

## Article

# MEM: An Algorithm for the Reliable Detection of Microsatellite Instability (MSI) on a Small NGS Panel in Colorectal Cancer

Guillaume Herbreteau <sup>1,\*</sup>, Fabrice Airaud <sup>2</sup>, Elise Pierre-Noël <sup>1,2</sup>, Audrey Vallée <sup>1</sup> , Stéphane Bézieau <sup>2</sup>, Sandrine Théoleyre <sup>1</sup>, Hélène Blons <sup>3</sup>, Simon Garinet <sup>3</sup> and Marc Guillaume Denis <sup>1</sup> 

- <sup>1</sup> Laboratoire de Biochimie et Plateforme de Génétique Moléculaire des Cancers, CHU Nantes (Nantes University Hospital), 44093 Nantes, France; elise.pierrenoel@chu-nantes.fr (E.P.-N.); audrey.vallee@chu-nantes.fr (A.V.); sandrine.charpentier@chu-nantes.fr (S.T.); marc.denis@chu-nantes.fr (M.G.D.)
  - <sup>2</sup> Service de Génétique Médicale, CHU Nantes (Nantes University Hospital), 44093 Nantes, France; fabrice.airaud@chu-nantes.fr (F.A.); stephane.bezieau@chu-nantes.fr (S.B.)
  - <sup>3</sup> Plateforme de Génétique Moléculaire des Cancers, Hôpital Européen Georges Pompidou, Assistance Publique des Hôpitaux de Paris (Greater Paris University Hospitals), 75015 Paris, France; helene.blons@aphp.fr (H.B.); simon.garinet@aphp.fr (S.G.)
- \* Correspondence: guillaume.herbreteau@chu-nantes.fr



**Citation:** Herbreteau, G.; Airaud, F.; Pierre-Noël, E.; Vallée, A.; Bézieau, S.; Théoleyre, S.; Blons, H.; Garinet, S.; Denis, M.G. MEM: An Algorithm for the Reliable Detection of Microsatellite Instability (MSI) on a Small NGS Panel in Colorectal Cancer. *Cancers* **2021**, *13*, 4203. <https://doi.org/10.3390/cancers13164203>

Academic Editor: Constantin N. Baxevanis

Received: 13 July 2021

Accepted: 18 August 2021

Published: 20 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Microsatellite instability (MSI) assessment has become a major issue in the management of colorectal cancer, with the recent approval of anti-PD1 immunotherapies in MSI-metastatic colorectal cancer. The reference PCR method (MSI-PCR) can be costly, time and tissue-consuming. However, NGS could facilitate the assessment of MSI status while simultaneously screening for targetable oncogenic mutations (KRAS, NRAS, BRAF) for any colorectal cancer, but the algorithms developed to date use a large number of microsatellites that have not been approved by international guidelines and which are generally incompatible with small NGS panels. We present the MEM algorithm, which mimics the interpretation of MSI-PCR data by a human operator to reliably assess MSI status using only five validated microsatellites (BAT-25, BAT-26, NR-21, NR-24 and NR-27). We demonstrated that the MEM algorithm was in perfect agreement with MSI-PCR results, in terms of both MSI status and individual microsatellite status, in a cohort of 146 patients.

**Abstract:** Purpose: MEM is an NGS algorithm that uses Expectation-Maximisation to detect the presence of unstable alleles from the NGS sequences of five microsatellites (BAT-25, BAT-26, NR-21, NR-24 and NR-27). The purpose of this study was to compare the MEM algorithm with a reference PCR method (MSI-PCR) and Mismatch Repair protein immunohistochemistry (MMR-IHC). Methods: FFPE colorectal cancer samples from 146 patients were analysed in parallel by MSI-PCR and NGS using the MEM algorithm. MMR-IHC results were available for 133 samples. Serial dilutions of an MSI positive control were performed to estimate the limit of detection. Results: the MEM algorithm was able to detect unstable alleles of each microsatellite with up to a 5% allelic fraction. Of the 146 samples, 28 (19.2%) were MSI in MSI-PCR. MEM algorithm results were in perfect agreement with those of MSI-PCR, at both MSI status and individual microsatellite level (Cohen's kappa = 1). A high level of agreement was noted between MSI-PCR/MEM algorithm results and MMR-IHC results (Cohen's kappa = 0.931). Conclusion: the MEM algorithm can determine the MSI status of colorectal cancer samples on a small NGS panel, using only five microsatellites approved by international guidelines, and can be combined with screening for targetable mutations.

