AMERICAN SOCIETY FOR MICROBIOLOGY | Molecular and Cellular Biology®

RESEARCH ARTICLE

The Human DNA Mismatch Repair Protein MSH3 Contains Nuclear Localization and Export Signals That Enable Nuclear-Cytosolic Shuttling in Response to Inflammation

Stephanie S. Tseng-Rogenski,[a] Koji Munakata,[a] Daniel Y. Choi,[a] Paul K. Martin,[a] Supal Mehta,[a] Minoru Koi,[a] Wei Zheng,[b] Yang Zhang,[b] John M. Carethers[a,c,d]

[a]Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA
[b]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA
[c]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA
[d]Rogel Cancer Center, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT Inactivation of DNA mismatch repair propels colorectal cancer (CRC) tumorigenesis. CRCs exhibiting elevated microsatellite alterations at selected tetranucleotide repeats (EMAST) show reduced nuclear MutS homolog 3 (MSH3) expression with surrounding inflammation and portend poor patient outcomes. MSH3 reversibly exits from the nucleus to the cytosol in response to the proinflammatory cytokine interleukin-6 (IL-6), suggesting that MSH3 may be a shuttling protein. In this study, we manipulated three putative nuclear localization (NLS1 to -3) and two potential nuclear export signals (NES1 and -2) within MSH3. We found that both NLS1 and NLS2 possess nuclear import function, with NLS1 responsible for nuclear localization within full-length MSH3. We also found that NES1 and NES2 work synergistically to maximize nuclear export, with both being required for IL-6-induced MSH3 export. We examined a 27-bp deletion (Δ27bp) within the polymorphic exon 1 that occurs frequently in human CRC cells and neighbors NLS1. With oxidative stress, MSH3 with this deletion (Δ27bp MSH3) localizes to the cytoplasm, suggesting that NLS1 function in Δ27bp MSH3 is compromised. Overall, MSH3's shuttling in response to inflammation enables accumulation in the cytoplasm; reduced nuclear MSH3 increases EMAST and DNA damage. We suggest that polymorphic sequences adjacent to NLS1 may enhance cytosolic retention, which has clinical implications for inflammation-associated neoplastic processes.

KEYWORDS EMAST, IL-6, MSH3, colorectal cancer, inflammation, mismatch repair, nuclear import/export, subcellular localization

DNA mismatch repair (MMR) is an evolutionarily conserved process that corrects mispaired nucleotides and microsatellite slippage mistakes after the replication of DNA in the nuclei of cells. Several proteins take part to effectuate human MMR, including two DNA recognition complexes: MutSα, consisting of a heterodimer of MSH2 and MSH6, and MutSβ, consisting of a heterodimer of MSH2 and MSH3, as well as a signaling complex consisting of MLH1 and PMS2. MutSα is capable of recognizing single nucleotide mispairs as well as slippage mistakes at mono- or dinucleotide microsatellite sequences, whereas MutSβ recognizes slippage mistakes at dinucleotide or longer microsatellite sequences and does not repair mononucleotide slippages (1–3). The unique repair aspects of the two MMR recognition complexes predict the type of microsatellite instability observed within human tissues. For instance, loss of function of MSH2, MLH1, or PMS2 would cause mono-, di-, tri-, and tetranucleotide instability, whereas isolated loss of MSH6 would cause mono- and some dinucleotide instability

Citation Tseng-Rogenski SS, Munakata K, Choi DY, Martin PK, Mehta S, Koi M, Zheng W, Zhang Y, Carethers JM. 2020. The human DNA mismatch repair protein MSH3 contains nuclear localization and export signals that enable nuclear-cytosolic shuttling in response to inflammation. Mol Cell Biol 40:e00029-20. https://doi.org/10.1128/MCB.00029-20.

Address correspondence to John M. Carethers, jcarethe@umich.edu.

Received 21 January 2020
Returned for modification 29 February 2020
Accepted 7 April 2020
Accepted manuscript posted online 13 April 2020
Published 15 June 2020

only. Isolated loss of MSH3 leads to di-, tri-, and tetranucleotide instability and would lack mononucleotide instability (3, 4). Classic MSI-High was defined by a National Cancer Institute (NCI) consensus panel as at least 2 frameshifts identified with the use of a panel of at least 5 microsatellite markers that detect mono- or dinucleotide instability (5). More recently, elevated microsatellite alterations at selected tetranucleotide repeats (EMAST) has been observed among several cancers in relation to inflammation (4, 6–8). The detection of EMAST (in the absence of detection of MSI-High) corresponds to isolated MSH3 dysfunction, matching its expected MMR biochemical repair profile (4, 9–11). Additionally, MSI-Low, defined as one marker showing frameshift among the mono- and dinucleotide NCI consensus panel, correlates with isolated MSH3 dysfunction as well due to essentially all MSI-Low assays showing only dinucleotide frameshifts instead of any mononucleotide frameshifts (4, 12, 13).

In addition to MSH3's role in MMR, MSH3 is involved in double-strand break (DSB) repair via homologous recombination, something unique among the MMR proteins (4, 14, 15). Thus, loss of MSH3 function would generate a complex repair defect, one involving MMR (detected by EMAST) and one involving DSBs (4). It is not clear, at this time, how the proportion of contribution of each repair defect drives the progression of human disease, but misrepair of DSBs can alter patient outcome from colorectal cancer (CRC) metastases (16). MSH3 may also be important in generating extrachromosomal circular DNA (eccDNA; also known as microDNA) containing repeat sequences that range from 4 to 8 bp (17). MicroDNAs can potentially produce short RNAs or truncated proteins that may impact the production or function of proteins generated from genomic DNA (17).

There is little known regarding regulatory mechanisms for inactivation of MSH3. Mutation of *MSH3* is not a contributory germ line factor causing Lynch syndrome, for which germ line mutations in *MSH2*, *MLH1*, *PMS2*, and *MSH6* (and *EPCAM*) have been described (18, 19). Biallelic germ line mutation of *MSH3* has been reported and causes an oligopolyposis syndrome, with normal tissues displaying EMAST due to absence of both *MSH3* alleles and failing to allow expression of MSH3; this syndrome is extremely rare (20). MSH3 can be inactivated by somatic frameshift mutation due to an intrinsic coding polyadenine microsatellite sequence; frameshift mutation occurs in ~30% of sporadic MSI-High CRCs due to prior inactivation of MLH1 (caused by hypermethylation of the *MLH1* promoter) (9, 21). This mechanism for MSH3 inactivation would be estimated to occur in ~5% or less of all sporadic CRCs, and the biological consequences of the addition of MSH3 dysfunction on top of MLH1 inactivation are not clear (9, 22). MSH3 is not known to be inactivated by methylation, and there are only rare reports of any other nonframeshift mutations (22). MSH3 can be commonly inactivated by inflammation that induces a nuclear-to-cytosolic shift for the protein (23, 24). This mechanism for MSH3 inactivation is suspected to account for ~50% of all sporadic cancers, since samples exhibit reduced nuclear MSH3 expression. If true, it means that somatic acquired MSH3-deficiency is the most common MMR defect among CRCs (4, 9, 22). MSH3's cytosolic shift is triggered by interleukin-6 (IL-6)/JAK/STAT3 signaling and oxidative stress, and when MSH3 is out of the nucleus, EMAST is detected in genomic DNA (24, 25). This observation is consistent with the histological finding of reduced nuclear MSH3 immunostaining among cancers that manifest EMAST (6, 8) as well as the association of higher IL-6 immunostaining among EMAST CRCs (23). Furthermore, there are multiple lines of evidence demonstrating that the presence of EMAST (or inflammation-associated microsatellite alterations) correlates with (i) the presence of inflammation, (ii) advanced CRC stage, (iii) progressive histological advancement of neoplasia in the colon, (iv) poor patient prognosis and survival, and (v) African American race, a group that presents with higher incidence and mortality from CRC than do Caucasians (4, 8, 13, 26–28). Thus, there are clinical consequences and altered patient outcomes related to the presence of EMAST and isolated MSH3 dysfunction. Understanding the mechanisms for MSH3's ability to shuttle to the cytosol with inflammation will be instrumental in potentially preventing MSH3-deficient DNA damage.

In this study, we examined the structure of the MSH3 protein and manipulated

TABLE 1 Primers used in this study

Name of primer	Sequence[a]
NLS1 Forward	5′-GGCCACCATGGACAGAAGAAAGAAGAGACCATTGGAAAATGATGGGCCTGTTAAAAAGAAAGTAAAGAAAGTCCAACAAAAGG-3′
NLS1 Reverse	5′-GATCCCTTTTGTTGGACTTTCTTTACTTTCTTTTTAACAGGCCCATCATTTTCCAATGGTCTCTTCTTTCTTCTGTCCATGGTGGCCTGCA-3′
NLS1 Full mutation Forward	5′-GGCCACCATGGAC*GCCGCCGCCGCCGCCCC*ATTGGAAAATGATGGGCCTGTT*GCCGCCGCCG*TA*GCCGCCG*TCCAACAA*GCCG*-3′
NLS1 Full mutation Reverse	5′-GATCCGGCTTGTTGGACGGCGGCTACGGCGGCGGCAACAGGCCCATCATTTTCCAATGGGGCGGCGGCGGCGGCGTCCATGGTGGCCTGCA-3′
NLS2 Forward	5′-GGCCACCATGGTTATAAAAAAGAGGAAGGATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCATTTGCAAGAAATACGAAAAATACTAAAAG-3′
NLS2 Reverse	5′-TTTTAGTATTTTTCGTATTTCTTGCAAATGCATTCGGATCTCGTCAATAACACCTTGAATTTCATCCTTCCTCTTTTTTATAACCATGGCA-3′
NLS2 Full mutation Forward	5′-GGCCACCATGGTTATAAAA*GCCGCCGCC*GATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCATTTGCAAGAAATA*GCCGCC*ATACTA*GCCG*-3′
NLS2 Full mutation Reverse	5′-GATCC*GGC*TAGTAT*GGCGGC*TATTTCTTGCAAATGCATTCGGATCTCGTCAATAACACCTTGAATTTCATC*GGCGG CGGC*TTTTATAACCATGGTGGCCTGCA-3′
NLS3 Forward	5′-GGCCACCATGATAAATACGAAAAGAAAGAGACTCAAGTATTTTGCAAAGTTATGGG-3′
NLS3 Reverse	5′-GATCCCCATAACTTTGCAAAATACTTGAGTCTCTTTCTTTTCGTATTTATCATGGTGGCCTGCA-3′
NLS1 deletion	5′-CCGCCGCACATAGCTACAGAAATTGAAGGAGGAAGTGAT-3′
NLS2 deletion	5′-AAAGACCTTTCTGACTTCCCTTTAAATCCTTCTGCACAA-3′
NES1 Forward	5′-GGCCACCATGACTGATTTTGATGATATCAGTCTTG-3′
NES1 Reverse	5′-GATCCAAGACTGATATCATCAAAATCAGTCATGGTGGCCTGCA-3′
NES2 Forward	5′-GGCCACCATGATCAGAGATGTGAAATCCTTAACCCTGG-3′
NES2 Reverse	5′-GATCCCAGGGTTAAGGATTTCACATCTCTGATCATGGTGGCCTGCA-3′
NES1+NES2 Forward	5′-GGCCACCATGACTGATTTTGATGATATCAGTCTTATTTCGATCAGAGATGTGAAATCCTTAACCCTGG-3′
NES1+NES2 Reverse	5′-GATCCCAGGGTTAAGGATTTCACATCTCTGATCGAAATAAGACTGATATCATCAAAATCAGTCATGGTGGCCTGCA-3′
NES1 deletion	5′-GCAGTTCTGCCAAAATGTCTACACGCAAAGAATGCA-3′
NES2 deletion	5′-GCTACACTTGAGTATTTCTTTGTCACCCATTATCCG-3′
Mutagenesis XmaI to AgeI selection	5′-TGCAAAAAGCTACCGGTAGCTTGTATAT-3′

[a]The sequences corresponding to the signals are underlined. The sequences mutated to examine if mutations could abolish the signal's function are in italics.

putative nuclear localization signal (NLS) and nuclear export signal (NES) sequences to identify bona fide signals that allow MSH3 to shuttle between the nucleus and cytosol. We further examined a polymorphism that we identified in close proximity to the bona fide NLS and its effect on MSH3's nuclear localization, as well the polymorphism's frequency among a common IL-6-enriched human inflammatory condition.

