagen W, et al. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. *Cancer Res* 2005;65:8662–70.
19. Woerner SM, Kloor M, von Knebel Doeberitz M, Gebert JF. Microsatellite instability in the development of DNA mismatch repair deficient tumors. *Cancer Biomark* 2006;2:69–86.
20. Sima Salahshor UK, Pahlman L, Glimelius B, Lindmark G, Lindblom A. Colorectal cancer with and without microsatellite instability involves different genes. *Genes Chromosomes Cancer* 1999;26:247–52.
21. Chang DK, Metzgar D, Wills C, Boland CR. Microsatellites in the eukaryotic DNA mismatch repair genes as modulators of evolutionary mutation rate. *Genome Res* 2001;11:1145–6.
22. Peled JU, Kuang FL, Iglesias-Ussel MD, et al. The biochemistry of somatic hypermutation. *Annu Rev Immunol* 2008;26:481–511.
23. Gangwal K, Sankar S, Hollenhorst PC, et al. Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proc Natl Acad Sci U S A* 2008;105:10149–54.
24. Brandt B, Meyer-Staeckling S, Schmidt H, Agelopoulos K, Buerger H. Mechanisms of EGFR gene transcription modulation: relationship to cancer risk and therapy response. *Clin Cancer Res* 2006;12:7252–60.
25. Loeb LA, Bielas JH, Beckman RA. Cancers exhibit a mutator phenotype: clinical implications. *Cancer Res* 2008;68:3551–7.
26. Sammalkorpi H, Alhopuro P, Lehtonen R, et al. Background mutation frequency in microsatellite-unstable colorectal cancer. *Cancer Res* 2007;67:5691–8.