**Keywords:** microsatellite instability; deficient mismatch repair system; colorectal cancer; NGS; expectation-maximisation algorithm

## 1. Introduction

Microsatellite instability (MSI) is the molecular consequence of a deficient mismatch repair system (dMMR) [1]. Microsatellites are DNA sequences formed by the continuous repetition of

patterns comprising 1 to 6 nucleotides. During replication, the number of repetitions (and therefore the length) of a microsatellite may vary due to slipped strand mispairing. The mismatch repair system (MMR) is involved in repairing these replication slippage errors, notably for mono-nucleotide microsatellites [2]. dMMR triggers a variation in the length of microsatellites and genetic instability associated with the onset of many cancers: Lynch syndrome, a key cancer predisposition, is linked to constitutional dMMR, while sporadic dMMR is found in 15% to 20% of colorectal and gastric cancers, in 20% to 30% of endometrial cancers and most solid cancers are likely to exhibit sporadic dMMR at lower frequencies [3,4].

dMMR is usually demonstrated by identifying the loss of expression of at least one of the MMR proteins (MLH1, MSH2, MSH6 or PMS2) by immunohistochemistry (MMR-IHC). dMMR can also be demonstrated by revealing its consequence, microsatellite instability (MSI), using molecular biology methods (MSI-PCR) [5,6]. MSI-PCR requires the amplification of a locus comprising a microsatellite of interest by PCR and then analyses the length of the amplicons generated, usually by electrophoresis of the amplification products on a capillary sequencer [7]. The analysis of five poly-A mononucleotide microsatellites (BAT-25, BAT-26, NR-21, NR-24 and NR-27; Supplementary Table S1) are recommended by the revised Bethesda guidelines and the ESMO guidelines for mCRC MMR status determination [5,6]. The MSI phenotype is defined by the instability of at least two of the five microsatellites, while the stability of microsatellites (MSS) is defined by the instability of zero or one microsatellite [6].

Several studies have recently shown that dMMR is predictive of a good response to immune checkpoints inhibitors in several cancer types. Indeed, genetic instability induced by dMMR is likely to generate an increased number of neo-antigens, which can confer better tumour immunogenicity [8]. Pembrolizumab, an anti-PD1 immunotherapy, showed superior efficacy to chemotherapy as a first-line treatment for dMMR metastatic colorectal cancer (dMMR-mCRC) in the phase III KEYNOTE-177 trial [9]. In addition, pembrolizumab has proved effective as a second-line treatment for other types of dMMR cancers (mainly endometrial, gastric, cholangiocarcinoma and pancreatic cancers) in the Phase II KEYNOTE-158 trial [10]. In 2017, the US Food and Drug Administration (FDA) granted accelerated approval to pembrolizumab for the first-line treatment of dMMR-mCRC and as a second-line treatment for any unresectable or metastatic dMMR solid cancer when no other treatment options were available [11]. Another anti-PD1 immunotherapy, nivolumab, has also proved effective alone or in combination with ipilimumab (anti-CTLA4 immunotherapy) as a second-line treatment for dMMR-mCRC in the Phase II CheckMate-142 trial [12,13]. Nivolumab and nivolumab-ipilimumab combination therapy have been approved by the FDA as second-line treatments for dMMR-mCRC, and a phase III trial is currently underway (CheckMate-8HW [14]).

With the development of these new therapeutic options, the determination of MMR status has become a critical point in cancer management. However, costs, time and tissue consumption relating to MMR-IHC and MSI-PCR analyses may represent limiting factors, particularly with the pembrolizumab site-agnostic indication. In addition, MSI-PCR and MMR-IHC are rarely performed on tumours with low dMMR prevalence (<1%), although the impact on treatment could be significant. The determination of MSI status by Next-Generation Sequencing (MSI-NGS) seems an alternative of choice since it allows simultaneous high-throughput analysis of numerous samples and can be coupled with the search for somatic mutations of theranostic interest, which is already carried out in routine practice [15]. Many MSI-NGS algorithms have been developed [16–28]; however, they require sequencing of large panels of microsatellites that have not been approved by international guidelines, and their diagnostic performance is often imperfect.

In this context, we present a novel MSI-NGS algorithm called MEM (MSI assessment by Expectation-Maximisation algorithm) to determine MSI status by analysing only the NGS sequencing data of the five microsatellites validated by the Bethesda and ESMO international guidelines (BAT-25, BAT-26, NR-21, NR-24 and NR-27), using a method that mimics the interpretation of MSI-PCR data by a human operator.

