Looking beyond the cytogenetics in haematological malignancies: decoding the role of tandem repeats in DNA repair genes

Priyanjali Bhattacharya1 · Trupti N. Patel1

Received: 26 April 2022 / Accepted: 30 June 2022 / Published online: 12 September 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Background In cancer research, one of the most significant findings was to characterize the DNA repair deficiency as carcinogenic. Amongst the various repair mechanisms, mismatch repair (MMR) and direct reversal of DNA damage systems are designated as multilevel safeguards in the human genome. Defects in these elevate the rate of mutations and results in dire consequences like cancer. Of the several molecular signatures in human genome, tandem repeats (TRs) appear at various frequencies in the exonic, intronic, and regulatory regions of the DNA. Hypervariability among these repeats in the coding and non-coding regions of the genes is well characterized for solid tumours, but its significance in haematologic malignancies remains to be explored. The purpose of our study was to elucidate the role of nucleotide repeat instability in the coding and non-coding regions of 10 different repair genes in myeloid and lymphoid cell lines compared to the control samples.

Methods and Results We selected MMR deficient extensively studied microsatellite instable colorectal cancer (HCT116), and MMR proficient breast cancer (MCF-7) cells along with underemphasized haematologic cancer cell lines to decipher the hypermutability of tandem repeats. A statistically significant TR variation was observed for MSH2 and MSH6 genes in 4 and 3 of the 6 cell lines respectively. KG1 (AML) and Daudi (Burkitt’s lymphoma) were found to have compromised DNA repair competency with highly unstable nucleotide repeats.

Conclusion Taken together, the results suggest that mutable TRs in intronic and non-intronic regions of repair genes in blood cancer might have a tumorigenic role. Since this is a pilot study on cell lines, high throughput research in large cohorts can be undertaken to reveal novel diagnostic markers for unexplained blood cancer patients with normal karyotypes or otherwise with karyotypic defects.

Keywords MMR · DRD · Coding and non-coding regions · Tandem repeats · Blood cancer cell lines

Introduction

The genetic characterization of haematologic malignancies, for classification, diagnosis, prognosis, and therapeutic interventions, are constantly evolving due to increasing genomic variation as supplemented by high throughput data. Numerical and structural chromosomal instabilities (CINs), resulting in a ‘chaotic genome’ are established hallmarks of blood cancer. However, the role of tandem repeat (TR) variations implicating molecular changes in haematologic cancers are highly argued [1, 2]. Normally, DNA repeats are grouped into two major classes—the tandem repeats (satellite DNA) and the interspersed repeats. While the mono-, di-, tri-, tetra-, penta-, and hexa-repeats are clustered; long, and short interspersed retrotransposable element (LINE and SINE) are scattered all over the genome. The tandem repeats are further classified into macrosatellites and microsatellites. The microsatellites, span the genome in clusters and are more prevalent in non-coding regions than coding regions of the DNA [3]. Besides their evolutionary relevance, the repetitive sequences have gained importance in cancer research as they display hypervariability in the repeat number [4, 5]. While the mutations in the coding repeats result in abnormal gene products, repeat instabilities in the dark DNA can cause genomic fragility yielding alter reading frames or splicing sites variants [6]. Harms causes by DNA repeats variations can be dictated by two opposing mechanisms—(1) any variation in repeat number for genes having TRs in critical positions may directly or indirectly cause functional loss.
The repeat number variations and corresponding shyness of function may cause onset of genomic instability which is critical for tumour evolution; (2) upon inactivation of repair mechanisms, increased heterogeneity is observed at simple repetitive DNA (e.g., mono- and di-nucleotides) in bacteria, yeast, and mammals. The association of defective mismatch repair (MMR) and an elevated instability of DNA repeats is particularly strong in solid tumours, but less predominant in blood cancers [7].

The DNA repair mechanisms are well placed in the eukaryotic machinery as they are dedicated to protect genomic integrity either by tolerating or removing the damage simultaneously reducing the deleterious consequences. Normally nine genes (MSH2, MSH3, MSH4, MSH5, MSH6, MLH1, MLH3, PMS1, and PMS2) constitute an intact Mismatch Repair (MMR) system responsible for correcting misaligned bases, insertion–deletion loops (INDELs), regulating–recombination, meiotic crossover, and diversification of antibody [8]. In eukaryotes, the MutS (MSH2, MSH3, MSH4, MSH5, and MSH6) and the MutL (MLH1, MLH3, PMS1, and PMS2) components are the key players that form different heterodimeric complexes fixing the DNA damage (DRD). These complexes namely, MutSα (MSH2–MSH6), MutSβ (MSH2–MSH3), MutLα (MLH1–PMS2), MutLβ (MLH1–PMS1), MutLγ (MLH1–MLH3) are involved in ‘MMR-associated genome maintenance’. The MutS- α and β recognize the mismatches while the MutL- α, β, and γ coordinate the downstream processes of mismatch repair. The MSH4 and 5 do not directly involve in MMR functions but play an important role during meiosis, and generation of immunoglobulin diversity. An abnormal MMR elevates the rate of mutation up to 1000 folds in a cell, which lead to genomic instability and fragile genome [9].

Another single step repair system—Direct Reversal of DRD specifically acts on pyrimidine dimers and alkylated guanine residues. The O6-methylguanine (O6meG) methyltransferase (MGMT) is a suicidal enzyme that removes the methyl adducts from the 6th position of guanine and in the process inactivates itself. Mutations resulting in complete or partial loss of functional MGMT are evident in lung, gastric, breast, and hepatobiliary cancers [10]. Loss of MGMT function is usually charted to promoter hypermethylation or nucleotide transversions through mispairing between O6meG and thymine (T) which is also marked by the MMR machinery [11]. However, it would be interesting to explore hypervariability in the repeat regions of MGMT gene. As evident, malfunctioned DNA repair systems pose a threat to the normal genome, giving us an opportunity to reflect on affects of tandem repeat contractions or expansions in various genes.

In cancer, the significance of repeat sequence variations has emerged in disease onset, progression, outcome, therapeutic response, and overall survival. Microsatellite instability was first identified in hereditary nonpolyposis colorectal cancer (HNPCC) and later in tumors of ovary, stomach, breast, skin, brain, endometrium, small intestine, urinary, and hepatobiliary tract for functionally impaired MMR genes [12]. Nevertheless, their role in haematologic cancers with varying degree of molecular characteristics and clinical behaviour, still remains debatable. Classically, blood cancers follow the pattern of chromosomal instability which triggers the early events of tumorigenesis conferring accumulation of mutations, aggressive phenotype, and therapeutic resistance during clonal evolution of cancer. In due course, with compromised DNA repair components, the tumours shift to molecular events including mutability of tandem repeats and loss of heterozygosity. However, here we propose an alternative hypothesis that in certain complex cases of haematologic malignancies, CINs alone may not be the prominent cause. Early events of repeat polymorphisms in the intronic and exonic regions of DNA may lead to impaired repair mechanisms causing hematopoietic disturbances followed by cancer development. A few research groups have confirmed presence of tandem repeat variation in myeloid and lymphoid cancers while others have found it to be rare or absent in blood cancers [13–16]. These incongruencies motivated us to design a pilot study to investigate tandem repeat instability in MMR and MGMT gene structures in blood cancer cell lines.

