Microsatellite instability as a driving force for cancer progression

Nives Pećina-Šlaus 1,2, Anja Bukovac 1,2, Iva Salamon 3, Anja Kafka 1,2

1 Laboratory of Neuro-oncology, Croatian Institute for Brain Research, School of Medicine University of Zagreb, Salata 12, Zagreb, Croatia
2 Department of Biology, School of Medicine, University of Zagreb, Salata 3, Zagreb, Croatia
3 University of Zagreb School of Medicine, Croatian Institute for Brain Research, Laboratory for Stem Cells, Šalata 12, HR-10000 Zagreb, Croatia

*Corresponding author:
Nives Pećina-Šlaus
Laboratory of Neurooncology
Croatian Institute for Brain Research
School of Medicine, University of Zagreb
Salata 12, HR-10000 Zagreb, Croatia
Phone: +385 1 46 21 140
Fax:+385 1 45 90199; +385 1 49 20 050
Email: nina@mef.hr

Hypothesis

Abstract
Impaired cellular DNA repair mechanisms are greatly involved in cancer initiation and progression. In the present paper we would like to draw attention to microsatellite instability (MSI) - a part of genomic instability that characterizes cancer cell and to propose its role in cancer initiation and the promotion of cancer progression. We believe that malfunctioning of mismatch DNA repair (MMR) plays an important role in cancer, sometimes as the main driving force and sometimes as an influential bystander. Appearance of MSI as a consequence of MMR deficiency has many potential clinical implications. It can be used for assessment of cancer risk, precision grading, and prognosis as
well as for the explanation why some cancers become resistant to chemotherapy.

**Keywords:** microsatellite instability, genomic instability, cancer, MSI, MMR

**Introduction**

Cancer is caused by mutations in our genome and can be considered a genetic disorder in which the normal control of cell growth is impaired. Genetic basis of cancer are overwhelming because they include changes of genes intrinsic to a variety of vital cellular processes like proliferation, apoptosis, differentiation, angiogenesis, cell motility as well as immune system. Finally, one must not forget that impaired cellular DNA repair mechanisms are also very much involved in cancer initiation and progression. Molecular and genetic basis of tumor development although still not completely understood, are nowadays explained in many aspects. We understand now that cancer is not a single disease but rather a collection of diseases with specific genetic profiles. Furthermore, we know that malfunctioning of several and specific signal transduction pathways are responsible for tumor formation and development. Novel scientific papers, especially those that bring results on system biology of tumors and large-scale analyses, often refer to the collection of the observed genetic changes as cancer genome landscapes (Vogelstein et al., 2013). Cancer genetics or better said genomics is now one of the fastest expanding medical specialties today and vast knowledge has been gathered in order to develop novel and efficient diagnostic and therapeutic methods.

Since a lot of hard work and data mining lies behind any scientific discovery, results obtained by hypothesis-free research are in our opinion as much important and valuable as results of studies based on sound and inventive hypotheses. Nevertheless, today maybe more than ever, the hypotheses in cancer research are needed to direct large-scale analyses but also to correctly interpret the abundant data.

Here we draw attention to microsatellite instability - a part of genomic instability that characterizes cancer cell - and propose a role for it in cancer initiation and the promotion of progression. We believe that malfunctioning of MMR repair plays an important role in cancer, sometimes as the main driving force and sometimes as an influential bystander.

