Comparing patterns of microsatellite instability in mismatch repair deficient tumours

Aaron Welson
Supervisor: Mauro Santibanez Koref

Abstract

testtest

1 Introduction

1.1 Mismatch Repair System

Polymerases, which replicate DNA during cell division, are not error-free. Eukaryotic polymerases are estimated to make one mistake for every 10^5 nucelotides (1). Fortunately (some) polymerases come equipped with proofreading capabilities, and beyond that there exists the mismatch repair (MMR) system; a set of evolutionarily conserved proteins working together to find and fix these mistakes. Specifically, the MMR system deals with mispaired bases or insertion/deletion (indel) loops formed by polymerase slippage events (2).

The simplest and most well-studied MMR system exists in *E.coli*. The main players consists of three proteins: MutS, MutL, and MutH (3). Briefly, MutS binds to the location of a mismatch/indel loop, then MutL is recruited which in turn recruits MutH. MutH has endonuclease activity and forms a nick in the newly synthesized daughter strand which allows exonuclease enzymes to begin breaking down the strand towards the direction of the mismatch site. The fix is complete after DNA polymerase resynthesizes the daughter strand, and DNA ligase ligates the nick.

In humans, the MMR system consists of the proteins: MSH2, MSH3, MSH6, PMS1, PMS2, MLH1, and MLH3, which pair up with each other to create MutS and MutL-related proteins (MutH homologs have not been detected). MutS homologs are built by pairing MSH2 with either MSH6 or MSH3. For MutL, MLH1 is paired with either PMS2, PMS1, or MLH3 (3). The two MutS homologs, MSH2-MSH3 (also called MutSβ) and MSH2-MSH6 (MutSα) both recognise and bind to mistakes in DNA, but with differing preferences. Mispaired bases are mostly recognized by MSH2-MSH6 while large indels are mostly recognized by MSH2-MSH3. Small indels are recognized by both MSH2-MSH6 and MSH2-MSH3 (2). MSH2-MSH6 is more important for MMR as evidenced by it being 10 times more abundant than MSH2-MSH3 (4). As for the MutL homologs, MLH1-PMS2 (MutLα) is the most important for MMR activity. It is recruited by MSH2-MSH6 (or MSH2-MSH3) to the error site where it uses its endonuclease activity to create nicks around the region of the error site. Similar to *E.coli*, Exo1 is recruited to the nick where it digests the error-containing strand. DNA polymerase resynthesizes the strand and the nick is ligated by DNA ligase (2). Not much is known about the other two homologs : MLH1-PMS1 (MutLβ) and MLH1-MLH3 (MutLγ). While there is a functional overlap between MLH3 and PMS2 (5), PMS1's role in MMR is still unclear. See Figure 1 for an overview.

The MMR system also plays a role in the DNA damage (e.g, from methylating agents) response, initiating an apoptotic cascade upon detection of an unnatural DNA adduct (6). It also plays a role in preventing homeologous recombination (7), and more. However, these "non-canonical" functions will not be discussed in depth here.

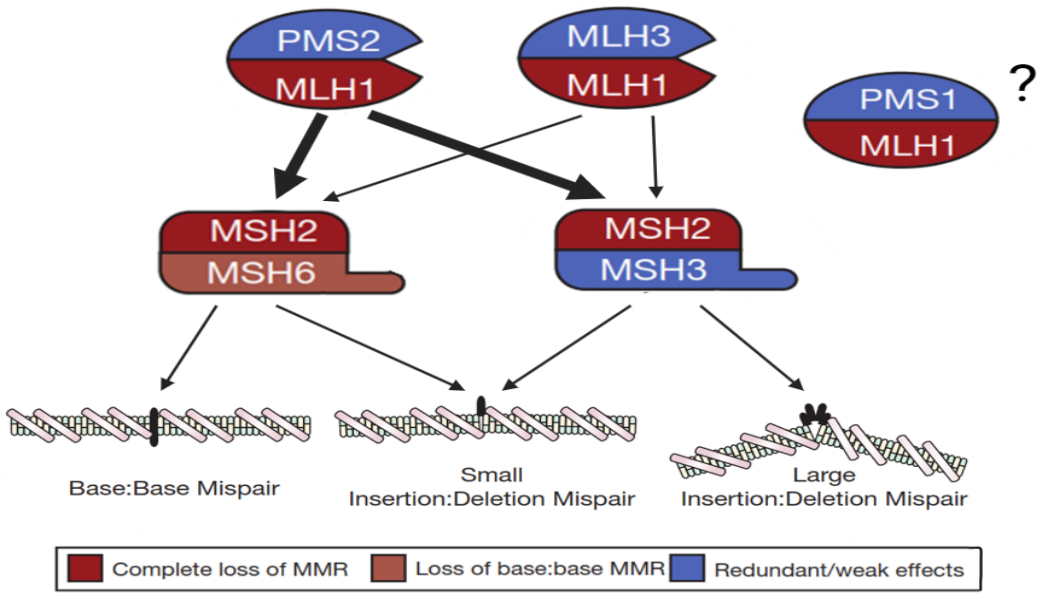

Figure 1: Overview of the human MMR system. Error recognition is done by the MutS homologs: MSH2-MSH6 and MSH2-MSH3 with their corresponding preferences for error type indicated by the arrows. The MutL homologs: MLH1-PMS2 and MLH1-MLH3 are recruited to the error site, with MLH1-PMS2 functioning as the major exonuclease in the MMR process. Nothing conclusive can be said about MLH1-PMS1 yet. Color indicates the effect of protein loss. Figure adapted from (2).

1.2 MMR in Cancer

The mutator hypothesis states that an increased mutational burden plays an important role in cancer initiation and progression. This would account for the high levels of heterogeneity within tumors, and their ability to gain resistance to therapy in a selection-based fashion (8). If mispaired bases and indel loops remain unfixed until the next replication cycle, one of the daughter cells will acquire a base/indel mutation. Thus, cells which are MMR deficicent (MMRd) are more susceptible to gaining mutations, and hence becoming cancerous (by the mutator hypothesis). Although MMRd is rarely present in most cancers, it plays a significant role in colorectal and endometrial cancers (CRC and EC), appearing in 15% and 30% of cases respectively (9).