RESULTS

Identification of putative MSH3 NLSs and NESs. As MSH3 is a nuclear DNA repair protein, we submitted its entire amino acid sequence to an NLS mapper program (see Materials and Methods) to identify potential NLSs. Six predicted monopartite and 52 bipartite NLSs were identified, and we further tested the top three candidates, including one monopartite and two bipartite signals. Directionally from the amino to carboxyl terminus of MSH3, we named the signals NLS1 (^{84}DRRKKROLENDGPVKKKVKKVQQK^{107}; bipartite), NLS2 (^{713}KKRKDEIQGVIDEIRMHLQEIRKILK^{738}; bipartite), and NLS3 (^{1097}INTKRKRLKYFAKLW^{1111}; monopartite). Similarly, we also examined the entire MSH3 sequence via an NES mapper program (also see Materials and Methods), which revealed two regions that may possess nuclear export function. Directionally from the amino to carboxyl terminus of MSH3, we named them NES1 (^{167}TDFDDISL^{174}) and NES2 (^{998}IRDVKSLTL^{1006}). We designed primers to amplify and subsequently performed site-directed mutagenesis at each of the putative NLS and NES sequences (Table 1). The amino acid sequences predicted to be critical for their nuclear localization or export functions were examined as the wild type (WT) to determine functionality and then mutated and/or deleted to examine if the mutations affected or abolished nuclear localization or export functions.

Initial testing and characterization of human MSH3 NLSs. To examine if any of the putative MSH3 NLSs are functional, we generated reporter constructs where each individual NLS, both WT and fully mutated versions (in which all lysine [K] and arginine [R] residues predicted to be critical amino acids in the putative signal were mutated to

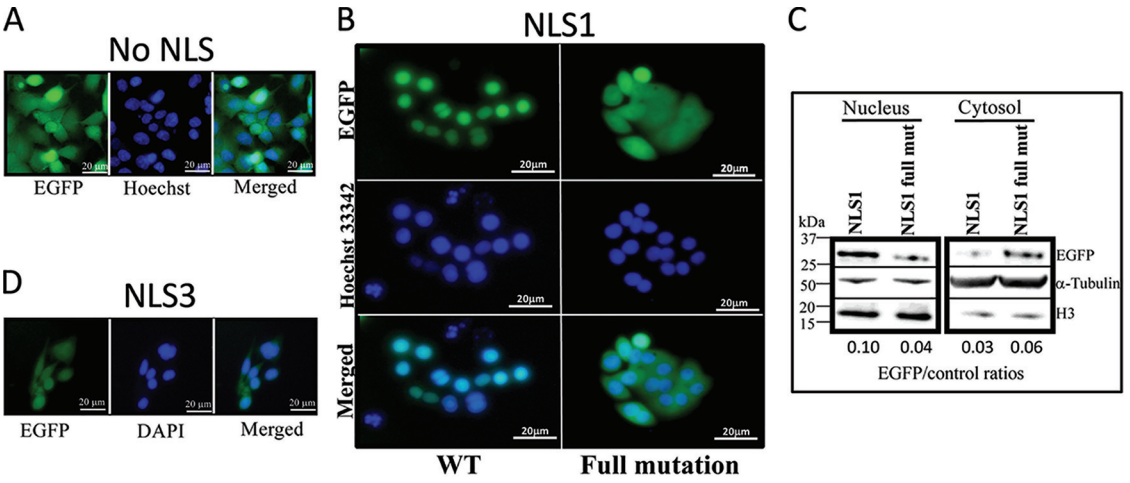

FIG 1 MSH3 NLS1 is a functional nuclear localization signal. (A) EGFP alone (control) was localized throughout the cell. (B) Fluorescence microscopy demonstrating that WT NLS1 labeled with EGFP was able to direct EGFP into the nucleus. The nuclear localization function is disrupted upon mutation of NLS1. The NLS1-EGFP was observed in the nucleus exclusively, and the mutated NLS1-EGFP was dispersed throughout the cells in all of the cells expressing EGFP (transfection efficiencies were about 70%). (C) Western blotting of whole-cell lysates separated into nuclear and cytosolic fractions. WT NLS1 (DRRKKRPLENDGPVKKKVKKVQQKEGGSDLG) localized exclusively in the nucleus, while mutant NLS1 (DAAAAAPLENDGPVAAAVAAVQQAEGGSDLG) was predominantly in the cytoplasm. HCT116 and HCT116+Ch3 cells were used as host cells (HCT116 shown). Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively. (D) EGFP led by WT NLS3 displayed diffused signal throughout the cell, indicating that NLS3 is not a functional nuclear localization signal. DAPI, 4′,6-diamidino-2-phenylindole.

alanine [A]) were fused to enhanced green fluorescent protein (EGFP) in the same reading frame, allowing determination of EGFP localization via fluorescence microscopy in real time. The EGFP was not led by any signaling sequences; thus, its nuclear entry heavily relies on a functional NLS. EGFP without an NLS is distributed throughout the cell (Fig. 1A). As a stand-alone sequence, WT NLS1 became located in the nucleus exclusively. In contrast, mutated NLS1 demonstrated dispersion within the cell's cytosol (Fig. 1B). To supplement our microscopic studies, we fractionated total cell lysates into nuclear and cytosolic fractions, with α-tubulin and histone 3 (H3) used as cytosolic and nuclear protein purification markers, respectively. Consistent with the microscopic results, WT NLS1 localized with the nuclear fraction, whereas the majority of mutated NLS1 localized to the cytosolic fraction (Fig. 1C). In examining NLS2 as a stand-alone sequence fused with EGFP, marked fractions of WT NLS2-EGFP localized to the nucleus, suggesting that this sequence possesses nuclear localization function which is weaker than that of NLS1 (Fig. 2). Mutated NLS2 demonstrated less nuclear localization than WT NLS2, with relative enrichment in the cytosol (Fig. 2). Overall, NLS2 may be a partially capable nuclear localization signal. As a stand-alone sequence fused with EGFP, WT NLS3 distributed throughout the cytosol, indicating that NLS3 does not possess detectable level of nuclear localization function (Fig. 1D). Thus, we did not further pursue NLS3 experimentation.

Since nuclear localization signals are not required for prokaryotic cells, we predicted that sequences responsible for nuclear localization of MSH3 may not exist in *Escherichia coli* MSH3 (MutS). We compared *E. coli* and human MSH3 using the publicly available Basic Local Alignment Search Tool (BLAST) program to align the amino acid sequences, and indeed, *E. coli* MutS does not contain either NLS1 or NLS2 sequences. NLS1 is highly conserved among primates (Table 2) but is not present in other species, such as *Xenopus* and/or yeast. NLS2 is also highly conserved among higher eukaryotic species. However, budding yeast (*Saccharomyces cerevisiae*) does not contain a sequence similar to NLS2. Overall, these experiments and observations indicate that NLS1 is the more potent nuclear localization signal over NLS2 (or NLS3).

The second cluster of K residues within NLS1 is the component critical for its nuclear import function. NLS1 is a bipartite signal, containing two clusters of K/R residues. When the NLS1 sequence is aligned with its corresponding sequences in

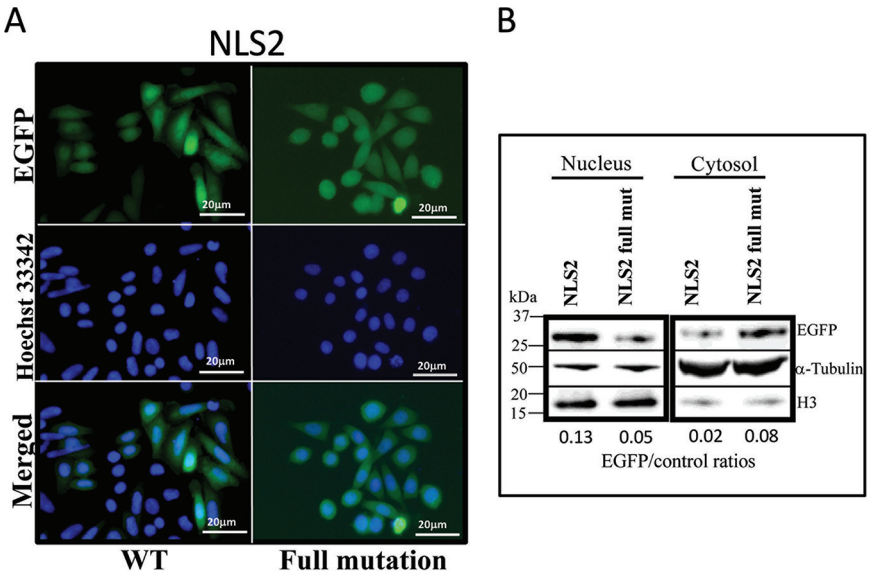

FIG 2 MSH3 NLS2 exhibits a detectable level of nuclear localization function. (A) Fluorescence microscopy demonstrating that WT NLS2 labeled with EGFP was able to direct some, but not all, EGFP into the nucleus. Transfection efficiencies were about 70%, and all EGFP-expressing cells displayed the same pattern for the signal. (B) Western blotting of whole-cell lysates separated into nuclear and cytosolic fractions. HCT116 and HCT116+Ch3 cells were host cells (HCT116 shown). MSH3 NLS2 (KKRKDEIQGVID EIRMHLQEIRKILK) was not as strong of signal as NLS1 but possessed a detectable level of nuclear localization function, in contrast to the mutated NLS2 (AAAADEIQGVIDEIAMHLQEIAAILA). Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively.

mouse and rat, all three homologs contain two clusters of K/R residues (Fig. 3A). We examined if NLS1 is truly a bipartite signal or if there is a dominant cluster function within the bipartite signal. Reporter constructs containing only the first cluster (NLS1-A) and/or the second cluster (NLS1-B) fused with EGFP were generated and tested. As a stand-alone sequence, NLS1-B appears to be far more critical for its nuclear import function than NLS1-A (Fig. 3B). Nevertheless, NLS1-B was not exclusively localized to the nucleus like WT NLS1 (Fig. 1B and C), suggesting that NLS1-A complements and completes the NLS1-B cluster to fully possess nuclear localization function. Another possibility is that longer sequences are needed in the construct containing NLS1-B to properly fold and present the signal. To distinguish this, we generated two new constructs containing the entire length of the NLS1 but with mutations on the first (NLS1-A mut) or second (NLS1-B mut) cluster separately. Our results clearly demonstrate that the second cluster of K residues is the critical component of the signal, with the first cluster of K/R residues required for exclusive nuclear localization of NLS1

TABLE 2 MSH3 NLS1 sequences are highly conserved among mammals[a]

Species	Amino acid sequence	Homology, no./total (%)
Human	DRRKKRPLENDGPVKKKVKKVQQK	
Gorilla	DRRKKRPLENDGPVKKKVKKVQQK	24/24 (100)
Chimpanzee	RRKKRPLENDGPVKKKVKKVQQK	23/24 (96)
Macaque	DRSKKRPLENDGPVKKKVKKAQQK	22/24 (91)
Marmoset	DRSRKRPLENDGPVKKKVKKVQEK	21/24 (88)
Rhinoceros	DRSKKRPLENDGPVKKKAKKVQEK	21/24 (88)
Horse	DRSKKRPLENDGPVKKKAKKVQEK	21/24 (88)
Dolphin	DRSKKRPLESDGPVKKKAKKVQEK	20/24 (83)
Killer whale	DRSKKRPLESDGPVKKKAKKVQEK	20/24 (83)
Mole	DRSKKRPLESDGPVKKKAKKVQEK	20/24 (83)
Squirrel monkey	DRSRKRPLENDGPVKKKVKKTQEK	20/24 (83)
Gibbon	DRSKKRPXENDGPARKKVKKVQQK	20/24 (83)
Tree shrew	DRIKKRPLENDGPVKKKAKKTQEK	20/24 (83)

[a]Listed are the top 12 that we identified.