**Microsatellite Instability**

Repetitive DNA sequences are the intrinsic part of our genome, polymorphic by their nature (Ionov et al., 1993; Payseur et al., 2011) and are called microsatellites (MS). They are short repeated sequences, also known as simple sequence repeats (SSRs), which are highly variable in base-pair lengths stretching from mononucleotides to even hexanucleotides, with a variable tendency of being repeated usually between 5 and 50 times. Very common microsatellites are tandem repeats consisting of dinucleotides and therefore also called short tandem repeats (STRs). The most
frequently occurring repeats throughout our genome are mononucleotide repeats; these are the ones with the highest frequency of microsatellite instability (Payseur et al., 2011; Watson et al., 2014). In addition, microsatellites are omnipresent in both prokaryotic and eukaryotic genomes with attributed tendency of having higher rates of mutation compared to other parts of the genome. Suggested mechanisms that are causing high abundance of microsatellites across the genome are DNA polymerase slippage during replication, unequal crossing overs and deficient repair processes (Oliveira et al., 2006). From all the above, one can conclude that microsatellites are indeed contributing to each individual’s DNA fingerprint and are thus used as polymorphic genetic markers. The length alterations within microsatellites are a reflection of deficient DNA mismatch repair system (MMR) (Schmidt and Pearson, 2016). This type of genomic instability that is characteristic of tumor cells is termed microsatellite instability (MSI). Since DNA synthesis is known to be an error-prone process resulting in spontaneous base-base mismatches and insertion-deletion DNA loops, the evolution ensured maintenance of genomic stability by the introduction of MMR system (Kunkel, 2009). Since genetic errors may pass unnoticed by the proofreading activity of DNA polymerase, MMR maintains genome stability mainly by correcting base-base and small insertion-deletion mispairs generated during DNA replication. Therefore, MMR genes acted in this way as tumor suppressor genes. In sporadic, non-heritable cancers (colorectal, ovarian, gastric, endometrial or prostate cancer), MSI arises either as a result of somatic mutation that causes inactivation of one of the MMR’s genes, epigenetic silencing of MMR gene expression, or downregulation of MMR genes by microRNAs (Gelsomino et al., 2016). Hypermethylation of the MLH1 promoter is very often associated with MSI in many human cancers. In hereditary cases, the causes of ineffective MMR system are usually germline mutations in one of the mismatch repair genes. Although, MSI might work as an oncogenic driver, it is noteworthy that not all microsatellite sequences are prone to display MSI phenotype. The distribution of microsatellites throughout the genome is not evenly divided (Tóth et al., 2000; Oliveira et al., 2006). Coding and non-coding regions differ significantly in their microsatellite distribution. Coding regions, transcription factor binding sites, and evolutionarily conserved genomic regions (CpG islands), comprise a lower number and less unstable microsatellites than intragenic regions, introns and splice sites. This verification is logical, as coding and conserved regions would otherwise be significantly altered due to the high microsatellite mutation rates and consequent introduction of MSI, which would lead to disturbances or complete loss of protein function. To the contrary, microsatellite spots within introns or untranslated regions are involved in modulating gene expression or influencing transcription and gene splicing. Exceptions are trinucleotide microsatellites with repeats in multiples of 3, which can evenly be found in both regions. The selection pressure against reading-frame mutations in a gene limits the frequency of microsatellites in coding regions.
However, trinucleotide repeats and their expansion in the coding region will not change the reading frame. Especially interesting microsatellites (simple sequence repeats) are trinucleotide repeats because of their role in several human neurodegenerative disorders, for example, fragile X syndrome and Huntington’s disease. Hence, the expansions of trinucleotide repeats are responsible for these genetic conditions.

Regarding the MSI frequency, the general observation is that intrinsic length and repeat composition of microsatellites are proportional with MSI frequency and might explain inheritance of MSI at particular locus. Furthermore, the frequency and type of genomic microsatellites are taxon-dependent, which also varies per species. Thus human microsatellites contain many more repeats compared to chimpanzees (Oliveira et al., 2006).

Recently a new phenomenon in genomic instability field was reported regarding elevated microsatellite alterations at selected tetranucleotide repeats, referred to as EMAST (Watson et al., 2014). The functional role of this specific form of instability remains unknown. Although tetranucleotide repeats occur at lower frequencies than mononucleotide repeats in our genome, they show higher mutation rates (Payseur et al., 2011; Watson et al., 2014). In most investigated cancers, EMAST instability occurs at loci with AAAGn or ATAGn repeats (Watson et al., 2014).