The two main causes of MMRd are: 1) biallelic hypermethylation of MLH1 leading to epigenetic silencing, 2) germline mutation of MMR proteins. Other causes include somatic MMR mutations, but these are more rare (10). The first option leads to the development of sporadic MMRd tumors, the second characterizes a heritable disease known as Lynch syndrome (LS). Specifically, LS is characterized by an autosomal dominant germline mutation in one of : MLH1, MSH2, MSH6, PMS2 (or rarely a deletion in EPCAM which leads to the epigenetic silencing of MSH2) (11). MLH1 and MSH2 mutations are the most common, both accounting for 60-80% of LS cases (11). LS can eventually lead to MMRd if additional events, e.g. mutations or promoter hypermethylation, were to inactivate the remaining wild-type gene, a mechanism known as the "two-hit hypothesis" (12). LS results in a predisposition to various cancers, mainly colorectal cancer (CRC) and EC with a lifetime risk of 90% and 40% respectively (13). Other cancers (prostate, pancreatic, stomach, etc...) are also associated with LS. Alternatively, when a germline MMR mutation is present in both alleles this leads to a condition known as constitutional MMRd (CMMRD). It is an extremely rare (1:1,000,000) childhood cancer predisposition syndrome (14).

1.3 Microsatellite Instability

Microsatellites are regions within the genome consisting of short, tandemly repeated DNA sequences (usually 1-6bp in length). These regions are susceptible to length mutations because indel loops may form via strand slippage (15) (Figure 2). Normally, MMR prevents these mutations from taking place so that there is low variation in microsatellite lengths between cells from the same sample, this is known as microsatellite stability (MSS). Thus, a high variation in microsatellite lengths, dubbed microsatellite instability (MSI), can be used as an indicator of MMRd. In fact, there is no clinical distinction between MMRd and MSI and these terms are often used interchangably or together in literature.

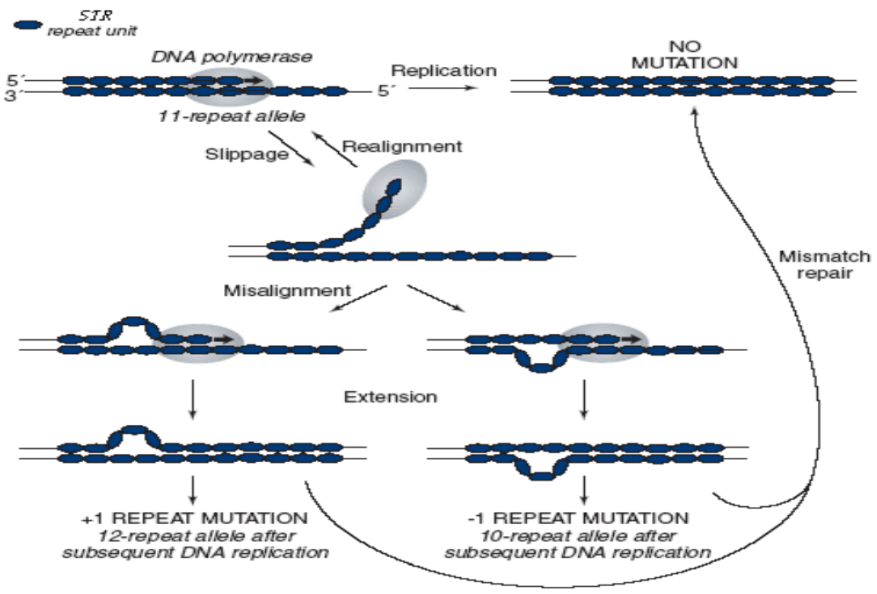

Figure 2: Mechanism of microsatellite length mutation. Polymerase slippage during DNA replication causes misalignment in either the template or the daughter strand. Consequently, the daughter strand may synthesize one additional or one fewer repeat unit than required, respectively. If left uncorrected by MMR, this misalignment leads to an indel mutation during subsequent replications. Figure taken from (16).

1.4 Clinical Significance

Despite the increased mutational burden, patients with MMRd/MSI tumors tend to have a better prognosis compared to their MSS counterparts. This was the case in Gryfe et al's analysis (17) of 607 CRC patients. Notably, metastasis is reduced in MMRd/MSI patients (odds ratio, $0.49; p = 0.02$). It is thought that MSI leads to an increased occurence of frameshift mutations which generates neoantigens that can be easily recognized by cytotoxic T-cells, resulting in an anti-tumor immune response (18). Usually, MMRd/MSI tumors suppress this immune activity through overexpression of immune checkpoint proteins like PDL1 which interacts with PD-1 receptors on exhausted T-cells, attenuating their activity (19). This mechanism can be prevented with antibodies against PD-1 like pembrolizumab, whose effectivity against MMRd/MSI CRC has been proven in late-stage clinical trials (20). The same treatment is not applicable to MSS tumors since they rarely express immune checkpoint proteins (21).

Conversely, MMRd/MSI tumors are resistant to 5-FU therapy but effective in MSS tumors (22). This is because MutSα binds 5-FU incorporated into DNA with high affinity and triggers apoptosis, this of course would not happen in MMRd tumors (23). MSI/MSS classification in CRC and EC tumors is thus important for tailoring personalized treatment strategies as well as informing patient prognosis.

1.5 Testing for MSI/MMRd

Currently there are two widely used techniques to screen for tumors with MSI/MMRd (24): 1) Immunohistochemistry (IHC), which uses antibodies against the 4 major MMR proteins to measure their expression, a lack of which would suggest MMRd. 2) MSI testing, which involves PCR amplification of specific microsatellite markers. The frequencies of the various length alleles are then compared between normal and tumor samples from the same individual. Alternatively, quasi-monomorphic microsatellite markers may be used. These are microsatellites whose major (also known as reference) allele is present with frequency ≥ 0.95 in the population (25). This means that PCR on normal tissues is unnecessary since the reference allele is assumed to be the same as seen in the hg19 library. Usually, allele lengths are referred to by the difference in length to the reference allele. For example, the relative allele has length zero, a 1bp deletion results in an allele of length -1, and so on.

Although both techniques are concordant in CRC (99.6% concordance) (26), MSI testing has a higher false negative rate in EC specifically (27). However, IHC is also problematic because some mutations may result in non-functional, but otherwise stable and antigenic MMR proteins. This will lead to a false negative result. These constitute about 6% of MSI/MMRd cases (28).