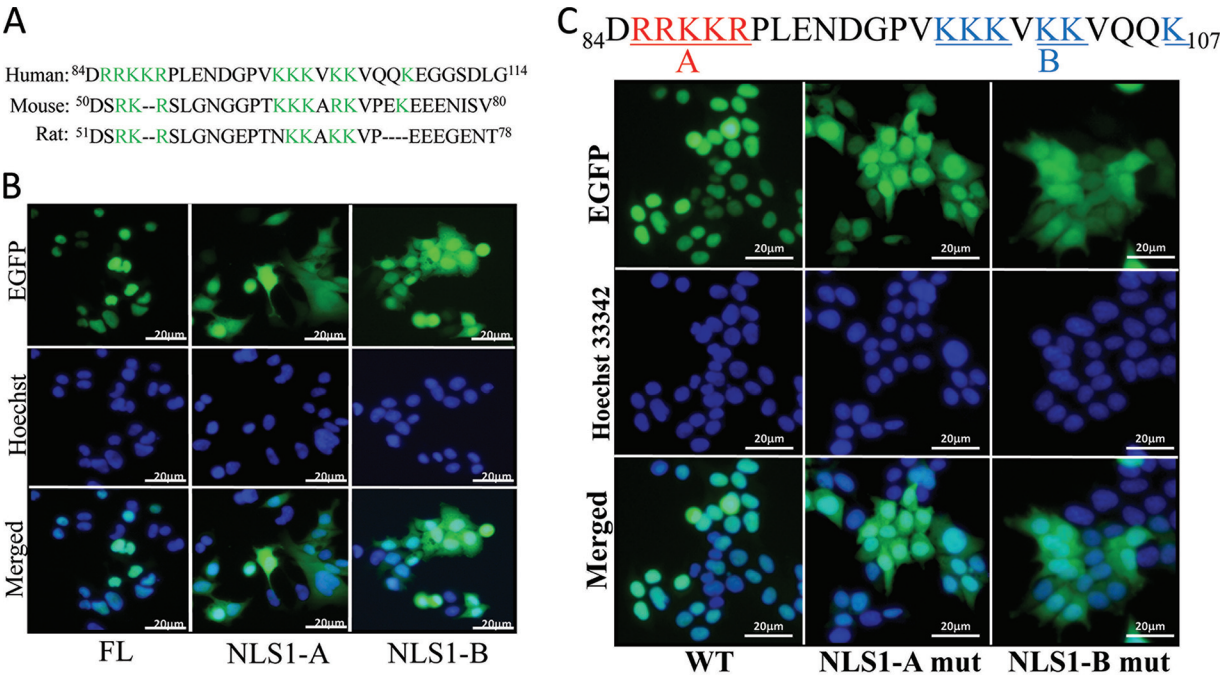

FIG 3 The second cluster of conserved lysine (K) residues is critical components of MSH3 NLS1. (A) Alignment of MSH3 NLS sequences with their corresponding sequences in mouse and rat. All three contain two clusters of K/R residues separated by eight or nine amino acids, presumably bipartite NLSs. (B) NLS1 is a bipartite signal (NLS1-A and/or NLS1-B). Each contains a cluster of components critical for the import function (K/R residues). EGFP led by NLS1-B demonstrated a higher level of nuclear signal enrichment than that by NLS1-A by fluorescence microscopy. (C) Fluorescence microscopy after transfection of constructs carrying WT NLS1, mutation within the first cluster of K/R (NLS1-A mut), and/or mutations on the second cluster (NLS1-B mut). Mutations on the B (second) cluster of lysines significantly impaired nuclear import function. HCT116 and HCT116+Ch3 cells were used for the study (HCT116+Ch3 shown). All EGFP-expressing cells exhibited similar localization patterns within the same experimental group. Transfection efficiencies were about 70% in all groups.

(Fig. 3C). Altogether, NLS1 is a bipartite signal, and the second cluster of K/R residues is more critical than the first cluster of K/R for its nuclear import function.

NLS1 is responsible for nuclear localization of MSH3. Since both NLS1 and NLS2 possess detectable levels of nuclear localization function as stand-alone sequences, we examined both sequences and their function within the context of FL MSH3. We generated a reporter construct that expresses FL MSH3 tagged with FLAG. We deleted NLS1 or NLS2 individually (Fig. 4A and B) to assess if MSH3 could enter into the nucleus without each signal. FL MSH3 (containing both WT NLS1 and WT NLS2) expressed through our reporter construct was able to enter into the nucleus as expected. Deletion of NLS1, but not NLS2, prevented FL MSH3 from entering the nucleus (Fig. 4C and D). Altogether, our results indicate that as an independent signal, both NLS1 and NLS2 possess detectable levels of nuclear localization function, with NLS1 $>$ NLS2. However, within the context of FL MSH3, NLS1 is the sole and bona fide signal permitting nuclear localization for MSH3.

Initial testing and characterization of putative human MSH3 NESs. Similar to our studies for NLSs, GFP was used as a reporter protein to evaluate the two potential NESs. We adopted an established system designed to test nuclear exit function to examine our NESs. In this system, GFP is tagged with a functional NLS and NES (NES3, derived from HIV) (see Materials and Methods). We inserted individual human MSH3 NESs and/or combined NESs to replace the NES3 on the construct. Compared to NES1 or NES2 alone, GFP fused with both NES1 and NES2 resulted in increased cytosolic expression suggesting synergism; such nuclear export function was abolished by mutations within both NES1 and NES2 (Fig. 5A and B). GFP was observed predominately in the nucleus led by a functional NLS only (Fig. 5C, top row). The insertion of a peptide containing scramble sequences resulted in a slight leakage of GFP into the cytoplasm (Fig. 5C, bottom row). This observation suggests that the extra sequence may have

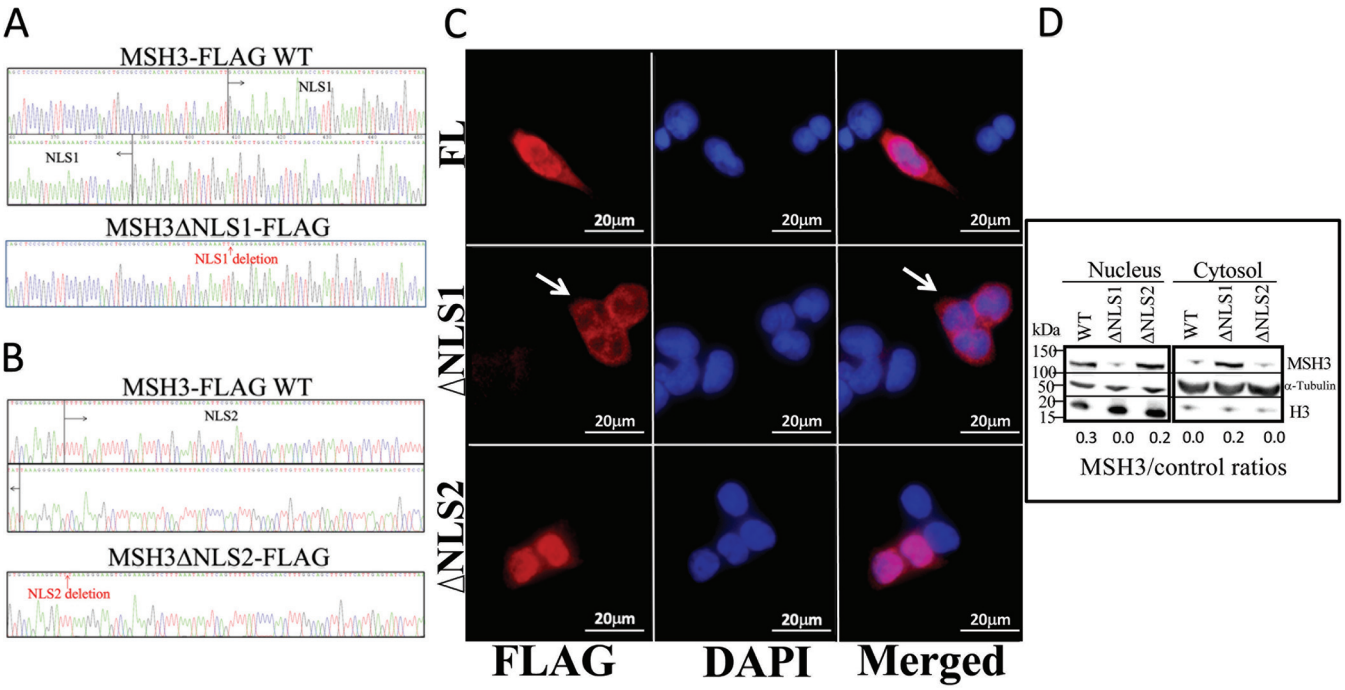

FIG 4 MSH3 NLS1, not NLS2, is responsible for the nuclear localization of MSH3. (A and B) Sequencing chromatograms of FL, ΔNLS1, and ΔNLS2 sequences. The expression constructs containing FL, ΔNLS1, and/or ΔNLS2 were sequenced. The deleted sequences corresponding to NLS1 (A) and NLS2 (B) are indicated. (C and D) An expression vector carrying full-length MSH3-FLAG (FL) was generated. Subsequently, the NLS1 and NLS2 sequences were removed from the FL MSH3 on the expression construct individually. These constructs were transfected into cells to examine subcellular localization of the FL MSH3 and mutant MSH3 without NLS1 (ΔNLS1) and/or NLS2 (ΔNLS2) by immunofluorescence microscopy (C) and nuclear-cytosolic fractionation and Western blotting (D). FL MSH3 and ΔNLS2 were able to enter into the nucleus, while ΔNLS1 was not (emphasized by arrows), demonstrating that NLS1 is the bona fide signal directing MSH3 nuclear localization. All cells expressing MSH3-FLAG displayed the same localization pattern in the same experimental group. Transfection efficiencies were about 15% in all groups. Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively. SW480 and HT29 cells were used (SW480 shown).

impacted the function of the NLS. To assess the effects of IL-6, we examined the posttreatment localization of NES1-NES2-GFP. IL-6 did not enhance NES1-NES2-GFP localization into the cytosol as effectively as it did for the native endogenous MSH3 (Fig. 5D and E). Our result is likely due to the NES1-NES2-GFP being expressed from stand-alone sequences and not in the context of FL MSH3, where other protein interactions for shuttling may take place. Both NES1 and NES2 sequences are highly conserved among multiple species (Table 3).

Both NES1 and NES2 are responsible for MSH3 nuclear export in response to the proinflammatory cytokine IL-6. We have previously demonstrated that IL-6 through its signaling pathway causes MSH3 to translocate from the nucleus to the cytosol (24). To examine if NES1 and/or NES2 facilitates MSH3's nuclear-to-cytoplasmic translocation upon IL-6 treatment, we deleted NES1 or NES2 individually as well as both NES1 and NES2 from FL MSH3 (Fig. 6A and B). All WT and mutated NES versions of proteins were localized to the nucleus in the absence of IL-6 (Fig. 6D). With IL-6 treatment, only FL MSH3-FLAG containing both WT NES1 and WT NES2 was able to exit the nucleus and move into the cytoplasm (Fig. 6C and D). MSH3-FLAG containing only NES1 and/or NES2 stayed in the nucleus upon IL-6 treatment (Fig. 6E). These results demonstrate that both NES1 and NES2 are required for MSH3 nuclear export upon IL-6 treatment.