Materials and methods

Procurement of cell lines and control samples

To study the repeat number variations, four blood-cancer cell lines—K562 of CML (Chronic Myeloid Leukaemia), KG-1 of AML (Acute Myeloid Leukaemia), Molt-4 of T-ALL (T-cell acute lymphoblastic leukaemia), and Daudi of Burkitt’s lymphoma were selected. Of these, K562, Daudi, and KG-1 are profiled as MMR proficient while Molt-4 is known to show mutation, in any of the MutSα (MSH2/MSH6) or MutLα (MLH1/PMS2) components. The MMR deficient HCT116 colorectal cancer (CRC), and proficient MCF-7 breast cancer cell lines were also used in the study. Five control DNA were used along with all the above cancer cell lines. The cytogenetic and mismatch repair status (proficient/deficient) of cell lines are provided in Supplementary Table 1. All cell lines were procured from National Centre for Cell Sciences, Pune, India except CRC cell line which was a kind courtesy of collaborating laboratory. The procured cell lines were maintained under proper culturing conditions until 80% confluency was achieved. The passage numbers, growth media, and growing conditions of the cell lines are mentioned in Supplementary Table 2. The blood
was collected from 5 healthy individuals with informed consent as control samples.

**Extraction and quantification of DNA**

The DNA from six cell lines and five control samples (2 males and 3 females) was extracted using HiPura Mammalian Genomic DNA extraction Kit (MB-506) with slight deviation from the manufacturer’s protocol. The isolated DNA was run on 0.8% Agarose gel using Tris Borate EDTA buffer and further analysed using NanoDrop spectrophotometer (Thermo Fisher Scientific).

**Designing of primer for gene-sequence specific tandem repeats**

Of the total 10 genes, 8 primers were designed while 2 primer sequences were obtained from literature search. The markers of MSH2 (BAT26- Bethesda panel marker), and MLH1 (D3S1611) were selected from peer-reviewed literature though these have either not been explored or explored with inconsistent outcomes for haematologic cancers [17, 18]. For the other genes—MSH3, MSH4, MSH5, MSH6, MLH3, PMS1, PMS2, and MGMT, repeat clusters were identified from the gene sequence (NCBI) and primers were designed (Primer3Plus—http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) for thereof. As anticipated the location of the repeats for MSH3, MSH4, MSH5, MLH1, PMS1, and MGMT genes was in the intronic regions. The markers for MSH2, MSH6, and MLH3 were located at exon–intron junction while PMS2 was an exonic marker. The primer pairs were finalized by validating through NBLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) and in silico PCR tools (https://genome.ucsc.edu/cgi-bin/hgPcr). The MSH2 and MSH5 markers were enriched with mononucleotide (A), repeats and that of MSH3, MSH4, MLH1, MLH3, PMS1, and MGMT were dinucleotide (CA), repeat markers. These repeat sequences are categorized as perfect microsatellites. While the marker of PMS2 [(CA)6(CA)6(CA)18] was an imperfect microsatellite repeat, the marker of MSH6 [(A)20(CA)20(CA)20(T)10] was a complex microsatellite repeat (Supplementary Table 3). Initially untagged primers were obtained for pilot runs followed by fluorescently tagged (HEX) primers, from Eurofins India Pvt. Ltd. The detailed information of primers is documented in Table 1.

**Polymerase chain reaction (PCR)**

Amplification was carried out using TaKaRa Clonetech R050A PrimeSTAR GXL DNA polymerase kit. Each experiment was carried out using 30 µl of reaction mixture [12 µl 5X GXL Buffer, 2.5 mM dNTP, 1.25 U/µl Primestar GXL DNA Polymerase and, 10 pM each of forward and reverse primer along with 50 ng DNA] for three replicates and the experiments were repeated thrice. The samples were amplified using Eppendorf Mastercycler Nexus PCR model for 35 cycles each. The initial denaturation temperature and final extension temperature were kept as 96 °C for 6 min and 72 °C for 7 min respectively while annealing temperatures were adjusted as mentioned in Table 1. Amplified products were analysed using agarose gel electrophoresis and further subjected to fragment analysis.

**Capillary electrophoresis-based fragment analysis**

The PCR products were pooled from triplicate samples for fragment analysis that was carried out using Applied Biosystems (ABI) 3730 XL genetic analyser. The reaction mixture consisted of PCR product (0.5 µl), Liz-500 size standard (orange colour dye) (0.5 µl), and HI-DI formamide (9 µl). The mix was denatured at 95 °C for 5 min and snap-frozen for 2–3 min before loading it on the sample tray. The fluorescence intensity was recorded as function of wavelength and time and the data was visualized using GeneMapper software (Thermo fisher scientific).

**Statistical analysis**

The two sample F-test was performed using MATLAB software (version R2018a) to detect the significance of left or right shift in the peaks of ten repeat-rich markers. The null hypothesis was accepted if the output generated the hypothesis value as h = 0. Otherwise, for h = 1, the alternative hypothesis was considered.

**Analysis of splice sites for individual markers**

The constitutive or cryptic splice sites of ten repeat rich markers were analysed using ASSP (Alternative Splice Site Prediction) web-tool (http://wangcomputing.com/assp/index.html). It helped to predict possible shifts in alternative splicing due to hypermutability, which is otherwise critical for regulation of gene expression. The tool provided information regarding frequency of GC nucleotide and presence of constitutive or cryptic splice site along with a predictive score.