1.6 Aims

There is a wide spectrum of variation between MSI/MMRd tumors which may be relevant either clinically or diagnostically. For instance, MSI/MMRd may be separated into cases which are sporadic or hereditary (LS). LS can be further subdivided based on the affected gene. There is plenty of evidence to suggest that these MSI/MMRd subtypes behave differently. For example, the penetrance of CRC in LS is dependent on the affected gene: MLH1;48%, MSH2;44%, MSH6;16%, PMS2;3% (11). In line with its weak penetrance, MSI phenotype in MSH6 mutants is less striking than in MLH1 and MSH2 (29). Clinically, mutational MSI/MMRd tumors respond better to PD1-inhibitors compared to epigenetic MSI/MMRd tumors (30).

These differences suggest that MSI/MMRd tumors may have distinct MSI profiles based on subtype. In this paper, patterns of microsatellite instability were compared between LS and sporadic MSI/MMRd tumors in CRC and EC. The possibility of using these differences to infer the genetic background of MSI/MMRd tumors was also explored.

2 Methods and Data

2.1 Data

CRC and EC cohort data were obtained from publically accessible sources (31, 32, 33, 34). This data included MSS/MSI status of each tumor as determined by the Promega MSI 1.2 assay (35). For the purposes of analysis, the reported MSS/MSI status was assumed to be true. Furthermore, MSI tumors were screened for LS via germline genetic testing, allowing further separation into Lynch and sporadic cases. The data also included PCR read counts (for each tumor sample) of 24 quasi-monomorphic mononucleotide microsatellite markers (Table 1) (36).

After quality control, the CRC cohort included 85 MSI (37 Lynch, 48 Sporadic) and 47 MSS samples. For EC, there were 196 MSI (38 Lynch and 158 sporadic) and 157 MSS samples. Lynch patients were further subdivided based on the affected gene (Table 2). In this paper, MLH1-LS tumors are referred to as just MLH1 tumors and respectively for the other three genes. Similarly, sporadic MSI/MMRd tumors are referred to as just sporadic tumors.

DEPDC2	GM01	GM07	GM09
GM11	GM14	GM17	GM22
GM26	GM29	IM16	IM49
LR10	LR11	LR17	LR20
LR24	LR36	LR40	LR44
LR46	LR48	LR49	LR52

Table 1: Quasi-monomorphic microsatellite markers used for analysis. For more information about these markers see Redford et al (36).

MLH1		**MSH2**		**MSH6**		**PMS2**	
CRC	**EC**	**CRC**	**EC**	**CRC**	**EC**	**CRC**	**EC**
13	8	19	14	3	13	2	3

Table 2: Lynch subtypes (according to affected gene) in CRC and EC cohort after quality control.

2.2 Data pre-processing

For each marker, the reads were grouped by allele lengths (this grouping is a marker-allele combination). These allele lengths were standardized relative to the length observed in the hg19 library (as described in the introduction). For each marker-allele combination, the allele frequency (AF) is calculated by dividing the allele read count with the total marker read count. The variant frequency (VF) is also calculated. This is a similar measure that excludes the reference read count from the total marker read count. VF is hypothesized to provide a more accurate representation of allele distribution since infiltrating lymphocytes may contaminate tumor samples (especially in MSI/MMRd) (37), thus artificially inflating the reference read count .

For data analysis, each sample is characterized by a vector of AF and/or VF values of each marker-allele combination. Furthermore, only alleles with lengths between -2 and 1 were considered as other variants do not have sufficient read

counts (at least 75% of reads are zero). Unless otherwise specified, data analysis was done separately for AF and VF, and separately for CRC and EC.

2.3 Principal Component Analysis (PCA)

PCA was performed on the AF vectors of each sample. This technique provides a representation (emphasizing variance) of the datapoints in a low-dimensional setting, allowing observations of how AF influences variation in the dataset. This was used to gain an initial understanding of how AF profiles relate to tumor subtypes in CRC and EC.

2.4 Ratio Analysis

This analysis involves pairwise comparisons of subtypes in CRC/EC. 7 pairwise comparisons were included : MSI/MSS, Lynch/sporadic, MLH1/sporadic, MSH2/sporadic, MSH6/sporadic. PMS2 was excluded due to low sample size.

For each marker-allele combination, a ratio is calculated using the formula :

$$r = \frac{x}{x+y},$$

where x, y are the median values for the frequencies (AF or VF) in the pairwise comparison. For example, in the MSI/MSS comparison x is the median frequency of the MSI samples, and likewise for y with the MSS samples. Only marker-allele combinations where both $x, y > 0$ were included to avoid the creation of extreme data points.

Theoretically we expect $x, y > 0$ anyway and instances where a zero is observed is the result of insufficient sequencing depth. This justifies the exclusion of these extreme data points. These ratios were grouped by allele length and their distribution was visualized using boxplots. A one-sample Wilcoxon test was used to assess statistical significance. If there is no subtype difference, then we would expect the ratios to not deviate from 0.5; this is the null hypothesis. Alternatively, clustering of ratios above 0.5 would indicate that the corresponding allele is more frequent in the first subtype over the second subtype (and vice versa for $r < 0.5$). Using the MSI/MSS comparison as an example, if the ratios in the -1 allele group cluster above 0.5, this would indicate that 1bp microsatellite deletions are more prevalent in MSI than MSS.

2.5 Classification

In CRC, LS/sporadic tumor discrimination is done via somatic BRAF gene testing, where the presence of BRAF.1779T>A is strong evidence for sporadic origin (10). The performance of this method was compared to a random forest model which was trained on AF and VF values. BRAF genotype data was also later included as part of the training data to see if the model could be improved.

For EC, LS/sporadic tumor discrimination involves MLH1 promoter methylation analysis (10). However, this data was not available so a comparison of methods could not be made.

Only AF and VF values from significant marker-allele combinations (selected based on p-values from a Wilcoxon test) were chosen for training the random forest model. Specifically, 10 were chosen based on significance in AF, alongside another 10 based on VF. Leave-one-out-cross-validation (LOOCV) (38) was then used to estimate performance of the model. Because the prediction is made on a sample that is not used in training, LOOCV is a reliable and unbiased way to measure performance.