A 27-bp deletion polymorphism within MSH3 exon 1 near NLS1 alters sensitivity of MSH3 to inflammation and oxidative stress. Based on the literature and information available from the National Center for Biotechnology Information (NCBI) website, some sequences within exon 1 of *MSH3* appear to be polymorphic (29). Specifically, *MSH3*'s exon 1 contains six repeats of an imperfect 9-bp tandem repeat (G[C]CCG[C]CAGCC), representing an area heavily enriched in alanine and proline (P)

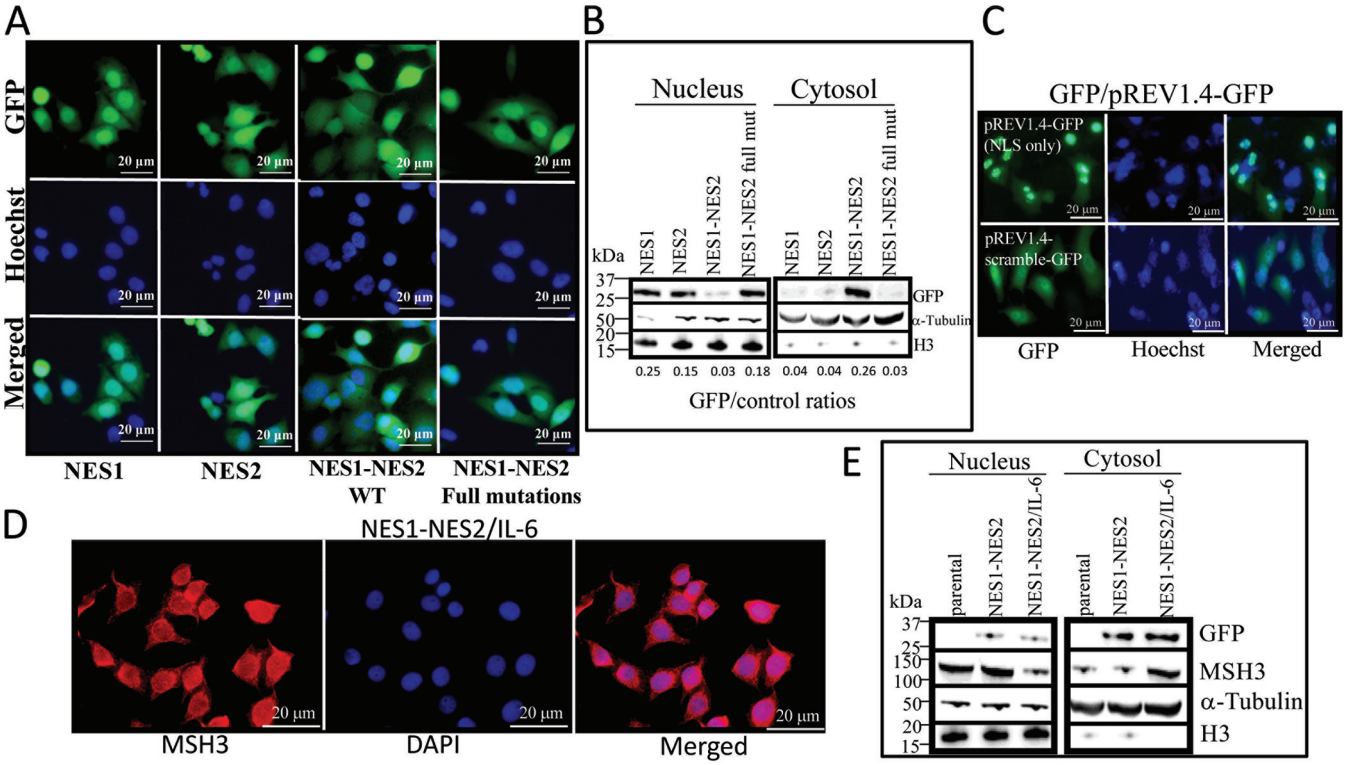

FIG 5 NES1 and NES2 from MSH3 appear to be synergistic in exhibiting detectable levels of nuclear export function to the cytosol. GFP fused with NES1 alone, NES2 alone, and/or NES1 and NES2 (NES1-NES2) were expressed in HCT116 cells to monitor subcellular localization of GFP by fluorescence microscopy (A) and nuclear-cytosolic fractionation and Western blotting (B). There was more cytosolic GFP when it was fused with NES1-NES2, indicating that the export function is greater when NES1 and NES2 are both present. GFP was observed in a similar pattern in all GFP-expressing cells within the same group. Transfection efficiencies were about 70% in all groups. (C) GFP controls led by a functional NLS alone was localized predominately in the nucleus (top), and the GFP led by a functional NLS and a scrambled peptide had significant leakage of GFP into the cytoplasm (bottom). (D and E) Endogenous MSH3 in cells expressing NES1-NES2-EGFP was able to shift into the cytoplasm more efficiently than EGFP in response to IL-6 treatment. (D) A portion of cells transfected with NES1-NES2-GFP constructs were stained with anti-MSH3 antibody to monitor its localization upon IL-6 treatment. (E) Whole-cell lysates were fractionated into nuclear and cytosolic fractions to examine the locations of the proteins of interest. Although IL-6 treatment caused only a very moderate shift of GFP into the cytoplasm, more significant shift was observed for endogenous MSH3. This result indicates that these cells were capable of supporting MSH3 shift and may suggest that other sequences within the full-length MSH3 are required for more effective shift of MSH3. Shown are experiments with HCT116 cells.

($[A]_{12}$PPAPPAPA; here called the poly-Ala tract). However, three to seven repeats, ranging from deletion of three copies to insertion of one copy, have been reported (29). Further inspection of this polymorphic sequence shows close proximity to MSH3 NLS1 (codons 51 to 77 in exon 1). Modification of the number of repeat sequences causes variation in the length of the poly-Ala tract. Poly-Ala sequences are proposed to be important for protein metabolism and function, including protein folding, and alterations in length (predominately expansions) are associated with human disease (29). We wondered if length deletions within this poly-Ala tract would have any impact on the function of MSH3 NLS1, given its proximity. We found the MSH3 exon 1 poly-Ala tract region to be highly polymorphic in human cancer cell lines, and we initially genotyped alleles from 4 cell lines: SW480 as Δ27bp/Δ27bp (deletion of 27 bp), HeLa as Δ27bp/I9bp/I18bp (insertion of 9 bp or 18 bp), HT29 as FL/I9bp, and HL60R as FL/FL (Fig. 7A to

TABLE 3 Sequences of NES1 and NES2 are highly conserved[a]

NES	Species
NES1 (TDFDDISL)	Chimpanzee, macaque, orangutan, bat, tree shrew, elephant, manatee, squirrel monkey, gibbon, marmoset
NES2 (IRDVKSLTL)	Chimpanzee, dog, orangutan, ferret, tree shrew, elephant, yak, turtle, boar, hamster

[a]Listed are 10 species for each NES as examples. These species contain sequences that are 100% identical to the sequences in humans.

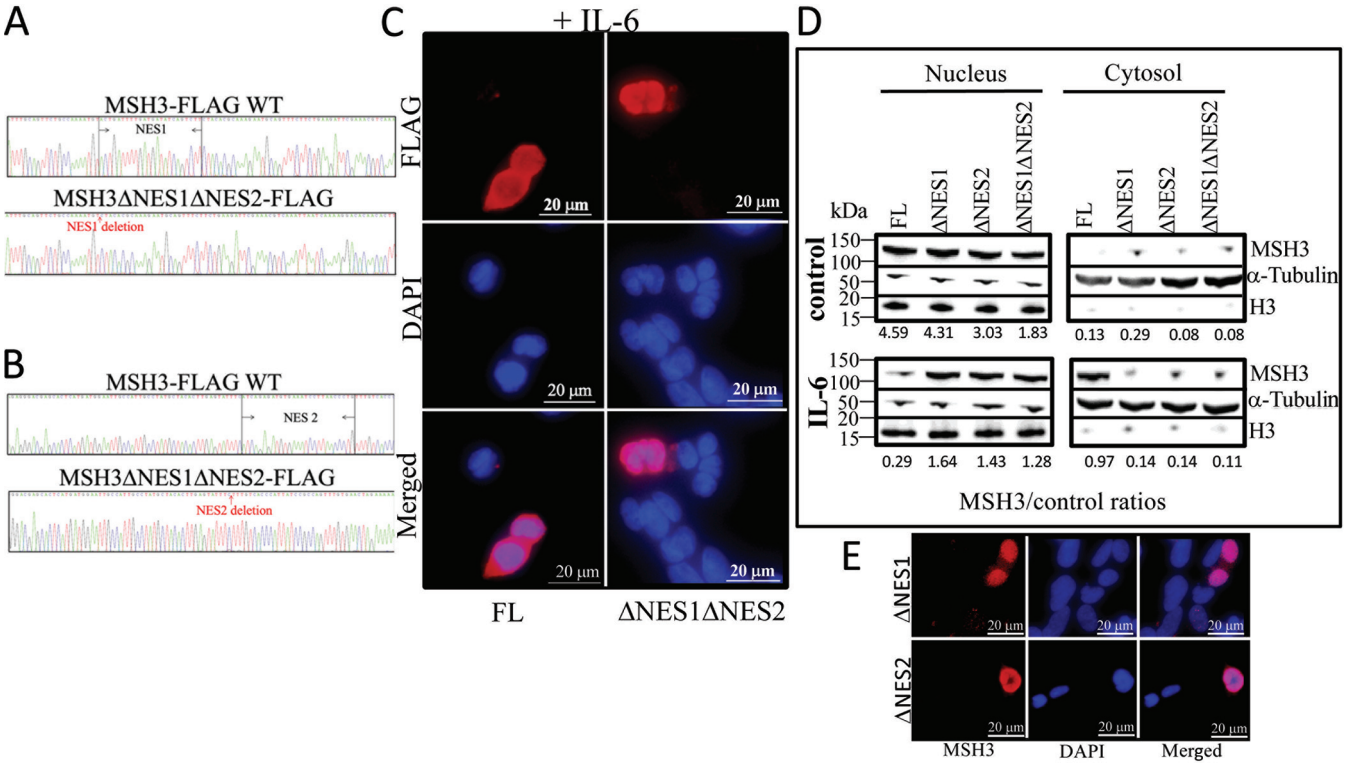

FIG 6 MSH3 NES1 and NES2 work synergistically to export MSH3 into the cytoplasm upon IL-6 treatment. (A and B) NES1 and NES2 were deleted individually and/or simultaneously from the reporter construct expressing FL MSH3-FLAG and sequenced. These expression constructs were electroporated into cells to express MSH3-FLAG protein to determine subcellular localization with and/or without IL-6 treatment. (C) Immunofluorescene microscopy demonstrating FL MSH3-FLAG shifts into the cytoplasm in response to IL-6 treatment (50 to 75% of MSH3-FLAG-expressing cells exhibited the shift upon IL-6 treatment). Deletion of NES1 and NES2 (ΔNES1ΔNES2) causes MSH3-FLAG to remain predominantly in the nucleus despite IL-6 treatment (with all MSH3ΔNES1ΔNES2-FLAG-expressing cells displaying the same pattern). Transfection efficiencies were about 15%. (D) Nuclear-cytosolic fractionation and Western blotting demonstrating that in the absence of IL-6, all MSH3-FLAG protein remains in the nucleus. Upon IL-6 treatment, only FL MSH3-FLAG protein containing both WT NES1 and WT NES2 was able to exit the nucleus into the cytoplasm. Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively. SW480 and HT29 cells were used (SW480 shown). (E) Expression constructs carrying MSH3ΔNES1-FLAG (ΔNES1) and/or MSH3ΔNES2-FLAG (ΔNES2) were transfected into SW480 cells, followed by IL-6 treatment. Without NES1 or NES2, MSH3 is unable to shift into the cytoplasm in response to IL-6 treatment.

E). With IL-6 treatment, cells containing either FL or Δ27bp MSH3 demonstrate endogenous MSH3 translocating from the nucleus to the cytoplasm. However, with H_2O_2 treatment (simulating oxidative stress), only Δ27bp MSH3, and not FL MSH3, moves to the cytosol (Fig. 7F to H). This suggests that Δ27bp MSH3 may have greater tendency to accumulate in the cytosol than FL MSH3. To assess this further, we deleted the same 27 bp within exon 1 of MSH3 (Δ27bp MSH3) on our reporter construct MSH3-FLAG and conducted transient-transfection experiments to examine the effects of H_2O_2 and/or IL-6 on MSH3 localization. Similar to the observations with endogenous FL and Δ27bp MSH3, transfected Δ27bp MSH3-FLAG translocated to the cytosol with both IL-6 and H_2O_2, while FL only responded to IL-6 (Fig. 8A and B). When transfected cells were treated with various doses of IL-6, Δ27bp MSH3-FLAG appeared to be more sensitive to nuclear-to-cytoplasmic shift with lower dosage of IL-6 than FL MSH3-FLAG (Fig. 8C). Detailed computer modeling examining the N terminus of FL and Δ27bp MSH3 fused with the published crystal structure of the remainder of FL MSH3 sequences was performed. The modeling showed NLS1 at the tip of a loop that is exposed, making it accessible for the docking of MSH3 onto the nuclear pore complex, whereas Δ27bp MSH3 through shortening of the alanine and proline repeats moves NLS1 to a less accessible location on the side of the N-terminal loop of MSH3 (Fig. 9), predicted to result in a slower nuclear entry for Δ27bp MSH3. These results indicate that the polymorphic Δ27bp MSH3 behaves differently than its FL counterpart and is more sensitive to a cytoplasmic shift with oxidative stress and inflammation likely due to