**Results**

**Profile of tandem repeat instability in cancer cell lines**

The PCR amplified products were run on agarose gel to confirm the amplification of repeat-rich regions prior to
### Table 1  Details of primer pairs along with position, annealing temperature, number of repeats, product and allele sizes for ten repair genes

| Genes          | Position     | Location | Primer sequences (5'→3')       | Annealing temperature (°C) | Number of repeats | In silico product size (bp) | Allele size range (bp) |
|----------------|--------------|----------|-------------------------------|-----------------------------|-------------------|-----------------------------|------------------------|
| MSH2 (BAT26)   | 2p21-p16.3   | E: 5     | I: 5–6                        | 62.3                        | (A)_{26}          | 126                         | 118–119                |
| MSH3           | 5q14.1       | I: 19–20 | F: TTAGAGAAAGCTGTCAGGA        | 61.3                        | (AT)_{12}         | 153                         | 158–178                |
| MSH4           | 1p31.1       | I: 19–20 | F: TGTGGGTTTAGTGTGTTGAGCT    | 62.5                        | (AT)_{10}         | 211                         | 158–221                |
| MSH5           | 6p21.33      | I: 10–11 | F: GCTGAGGCGAGAAAATCATCTGAC  | 54.0                        | (A)_{16}          | 150                         | 144–148                |
| MSH6           | 2p16.3       | I: 9–10  | E: 10                         | 61.0                        | (A)_{n}C(A)_{n}C(T)_{18} | 219                         | 216–217                |
| MLH1 (D3S1611) | 3p22.2       | I: 13–14 | F: CCAAGGCTGCAGTTGAGGCA      | 60.1                        | (CA)_{13}         | 262                         | 258–262                |
| MLH3           | 14q24.3      | E: 5     | I: 5–6                        | 65.0                        | (AT)_{18}         | 215                         | 225–229                |
| PMS1           | 2q32.2       | I: 1–2  | F: GGAAGGGGCCTAATCTACCTG     | 62.3                        | (AT)_{16}         | 157                         | 155–165                |
| PMS2           | 7p22.1       | E: 15    | F: CTTTACCTGGAGTTACCTG        | 55.2                        | (A)_{n}C(A)_{n}C(A)_{n}C(A)_{13} | 218                         | 218                    |
| MGMT           | 10q26.3      | I: 2–3  | F: CACACATGGTTAGCTACCTG    | 64.1                        | (CA)_{18}         | 209                         | 214–216                |

*E, Exon; (…)_n, Repeat; I Intron; F Forward primer; R, reverse primer; R*, 5′ end of reverse primer was tagged with fluorescent dye HEX*
fragment analysis (Supplementary Fig. 1). The instability profiles of six cell lines for 10 repeat-rich markers were compared with five healthy controls. This was detected by the left or right shift in the electropherogram, generated by GeneMapper software indicating contraction or expansion of nucleotide repeats respectively. The markers at different loci in PCR amplified product exhibited fine peak resolution that helped to calculate the range of allele size. Additionally, no stutter peaks hindered the recognition between the homozygotes and heterozygotes. The product size of each marker is non-comparable with respective allelic size in homozygous and heterozygous conditions (Supplementary Figs. 2–11). However, for selective repair genes, low to high alterations in perfect, imperfect, and complex repeat sequences were evident in six cell lines. This established possible changes in tandem repeats of mismatch repair and MGMT genes. The graphical representation of loss and/or gain of nucleotides is depicted in Fig. 1.

**Tandem repeat instability for genes of MutS complex of MMR**

- **MSH2**: In MSH2, all the cell lines and control samples were homozygous for BAT26 marker which encompasses an exon–intron junction. The allele size of BAT26 for control samples ranged between 118 and 119 bp. In cell lines, the left shift was noted at—114 bp for MOLT-4, 116 bp for Daudi, 102 bp for KG-1, and 100 bp for HCT116, indicating loss of repeats for approximately 4 bp (Molt-4), 2 bp (Daudi), 16 bp (KG-1), and 18 bp (HCT116). No variation was seen for K562 of CML and MCF-7 of breast cancer (Fig. 2A). The contraction of mono-A tract was found to be statistically significant for MSH2 gene (p < 0.001).

- **MSH3**: For intronic MSH3 marker, one control was homozygous, while four controls were heterozygous. The homozygous control showed allele size at 176 bp for dinucleotide (AT) rich marker. In three heterozygous controls, the size of allele 1 was at 160 bp and that of allele 2 was at 178 bp. However, one heterozygous control showed fragments at 151 bp and 158 bp for both alleles. In deep intronic repeats, large variations of tandem repeats occur in normal genomes. Hence, no statistical significance was reflected here and large cohorts of healthy and diseased samples need to be studied for comparable data. The homozygous KG-1 of AML, Daudi of Burkitt lymphoma, and HCT116 of CRC showed maximum and minimum loss of (AT) repeats for ~16 bp, ~18 bp, and ~2 bp respectively. The allele 1 of heterozygous MCF-7 (breast cancer) and Molt-4 (T-ALL) conferred loss of ~10 bp and ~2 bp of repeats individually. The allele 2 for these two cell lines

![Graph representing tandem repeat instability for 10 marker in six different cell lines. Not all the cell lines showed variability for each of the markers](image)
showed ~1–2 bp loss which was disregarded. Both gain of ~6 bp and loss of ~2 bp repeats were observed for heterozygous K562 of CML. Overall, the repeat sequence variation was not statistically significant (Fig. 2B).

- **MSH4**: The dinucleotide intronic MSH4 marker was found to be homozygous for two controls and heterozygous for three controls. The allele size of homozygous samples was found at 161 bp and 205 bp. From heterozygous controls, one control showed fragment in the range of 160–182 bp while the other two showed fragments at 203–215 bp, and 203–221 bp respectively. As seen in MSH3, the deep intronic MSH4 too showed tandem repeat variations in controls. Due to discrepancy in heterozygous populace, the range of mutability could not be defined. Excluding Burkitt lymphoma cell line Daudi, all other cell lines were homozygous for this marker. Approximately, 1 bp loss of dinucleotide (AT) repeats were found in—K562 of CML, KG-1 of AML, and MCF-7 of breast cancer. Both gain and loss of repeat was evident in Daudi that ranged between ~7 and ~1 bp for allele 1 and 2 respectively. The contraction and expansion of nucleotide repeats were not statistically significant (Fig. 2C).

- **MSH5**: All the controls and cell lines were homozygous for intronic mono-A MSH5 marker which was located at intron–exon junction. In controls, the allele size ranged between 216 and 217 bp. In CML, AML, and CRC cell lines, maximum contraction in complex repeat tract accompanied by poly-A and T nucleotides were observed for ~56 bp, ~66 bp, and ~55 bp respectively. The cell lines of T-ALL, Burkitt lymphoma, and breast cancer were microsatellite stable. The alterations in repeat sequences were statistically significant (p < 0.001) (Fig. 2E).

Tandem repeat instability for genes of MutL complex of MMR

- **MLH1**: The intronic D3S1611 marker of MLH1 was homozygous for one control and heterozygous for four controls. The fragments were observed at 252 bp and 262 bp for allele 1 and 2 of heterozygous control. The heterozygous cell lines—KG-1, and HCT116 showed moderate to high expansion of dinucleotide (CA) repeat for ~10 bp/2 bp, and ~6 bp/4 bp in allele 1 and 2 respec-
tively. The gain of ~6 bp and loss of ~2 bp repeats was observed in both alleles of Daudi. The homozygous K562, Molt-4, and MCF-7 did not show hypervariability for this marker. The alterations seen were not statistically significant (Fig. 3A).