## 2. Materials and Methods

### 2.1. MEM Algorithm

In MSI-PCR, a microsatellite is considered stable if its length distribution is comparable to the reference distribution and unstable if its length distribution corresponds to a mixture model, i.e., the mixture of several sub-distributions of different mean lengths (generally a sub-distribution corresponding to the length of the stable allele, similar to the reference distribution, and one or more sub-distributions of different mean lengths, corresponding to the unstable alleles).

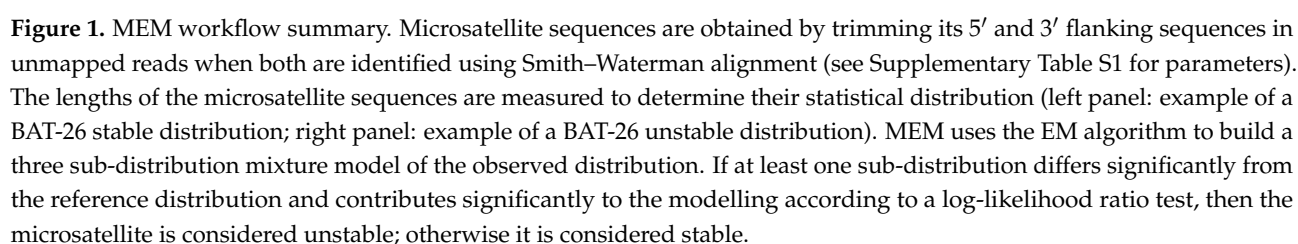
MEM is a Java-based bioinformatics algorithm built on CLC Genomics Workbench 20 (QIAGEN, Hilden, Germany), available at <https://github.com/MGPC-Nantes/MEM> (accessed on 19 August 2021). MEM attempts to closely replicate the MSI-PCR interpretation method. Since the amplification steps during NGS library preparation induce replication slippage of microsatellite sequences as with MSI-PCR, MEM identifies the stable or unstable nature of each microsatellite by (i) determining the length distribution of microsatellite sequences, without post-analytical bias related to filtering or alignment of the sequencing data, and (ii) determining whether the observed distribution corresponds to a mixture model whose sub-distributions differ from the reference distribution (Figure 1; see Appendix A for a detailed description).

The first step of the MEM analysis is to identify, for each microsatellite, the 5' and 3' flanking sequences of the microsatellite from unmapped and quality unfiltered, paired-end reads using Smith–Waterman alignment (Supplementary Table S1). If both 5' and 3' flanking sequences are identified in a read, then that read is trimmed from these sequences, and the resulting sequence is retained if it contains a homopolymeric sequence (otherwise, the read is excluded from the analysis). The microsatellite length distribution is then determined by measuring the length of each trimmed sequence.

The reference length distribution of each microsatellite was determined with the same method, using unaligned reads from merged FASTQ of 36 MSS samples showing the stability of all five microsatellites in MSI-PCR and expressing MLH1, MSH2, MSH6 and PMS2 in MMR-IHC.

To determine whether a microsatellite is unstable, MEM builds a model of the observed length distribution using a three sub-distribution mixture model (assuming one sub-distribution for the stable allele and up to two sub-distributions for two unstable alleles). The “shape” of each sub-distribution  $n$  is extrapolated from the reference length distribution previously defined for the microsatellite on MSS samples, and each sub-distribution is defined by two parameters: its expected value  $m_n$  and the proportion  $P_n$  of the mixture model represented by that sub-distribution. MEM uses the Expectation-Maximisation algorithm to determine a vector of parameters  $\{m_1, m_2, m_3, P_1, P_2, P_3\}$ , defining a mixture model approximating the observed length distribution with maximum likelihood. If the expected value  $m_n$  of a sub-distribution  $n$  is equal to the mean of the reference distribution, plus or minus 10%, then this sub-distribution is considered to be associated with a stable allele; otherwise, it is considered to be potentially associated with an unstable allele.