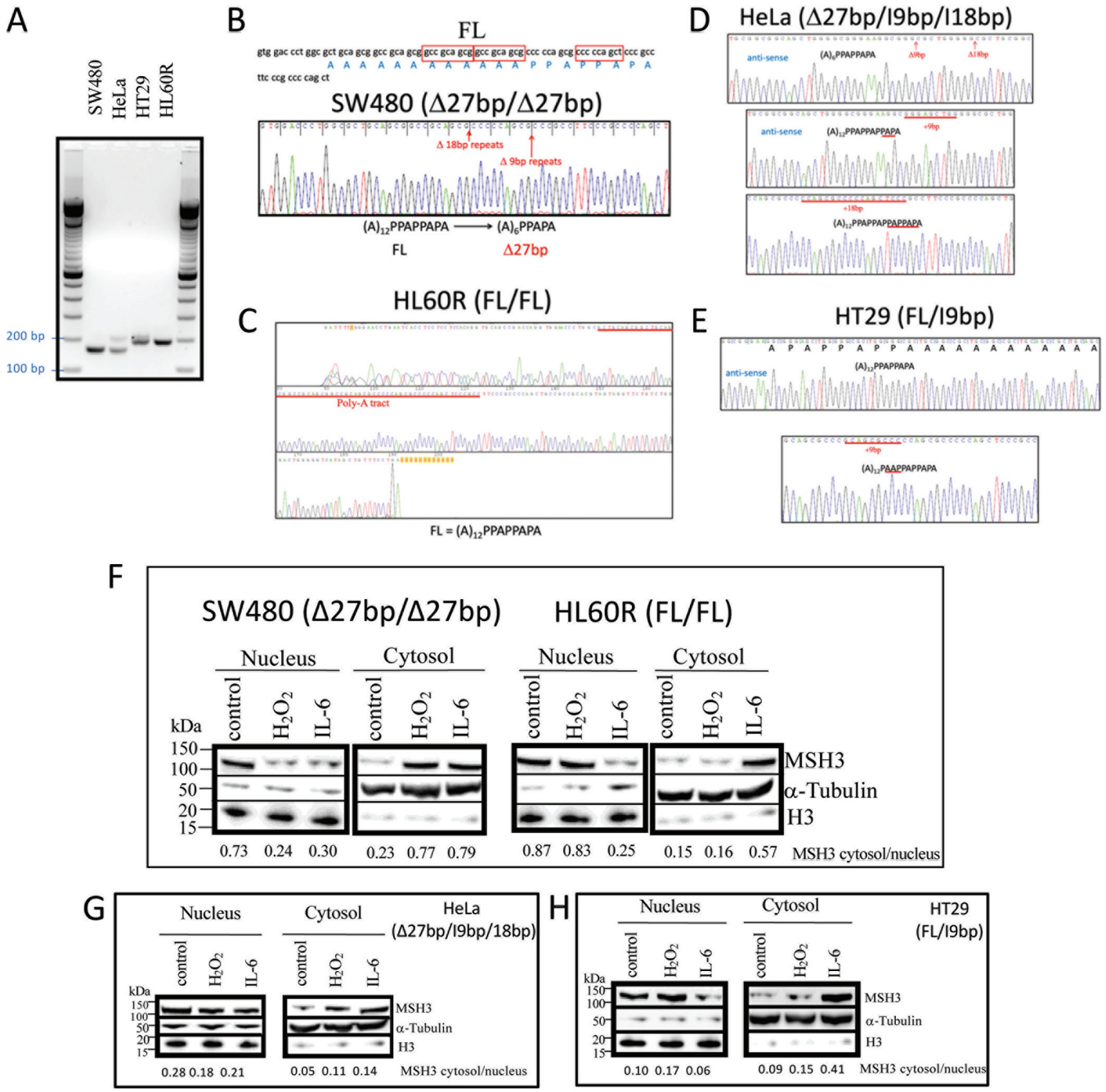

FIG 7 Human cancer cell lines exhibit polymorphisms within the sequence corresponding to the poly-Ala tract region near MSH3 NLS1. (A) Primers were designed to amplify a DNA fragment of 189 bp (FL) or 162 bp (Δ27bp) corresponding to the poly-Ala tract within MSH3 exon 1, allowing genotyping of cancer cells for the polymorphism. (B to E) PCR products amplifying the sequences corresponding to the poly-Ala tract were directly submitted for DNA sequencing for SW480 (B) and HL60R (C). Because of the heterozygous status, PCR products from HeLa (D) and HT29 (E) cells were subjected to TA cloning before DNA sequencing. (F) Nuclear-cytosolic fractionation and Western blotting for native MSH3 in SW480$^{\Delta 27bp/\Delta 27bp}$ demonstrates cytosolic shift with either H_2O_2 or IL-6 treatment, but native MSH3 in HL60R$^{FL/FL}$ shifts only with IL-6 and not H_2O_2, suggesting that the polymorphism may affect MSH3 sublocalization based on the stimulus. (G and H) MSH3 in HeLa cells appeared to shift with both H2O2 and IL-6 treatments (G), while MSH3 in HT29 shifted only with IL-6 treatment (H). Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively.

accessibility of MSH3's NLS1, which may have important clinical implications for patients with MSH3 poly-Ala tract polymorphisms.

Diseased tissues from UC patients contain Δ27bp MSH3 and correlate with EMAST. Based on data from the UCSC database (https://genome.ucsc.edu), the frequency of *Δ27bp MSH3* in a general, healthy population is very low (0.0012%), yet we

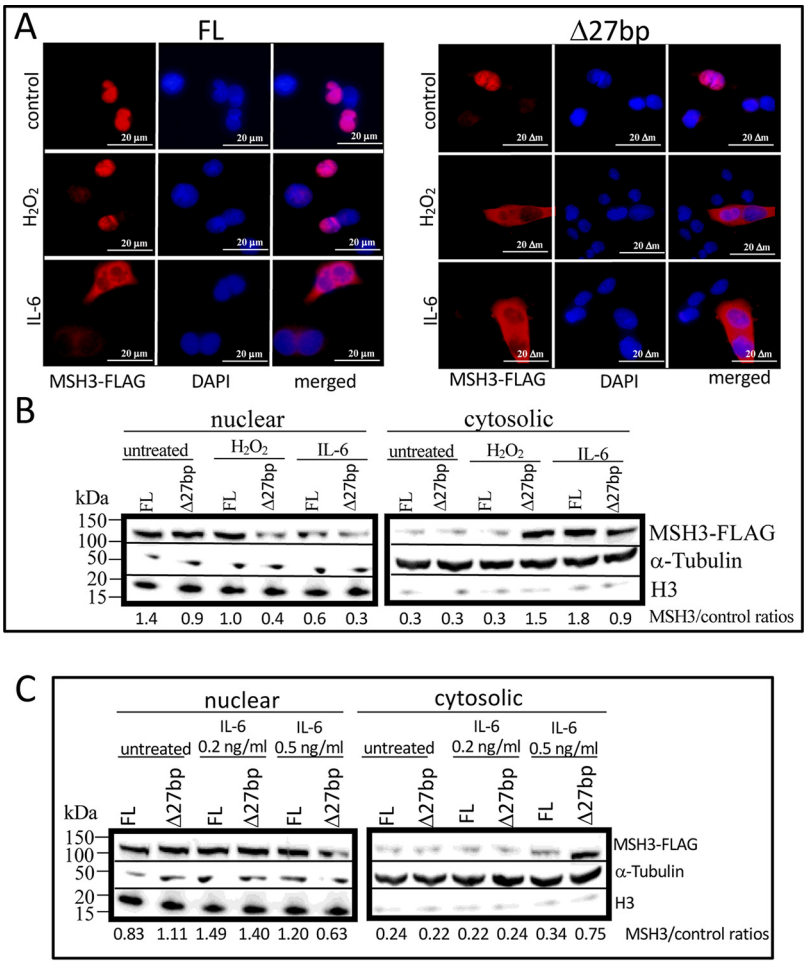

FIG 8 FL and Δ27bp MSH3 respond differentially to H_2O_2 and IL-6 treatment for MSH3 subcellular localization. (A) Immunofluorescence microscopy of SW480 cells containing FL MSH3-FLAG (FL) or Δ27bp MSH3-FLAG (Δ27bp) with and/or without H_2O_2 or IL-6 treatment. (B) Nuclear-cytosolic fractionation and Western blotting of SW480 cells containing FL MSH3-FLAG or Δ27bp MSH3-FLAG with and/or without H_2O_2 or IL-6 treatment. Note that H_2O_2 solely shifted Δ27bp MSH3-FLAG (50 to 70% of transfected cells), and IL-6 shifted both FL MSH3-FLAG (50 to 75% of transfected cells) and Δ27bp MSH3-FLAG (60 to 90% of transfected cells), suggesting that two mechanisms, oxidative stress and IL-6 signaling, can cause Δ27 bp MSH3-FLAG to accumulate in the cytosol. Tubulin and histone H3 proteins were used as cytosolic and nuclear protein markers, respectively. Transfection efficiencies were about 15% in all groups. (C) Δ27bp MSH3 translocates into the cytoplasm more effectively than FL MSH3 upon IL-6 treatment. SW480 cells were transfected with the expression construct expressing FL MSH3-FLAG and/or Δ27bp MSH3-FLAG, followed by treatment with different doses of IL-6. Total cell lysates were fractionated into nuclear and cytosolic fractions to monitor the localization of MSH3. More cytosolic Δ27bp MSH3 was detected upon IL-6 treatment.

identified the Δ*27bp MSH3* polymorphism in 6 of 11 (54.5%) human cancer cell lines we screened (Fig. 10A), raising the possibility that such deletion could be selected for during tumorigenesis. We showed that a significant portion of colon tissue samples from patients with ulcerative colitis (UC) exhibit EMAST (10). Because long-term UC is a risk factor for CRC development and UC tissues demonstrate high levels of IL-6 and oxidative stress, we hypothesized that the Δ*27bp MSH3* polymorphism could be a factor among these patient samples. Eight of 42 (19.0%) UC samples exhibited Δ*27bp MSH3*, 15,833-fold higher than expected based on the general population. Six of eight (75%) Δ*27bp MSH3* samples were EMAST positive, suggesting that this genotype could be more strongly associated with EMAST formation ($P = 0.049$) (Fig. 10B). Repeated inflammatory insults could select for shorter polymorphisms of MSH3's poly-Ala tract to drive EMAST, other DNA damage, and carcinogenesis, as Δ27bp MSH3 is more respon-

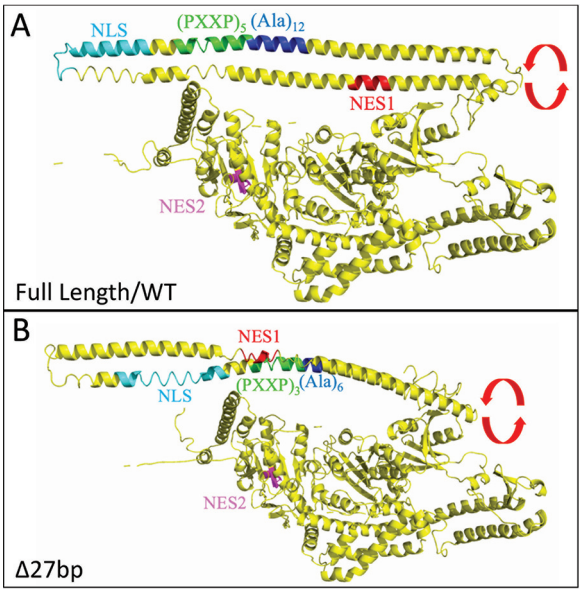

FIG 9 Computer modeling showing the N terminus of the full-length/wild-type MSH3 protein (A) and Δ27bp MSH3 protein (B), fused with the published crystal structure of MSH3. The predicted N-terminal structures can be rotated in the indicated direction (arrows). Note that FL MSH3 has its NLS1 at the tip of a loop that is exposed, whereas Δ27bp MSH3 through shortening of the alanine (Ala) and proline (PXXP) repeats moves the NLS1 to a less accessible location on the side of the N-terminal loop of MSH3. The locations of NES1 and NES2 are also shown.

sive to oxidative stress and inflammation to exit the nucleus and be retained in the cytosol, where it cannot repair DNA.