- **MLH3**: For MLH3 marker, located at exon–intron junction, was found to be homozygous for three control samples with fragments between 199 and 200 bp. The two heterozygous controls showed fragments at 195 bp/225 bp and at 225 bp/229 bp for both the alleles. The allele size 225 bp and 229 bp strikingly varied from the rest, indicating possible ambiguity in these controls, a common occurrence for microsatellite repeats. However, considering the allelic positions which are frequent, we can draw a comparison in control and cancer cells. The gain and loss of nucleotide repeats were thus evident in the heterozygous AML for ~4 bp and ~2 bp at allele 1 and 2 respectively. The breast cancer cells showed expansion of repeat for ~26 bp at allele 1. No instability was noted in CML, T-ALL, Burkitt lymphoma, and CRC cells here. The variations were statistically insignificant using F-test (Fig. 3B).

- **PMS1**: The intronic dinucleotide PMS1 marker was homozygous for three controls showing fragments at 155–157 bp and heterozygous for two controls with allele size ranged between 153–165 and 144–157 bp. Here again, the range of heterozygous alleles were considerably different for one control. Contraction of (AT) repeats were observed in homozygous Molt-4 of T-ALL for ~6 bp, and Daudi of Burkitt lymphoma for ~13 bp. The KG-1 of AML showed expansion of repeats for ~8 bp. Both loss and gain of repeats were observed in heterozygous HCT116 of CRC for ~9 bp and ~10 bp in allele 1 and 2 respectively which is atypical from the control which has a shifted range rather than gain and loss. The K562 of CML, and MCF-7 of breast cancer were microsatellite stable. The changes observed were not statistically significant (Fig. 3C).

- **PMS2**: The exonic marker of PMS2 was accompanied by imperfect repetitive sequences of poly-A nucleotides. All the cell lines and controls were homozygous for this marker. The control samples showed fragment at 218 bp while a shift of single bp repeat was observed only in three cell lines—K562, KG-1, and MCF-7, which may have no statistical significance for the exonic position of the marker. The Daudi, Molt-4, and HCT116 did not confer instability (Fig. 3D).

**Repeat variations in MGMT gene of direct reversal of DNA damage**

- **MGMT**: For intronic dinucleotide MGMT marker, three homozygous controls showed fragment at 214 bp and 216 bp, while the allele size of two heterozygous controls was observed at 206–214/216 bp for allele 1 and 2. The cell lines were homozygous for this marker except

---

![Fig. 3 A–E](image-url) Red arrow indicates contraction of repeat sequences, blue arrow indicates expansion while black arrow indicates both. A MLH1 marker depicting loss and gain of ~2 bp and 6 bp (CA) repeats in Daudi cell line; B MLH3 marker depicting loss and gain of ~2 bp and 4 bp (AT) repeats in KG-1 cell line; C PMS1 marker depicting gain of ~8 bp (AT) repeats in KG-1 cell line; D PMS2 marker depicting loss of ~1 bp imperfect repeats in MCF-7 cell line; E MGMT marker depicting gain of ~6 bp (CA) repeats in Daudi cell line. For details see Supplementary Figs. 7–11
Daudi which was heterozygous with an approximate 6 bp gains in dinucleotide repeats. The K562, and Molt-4 did not show CA-repeat variations. Contraction in repeat sequence was observed in cell lines MCF-7, KG-1, and HCT116 for approximately 1, 2 and 8 basepairs respectively. (Fig. 3E).

The major shortcomings here was the inability to define the range of heterozygous alleles in the control population due to small sample size. This drawback will be remedied in our future study using larger sample numbers for controls and patients, scanning whole gene regions using high throughput technique for these 10 genes. However, to the credit of this study, haematologic cell lines KG-1, and Daudi were distinctly unstable for more than 5 out of 10 markers analysed. Polymorphic repeat were observed in HCT116 and MCF-7, as consistent with previous findings [19, 20]. The total percentage accounting for hypermutability from all markers in four blood cancer cell lines was – 100% in KG-1, 70% in Daudi, 50% in K562, and 30% in Molt-4 (Supplementary Table 4). The highest tandem repeat fluctuations were recorded for MSH3 marker as observed in all four blood cancer cell lines. This was followed by MSH2 and PMS1 for Molt-4, Daudi, and KG-1; MSH4 and MSH5 for KG-1, Daudi, and K562. The genes MSH6 (K562, KG-1), MLH1 (Daudi, KG-1), PMS2 (KG-1, K562) and MGMT (Daudi, KG-1) were varying in two cell lines with least changing MLH3 in only one cell line, KG-1. The HCT116 and MCF-7 cell lines showed repeat variability in seven (MSH2, MSH3, MSH4, MSH6, MLH1, PMS1, MGMT), and five genes (MSH3, MSH5, MLH3, PMS1, and MGMT) respectively which was a good comparable scale to emphasize our hypothesis that fluctuations in blood cancer cells also occur at the same rate as solid tumour cells.

**Prediction of splice sites in individual marker**

Most of repeat rich markers were found to carry cryptic donors and acceptors. Activation of these cryptic splice sites due to change in repeat number may cause unfavourable consequences. Likewise, classical splice sites may also get disrupted due to repeat variations in constitutive regions. Here, the ‘intronic GC’ represents the frequency with which GC nucleotides occurs around the splice site and is closely associated with site recognition and usage [21]. The confidence for a constitutive or cryptic splice site always ranges between 0.000 < score < 1.000, while for an unclassified splice site, the score is 0.000 (Supplementary Table 5). The BAT26 marker of MSH2 was predicted to carry a cryptic donor with a high confidence of 0.885 and an unclassified splice acceptor. For MSH3 marker, one cryptic donor with score of 0.852 was identified. In MSH4 marker, four cryptic donor (score: 0.258–0.867) and one acceptor (score: 0.727) were recognized. In MSH5, one cryptic splice donor with score 0.974, and acceptor with score 0.267 were found. In MSH6 marker, one cryptic donor (score: 0.944) was predicted. The MLH1 marker was enriched with two cryptic donor (score: 0.789–0.900) and four cryptic acceptors (score: 0.523–0.955). In MLH3 marker, one constitutive donor with score 0.940, and one cryptic acceptor with score 0.185 were identified. For PMS1 marker, only one constitutive donor with score 0.274 was predicted. Since the PMS2 marker is exonic, the ASSP did not generate any predictive result. In MGMT marker, one constitutive donor (score: 0.451) and two each of cryptic donors (score: 0.764–0.926) and acceptors (score: 0.437–0.521) were recognized. In view of the confidence of ASSP prediction, the cryptic donors in MSH2 and MSH6 markers were identified with 88.5% and 94% confidence which ascertains the definite presence of splice sites and correlates to repeat displacement. The presence of cryptic donor in MSH3, MSH5 and constitutive donor in MLH3 was identified with 83%, 97%, and 94% confidence respectively. The splice sites harbouring donors and acceptors for MLH1 was identified with high confidence and that of MSH4, MGMT were recognized with moderate confidence. In PMS1, the confidence was found to be lowest which is 27%.