MSI has usually been detected by conventional methods, which include MSI-PCR followed by gel or capillary electrophoresis and/or immunohistochemical analysis of MMR proteins as a part of clinical testing (Lindor et al., 2002; Modica et al., 2007). To discover microsatellite instability (MSI) in tumor DNA, samples can be visualized on PAGE or Spreadex EL gels (Elchrom Scientific, ALLaboratechnik, AL-Diagnostic GmbH, Amstetten, Austria) stained with SYBR Gold, (Invitrogen, Molecular Probes, Eugene, OR, USA) (Figure 1.). Samples are considered MSI-positive if additional DNA bands are present, or the existing DNA band had shifted position in comparison to bands of autologous blood tissue indicating improper number of repeats, and thus impaired cellular DNA mismatch repair in tumor cells.
Mismatch Repair Genes and Proteins

The post-replicated DNA strand is checked by MMR system that consists of a group of proteins specialized in recognizing mispaired bases and small loops of insertion or deletion. MMR proteins usually function by forming heterodimers. MMR machinery consists of 8 genes in the human that code components of MMR: hMLH1 (3p21), hPMS1 (hMLH2) (2q31.1), hMLH3 (14q24.3), hPMS2 (hMLH4) (7q22.2), hMSH2 (2p21), hMSH3 (5q14.1), hMSH5 (6p21.3), hMSH6 (2p16) (Amaral-Silva et al. 2016; Lipkin et al. 2000; Clark et al. 2013). Variations and alterations in DNA repair genes are important factors for specific tumor susceptibility. Loss of proper functioning of DNA damage repair genes and proteins, whether through mutations or loss of their expression, is strongly correlated to
the introduction of genomic instability and consequent tumorigenesis (Clark et al 2013).

Mismatch repair system was originally discovered in *Escherichia coli* (Modrich, 2016) where mismatches induce a repair reaction upon transformation into the *E. coli* cell. Later, 4 *E. coli* genes were identified: MutS, MutL, MutH, and uvrD, and it was found that inactivation of any of these genes increased mutation rate in the *E. coli* 50 to 100 fold (Modrich, 2016; Hanaoka and Sugasawa, 2016). Briefly, MutS is responsible for mismatch recognition, MutH strand incision at an unmethylated sequence, MutL is recruited to the MutS-mismatch complex for match-making and uvrD for strand excision. DNA methylation would assure that only the newly mutated DNA strand is going to be cleaved. The basic mechanisms and proteins involved in the processes of MMR are highly conserved from bacteria to man. In humans, multiple MSH and MLH/PMS proteins functioning as heterodimers have evolved. In this way, specific functions within MMR repair have diversified according to the combination of proteins in the dimers. We will briefly describe the basic functions of human MMR genes (Itkonen et al., 2016).

**hMLH1**

This gene is a human homolog of the *E. coli* DNA mismatch repair gene mutL. Initial recognition of mismatches performed by heterodimers MSH2-MSH6 and/or MSH2-MSH3 is followed by binding of second protein heterodimer homologs of MutL, namely heterodimers MLH1-PMS2, or MLH1-MLH2, or MLH1-MLH3, which are recruited to the heteroduplex. Whether these 3 heterodimers function independently or jointly within the single pathway is not known. Nevertheless, the formed complex is sufficient to activate endonuclease activity of PMS2 and enable initiation of repair of the mismatch defect. Single-strand breaks are going to be introduced near the mismatches and let the exonuclease EXO1 degrade the strand together with the mismatch. MLH1-PMS2 interacts physically with the clamp loader subunits of DNA polymerase III, suggesting that it may seize the enzyme to the site of the MMR (Fishel, 2015). MLH1 gene was identified as a locus frequently mutated in Lynch syndrome. Moreover, the gene is often epigenetically silenced by CpG methylation within its promoter region, then MLH1 polypeptide is not produced, and mismatch repair is defective.

**hMLH2 (PMS1)**

This is yet another homolog of the *E. coli*’s mutL gene, also known by the name post-meiotic segregation increased 1. As mentioned above, the protein can form heterodimers with MLH1. Intracellular levels of PMS1 seem to be lower than that of PMS2. PMS1 and PMS2 also compete for interaction with MLH1 (Cannavo et al., 2007). Mutations in this gene cause Lynch Syndrome type 3 either alone or in combination with mutations in other genes involved in MMR.
hMLH3
Yet another member of the family of mutL homologs of DNA mismatch repair genes whose protein functions as a heterodimer with other family members. Mammalian MLH3 is predicted to participate primarily in repair of DNA insertion/deletion loops, and its disruption is predicted to manifest MSI as short repetitive sequence length instability (Lipkin et al., 2000). Somatic mutations in MLH3 frequently occur in tumors with microsatellite instability. Diseases associated with this gene include colorectal cancer, hereditary non-polyposis colorectal cancer type 7 (HNPCC7), endometrial cancer and low-grade gliomas (Duraturo et al., 2016).