DISCUSSION

Following our observations that MSH3 could shift from the nucleus into the cytoplasm in cells exposed to elevated oxidative stress and/or inflammation (23, 24), we investigated the possibility that MSH3 is a shuttling protein. Utilizing bioinformatic

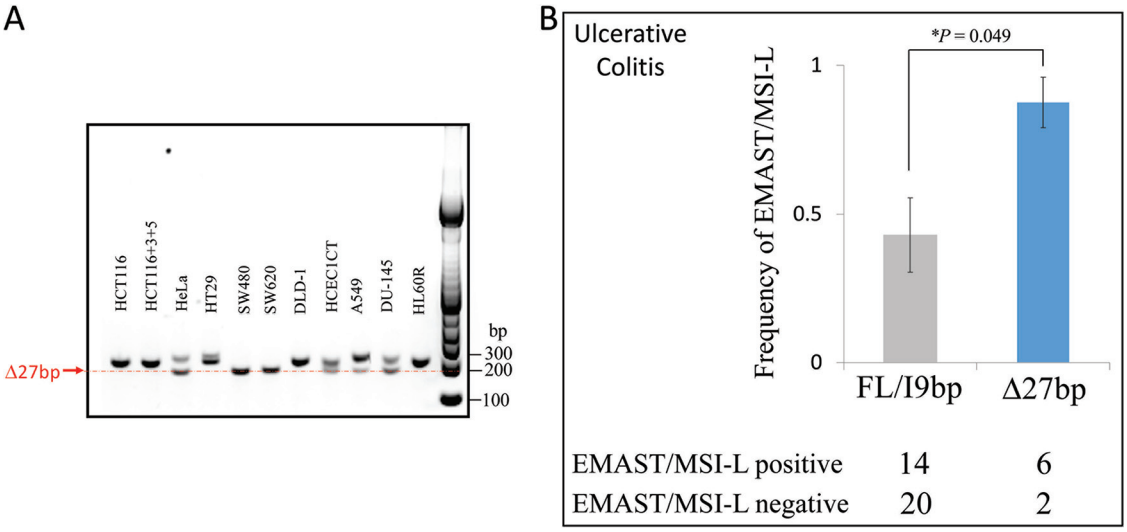

FIG 10 Genotyping of exon 1 MSH3 polymorphism in human cancer cells and ulcerative colitis (UC). (A) PCR products amplified from the poly-A tract sequences within MSH3 exon 1 from 11 human cancer cell lines revealed that this region is highly polymorphic. Six out of 11 lines carry the Δ27bp deletion (54.5%), starkly contrasting with 0.0012% detected in a general population (https://genome.ucsc.edu). (B) Eight of 42 UC samples (19%) genotyped for Δ27bp MSH3, with 6 of 8 (75%) associated with EMAST detected in the UC tissue ($P = 0.049$). The Z-test was utilized for the statistical analysis. *P* value was computed using the 5000 Monte Carlo method (99% confidence interval [CI] is 0.041 to 0.056).

analyses, we identified potential NLSs and NESs within the human MSH3 protein sequence. We first tested these potential signals as individual sequences outside the context of FL MSH3 to assess the ability to direct the import and/or export of a reporter protein. Subsequently, the signals were examined within the context of FL MSH3 to test if the signals are directly responsible for import and/or export of MSH3. Our studies clearly demonstrate the following: (i) as individual independent signals, both NLS1 and NLS2 possess detectable levels of nuclear import function, with NLS1 > NLS2; (ii) within the context of FL MSH3 protein, NLS1 is responsible for the nuclear import of MSH3 (i.e., it is the bona fide MSH3 NLS); (iii) MSH3 NLS1 is a bipartite nuclear localization signal, where the second cluster of K residues is more critical than the first cluster of K/R residues; and (iv) NES1 and NES2 are synergistic for their export function with the proinflammatory cytokine IL-6. With our findings, we conclude that MSH3 exits the nucleus via its NES1 and NES2 signals in response to inflammation (resulting in EMAST formation and lack of DNA repair) (24) and via its NLS1 signal reenters the nucleus, where MSH3 can then perform its DNA repair functions.

At present, there is no known role for MSH3 in the cytosol. MSH3 is a nuclear protein that participates in DNA MMR (preventing EMAST) and assists in homologous recombination repair for DNA double-strand breaks (4, 14, 15). In addition to MSH3, other DNA MMR proteins possess NLSs, including ones directing the nuclear import of MLH1, PMS2 (https://genome.ucsc.edu), and MSH6 (30, 31), that guide nuclear localization after protein synthesis. Signals responsible for MLH1 and PMS2 export have been identified as well (https://genome.ucsc.edu). MLH1 has been shown to be exported via CRM-1/exportin pathway (32). Consequently, it was suspected that MutLα (MLH1/PMS2 heterodimer) shuttles between the nucleus and cytoplasm and that such shuttling behavior may be required to signal DNA damage to the downstream pathways (32). However, MSH3 is the only MMR protein that shuttles to the cytosol in response to oxidative stress or inflammation (23, 24). In this study, we identified the NLS responsible for nuclear localization of MSH3. Not surprisingly, *E. coli* MutS lacks the sequences of the first ~200 amino acids of the human version and does not carry sequences similar to the MSH3 NLS. Interestingly, MSH3 homologs in lower eukaryotic cells, such as yeasts, do not harbor similar sequences either. The MSH3 NLS sequences, however, are present in higher eukaryotic cells and are highly conserved among primates (Table 2). Sequences similar to MSH3 NES1 do not exist in *E. coli* but may be identified in eukaryotic cells, and NES2 sequences may be located even in *E. coli*. These observations suggest that MSH3 NLS and NES1 sequences as well as the rest of its N-terminal sequences were added during evolutionary changes as a more sophisticated regulatory system that might be required to adequately meet the needs of higher eukaryotic cells.

We constructed a working model illustrating MSH3's normal cellular function and its shift in response to oxidative stress and inflammation (Fig. 11). We propose that under normal cellular conditions, the majority of MSH3 protein resides in the nucleus to repair DNA, with only a tiny fraction existing in the cytoplasm. Within a highly inflammatory microenvironment or potentially with high levels of oxidative stress, nuclear MSH3 dissociates from MSH2 and shifts into the cytoplasm, removing the ability to repair DNA lesions and allowing generation of mutations. It is not clear what might cause dissociation of MSH3 from MSH2. One hypothesis is that oxidative stress generated by reactive oxygen species (ROS) leads to posttranslational modification and subsequent MSH3 conformational changes, which causes the dissociation, and we are actively investigating this possibility. Alternatively, an MSH3-interacting protein, whose expression might be induced by IL-6 (or oxidative stress), may bind to MSH3 after MSH3 dissociates from MSH2. Once MSH3 dissociates from MSH2, it may undergo a conformational change reducing NLS1 accessibility but increasing NESs site exposure, resulting in MSH3's exit from the nucleus. As to why oxidative stress translocates MSH3 out of the nucleus, one hypothesis is that more MSH2 could be available to pair with MSH6 (forming MutSα) to repair DNA damaged by ROS. As MSH3 accumulates into the cytoplasm, DNA repair processes that require MSH3 fail in the nucleus, allowing mutations that may ultimately contribute to cancer progression.

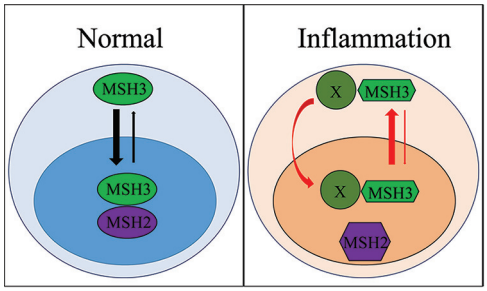

FIG 11 Working model for MSH3 shuttling in normal versus inflammatory conditions. Under normal or noninflammatory conditions (left), a small fraction of MSH3 may exit the nucleus into the cytoplasm but quickly returns to the nucleus. Upon IL-6 treatment (inflammatory condition [right]), MSH2 and MSH3 may undergo conformational changes, leading to disassociation. MSH3 forms a new complex with a potentially new partner(s) (X) to facilitate MSH3's exit from the nucleus into the cytoplasm. Protein(s) X may or may not recycle back to the nucleus, while MSH3 is unable to return to the nucleus in its new conformation, resulting in cytosolic accumulation. MSH3 cytosolic retention can also be an imbalance between nuclear export and import (straight red arrow).

Among human clinical specimens, MSH3 function can be compromised by (i) mutation (typically frameshift mutation at its mononucleotide coding microsatellite as a consequence of *MLH1* hypermethylation and subsequent loss of expression) observed in ~30% of MSI-High CRCs (9, 19) and (ii) cytosolic accumulation due to inflammation and/or oxidative stress, which occurs in 50% of all CRCs (24). Based on the data in this report, we believe that relative changes in the nuclear import/export shuttling of MSH3 are a principal mechanism leading to cytosolic accumulation. We are actively investigating if the MSH3 cytosolic accumulation is a result of altered import/export rates alone or in combination with other mechanisms, such as MSH3 cytosolic sequestration. With cytosolic accumulation of MSH3, inflammation-induced microsatellite alterations or EMAST is observed, with frameshift mutations occurring in genomic di-, tri-, and tetranucleotide microsatellites caused by the absence of MSH3 function (3, 26). There currently is no evidence for methylation of MSH3 as a way to inactivate its expression, and there are only rare reports of missense mutations outside of its coding $(A)_8$ microsatellite, with none involving these newly mapped NLS1 or the NESs (21, 22). Thus, cytosolic accumulation of MSH3 in response to inflammation appears to be the major mechanism for removing MSH3 from the nucleus in CRCs, and understanding how MSH3 shuttles between the nucleus and cytosol becomes important. One remaining question is if dysfunction of NESs will then trap MSH3 in the nucleus, mitigating EMAST formation within inflammatory environment. Since NES1 may be critical for the interaction with MSH2, we are exploring ways to target delete NES2 to examine if MSH3 will remain in the nucleus upon IL-6 treatment. A key finding from our data is the variation of cytosolic accumulation of MSH3 due to polymorphisms within exon 1, located near NLS1. We demonstrate that cells containing a Δ27bp MSH3 protein allow translocation to the cytosol under IL-6 and oxidative stress conditions and that the protein appears to have more difficulty in returning to the nucleus, allowing accumulation in the cytosol. The presence of this rare polymorphism in a clinical setting may have untoward consequences, with easier and perhaps prolonged MSH3 dysfunction. Colorectal cancer cells might self-select out for this polymorphism (~55% of cell lines in our study), allowing the cancer cells more opportunity to accumulate mutations. With our examination of UC tissue that is in an enriched IL-6 microenvironment, Δ27bp MSH3 also seemed self-selected for, with ~15,000 higher than expected proportion than for the general population (https://genome.ucsc.edu). Although additional studies are needed, we suggest that compromised function of MSH3's NLS1 could have profound clinical effects that may be important in carcinogenesis.

Recently, Watson et al. reported that EMAST is not associated with MSH3 nuclear expression loss among stage I to III colon cancers, based on the data they collected from 152 patient samples (33). There are major problems with how the data were

interpreted in that report. The first and largest issue is that 42 of 50 of their EMAST-positive samples are also MSI-H, which means that the EMAST in these 42 samples were due to MLH1 (or MSH2 or PMS2) deficiency. The fact that the MSI mononucleotide repeats markers used in their study showed frameshifts is indicative that the characterization for EMAST in these samples are not the results of sole MSH3 deficiency. Studies using MSH3-deficient cells in which MSH3 expression was restored via chromosome complementation (6, 10) and/or using MSH3-proficient cells in which MSH3 was knocked down by small hairpin RNA (shRNA) (25) clearly showed that loss of MSH3 leads to EMAST and MSI-L (specifically among dinucleotide mutations but excluding mononucleotide frameshifts). Recently, our laboratory has generated a MSH3 knockout cell line which, as expected, exhibited frameshift mutations at di- and/or tetranucleotide microsatellite repeats, with no mutations detected for mononucleotide repeats (unpublished data). Consistent with our results, tumors from patients with biallelic MSH3 germ line mutations (complete loss of nuclear MSH3 in normal and tumor tissues) are indeed EMAST positive (20). All of these results conclusively demonstrate that MSH3 deficiency is the driver for EMAST formation. Among sporadic CRCs, the expectation of near correlation between loss of MSH3 nuclear expression with EMAST would not be perfect. Sensitive PCR-based tests for frameshift detection at microsatellites including tetranucleotides will demonstrate the results of common and/or specific DNA MMR protein deficiencies. Specifically, an EMAST-positive tumor (without mononucleotide frameshifts) only demonstrates that host cells have experienced MSH3 deficiency, but it does not indicate when it happened during the tumor's pathogenesis. As we reported, MSH3 may exit the nucleus under high oxidative stress and/or inflammation, leading to EMAST formation. However, this is a reversible process, which means MSH3 can return into the nucleus when the stress or inflammation has dissipated (24), with cells acquiring and maintaining di- and tetranucleotide frameshifts. Consequently, it is critically important to eliminate MSI-H samples from studies aiming to investigate inflammation-associated microsatellite alterations or EMAST. MSH3 immunohistochemistry (IHC) is not as sensitive as PCR-based microsatellite assays, with a certain percentage of cells showing loss of nuclear MSH3 (10). It is, however, imperative to bear in mind that EMAST analysis reveals if cells have ever experienced MSH3 deficiency, while MSH3 IHC shows the MSH3 expression at the time when the sample was harvested.