**Discussion**

A preliminary investigation was carried out based on the premise that, microsatellite repeat variations are crucial to haematologic malignancies and that sometimes cytogenetic anomalies may not take precedence especially in atypical cases of blood cancer. The hotspots in short repetitive regions were decoded from the vastness of gene regions particularly, the dark DNA. Since the cells were maintained at low passage, additional changes in their cellular and molecular characteristics can be ruled out. Various forms of genomic instabilities are documented in cancers of the blood and other tissues. Tumors driven by defective DNA repair exhibit hypervariability in length of tandem repeats that results in functional disturbances. The diversity in repeat sequences of satellite DNA was higher for haematologic KG-1 (10 markers) and Daudi (7 markers) cell lines along with HCT116 (7 markers) and MCF-7 (5 markers). Such outcomes confirm susceptibility to repair deficiency in cancer, irrespective to its tissue of origin. K562, and Molt-4 manifested instability in 5 and 3 repair genes respectively. The degree of instability was statistically significant for markers of MSH2, and MSH6. Interestingly, though the BAT26 in MSH2 is known for its instability in solid tumors, MSH6 a novel marker, showed a similar trend in most cell lines here. The Table 2 consolidates the position, repeats, zygosity,
observed allele sizes, and loss/gain in repeat sequences of ten markers in control samples and cell lines.

A large contraction of approximate 16 to 18 bp was noted in poly-A tract of BAT26 for AML (~16 bp), and CRC (~18 bp) cells. This marker spans the exon–intron junction of MSH2 and accompanies a cryptic donor and unclassified splice acceptor as predicted by ASSP tool. Hence, this loss of nucleotide repeat at 5th exon–intron junction may change the spliceosome binding at the donor-splice site and simultaneously affect the conserved reading frame. As a part of Bethesda Panel, BAT26 is confirmed to be replication error positive (RER+) and it impairs the functional rightness of MMR in colorectal and breast cancers [22–24]. Opposing opinions on the impact of hypervariability in this marker exist for blood cancers. Few investigations suggest it to be insensitive for microsatellite instability in AML [16, 25] while others propose MSI as a distinct feature especially in de novo non-mutant FLT3 AML patients [26]. Here, FLT3 normal cell line KG-1 here showed a very much similar trend probably downregulating MSH2. We suggest that BAT26 could serve as a potential marker in primary AML cases with unaltered FLT3 status and complex cytogenetics.

There is no literature available citing tandem repeat variations in MSH3 which inspired us to design a primer