Additionally, volcano plots were made to visualize the distribution of these significant marker-allele combinations. The directionality of the significance is given by the one-directional AUC value (x-axis of the volcano plot). A low AUC indicates that the marker-allele is more frequent in LS than sporadic, and vice versa.

2.6 Software

3 Results

3.1 PCA

In CRC, a clear separation between MSI and MSS tumors is observed along the PC1 axis (Figure 3A). This indicates that AF profiles differ between these two groups. Focusing on the MSI samples (Figure 3B) there is no clear between Lynch and sporadic tumors. However, Lynch tumors tend to cluster more towards the left compared to sporadic cases.

This is evident by their centroid (aggregate position of data points) PC1 coordinate (LS;0.227, sporadic;0.388). The closer proximity of Lynch tumors to MSS tumors in the PCA plot suggests that these two groups are more similar (in terms of their AF profile) compared to sporadic tumors, i.e, the degree of microsatellite instability is less severe in Lynch than sporadic tumors. Within the Lynch subtypes (Figure 3C) it can be seen that MSH2 and MSH6 samples are what contributes to the leftward grouping in Lynch whereas the other genes have a more varied spread.

In EC, the separation between MSI and MSS samples is less clear (Figure 2A). Unlike CRC, a considerable number of MSI samples overlap with the MSS cluster, this may indicate a low degree of microsatellite instability within these samples. Like CRC, Lynch samples in EC also tends to cluster leftwards (centroid PC1 coordinate = 0.109) compared to sporadic samples (0.278), with MLH1 and MSH6 mainly contributing to this leftward grouping.

3.2 Ratio Analysis

Unlike the PCA plots, the ratio analysis allows us to compare AF and VF values in CRC/EC subtypes on a per-allele basis. In MSI/MSS (Figure 3A:D), deletions (negative alleles) tend to have ratios significantly greater than 0.5 and vice versa for insertions (positive allele). This suggests that the instability in MSI tumors is mainly due to deletions instead of insertions. In fact, the results seem to suggest that insertions occur more frequently in MSS than MSI tumors. The frequency of the reference allele is also decreased in MSI compared to MSS tumors. Thus, MSI can be characterized by the pattern: $r > 0.5$ in negative alleles and $r < 0.5$ in non-negative alleles. The MSI pattern is more pronounced/visible in CRC compared to EC, at least for the AF ratios.

In Lynch/sporadic (3E:H), deletions (-2 in CRC, -1 in EC) have ratios significantly less than 0.5 and vice versa for insertions, i.e, LS tumors present with less deletions but more insertions when compared to sporadic tumors. This is a reversal of the MSI pattern, suggesting that MSI in LS tumors is less severe compared to sporadic tumors.

Looking at Figure 4, it is clear that Lynch subtypes behave differently based on cancer type. This is most clear with MSH2 (4E:H) where the MSI pattern is observed in EC, but the reverse is seen in CRC. Furthermore, The ratios in the MLH1/sporadic comparison are mostly nonsignificant in CRC (4A:B), but the same cannot be said about EC (4C:D). Meanwhile MSH6 behaves consistently in CRC and EC, where the "reverse MSI pattern" is observed in both CRC and EC (4I:L).

There does not seem to be much difference between AF and VF except that switching to VF tends to exaggerate ratio values in insertions whilst diminishing them in deletions.

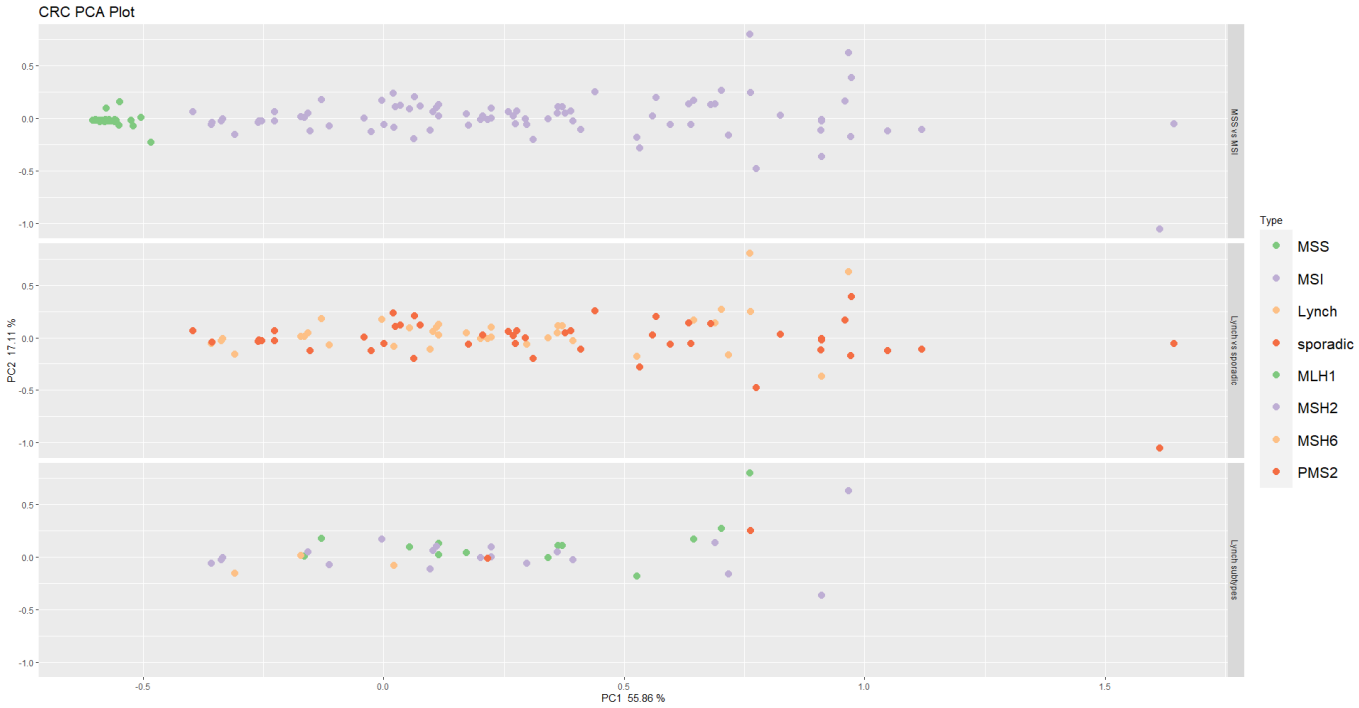

Figure 3: **PCA plots of tumor AFs in CRC**. All three plots are the same principal components projection. **A**, MSS and MSI comparison. **B**, MSI samples with color indicating Lynch or sporadic cases. **C**, LS samples colored by affected gene.