In our experience, a small cytosolic pool of MSH3 is consistently observed. It is not clear as to why MSH3 evolved to shuttle into the cytoplasm. We subjected the MSH3 NLS1 sequences to the BLAST (Basic Local Alignment Search Tool) program against the human genome database, searching for proteins that share the identical or a similar NLS. The search yielded two proteins, NF90 and ILF3, which are encoded by variants of the same gene (Fig. 12A). Interestingly, both are nuclear proteins that shuttle between the nucleus and the cytoplasm. Both are involved in multiple aspects of RNA metabolism, including splicing, transport, stability, and translation through direct interaction with a targeted RNA via their double-stranded RNA binding domains (DRBDs). Although their NESs do not share similarity with those of MSH3, the overall anatomies of NF90/ILF3 and MSH3 are intriguingly alike (Fig. 12B). Specifically, the sequence of the NES (NES1 for MSH3) overlaps with those required to interact with their partners in the nucleus (NF45/ILF2 for NF90/ILF3 and MSH2 for MSH3), pointing to a scenario in which their NESs are concealed via the interaction with their partner binding proteins to ensure nuclear localization. NF90 has been reported to shift into the cytoplasm from the nucleus during cell activation and/or coexpression with HIV Rev protein, presumably to stabilize target RNA and assist in retroviral gene expression and replication (34, 35). Considering the fact that MSH3 is able to bind DNA and shifts into cytoplasm upon IL-6 treatment, it is interesting to speculate that MSH3 may shuttle into the cytoplasm as a part of cellular defense mechanisms to detect invading pathogens that contain DNA. Further studies are ongoing to help identify if this is a possibility.

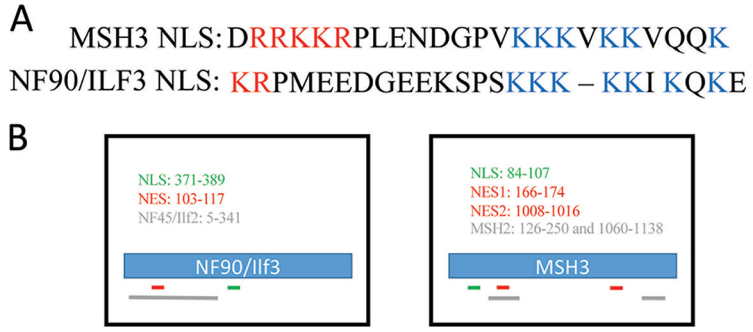

FIG 12 Homology between MSH3 and NF90/ILF3 NLSs. MSH3 NLS sequences was subjected to the BLAST program against the entire human genome sequences to search for proteins containing similar sequences. The proteins with the highest score were NF90/ILF3, which are shuttling proteins involved in many aspects of RNA metabolism. (A) Sequences of MSH3 and NF90/ILF3 NLSs. (B) In addition to the similarity in their NLS sequences, the arrangement of their protein signals critical for shuttling behavior resemble each other, specifically the overlapping of NES with the sequences interacting with their nuclear partners (green bar, location of the NLS; red bar, location of the NES; gray bar, location[s] of the sequences critical for interacting with the partner).

MATERIALS AND METHODS

Cell lines and reagents. The human colon cancer (SW480 and HCT116), cervical cancer (HeLa), and promyelocytic leukemia (HL60R) cell lines were purchased from the American Type Culture Collection (ATCC; Manassas, VA). SW480 cells were cultured as described previously (24). HCT116, HCT116+Ch3, HCT116+Ch3+Ch5, HeLa, and HL60R cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1× penicillin-streptomycin (pen-strep), and 50 μg/ml of Geneticin (Invitrogen). Mouse anti-MSH3 was purchased from BD Biosciences (San Jose, CA). Rabbit anti-FLAG and anti-histone 3 antibodies were from Cell Signaling (Danvers, MA). Mouse anti-GFP antibody (clone GF28R) was purchased from Invitrogen. Anti-α-tubulin antibody was purchased from Sigma. Alexa Fluor 488-conjugated anti-mouse antibody and Alexa Fluor 589-conjugated anti-rabbit antibody were purchased from Invitrogen.

Identification of putative MSH3 NLSs and NESs. The human MSH3 (NCBI reference sequence: NM_002439.4) amino acid sequence was submitted to an NLS mapper program (http://nls-mapper.iab.keio.ac.jp), requesting the program to identify all potential NLSs. The three putative NLSs with the highest score were studied in this investigation (named NLS1, NLS2, and NLS3 directionally from N to C terminus). The MSH3 sequence was also subjected to an NES mapper program (http://www.cbs.dtu.dk/services/NetNES/), allowing us to identify two putative NESs (NES1 and NES2 directionally from N to C terminus).

Generation of reporter constructs containing MSH3 NLS and/or NES for nuclear import and/or export assays. Primers containing sequences of the predicted MSH3 NLSs and/or NES were synthesized (Table 1). Forward primers were designed to carry a PstI site 5′ overhang at the 5′ end and reverse primers a 3′ overhang of the BamHI site at the 3′ end. Equal amounts of the paired phosphorylated forward and reverse primers were combined, heated at 100°C for 5 min, and allowed to cool down gradually at room temperature for 2 h for primer annealing. Two microliters of the annealed primers was transferred into a new tube containing 198 μl of double-distilled water (d_2H_2O) to dilute the annealed primers to a final concentration of 1 μM, which was used to set up ligation reactions with pEGFP-N1 (Clontech; Mountain View, CA) for NLSs and other variants of NLS1. pRev1.4-(NES3)-GFP was used for NESs (a gift from Beric R. Henderson, University of Sydney). There are six random nucleotides between NES1 and NES2 sequences on the NES1-NES2-EGFP construct. An additional control, in which NES3 sequences (amino acid sequence QELNEEQAISHLHLQLVK) were replaced by random sequences of the same length (amino acid sequence ARFNHQAAIFQVKKANRL), was also included in the study. Ten to 20 colonies were subjected to DNA sequencing to screen for positive clones.

Generation of the MSH3-FLAG-expressing construct and its variations. A DNA fragment containing ~3.1 kb of human MSH3 cDNA from the 5′ end was removed from a commercially available human MSH3 cDNA-containing clone (OriGene Technologies, Inc., Rockville, MD) and subcloned onto pcDNA3.1 (Invitrogen) via BamHI sites. Approximately 0.9 kb of DNA corresponding to the 3′ end of MSH3 cDNA fragment was then amplified via PCR, which allowed us to tag MSH3 with FLAG and add a stop codon after the FLAG tag via the BsrGI and XhoI sites (forward, 5′-CTGCAGACCAACTGTACAAGAAGAAAGA-3′; reverse, 5′-CTAAGTCTCGAGTTATTTGTCGTCGTCTTTGTAGTCATGAAGAAGAGAAGTCTGTGTTTCTTCCAT-3′). The created MSH3-FLAG construct was sequenced to ensure that the entire MSH3 cDNA sequence was correct.

A Transformer site-directed mutagenesis kit (Clontech) was used to delete the sequences corresponding to the NLSs and/or NESs within the full-length MSH3-FLAG following manufacturer's instructions. The sequences for mutagenesis primers are summarized in Table 1. To make the construct expressing MSH3Δ27bp-FLAG, the first ~1.8-kb DNA fragment of MSH3 was amplified via reverse transcription-PCR (RT-PCR) using SW480 RNA, subcloned onto pGEM-T vector by TA cloning, and sequenced. An ~1.3-kb XcmI-HpaI fragment containing the 27-bp deletion was extracted to swap with the wild-type sequences to yield MSH3Δ27bp-FLAG-expressing construct.

Delivery of the synthesized constructs. The constructs were electroporated into the desired cell lines using Amaxa Cell Line Nucleofector kit V (Lanza; Cologne, Germany) following the manufacturer's instruction. Cells were allowed to recover for 24 to 48 h before further assessment. HCT116 and HCT116+Ch3 cells were the host cells to test the import/export function of the signals. SW480 and HCT116+Ch3+Ch5 cells were used to examine the import and/or export of the signals within the context of MSH3 proteins.

IFM. Immunofluorescence microscopy (IFM) was performed as previously described (24). For H_2O_2/IL-6 treatment, 1 million cells were subjected to electroporation before being seeded onto 4-chamber slides (Invitrogen) and incubated at 37°C and 5% CO_2 overnight, followed by serum starvation for 24 h before H_2O_2 (100 μM; Sigma-Aldrich) and/or IL-6 treatment (1 ng/ml, except for Fig. 8C [0.2 and 0.5 ng/ml]; R & D Systems, Minneapolis, MN) for 16 to 18 h at 37°C and 5% CO_2. Pictures were taken using an Olympus fluorescence microscope.

Whole-cell lysate fractionation and WB. Two hundred fifty thousand cells were seeded per well onto 6-well plates and incubated at 37°C and 5% CO_2 overnight. After IL-6 treatment (see above), whole-cell lysates were prepared from equal numbers of cells, fractionated into nuclear and cytosolic fractions, and then subjected to Western blotting (WB). When comparing the amount of the protein of interest in the nucleus versus cytoplasm, the blots containing proteins from both compartments were probed in the same container with each individual antibody and subjected to the same exposure time. Detection of the protein of interest by WB was performed as described previously (24). The intensity of the band was quantitated using ImageQuant LAS 4000 (GE Healthcare, Chicago, IL). To calculate the ratio of the protein of interest over the housekeeping protein, the band intensity of the protein of interest was divided by the band intensity of the housekeeping protein (H3 for nuclear fraction and tubulin for cytosolic fraction). In Fig. 1, for example, the band intensities for nuclear EGFP and H3 in the NLS1 full mutant were 4,176,221 and 96,341,720, separately. The ratio was calculated by dividing 4,176,221 by 96,341,720, which yielded 0.043348 (rounded to 0.04). For the cytosolic fraction of NLS1 full mutant, the band intensities for the cytosolic EGFP and tubulin were 5,032,710 and 89,831,170, respectively. The ratio of 0.056024 (rounded to 0.06) was obtained by dividing 5,032,710 by 89,831,170. All of the fractionation and WB experiments were repeated at least three times, leading to the same conclusions for each repeat experiment.

Examining MSH3 exon 1 polymorphism. Genomic DNA obtained from cancer cells or paraffin-embedded samples were used as templates for PCR. Forward primer (5′-TGAGCCGATTCTTCCAGTCTAC GGG-3′) and reverse primer (5′-CCCAGTCCCAGACAGAACCTACTA) were used for PCR (40 cycles). PCR products from cell lines were either directly sequenced or subjected to TA cloning before sequencing. 6-Fluorescein amidite (FAM)-labeled forward primers were used for fragment analysis for DNA from patient samples. PCR products from patient samples were sequenced to confirm that they were consistent with MSH3 exon 1 sequences. Patient sample use was approved by the University of Michigan Medical School institutional review board (IRB) and abided by the Declaration of Helsinki principles.

Computer modeling of wild-type and Δ27bp MSH3 proteins. Structure models of the wild-type and Δ27bp MSH3 N-terminal domains were predicted on the I-TASSER server (36). I-TASSER utilizes LOMETS (37) to search the query protein sequence against the PDB database by aligning the query sequence to the template structures to extract continuous fragments. These fragments are assembled into a full-length structure by a replica-exchange Monte Carlo (REMC) simulation guided by a composite force field consisting of template-derived distance restraints and knowledge-based energy terms calculated based on statistics of PDB structures. The REMC simulation produces tens of thousands of "decoy" conformations, which are clustered by pairwise structure similarity (38). The centroid of the largest cluster is refined at the atomic level (39) to obtain the final I-TASSER model. The final model quality is estimated by the predicted TM score. In this study, the estimated TM scores for the wild-type and Δ27bp MSH3 N-terminal domains were 0.56 and 0.53, respectively, which indicates they are reliable models (we view a TM score of $>$0.5 as a successful model) (40). Then, the wild-type and Δ27bp models for the N-terminal domains, together with the experimental structure of the MSH3 C-terminal domain (PDB ID 3THW), were assembled to obtain the full-length models on the AIDA server (41).