---

Table 2: Consolidate table showing ten repeat rich markers along with the fragment sizes as observed in homozygous and heterozygous controls and cell lines.

```plaintext

Control samples

| Allele change | A1 | A2 | Loss/ gain |
|---------------|----|----|------------|
| C1            | 1  | 1  | NA         |
| C2            | 1  | 1  | NA         |
| C3            | 1  | 1  | NA         |
| C4            | 1  | 1  | NA         |
| C5            | 1  | 1  | NA         |

Cell lines

| Allele change | A1 | A2 | Loss/ gain |
|---------------|----|----|------------|
| MC            | 1  | 1  | NA         |
| KG            | 1  | 1  | NA         |
| Da cell       | 1  | 1  | NA         |
| Mo cell       | 1  | 1  | NA         |
| KG            | 1  | 1  | NA         |

*NA* - not applicable, A1- allele 1; A2- allele 2; + blue—gain of repeat, — red—loss of repeat, bold- control samples with disparities in repeat

---

References

[22–24]...

---

Springer
within the deep intronic AT-rich region of this gene. A significant loss of dinucleotide tract, in AML and Burkitt’s lymphoma cell lines (16 and 18 bp respectively) was seen while, other cell lines showed minor loss and/or, gain which may have no effect on gene expression given the deep intrinsic location of the marker. The ASSP tool predicted presence of cryptic donor upstream to the AT-rich regions which though farfetched, may be recognized by spliceosome machinery generating a pseudo exon into mature transcript. Reduced expression of MSH3 was reported in AML earlier while deficiency of this protein made mice models susceptible to tumour development [27, 28]. The role of MSH3 mutations and reduced expression is reiterated in haematologic malignancies like the adult T-cell leukaemia and childhood ALL [29–31]. The dimerization of MSH3 with MSH2 (MutSβ) is specific for repair of INDELs, and intronic sequence perturbations may have overall negative outcome on mature mRNA, which calls for comprehensive research of MSH3 repeat variations in haematologic malignancies.

MSH4 gene is not studied extensively for tandem repeat instability in solid tumors or blood cancers. Here too, like MSH3, we designed a deep intronic AT-rich marker which showed no notable variation in repetitive sequences. The in silico analysis predicted that this approximate 30 nucleotide AT-rich region was accompanied by cryptic donors and acceptors. We observed the cell line Daudi to be moderately unstable for this gene. Though not likely, inefficient splicing might occur due to contraction and expansion of nucleotide repeats [32]. Yang and colleagues (2020) identified somatic mutations of MSH4 to be associated with microsatellite instability in metastatic bladder cancer [33] and proposed inclusion of this gene in IHC panel to refine the MSI testings. However, for the novel marker here, we suggest a comprehensive screening to generate valuable information on defective MMR that can surpass the prognostic barriers of haematologic cancers cannot be ruled out to eliminate emergence of MSI.

The MLH1 intronic marker D3S1611 showed prominent shift of CA-rich peaks in KG-1 of AML, and low to moderate changes in Daudi of Burkitt’s lymphoma, and HCT116 of CRC. This marker was earlier evaluated for MSI in cholangiocarcinoma wherein only one patient was microsatellite instable at D3S1611 loci. Later, the hypervariability in this marker was identified to influence MLH1 expression in solid tumors [38–40]. Here, we report the consequence of tandem repeat instability might deregulate MLH1 functions in myeloid and lymphoid malignancies. Being a central component of downstream processes of repair, loss of MLH1 collapses the MMR activity [41, 42]. It also carries cryptic donor and acceptor splice sites. The dinucleotide CA repeats are involved in regulation of splicing and can activate the 5’ splice sites. Hence, contraction and expansion of deep intrinsic repeats might abrogate mRNA splicing. We suggest that D3S1611 marker should be investigated further to evaluate the biological and clinical risk factors of lymphoid malignancies in correlation to damage repair and immunosuppression.

There was marginal loss and gain of (AT) repeats in KG-1 and MCF-7 cell lines respectively for MLH3 marker. In humans, the activity of MutLy involving MLH1 and MLH3 is minimal usually with functions in strand incision and endonuclease activity. Earlier a single base pair deletion in exonic mono-A repeat was identified in microsatellite instable CRC patients [43]. However, the mutator effect in MLH3 was found to be milder than other
MMR components. Mutations in this gene is majorly associated with Lynch syndrome [44] but rarely correlates to familial colorectal cancer. This observation was in accordance with our result, as here the CRC cell line HCT116 did not show any alterations across the repeat length. Somatic mutations of MLH3 are often found in microsatellite instable colon, rectal, and endometrial cancers [8]. A constitutive donor was identified at the immediate end of exon 5. This may implicate that variation in \((AT)_{18}\) might relocate

ble colon, rectal, and endometrial cancers [8] and being an integral part of mismatch repair system, there is rationale to consider MLH3 for an extensive analysis in myeloid and lymphoid malignancies.

The intronic AT rich marker of PMS1 gene showed low to moderate contraction and expansion of the repetitive tract in—Daudi of Burkitt lymphoma, KG-1 of AML, Molt-4 of T-ALL, and HCT116 of CRC. The region was identified to carry a constitutive donor. Hence, the repeat hypervariability might result in retention of mutated intron or impede early phase of transcriptional elongation. Loss of expression of PMS1 was observed in acute T-cell leukaemia cases [29]. Absence of PMS1 expression in ovarian cancer only occurs in absence of both MLH1 and MSH2 [45]. A mutant PMS1, alone or in association with other MMR genes, causes type 3 Lynch syndrome. The TR specific hypermutability of this gene has not been deciphered in blood cancers and the present study can serve as a primary setup to initiate a drive towards research on PMS1 in haematologic malignancies.

The exonic PMS2 marker showed 1 bp deletion in CML (K562), AML (KG-1), and breast cancer (MCF-7) cell lines. PMS2 in association with MLH1, edits the erroneous nucleotides and synthesizes error-free DNA chain at the site of mismatch. The exon 15 of PMS2 was identified as a pseudogene [46] wherein the current novel marker was located. Besides being a ‘silent relic’, many pseudogenes exhibit tissue-specific activation as they also regulate oncogenes and Tumour suppressor genes (TSGs) by acting as ‘microRNA decoy’ [47]. Minimal deregulation of this region mandates more research to decipher the extent of their involvement in blood cancer progression.

The degree of variability in intronic CA rich elements of MGMT was grossly insignificant. Disruption in pre-mRNA splicing could be a possibility due to activation of cryptic donors and acceptors adjacent to the dinucleotide repeats. Only 4% of somatic errors are responsible for loss of expression in MGMT, while promoter hypermethylation is most frequent phenomenon for the inactivation of MGMT, especially in MSI-low CRC related gene silencing [48, 49].

Here, we propose that large number of haematologic cancer samples should be studied for recurrence of tandem repeat length polymorphisms. Despite being low grade such repeats might set a novel foundation to outline a distinct molecular pathway in haematologic malignancies.