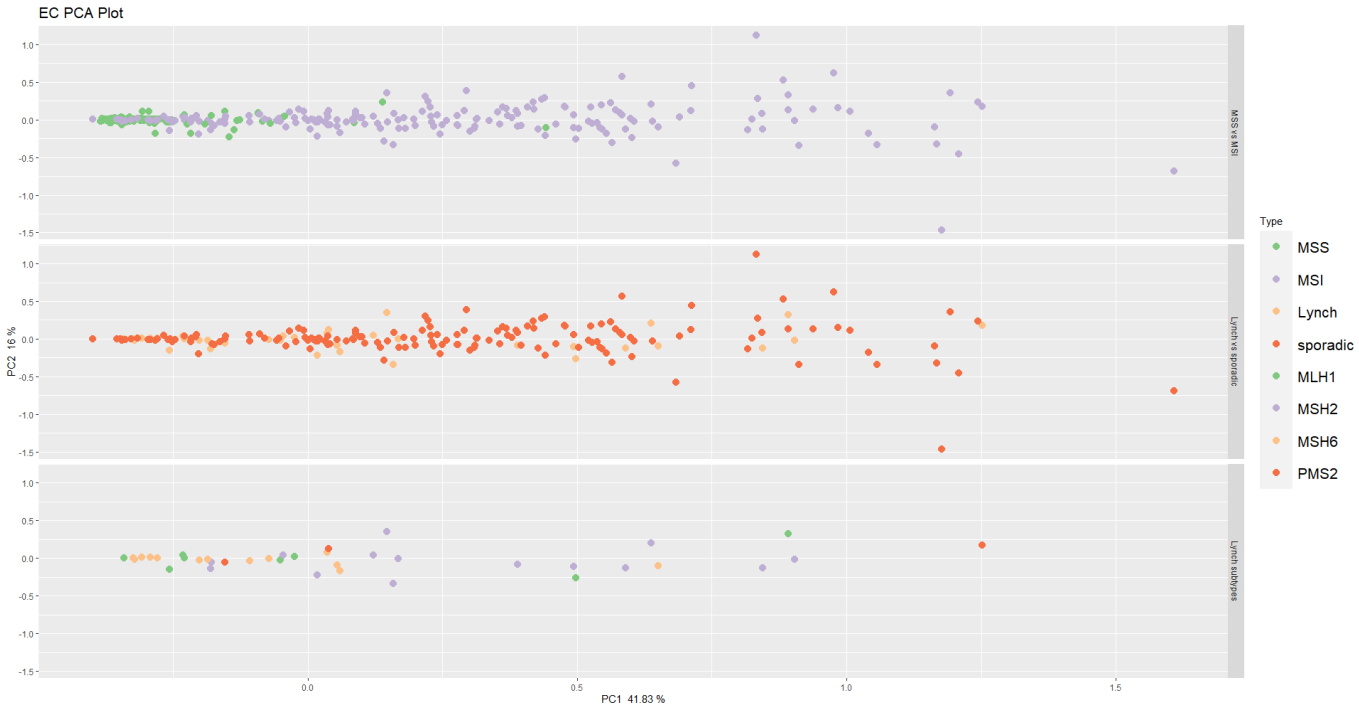

Figure 4: **PCA plot of tumor AFs in EC**. Same description as in Figure 1.