ACKNOWLEDGMENTS

We thank the DNA Sequencing Core at the University of Michigan Medical School for their assistance.

Funding was provided by the National Institutes of Health (DK067287, CA162147, and CA206010) and the A. Alfred Taubman Medical Research Institute of the University of Michigan.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

All authors have declared that no conflict of interest exists.

Author contributions are as follows: conceived and designed experiments, S.S.T.-R., M.K., and J.M.C.; performed experiments, S.S.T.-R., D.Y.C., K.M., P.K.M., and S.M.; analyzed data, S.S.T.-R., K.M., M.K., W.Z., Y.Z., and J.M.C.; contributed reagents, materials, and analysis tools, J.M.C.; and wrote and edited the manuscript, S.S.T.-R. and J.M.C.

REFERENCES

1. Boland CR, Goel A. 2010. Microsatellite instability in colorectal cancer. Gastroenterology 138:2073–2087. https://doi.org/10.1053/j.gastro.2009.12.064.
2. Grady WM, Carethers JM. 2008. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 135:1079–1099. https://doi.org/10.1053/j.gastro.2008.07.076.
3. Carethers JM. 2017. Microsatellite instability pathway and EMAST in colorectal cancer. Curr Colorectal Cancer Rep 13:73–80. https://doi.org/10.1007/s11888-017-0352-y.
4. Carethers JM, Koi M, Tseng-Rogenski S. 2015. EMAST is a form of microsatellite instability that is initiated by inflammation and modulates colorectal cancer progression. Genes (Basel) 6:185–205. https://doi.org/10.3390/genes6020185.
5. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. 1998. 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 58:5248–5257.
6. Haugen AC, Goel A, Yamada K, Marra G, Nguyen TP, Nagasaka T, Kanazawa S, Koike J, Kikuchi Y, Zhong X, Arita M, Shibuya K, Oshimura M, Hemmi H, Boland CR, Koi M. 2008. Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. Cancer Res 68:8465–8472. https://doi.org/10.1158/0008-5472.CAN-08-0002.
7. Watson MM, Berg M, Soreide K. 2014. Prevalence and implications of elevated microsatellite alterations at selected tetranucleotides in cancer. Br J Cancer 111:823–827. https://doi.org/10.1038/bjc.2014.167.
8. Devaraj B, Lee A, Cabrera BL, Miyai K, Luo L, Ramamoorthy S, Keku T, Sandler RS, McGuire K, Carethers JM. 2010. Relationship of EMAST and microsatellite instability among patients with rectal cancer. J Gastrointest Surg 14:1521–1528. https://doi.org/10.1007/s11605-010-1340-6.
9. Carethers JM, Jung BH. 2015. Genetics and genetic biomarkers in sporadic colorectal cancer. Gastroenterology 149:1177–1190. https://doi.org/10.1053/j.gastro.2015.06.047.
10. Munakata K, Koi M, Kitajima T, Tseng-Rogenski S, Uemura M, Matsuno H, Kawai K, Sekido Y, Mizushima T, Toiyama Y, Yamada T, Mano M, Mita E, Kusunoki M, Mori M, Carethers JM. 2019. Inflammation-associated microsatellite alterations caused by MSH3 dysfunction are prevalent in ulcerative colitis and increase with neoplastic advancement. Clin Transl Gastroenterol 10:e00105. https://doi.org/10.14309/ctg.0000000000000105.
11. Raeker MO, Pierre-Charles J, Carethers JM. 2020. Tetranucleotide microsatellite mutational behavior assessed in real time: implications for future microsatellite panels. Cell Mol Gastroenterol Hepatol 9:689–704. https://doi.org/10.1016/j.jcmgh.2020.01.006.
12. Hile SE, Shabashev S, Eckert KA. 2013. Tumor-specific microsatellite instability: do distinct mechanisms underline the MSI-L and EMAST phenotypes? Mutat Res 743–744:67–77. https://doi.org/10.1016/j.mrfmmm.2012.11.003.
13. Garcia M, Choi C, Kim HR, Daoud Y, Toiyama Y, Takahashi M, Goel A, Boland CR, Koi M. 2012. Association between recurrent metastasis from stage II and III primary colorectal tumors and moderate microsatellite instability. Gastroenterology 143:48–50. https://doi.org/10.1053/j.gastro.2012.03.034.
14. Van Oers JM, Edwards Y, Chahwan R, Zhang W, Smith C, Pechuan X, Schaetzlein S, Jin B, Wang Y, Bergman A, Scharff MD, Edelmann W. 2014. The MutSβ complex is a modulator of p53-driven tumorigenesis through its functions in both DNA double-strand break repair and mismatch repair. Oncogene 33:3939–3946. https://doi.org/10.1038/onc.2013.365.
15. Dietlein F, Thelen L, Jokic M, Jachimowicz RD, Ivan L, Knittel G, Leeser U, van Oers J, Edelmann W, Heukamp LC, Reinhardt HC. 2014. A functional cancer genomics screen identifies a druggable synthetic lethal interaction between MSH3 and PRKDC. Cancer Discov 4:592–605. https://doi.org/10.1158/2159-8290.CD-13-0907.
16. Koi M, Garcia M, Choi C, Kim H-R, Koike J, Hemmi H, Nagasaka T, Okugawa Y, Toiyama Y, Kitajima T, Imaoka H, Kusunoki M, Chen Y-H, Mukherjee B, Boland CR, Carethers JM. 2016. Microsatellite alterations with allelic loss on 9p24.2 signify less aggressive colorectal cancer metastasis. Gastroenterol 150:944–955. https://doi.org/10.1053/j.gastro.2015.12.032.
17. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, Pommier Y, Takeda S, Dutta A. 2015. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. Cell Rep 11:1749–1759. https://doi.org/10.1016/j.celrep.2015.05.020.
18. Carethers JM, Stoffel EM. 2015. Lynch syndrome and Lynch syndrome mimics: the growing complex landscape of hereditary colon cancer. World J Gastroenterol 21:9253–9261. https://doi.org/10.3748/wjg.v21.i31.9253.
19. Carethers JM. 2014. Differentiating Lynch-like from Lynch syndrome. Gastroenterology 146:602–604. https://doi.org/10.1053/j.gastro.2014.01.041.
20. Adam R, Spier I, Zhao B, Kloth M, Marquez J, Hinrichsen I, Kirfel J, Tafazzoli A, Horpaopan S, Uhlhaas S, Stienen D, Friedrichs N, Altmüller J, Laner A, Holzapfel S, Peters S, Kayser K, Thiele H, Holinski-Feder E, Marra G, Kristiansen G, Nöthen MM, Büttner R, Möslein G, Betz RC, Brieger A, Lifton RP, Aretz S. 2016. Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis. Am J Hum Genet 99:337–351. https://doi.org/10.1016/j.ajhg.2016.06.015.
21. Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–337. https://doi.org/10.1038/nature11252.
22. Carethers JM. 2016. Hereditary, sporadic and metastatic colorectal cancers are commonly driven by specific spectrums of defective DNA mismatch repair components. Trans Am Clin Climatol Assoc 127:81–97.
23. Tseng-Rogenski S, Chung H, Wilk MB, Zhang S, Iwaizumi M, Carethers JM. 2012. Oxidative stress induces nuclear-to-cytosolic shift of MSH3, a potential mechanism for EMAST in colorectal cancer cells. PLoS One 7:e50616. https://doi.org/10.1371/journal.pone.0050616.
24. Tseng-Rogenski S, Hamaya Y, Choi DY, Carethers JM. 2015. Interleukin-6 alters localization of hMSH3, leading to DNA mismatch repair defects in colorectal cancer cells. Gastroenterology 148:579–589. https://doi.org/10.1053/j.gastro.2014.11.027.
25. Campregher C, Schmid G, Ferk F, Knasmüller S, Khare V, Kortüm B, Dammann K, Lang M, Scharl T, Spittler A, Roig AI, Shay JW, Gerner C, Gasche C. 2012. MSH3-deficiency initiates EMAST without oncogenic transformation of human colon epithelial cells. PLoS One 7:e50541. https://doi.org/10.1371/journal.pone.0050541.
26. Koi M, Tseng-Rogenski SS, Carethers JM. 2018. Inflammation-associated microsatellite alterations: mechanisms and significance in the prognosis of patients with colorectal cancer. World J Gastrointest Oncol 10:1–14. https://doi.org/10.4251/wjgo.v10.i1.1.
27. Lee S-Y, Chung H, Devaraj B, Iwaizumi M, Han HS, Hwang D-Y, Seong MK, Jung BH, Carethers JM. 2010. Elevated microsatellite alterations at selected tetranucleotide repeats are associated with morphologies of colorectal neoplasia. Gastroenterology 139:1519–1525. https://doi.org/10.1053/j.gastro.2010.08.001.
28. Hamaya Y, Guarinos C, Tseng-Rogenski SS, Iwaizumi M, Das R, Jover R, Castells A, Llor X, Andreu M, Carethers JM. 2015. Efficacy of 5-fluorouracil adjuvant therapy for patients with EMAST-positive stage II/III colorectal cancers. PLoS One 10:e0127591. https://doi.org/10.1371/journal.pone.0127591.
29. Amiel J, Trochet D, Clement-Ziza M, Munnich A, Lyonnet S. 2004. Polyalanine expansions in humans. Hum Mol Genet 13:R235–R243. https://doi.org/10.1093/hmg/ddh251.
30. Wu X, Platt JL, Cascalho M. 2003. Dimerization of MLH1 and PMS2 limits nuclear localization of MutSalpha. Mol Cell Biol 23:3320–3328. https://doi.org/10.1128/mcb.23.9.3320-3328.2003.
31. Gassman NR, Clodfelter JE, McCauley AK, Bonin K, Salsbury FR, Scarpinato KD. 2011. Cooperative nuclear localization sequences land a novel role to the N-terminal region of MSH6. PLoS One 6:e17907. https://doi.org/10.1371/journal.pone.0017907.
32. Brieger A, Adam R, Passmann S, Plotz G, Zeuzem S, Trojan J. 2011. A CRM1-dependent export pathway is involved in the regulation of MutLα subcellular localization. Genes Chromosomes Cancer 50:59–70. https://doi.org/10.1002/gcc.20832.
33. Watson MM, Lea D, Hagland HR, Søreide K. 2019. Elevated microsatellite alterations at selected tetranucleotides (EMAST) is not attributed to MSH3 loss in stage I-III colon cancer: an automated, digitalized assessment by immunohistochemistry of whole slides and hot spots. Transl Oncol 12:1583–1588. https://doi.org/10.1016/j.tranon.2019.08.009.
34. Sandrine C, Rozenn B, Mélanie C, Aurélie F, Jean-Christophe L. 2015.

Ilf3 and NF90 functions in RNA biology. Wiley Interdiscip Rev RNA 6:243–256.
35. Li Y, Belshan M. 2016. NF45 and NF90 bind HIV RNA and modulate HIV gene expression. Viruses 8:47–63. https://doi.org/10.3390/v8020047.
36. Zheng W, Zhang C, Bell EW, Zhang Y. 2019. I-TASSER gateway: a protein structure and function prediction server powered by XSEDE. Future Gener Comput Syst 99:73–85. https://doi.org/10.1016/j.future.2019.04.011.
37. Zheng W, Zhang C, Wuyun Q, Pearce R, Li Y, Zhang Y. 2019. LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. Nucleic Acids Res 47:W429–W436. https://doi.org/10.1093/nar/gkz384.
38. Zhang Y, Skolnick J. 2004. SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 25:865–871. https://doi.org/10.1002/jcc.20011.
39. Zhang J, Liang Y, Zhang Y. 2011. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. Structure 19:1784–1795. https://doi.org/10.1016/j.str.2011.09.022.
40. Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26:889–895. https://doi.org/10.1093/bioinformatics/btq066.
41. Xu D, Jaroszewski L, Li Z, Godzik A. 2015. AIDA: ab initio domain assembly for automated multi-domain protein structure prediction and domain-domain interaction prediction. Bioinformatics 31:2098–2105. https://doi.org/10.1093/bioinformatics/btv092.