The ASSP analysis predicted percentage frequency of GC nucleotides in nine of ten repeat-rich markers (excluding PMS2). The GCs were found to be involved in splice-site recognition under normal condition. However, any contraction or expansion in intronic or exonic repeat regions lowers the capacity of splicing at the actual site. Such misrecognition can develop an abnormal transcript and translate into unwanted proteins [50]. The in silico splice site analysis also conferred presence of cryptic donors and acceptors. These, in general, are not utilized by spliceosome but upon mutation the cryptic splice site may get activated. Thus, large repeat expansion or deletion in the intronic regions as well as any change in the intron–exon junction, regulatory or exonic region might correlate with abnormal mRNA regulation.

The confidence with which the splice donors and acceptors were predicted for individual marker here, projected possible abnormal transcript formation ultimately affecting the repair mechanism in the genome. Though there are limitations in this study, owing to unavailability of control range for designed markers; nonetheless, the fact that three out of four blood cancer cell lines had repeat variations for most markers here, proves that genome stability was compromised.

Based on the research evidence it can be theorised that MMR deficiency may not only contribute to cancer onset in solid tumors but may also alter the landscape of haematologic malignancies by elevating the mutational rate and clonal evolution of cancer cells. The magnitude with which the repair genes showed intronic, exonic or regulatory TR fluctuations here, justifies misplaced splice sites and formation of irregular protein isoforms that undeniably change the inter and intra-functional networks to establish the disease state. While the MSI profiling and use of Bethesda markers are practiced routinely in diagnostics of solid tumors, the same approach needs to be contemplated for selective MMR deficient or therapy resistant blood cancer cases, essentially where the cytogenetics fails. In 2017, the FDA (US Food and Drug Administration) approved immune checkpoint blockade (ICB) therapy and use of Pembrolizumab for all MMR deficient or MSI-high solid tumors. Furthermore, three MMR inhibitors—methotrexate, FdCyd also known as 5-Fluoro-2′-deoxycytidine (to restore functionality), and Poly (to create synthetic lethality in cells deficient of MSH2/MLH1) along with some anti-PD-1 (Programmed death receptor) monoclonal antibodies such as—Avelumab, Durvalumab, Atezolizumab, and Nivolumab, alone or in combination are currently being investigated [51, 52]. We propose that the use of these medications can not only combat MMR infidelity, but also provide a survival advantage to the haematologic malignant patients as well. In future, drug related TRs variations in DDR genes including MGMT can also be investigated. Some facts of AML and MSI + phenotypes...
which we came a across while pondering over this query. DNA methyltransferase or histone deacetylase inhibitors are in fact a part of the therapy in AML. MSI + phenotype can occur in up to 50% of therapy related AML which is attributed to a loss of immunosurveillance resulting from transient or prolonged immunosuppression especially in cases of leukemia. Such mishaps in the genome occur post stem cell transplants or following chemo- or radio-therapy. It would be interesting to understand if repeat variations in DDR genes are linked to treatment resistance in future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-022-07761-y.

Acknowledgements We are extremely grateful to Penn State Cancer Institute, Hershey, PA 17033, USA for generous donation of HCT116 cell lines.

Funding We are thankful to Vellore Institute of Technology for providing the facilities and fund for research student without which the work would have not been completed.

Declarations

Conflict of interest The authors declares that they have no conflict of interest to disclose.

References

1. Yang G, Zheng RY, Jin ZS (2019) Correlations between microsatellite instability and the biological behaviour of tumours. J Cancer Res Clin Oncol 145:2891–2899. https://doi.org/10.1007/s00432-019-03053-4

2. Bhattacharya P, Patel TN (2018) Microsatellite instability and promoter hypermethylation of DNA repair genes in hematologic malignancies: a forthcoming direction toward diagnostics. Hematology 23:77–82. https://doi.org/10.1080/10245332.2017.1354428

3. Vieira ML, Santini L, Diniz AL, Munhoz Cde F (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 39:312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

4. Sulovari A, Li R, Audano PA, Vollger MR, Logsdon GA et al (2019) Human-specific tandem repeat expansion and differential gene expression during primate evolution. Proc Natl Acad Sci USA 116:23243–23253. https://doi.org/10.1073/pnas.1912175116

5. Bolton KA, Ross JP, Bowden NA, Holliday EG, Avery-Kiejda KA et al (2013) STaRRRT: a table of short tandem repeats in regulatory regions of the human genome. BMC Genomics 14:795. https://doi.org/10.1186/1471-2164-14-795

6. Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007. https://doi.org/10.1093/molbev/msi073

7. Li K, Luo H, Huang L, Luo LH, Zhu X (2020) Microsatellite instability: a review of what the oncologist should know. Cancer Cell Int 20:16. https://doi.org/10.1186/s12935-019-1091-8

8. Pečina-Slaus N, Kafka A, Salamon I, Bukovac A (2020) Mismatch repair pathway, genome stability and cancer. Front Mol Biosci 7:122. https://doi.org/10.3389/fmolb.2020.00122

9. Modrich P, Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65:101–133. https://doi.org/10.1146/annurev.bi.65.070196.000533

10. Chikan NA, Bukhari S, Shabir N, Amin A, Shaﬁ S, Quadri KA et al (2015) Atomic insight into the altered O6-methylguanine-DNA methyltransferase protein architecture in gastric cancer. PLoS ONE 10:e0127741. https://doi.org/10.1371/journal.pone.0127741

11. Rye PT, Delaney JC, Netrojananukul C, Sun DX, Liu JZ, Eissmann JM (2008) Mismatch repair proteins collaborate with methyltransferases in the repair of O(6)-methylguanine. DNA Repair (Amst) 7:170–176. https://doi.org/10.1016/j.dnarep.2007.09.003

12. Akagi K, Oki E, Taniguchi H, Nakatani K, Aoki D, Kuwata T et al (2020) Nationwide large-scale investigation of microsatellite instability status in more than 18,000 patients with various advanced solid cancers. J Clin Oncol 38:803–803. https://doi.org/10.1200/JCO.2020.84.8_suppl.803

13. Magalhães M, Farre L, Bittencourt A (2011) Microsatellite instability in the smoldering form of adult T-cell leukemia/lymphoma (ATL) in Brazil. Retrovirology 8:A148. https://doi.org/10.1186/1742-4690-8-S1-A148

14. Miyashita K, Fujii K, Taguchi K, Shimokawa M, Yoshiha MA, Abe Y et al (2017) A specific mode of microsatellite instability is a crucial biomarker in adult T-cell leukemia/lymphoma patients. J Cancer Res Clin Oncol 143:399–408. https://doi.org/10.1007/s00432-016-2294-1

15. Sanz-Vaqué L, Colomer D, Bosch F, López-Guillermo A, Dreyling MH, Bosch F et al (2001) Microsatellite instability analysis in typical and progressed mantle cell lymphoma and B-cell chronic lymphocytic leukemia. Haematologica 86:181–186

16. Walker CJ, Eifisd AF, Genutis LK, Bainazar M, Kohlschmidt J, Mrózek K et al (2017) No evidence for microsatellite instability in acute myeloid leukemia. Leukemia 31:1474–1476. https://doi.org/10.1038/leu.2017.97

17. Losso GM, Moraes Rda S, Gentili AC, Messias-Reason IT (2012) Microsatellite instability—MSI markers (BAT26, BAT25, D2S123, D5S346, D17S250) in rectal cancer. Arq Bras Cir Dig 25:240–244. https://doi.org/10.1590/1981-7718-4685-2012000400006