hMLH4 (PMS2)
This gene is better known as post-meiotic segregation increased 2 (PMS2); it encodes a fundamental element of mismatch repair. It is involved in major nuclear MMR system by forming heterodimers with MLH1. Together they interact with MSH2-MSH6 and/or MSH2-MSH3. The assembly of these ternary complexes will initiate endonuclease activity of PMS2. It introduces single-strand breaks near the mismatch and thus generates entry sites for the exonuclease EXO1 (Jeon et al., 2016). Alterations in this gene have been associated with several diseases, Lynch syndrome type 4, colorectal cancer, primitive neuroectodermal tumors and Turcot syndrome. MLH4 also interacts with p53 and p73, which indicates that it contributes to genome integrity, not only through DNA repair, but by enhancing DNA damage-induced apoptosis (Shimodaira et al., 2003).

hMSH2
This gene is a homolog of the E. coli mismatch repair gene mutS; it is the main MSH protein that corrects mismatches. It forms 2 different heterodimers, one with MSH6 that predominates in human cells, and another one with MSH3. Heterodimers MSH2-MSH6 (also named MutSα) and MSH2-MSH3 (MutSβ) bind to DNA mismatches and initiate DNA repair. Both MSH complexes form sliding clamps on DNA and slide until mismatches and other extrahelical lesions are recognized (Brown et al., 2016). Heterodimer MSH2-MSH6 recognizes single-base mismatches and dinucleotide insertion-deletion loops, whereas the heterodimer MSH2-MSH3 recognizes larger insertion-deletion loops up to 13 nucleotides in length. Subsequently, they will bind to the MLH1/PMS2 complexes, which will degrade the mutated stretch and initiate resynthesis.

hMSH3
A homolog of the E. coli mismatch repairs gene mutS. As already mentioned, its protein product, when bound to MSH2 in a heterodimer complex, finds DNA mismatches, thereby initiating DNA repair. The MSH2-MSH3 heterodimer bends the DNA helix and shields ~20 base-pairs. Diseases
associated with MSH3 defects, besides colorectal disease, include endometrial cancer and urinary bladder cancer (Kawakami et al., 2004; Yamamoto and Imai, 2015).

**hMSH5**

This MutS homolog forms an exclusive heterodimer with MSH4 (Clark et al., 2013). It functionally differs from other MutS homologs because it is primarily involved in meiotic recombination. Therefore, MSH4-MSH5 is abundantly and specifically present in reproductive tissue. This protein is abundantly expressed in the mammalian meiotic tissues functioning in recombination for crossing-over events and gene conversion during meiosis. Although hMSH5 has not been directly implicated in MMR events, it seems that it performs a number of different cellular roles. It is involved in immunoglobulin diversity, DNA damage response and double-strand break repair, and thus it is also important in mitotic cells. hMSH5 SNP loci have been associated with a variety of human diseases, including neoplasia.

**hMSH6**

MutS Homolog 6 is a member of the DNA mismatch repair machinery. The encoded protein undergoes heterodimerization with MSH2 to form a mismatch recognition complex (Itkonen et al., 2016). It functions as a molecular switch that exchanges ADP for ATP when a G/T mismatch is recognized. DNA mismatches are bound and dissociated. After establishing the amino acid sequence homology, MSH6 shows strong similarities to the human "G/T binding protein," (GTBP), also called p160, and these those are the same molecules. A highly conserved region of ~150 amino acids, called the Walker-A adenine nucleotide binding motif, exists in MSH6. Diseases associated with MSH6 alterations include Lynch syndrome type 5, colorectal cancer and endometrial cancer.