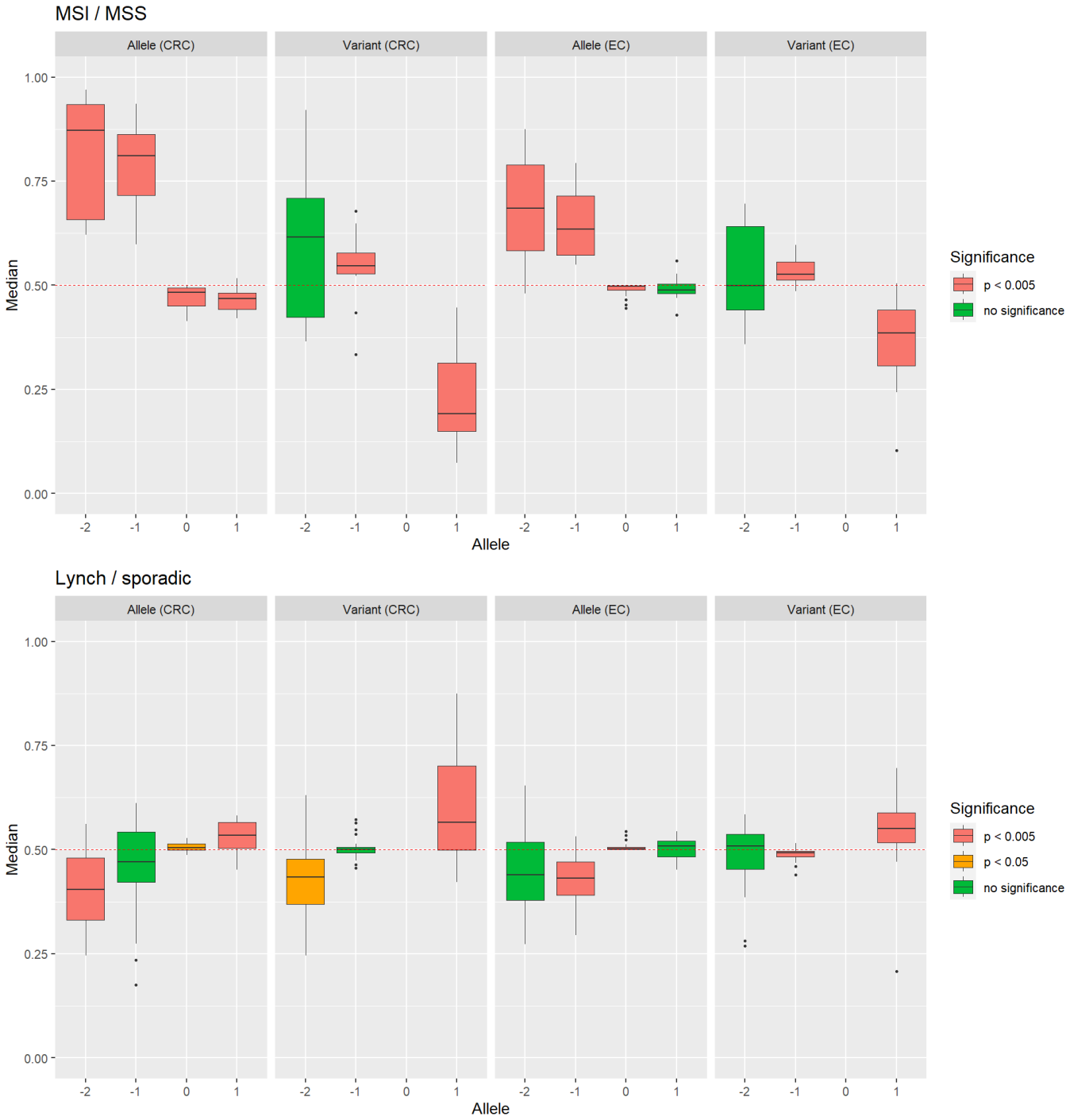

Figure 5: **Ratio plots of MSI/MSS and Lynch/sporadic in CRC and EC. A:D**, MSS/MSI ratios of AF and VF in CRC and EC grouped by allele. **E:H**, Lynch/sporadic ratios.

Figure 6: **Ratio plots of Lynch subtype / sporadic in CRC and EC. A:D**, MLH1/sporadic ratios of AF and VF in CRC and EC grouped by allele. **E:H**, MSH2/sporadic ratios. **I:L**, MSH6/sporadic ratios.

3.3 Classification

p-values are smaller in AF than VF in both CRC and EC (Tables 3-4), thus it appears that AF is more relevant for LS/sporadic discrimination than VF. Interestingly, the overlap in significant marker alleles between CRC and EC is weak (1/10 matching marker-alleles in the AF group, 2/10 in VF). This suggests that markers are differentially affected based on cancer type in MSI tumors.

The volcano plot (Figure 7) shows that significant marker-alleles are more abundant in CRC than EC. It is also clear that the direction of significance is distinct between the negative and non-negative alleles, in line with the boxplots from the ratio analysis (Figure 5E:H).

Unsurprisingly, performance of the random forest model is weaker in EC than CRC (Table 5). However, the random forest model still underperforms BRAF testing in CRC. When BRAF genotype is included as part of the training data, there is an improvement with specificity, but sensitivity is reduced (Table 5).

Marker	Allele	p value	By	Marker	Allele	p value	By
LR52	-1	2.837×10^{-4}	AF	GM07	1	2.559×10^{-4}	VF
GM07	1	3.366×10^{-4}	AF	GM14	1	2.131×10^{-3}	VF
LR24	1	3.985×10^{-4}	AF	LR52	-2	3.116×10^{-3}	VF
LR11	1	1.781×10^{-3}	AF	LR24	1	3.912×10^{-3}	VF
GM14	1	4.255×10^{-3}	AF	LR52	-1	4.889×10^{-3}	VF
GM11	0	8.149×10^{-3}	AF	LR44	-2	9.524×10^{-3}	VF
GM11	-1	8.365×10^{-3}	AF	GM11	1	1.497×10^{-2}	VF
IM49	-1	8.812×10^{-3}	AF	IM49	-1	2.955×10^{-2}	VF
GM14	-2	1.732×10^{-2}	AF	LR40	1	2.988×10^{-2}	VF
LR24	0	1.774×10^{-2}	AF	LR44	-1	3.303×10^{-2}	VF

Table 3: Selection of significant marker-allele combinations in CRC chosen for random forest training

Marker	Allele	p value	By	Marker	Allele	p value	By
LR52	0	2.885×10^{-3}	AF	GM11	-2	8.407×10^{-3}	VF
GM11	-2	5.066×10^{-3}	AF	GM07	-1	1.170×10^{-2}	VF
LR11	-1	1.068×10^{-2}	AF	LR11	1	1.292×10^{-2}	VF
LR52	-2	1.088×10^{-2}	AF	LR44	-1	1.787×10^{-2}	VF
GM07	-1	1.170×10^{-2}	AF	GM29	1	1.850×10^{-2}	VF
LR52	-1	1.437×10^{-2}	AF	LR11	-1	2.164×10^{-2}	VF
LR48	0	1.727×10^{-2}	AF	DEPDC2	-2	4.009×10^{-2}	VF
LR11	0	2.084×10^{-2}	AF	LR52	-2	4.040×10^{-2}	VF
LR10	-2	2.233×10^{-2}	AF	LR24	-1	4.085×10^{-2}	VF
GM07	0	2.343×10^{-2}	AF	LR10	-2	4.500×10^{-2}	VF

Table 4: Selection of significant marker-allele combinations in EC chosen for random forest training

CRC

Model	**Sensitivity**	**Specificity**
Random Forest	0.703	0.750
BRAF	0.972	0.766
Random Forest + BRAF	0.861	0.872

EC

Model	**Sensitivity**	**Specificity**
Random Forest	0.632	0.684

Table 5: Sensitivity and specificity of different models in testing for LS within a cohort of MSI/MMRd cases.