18. Liangswangwong U, Nitta T, Kashiwagi H, Kikukawa H, Kawanoto T, Todokoro T et al (2003) Infrequent microsatellite instability in liver fluke infection-associated intrahepatic cholangiocarcinomas from Thailand. Int J Cancer 107:375–380. https://doi.org/10.1002/ijc.11380

19. Lu Y, Soong TD, Elemento O (2013) A novel approach for characterizing microsatellite instability in cancer cells. PLoS ONE 8:e63056. https://doi.org/10.1371/journal.pone.0063056

20. Nugoli M, Chuchana P, Vendrell J, Orsetti B, Ursule L, Nguyen C et al (2003) Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications. BMC Cancer 3:13. https://doi.org/10.1186/1471-2407-3-13

21. Zhang J, Kuo CC, Chen L (2011) GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics 12:90. https://doi.org/10.1186/1471-2164-12-90

22. Evrard C, Tachon G, Randrian V, Karayan-Tapon L, Tougeron D (2013) V12007, a polyamine transporter, affects splicing through pre-mRNA secondary structures. BMC Genomics 12:90. https://doi.org/10.1186/1471-2164-12-90

23. Nojadeh JN, Behrouz Sharif S, Sakhinia E (2018) Microsatellite instability markers (BAT26, BAT25, D2S123, D5S346, D17S250) in rectal cancer. Arq Bras Cir Dig 25:240–244. https://doi.org/10.1590/1981-7718-4685-2012000400006

24. Adem C, Soderberg CL, Cunningham JM, Reynolds C, Sebo TJ, Marrone JM et al (2020) Mismatch repair gene DNA methylation in gastric cancer. PLoS ONE 10:e0127741. https://doi.org/10.1371/journal.pone.0127741

25. Rimsa LM, Kopecky KJ, Ruschulte J, Chen IM, Slovak ML, Karanes C et al (2020) Microsatellite instability is not a defining
genetic feature of acute myeloid leukaemogenesis in adults: results of a retrospective study of 132 patients and review of the literature. Leukemia 14:1044–1051. https://doi.org/10.1038/sj.leu.2401699

26. Nomdedéu JF, Perea G, Estivill C, Lasa A, Carnicer MJ, Bru-Net S et al (2005) Microsatellite instability is not an uncommon finding in adult de novo acute myeloid leukaemia. Ann Hematol 84:368–375. https://doi.org/10.1007/s00277-005-1035-3

27. Inokuchi K, Ikekima M, Watanabe A, Nakajima E, Orimo H, Nomura T et al (1995) Loss of expression of the human MSH3 gene in hematological malignancies. Biochem Biophys Res Comm 214:171–179. https://doi.org/10.1006/bbrc.1995.2271

28. Edelmann W, Umar A, Yang K, Heyer J, Kucherlapati M, Lia M et al (2000) The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res 60:803–807

29. Morimoto H, Tsukada J, Kominato Y, Tanaka Y (2005) Reduced expression of human mismatch repair genes in adult T-cell leukaemia. Am J Hematol 78:100–107. https://doi.org/10.1002/ajh.20259

30. Matheson EC, Hogarth LA, Case MC, Irving JA, Hall AG (2007) DHFR and MSH3 co-amplification in childhood acute lymphoblastic leukaemia, in vitro and in vivo. Carcinogenesis 28:1341–1346. https://doi.org/10.1093/carcin/bgl235

31. Nuuri K, Arroyo-Berdugo Y, Esposito MT (2020) Analysis of expression of MSH2 and MSH3 genes in Mixed Lineage Leukemia (MLL)–Acute Lymphoblastic Leukemia (ALL) human cell lines. J Young Investig 37:38–48

32. Rajewska M, Wegryn K, Konieczny I (2012) AT-rich region and repeated sequences—the essential elements of replication origins of bacterial replicons, FEMS. Microbiol Rev 36:408–434. https://doi.org/10.1111/j.1574-6976.2008.00130.x

33. Yang Y, Jain RK, Glenn ST, Xu B, Singh PK, Wei L et al (2020) Complete response to anti-PD-L1 antibody in a metastatic bladder cancer associated with novel MSH4 mutation and microsatellite instability. J Immunother Cancer 8:e000128. https://doi.org/10.1136/jitc-2019-000128

34. Clark N, Wu X, Her C (2013) MutS homologues hMSH4 and hMSH5: genetic variations, functions, and implications in human diseases. Curr Genomics 14:81–90. https://doi.org/10.2174/138902911340200020

35. Sekine H, Ferreira RC, Pan-Hammarström Q, Graham RR, Ziemma B, de Vries SS et al (2007) Role for Msh5 in the regulation of Ig class switch recombination. Proc Natl Acad Sci U S A 104:7193–7198. https://doi.org/10.1073/pnas.0700815104

36. de Valles-Báñez G, Esteve-Solé A, Piquer M, González-Navarro EA, Hernandez-Rodriguez J et al (2018) Evaluating the genetics of common variable immunodeficiency: monogenetic model and beyond. Front Immunol 9:636. https://doi.org/10.3389/fimmu.2018.00636

37. Graham RP, Kerr SE, Butz ML, Thibodeau SN, Halling KC, Smyrk TC et al (2015) Heterogenous MSH6 loss is a result of microsatellite instability within MSH6 and occurs in sporadic and hereditary colorectal and endometrial carcinomas. Am J Surg Pathol 39:1370–1376. https://doi.org/10.1097/PAS.0000000000000459

38. Ollikainen M, Hannelius U, Lindgren CM, Abdel-Rahman WM, Kere J, Peltomäki P (2007) Mechanisms of inactivation of MLH1 in hereditary nonpolyposis colorectal carcinoma: a novel approach. Oncogene 26:4541–4549. https://doi.org/10.1038/sj. onc.1210236

39. Hutter P, Wijnen J, Rey-Berthod C, Thiffault I, Verkuiljen P, Farber D et al (2002) An MLH1 haplotype is over-represented on chromosomes carrying an HNPCC predisposing mutation in MLH1. J Med Genet 39:323–327. https://doi.org/10.1136/jmg.39.5.323

40. Limpaiboon T, Krissadarak K, Sripa B, Jearanaikoon P, Bhuhsawadi V, Chau-in S et al (2002) Microsatellite alterations in liver fluke related cholangiocarcinoma are associated with poor prognosis. Cancer Lett 181:215–222. https://doi.org/10.1016/s0304-3835(02)00052-6

41. Leekema OC, de Miranda NF, Veelken H (2017) Germline mutations predisposing to diffuse large B-cell lymphoma. Blood Cancer J 7:e532. https://doi.org/10.1038/jbc.2017.15

42. Bakry D, Aronson M, Durro C, Rimawi H, Farah R, Alharbi QK et al (2014) Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. Eur J Cancer 50:987–996. https://doi.org/10.1016/j.ejca.2013.12.005

43. Loukola A, Vilikki S, Singh J, Launonen V, Aaltonen LA (2000) Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer. Am J Pathol 157:347–352. https://doi.org/10.1016/s0002-9440(06)4546-4

44. Bacher JW, Clipson L, Steffen LS, Halberg RB (2016) Microsatellite instability and its significance to hereditary and sporadic cancer. Microsatellite markers. IntechOpen, London. https://doi.org/10.5772/65065

45. Geisler JP, Goodheart MJ, Sood AK, Holmes RJ, Hatterman-Zogg M, Buller RE (2003) Mismatch repair gene expression defects contribute to microsatellite instability in ovarian carcinoma. Cancer 98:2199–2206. https://doi.org/10.1002/cncr.11770

46. Sugano K, Nakajima T, Sekine S, Taniguchi H, Saito S, Takanashi M et al (2011) Germline PMS2 mutation screened by mismatch repair protein immunohistochemistry of colorectal cancer in Japan. Cancer Sci 107:1677–1686. https://doi.org/10.1111/cas.13073

47. Pink RC, Wicks K, Caley DP, PunchEK, Jacobs L, Carter DR (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17:792–798. https://doi.org/10.1261/rna.265831

48. Whitehall VL, Walsh MD, Young J, Leggett BA, Jass JR (2001) Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. Cancer Res 61:827–830

49. Fox EJ, Leahy DT, Geraghty R, Mulcahy D, Hyland JM et al (2006) Mutually exclusive promoter hypermethylation patterns of hMLH1 and O6-methylguanine DNA methyltransferase in colorectal cancer. J Mol Diagn 8:68–75. https://doi.org/10.2353/jmoldx.2006.050084

50. Amit M, Donyo M, Hollandar D, Goren A, Kim E, Gelfman S et al (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep 1:543–556. https://doi.org/10.1016/j.celrep.2012.03.013

51. Zhao P, Li L, Jiang X, Li Q (2019) Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. J Hematol Oncol 12:54. https://doi.org/10.1186/s13045-019-0738-1

52. Kelley MR, Logsdon D, Fishel ML (2014) Targeting DNA repair pathways for cancer treatment: what’s new? Future Oncol 10:1215–1237. https://doi.org/10.2217/fon.14.60

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.