**Mutator hypothesis**

The mutator hypothesis has been postulated to explain the discrepancies between abnormally elevated mutation rates seen in tumor cells and the rate of mutations spontaneously occurring in normal cells. Substantial evidence now indicates that the mutation rate of normal human cells is much too low to account for the accumulation of the 100’s of genetic alterations occurring in tumors. The usual incidence rate of spontaneous somatic mutations over a single individual's life-time does not match to the rates of genetic changes seen in tumor cells, and this increased frequency is the result of genomic instability that characterizes tumors. We know that multiple genetic changes are required for carcinogenesis; nevertheless, the mutator hypothesis relates primarily to the defective functioning of MMR proteins that leads to the mutator phenotype, in which an elevated rate of mutation persists. The consequence of this hypermutation phenotype, characteristic for cancer
genomes, is nascent microsatellite instability. Mutations of mutator genes are referred to as mutator mutations (Fishel and Lee, 2016). A mutator gene is a gene whose mutation increases the level of genetic changes of the individual's genome, and cancer cells are thought to have a mutator phenotype, with an anticipated increase rate of genomic instability. The mutator gene can be any gene, but usually comes from the group whose products are involved in DNA repair mechanisms, or in genes whose products control the fidelity of DNA synthesis. Extensive sequencing of different tumors largely supports the mutator hypothesis (Fishel and Lee, 2016). The mutational findings that support the mutator phenotype of human tumors can be found in The Cancer Genome Atlas (TCGA), the mission of the project being to determine the number of mutations in each tumor. TCGA indicates that the number of observed mutations per tumor differs substantially, ranging from 500 to 100,000 (Loeb, 2016).

Although it is usually assumed that genetic instability is a later event in tumor evolution associated with progression, there are suggestions that it may be an early causative event in the formation of a specific tumor.

**Cancer related to MSI**

Many types of human tumors have the MSI phenotype; those with it show specificity in the clinical course of the disease, response to therapy, and survival outcomes (Clark et al., 2013; Moy et al., 2015; Thompson and Spurdle, 2015; Hause et al., 2016; Kubeček and Kopecký, 2016). The longest known involvement of MSI phenotype is in colon cancer, being especially causative for human hereditary non-polyposis colorectal cancer syndrome or Lynch syndrome (Hoang, et al., 1997; Umar, et al., 2004; Poulogiannis et al., 2010; Sameer et al., 2014). Ineffective MMR system due to germline mutations in one of the mismatch repair genes is responsible for the Lynch syndrome, a common autosomal dominant disorder characterized by early onset (<45 years) and the presence of neoplastic lesions showing MSI in endometrial, gastric, renal, ovarian and skin tissues (Hemminki et al., 1994; Pastrello et al., 2006).

MSI events identified and quantified when comparing patient’s normal and tumor DNA can be classified as MSI-High (MSI-H), MSI-Low (MSI-L), or Microsatellite-Stable (MSS) depending on the number of microsatellite markers showing MSI. The main specificity of MSI-H cancers is a gain in the numbers of microsatellite alleles and the overall number of unstable microsatellite loci. According to Bethesda guidelines (Umar et al., 2004), MSI-H refers to tumors in which MSI occurs in 2 or more of the 5 National Cancer Institute-recommended panels of microsatellite markers (BAT-25, BAT-26, D2S123, D5S346, and D17S250); MSI-L refers to tumors displaying changes in only one of the 5 microsatellite markers and MSS are those where no instability is detected. The consensus on which tumor belongs to which category has been soundly defined only for colon cancers, and this
classification is pertinent for predictive and prognostic purposes (Zhou et al., 1997; Brennetot et al., 2005; Watson et al., 2014). To simplify classification of molecular MSI subtypes of colorectal carcinoma, MSS and MSI-L tumors are usually grouped together into one category, with MSI-H being classified into another (Mokarram et al., 2014; Watson et al., 2014; Gatalica et al., 2016). Moreover, this subcategory of MSI-L cancers behaves quite similarly to MSS cancers according to the total number of identified MSI events, a finding supporting the idea of grouping them together. However, both variances in MSI amongst cancer types, as well as the MSI-genomic landscape, remain poorly investigated.

Reports indicate that tumors with MSI-H have a more positive prognosis compared to MSI-L or MSS tumors (Yamamoto and Imai, 2015). This somewhat paradoxical situation in that the patients diagnosed with MSI-H type of cancer and attributed with higher overall MSI events showed a tendency to survive much longer has tried to be explained by Hause et al. (2016). In this regard, it has been proposed that metastatic potential of a cancer is not exclusively dependent on MSI. This finding opened up a new hypotheses by which one tries to explain the longer-survival-higher-MSI-correlation in a way that cancer cells, frequent in overall MSI potential, produce larger numbers of mutated and truncated proteins of all kinds, therefore leading to an awakening of an already weakened immune system that manages to slow down tumor progression (Hause et al., 2016).