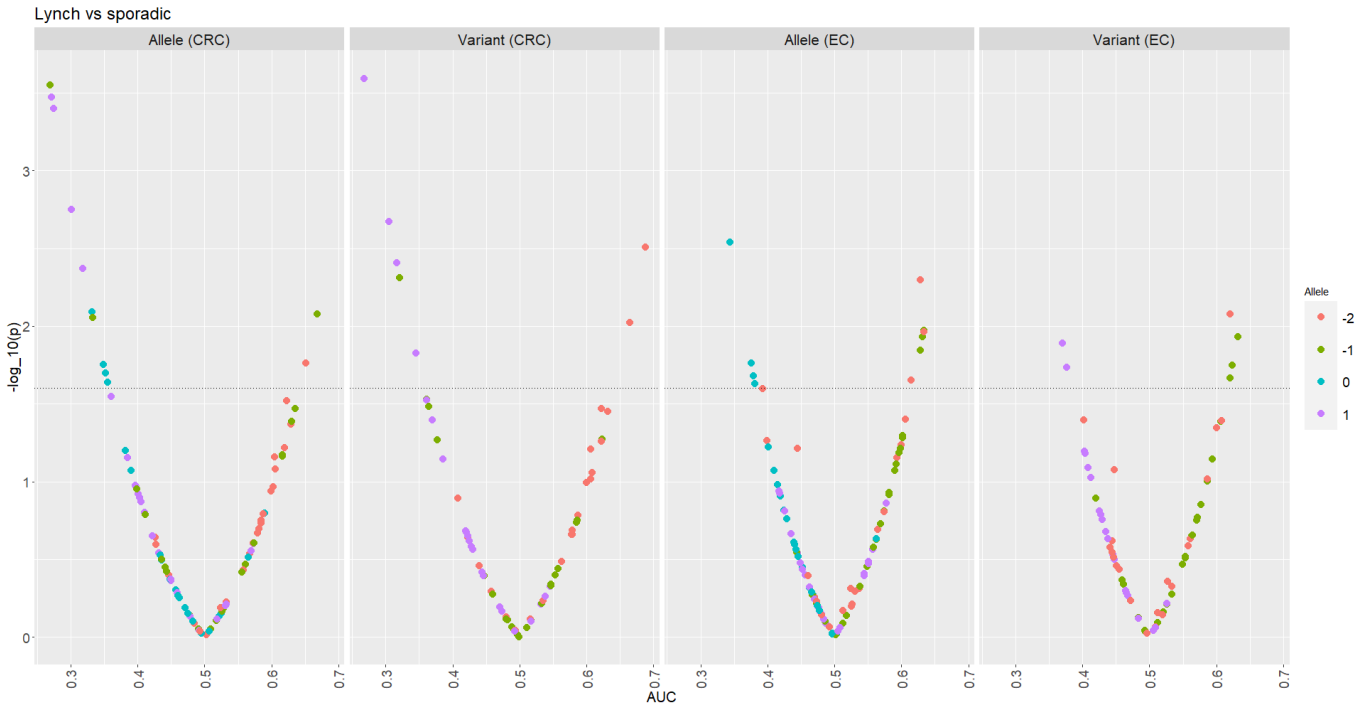

Figure 7: Volcano plot of significant marker-alleles in LS/sporadic tumor discrimination. Marker-alleles on the left side are more frequent in LS than sporadic, and vice versa. Threshold line indicates $p < 0.025$

4 Discussion

In this paper, MSI patterns of MSI/MMRd tumors in CRC and EC were analyzed. The results can be summarized into three main points: MSI is less severe in tumors from EC than CRC, MSI is less severe in LS compared to sporadic tumors, and that AF/VF patterns in LS differ depending on the affected MMR gene and cancer type.

Point 1 is made obvious by the noticable overlap between MSS and MSI tumors in the EC PCA plot (Figure 4A) which is not present in CRC (Figure 3A). AF profiles of EC MSI tumors in the overlap do not significantly differ from that of MSS tumors, but this obviously contradicts the definition of MSI. This raises the possibility of MSS/MSI misclassification within the EC dataset. Indeed, PCR-based MSI assays have a higher false negative than IHC in EC, but not in CRC (27).

It has been shown that MSI is less severe in EC than CRC. For example, Wang et al's paper (27) analyzed PCR length fragment differences in 5 microsatellite markers between paired tumor/normal tissue samples in MSI cancers. The mean difference in peak size (in bp) between the leftmost peaks on the electropherogram was -6bp in CRC, but -3bp in EC. They also found that a 30% dilution of EC tumor samples was enough to make MSI no longer observable, this effect was suggested to explain the higher false negative rate in EC.

References

[1] Kunkel TA. Evolving Views of DNA Replication (In)Fidelity. Cold Spring Harbor Symposia on Quantitative Biology. 2009 Jan;74(0):91–101.

[2] Graham V WJ, Putnam CD, Kolodner RD. DNA Mismatch Repair: Mechanisms and Cancer Genetics. In: Boffetta P, Hainaut P, editors. Encyclopedia of Cancer (Third Edition). third edition ed. Oxford: Academic Press; 2019. p. 530-8.

[3] Martin SA. Chapter 6 - The DNA mismatch repair pathway. In: Kelley MR, Fishel ML, editors. DNA Repair in Cancer Therapy (Second Edition). second edition ed. Boston: Academic Press; 2016. p. 151-77.

[4] Reyes GX, Schmidt TT, Kolodner RD, Hombauer H. New insights into the mechanism of DNA mismatch repair. Chromosoma. 2015 Apr;124(4):443–462.

[5] Chen PC, Dudley S, Hagen W, Dizon D, Paxton L, Reichow D, et al. Contributions by MutL Homologues Mlh3 and Pms2 to DNA Mismatch Repair and Tumor Suppression in the Mouse. Cancer Research. 2005 Oct;65(19):8662–8670.

[6] Wyatt MD, Pittman DL. Methylating Agents and DNA Repair Responses: Methylated Bases and Sources of Strand Breaks. Chemical Research in Toxicology. 2006 Oct;19(12):1580–1594.

[7] Spies M, Fishel R. Mismatch Repair during Homologous and Homeologous Recombination. Cold Spring Harbor Perspectives in Biology. 2015 Mar;7(3):a022657.

[8] Loeb LA. Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. Cancer Research. 2016 Apr;76(8):2057–2059.

[9] Baretti M, Le DT. DNA mismatch repair in cancer. Pharmacology and Therapeutics. 2018 Sep;189:45–62.

[10] Leclerc J, Vermaut C, Buisine MP. Diagnosis of Lynch Syndrome and Strategies to Distinguish Lynch-Related Tumors from Sporadic MSI/dMMR Tumors. Cancers. 2021 Jan;13(3):467.

[11] Gregory I, Laura V. Lynch Syndrome [Updated 2021 Feb 4];. In: Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 2004. Available from: `https://www.ncbi.nlm.nih.gov/books/NBK1211/`.

[12] Pećina-Šlaus N, Kafka A, Salamon I, Bukovac A. Mismatch Repair Pathway, Genome Stability and Cancer. Frontiers in Molecular Biosciences. 2020 Jun;7.