If all the sub-distributions are stable, or if the unstable sub-distributions represent less than 2% of the mixture model, then the microsatellite is considered stable. If there are one or more unstable sub-distributions representing more than 2% of the mixture model, then MEM compares the log-likelihood of the full mixture model to the log-likelihood of a mixture model, including only the stable sub-distributions, using a log-likelihood ratio test. If the complete model represents the observed length distribution significantly better than the model, including only the stable sub-distributions, then the microsatellite is considered unstable, and MEM quantifies the proportion of unstable alleles and their mean lengths from the parameters of the mixture model. Otherwise, MEM estimates the power of the log-likelihood ratio test for the number of sequences in the microsatellite: if the power of the test is greater than 80%, then the microsatellite is considered stable; otherwise the analysis of this microsatellite is considered to be non-contributory. Similar to MSI-PCR, the sample is then considered MSI if at least two out of five microsatellites are unstable.



## 2.2. Evaluation of the Limit of Detection

In order to evaluate MEM performance for the detection of instabilities in the presence of a small percentage of unstable alleles, DNA samples were reconstituted by diluting DNA from an MSI positive control with a high proportion of unstable alleles for the five microsatellites in MSI-PCR and the DNA from an MSI negative control. Six dilutions were prepared (1/4, 1/8, 1/12, 1/16, 1/20), then sequenced by NGS according to the protocol described below and analysed with the MEM algorithm.

## 2.3. Validation on Tumour Samples

NGS data from colorectal cancer samples sent to Nantes University Hospital between 1 January and 31 December 2019 for MSI assessment by MSI-PCR in conjunction with testing for KRAS, NRAS and BRAF by NGS were re-analysed retrospectively with the MEM algorithm.

The DNA of formalin-fixed paraffin-embedded (FFPE) colorectal cancer samples was extracted using the Maxwell<sup>®</sup> RSC RNA FFPE Kit. NGS libraries were produced from these DNA extracts using the QIAseq Targeted DNA Custom Panel (QIAGEN, Hilden, Germany) kit, an amplicon library construction kit based on Anchored Multiplex PCR (AMP) technology. This panel targeted, among others, KRAS, NRAS and BRAF genes and BAT-25, BAT-26, NR-21, NR-24 and NR-27 microsatellites (see Supplementary Table S1 for the genomic positions of the primers targeting microsatellites). NGS libraries were prepared according to the supplier's recommendations and then sequenced on a MiSeq sequencer (Illumina, San Diego, CA, USA).

The MSI-PCR analysis was carried out by the genetics laboratory of Nantes University Hospital using an in-house kit targeting BAT-25, BAT-26, NR-21, NR-22 and NR-24 microsatellites (see [29] for the list of primer sequences). Fragment analysis was performed on a 3500 xL Genetic Analyser capillary sequencer (Applied Biosystems, Waltham, MA, USA) and interpreted in comparison with a control sample stable for the five microsatellites, on GeneMapper software 5, by trained operators who were kept blinded to MEM results.

Both MEM algorithm results and MSI-PCR results were collected in an anonymized computer database together with the percentage of cancer cells in the sample, estimated by examining hematoxylin and eosin-stained tissue slides and the tumour expression of MLH1, MSH2, MSH6 and PMS2 evaluated by MMR-IHC, provided by the pathologist.

## 2.4. Ethical Aspects

No result related to MSI assessment by MEM was communicated for the purpose of patient management. The MSI-PCR or MMR-IHC results were not rechecked against MEM results. According to French and European legislation, the use of anonymous data does not require ethics committee approval. This study has been registered at Nantes Hospital by the Local Data Protection Officer under reference TS005-BIO.2019\_4.

# 3. Results

## 3.1. Limit of Detection

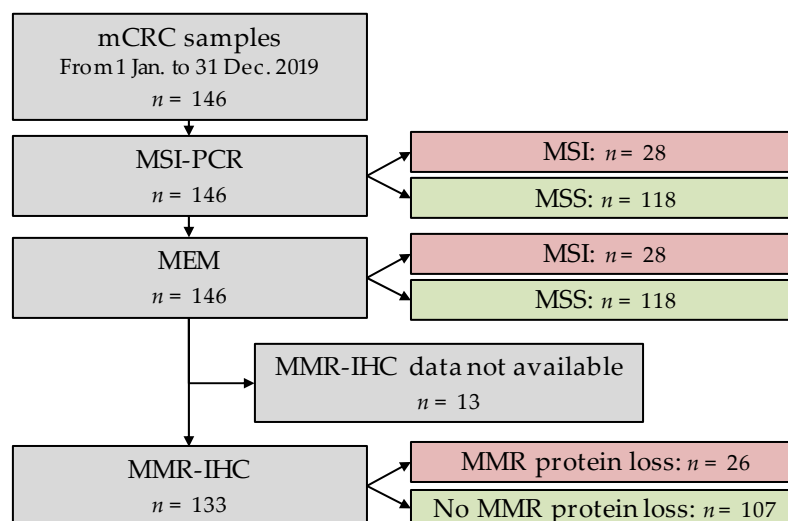
The MSI positive control used to assess the limit of detection showed 73%, 77%, 74%, 70% and 77% of unstable alleles for BAT-25, BAT-26, NR-21, NR-24 and NR-27, respectively, according to the MEM analysis. Among the samples prepared by diluting this MSI positive control with an MSS negative control, MEM identified MSI status for 1/4, 1/8, 1/12 and 1/16 dilutions, i.e., a limit of detection of approximately 5% of unstable alleles. For all microsatellites, the proportions of unstable alleles quantified by MEM were consistent with expected proportions, taking into account the dilution factors and the initial proportion of unstable alleles in the positive control ( $R^2 > 0.99$  for each microsatellite, data not shown).