Expression patterns of MutS homologues in a group of European colorectal carcinoma patients indicated that their tumors expressed of all MMR genes more poorly (Clark et al., 2013). Losing the function of MMR proteins either by mutation or reduced expression provides some evidence that correlates cancer development and genomic aberrations of all or the majority of MMR repair proteins.

In a formidable analysis, Hause et al. (2016) investigated the overall MSI frequency per particular unstable microsatellite loci. They used tumor exomes from the Cancer Genome Atlas (TCGA) Research Network to examine MSI across tumor types, and constructed a genomic classifier for MSI by examining 5,930 cancer exomes from 18 cancer types at >200,000 microsatellite loci. This comparative observation of MSI between cancer types confirmed that cancer exomes generally contain about 87 to 9,032 unstable microsatellites on the whole, but it has also revealed that the average number of unstable microsatellite loci varies considerably with cancer type, colon cancers displaying the highest number of loci and thyroid cancers the lowest. Moreover, their results indicated that MSI found in MSI-H and MSS specimens from the same cancer type share consistent MSI patterns. Their investigation of the MSI landscape through cancer-normal tissue comparison showed that frequently unstable microsatellites in MSI-H tumors were more often located within, or in close proximity to, the genes already known for their oncogenic potential, which may suggest that MSI contributes in generating cancer-driving mutations. This means that unstable microsatellites are
creating a foundation for future identification of novel candidate cancer-driving genes. This study also checked the overall MSI patterns per individual microsatellite loci, which showed that cancer types could be hierarchically grouped by the similar MSI signatures according to the similarity in MSI rates among groups.

MSI has been reported in many cancer types, including endometrial, ovarian, gastric, melanoma, prostate, lung, stomach and glioblastoma tumors (Leung et al., 1998; Alvino et al., 2000; Arai et al., 2016; Hause et al., 2016; Kubeček and Kopecký, 2016). Our own work on MSI in human intracranial meningiomas (Pećina-Šlaus et al., 2010, 2016a,b) has demonstrated the appearance of a low, but constant, frequency of MSI loci for the AXIN1 and CDH1 genes in meningioma. The results show the involvement of MMR machinery in meningioma in accord with the findings of others (Pykett et al. 1994; Sobrido et al. 2000; Chen et al., 2012).

Conclusions
It is now clear that MSI is more important in cancer than previously thought. We believe that MSI phenotype is an intrinsic part of the genetic profile of certain group of tumors, even within the same general diagnostic category. It seems that in some tumors MSI is a continuous characteristic rather than appearing later during tumor progression. Tomasetti and Vogelstein (2015) calculated that the total number of stem cell divisions within a tissue correlates with the lifetime risk of cancer incidence. We propose that the potential number of cell divisions of a specific tissue should also be taken into account when considering the role of MSI, and should be added together to the number of events of genomic instabilities observed for any specific tumor. It is unclear whether malfunctioning of MMR arises because it cannot keep up with the velocity of cell proliferation with an increased need for repair that accompanies the increased rate of DNA replication, or it represents an intrinsic root-cause of the genesis of a specific tumor. This point of view on the cumulative effect of MSI and number of cell divisions could help us explain and understand cancer progression more fully.

In conclusion, because of the indications on the continuous distribution of global instability in the cancer genome, the prognostic and predictive relevance of MSI in any specific type of cancer should be considered as indicative of patient survival. Ultimately, this might lead to development of novel cancer treatment, probably more oriented toward differences in responses to therapy.

Acknowledgements
Funded by grant number 6625 from Croatian Science Foundation.

Conflict of interest: The authors declare that they have no competing interests.
References


Gatalica Z, Vranic S, Xiu J, Swensen J and Reddy S. High microsatellite instability (MSI-H) colorectal


Pećina-Šlaus N, Kafka A, Vladušić T, Pećina, HI, Hraščan R. AXIN’s expression and localization in