[13] Sehgal R, Sheahan K, O'Connell P, Hanly A, Martin S, Winter D. Lynch Syndrome: An Updated Review. Genes. 2014 Jun;5(3):497–507.

[14] Wimmer K, Kratz CP, Vasen HFA, Caron O, Colas C, Entz-Werle N, et al. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). Journal of Medical Genetics. 2014 Apr;51(6):355–365.

[15] Li Y, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology. 2002 Dec;11(12):2453–2465.

[16] Fan H, Chu JY. A Brief Review of Short Tandem Repeat Mutation. Genomics, Proteomics & Bioinformatics. 2007;5(1):7–14.

[17] Gryfe R, Kim H, Hsieh ETK, Aronson MD, Holowaty EJ, Bull SB, et al. Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer. New England Journal of Medicine. 2000 Jan;342(2):69–77.

[18] Brown M, Velazquez L, Bicak M, Habib J, Holtzman S, Saleh M, et al. 138 Shared frameshift neoantigens are expressed throughout mismatch repair deficient cancer development and are recognized by tissue infiltrating T cells that are dysregulated in advanced lesions. In: Regular and Young Investigator Award Abstracts. BMJ Publishing Group Ltd; 2023. .

[19] Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The Vigorous Immune Microenvironment of Microsatellite Instable Colon Cancer Is Balanced by Multiple Counter-Inhibitory Checkpoints. Cancer Discovery. 2015 Jan;5(1):43–51.

[20] André T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. New England Journal of Medicine. 2020 Dec;383(23):2207–2218.

[21] Lee KS, Kwak Y, Ahn S, Shin E, Oh HK, Kim DW, et al. Prognostic implication of CD274 (PD-L1) protein expression in tumor-infiltrating immune cells for microsatellite unstable and stable colorectal cancer. Cancer Immunology, Immunotherapy. 2017 Apr;66(7):927–939.

[22] Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. New England Journal of Medicine. 2003 Jul;349(3):247–257.

[23] Iwaizumi M, Tseng-Rogenski S, Carethers JM. DNA mismatch repair proficiency executing 5-fluorouracil cytotoxicity in colorectal cancer cells. Cancer Biology & Therapy. 2011 Oct;12(8):756–764.

[24] Cohen SA, Pritchard CC, Jarvik GP. Lynch Syndrome: From Screening to Diagnosis to Treatment in the Era of Modern Molecular Oncology. Annual Review of Genomics and Human Genetics. 2019 Aug;20(1):293–307.

[25] Bando H, Okamoto W, Fukui T, Yamanaka T, Akagi K, Yoshino T. Utility of the quasi-monomorphic variation range in unresectable metastatic colorectal cancer patients. Cancer Science. 2018 Oct;109(11):3411–3415.

[26] Guyot D'Asnières De Salins A, Tachon G, Cohen R, Karayan-Tapon L, Junca A, Frouin E, et al. Discordance between immunochemistry of mismatch repair proteins and molecular testing of microsatellite instability in colorectal cancer. ESMO Open. 2021 Jun;6(3):100120.

[27] Wang Y, Shi C, Eisenberg R, Vnencak-Jones CL. Differences in Microsatellite Instability Profiles between Endometrioid and Colorectal Cancers. The Journal of Molecular Diagnostics. 2017 Jan;19(1):57–64.

[28] Hechtman JF, Rana S, Middha S, Stadler ZK, Latham A, Benayed R, et al. Retained mismatch repair protein expression occurs in approximately 6mutations in mismatch repair genes. Modern Pathology. 2020 May;33(5):871–879.

[29] Vilar E, Mork ME, Cuddy A, Borras E, Bannon SA, Taggart MW, et al. Role of microsatellite instability-low as a diagnostic biomarker of Lynch syndrome in colorectal cancer. Cancer Genetics. 2014 Oct;207(10–12):495–502.

[30] Chow RD, Michaels T, Bellone S, Hartwich TMP, Bonazzoli E, Iwasaki A, et al. Distinct Mechanisms of Mismatch-Repair Deficiency Delineate Two Modes of Response to Anti–PD-1 Immunotherapy in Endometrial Carcinoma. Cancer Discovery. 2022 Oct;13(2):312–331.

[31] Borthwick G. Finding the right dose of aspirin to prevent Lynch Syndrome cancers. ISRCTN [Internet]; 2015. Available from: `http://dx.doi.org/10.1186/ISRCTN16261285`.

[32] Pearlman R, Frankel WL, Swanson BJ, Jones D, Zhao W, Yilmaz A, et al. Prospective Statewide Study of Universal Screening for Hereditary Colorectal Cancer: The Ohio Colorectal Cancer Prevention Initiative. JCO Precision Oncology. 2021 Nov;(5):779–791.

[33] Levine MD, Pearlman R, Hampel H, Cosgrove C, Cohn D, Chassen A, et al. Up-Front Multigene Panel Testing for Cancer Susceptibility in Patients With Newly Diagnosed Endometrial Cancer: A Multicenter Prospective Study. JCO Precision Oncology. 2021 Nov;(5):1588–1602.

[34] Ryan N, Wall J, Crosbie EJ, Arends M, Bosse T, Arif S, et al. Lynch syndrome screening in gynaecological cancers: results of an international survey with recommendations for uniform reporting terminology for mismatch repair immunohistochemistry results. Histopathology. 2019 Sep;75(6):813–824.

[35] Wu S, Liu X, Wang J, Zhou W, Guan M, Liu Y, et al. DNA Mismatch Repair Deficiency Detection in Colorectal Cancer by a New Microsatellite Instability Analysis System. Interdisciplinary Sciences: Computational Life Sciences. 2020 Jan;12(2):145–154.

[36] Redford L, Alhilal G, Needham S, O'Brien O, Coaker J, Tyson J, et al. A novel panel of short mononucleotide repeats linked to informative polymorphisms enabling effective high volume low cost discrimination between mismatch repair deficient and proficient tumours. PLOS ONE. 2018 Aug;13(8):e0203052.

[37] Germano G, Amirouchene-Angelozzi N, Rospo G, Bardelli A. The Clinical Impact of the Genomic Landscape of Mismatch Repair–Deficient Cancers. Cancer Discovery. 2018 Dec;8(12):1518-28.

[38] Leave-One-Out Cross-Validation. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. Boston, MA: Springer US; 2011. p. 600-1.