## 3.2. MEM Algorithm vs. MSI-PCR Comparison

A total of 146 colorectal cancer samples were sent to Nantes University Hospital between 1 January and 31 December 2019, to test for KRAS, NRAS or BRAF mutations by



NGS and microsatellite instability by MSI-PCR. A total of 28 samples out of 146 (19.2%) were MSI according to MSI-PCR, and the remaining 118 samples were MSS (Figure 2). The percentage of cancer cells in samples exceeded 50% for 117 samples (80.1%), ranged from 25% to 50% for 28 samples (19.2%) and from 10% to 25% for a single sample (0.7%). No sample contained less than 10% cancer cells.



**Figure 2.** Study workflow.

NGS data obtained from these samples were re-analysed with the MEM algorithm. MEM conclusions were entirely consistent with those of MSI-PCR (Cohen's kappa = 1; Table 1). All MSI samples in MSI-PCR were identified as MSI by MEM (28/28), and all MSS samples in MSI-PCR were identified as MSS with MEM (118/118). Considering MSI-PCR as gold-standard, MEM had sensitivity, specificity and positive and negative predictive values of 100% for the determination of MSI status.

**Table 1.** Concordance between MSI-PCR results and MEM results for MSI status.

		MSI-PCR		Cohen's Kappa
		MSI	MSS	
MEM	MSI	28	0	1
	MSS	0	118	

In terms of each individual microsatellite, the MEM algorithm was also entirely consistent with the MSI-PCR results for BAT-25, BAT-26, NR-21 and NR-24 (Cohen's kappa = 1 in each case; Table 2). For a single MSI sample for which BAT-25, BAT-26, NR-21 and NR-22 were unstable with the MEM algorithm, MSI-PCR did not generate enough amplicons of the BAT-26 locus to allow interpretation ( $n = 145$  for BAT-26).

Concordance between MEM and MSI-PCR could not be established for NR-27 because the MSI-PCR analysis used did not evaluate this microsatellite. However, MEM results for NR-27 were strongly concordant with MSI status in MSI-PCR (Table 3). Among the 28 MSI samples, MEM identified NR-27 as unstable for 26 samples (92.9%), stable for 1 sample and non-contributory for 1 sample due to an overall constitutive lack of NR-27 coverage, probably due to deletion of the NR-27 locus (as this microsatellite was targeted by several primers in our panel). Among the 118 MSS samples, NR-27 was identified as stable in 116 cases (98.3%). NR-27 was unstable for 2 MSS samples, with both cases having an unstable allele with an average length of  $-4$  bp compared to the reference distribution and an allele frequency of approximately 50% according to MEM, raising suspicions of NR-27 polymorphism in these samples.

**Table 2.** Concordance between MSI-PCR and MEM results for BAT-25, BAT-26, NR-21 and NR-24.

			MSI-PCR		Cohen's Kappa
			Unstable	Stable	
BAT-25	MEM	Unstable	28	0	1
		Stable	0	118	
BAT-26	MEM	Unstable	27	0	1
		Stable	0	118	
NR-21	MEM	Unstable	28	0	1
		Stable	0	118	
NR-24	MEM	Unstable	26	0	1
		Stable	0	120	

**Table 3.** Concordance between MEM results for NR-27 and MSI status in MSI-PCR.

			MSI-PCR	
			MSI	MSS
NR-27	MEM	Unstable	26	2
		Stable	1	116

### 3.3. MEM vs. MMR-IHC Comparison

MMR-IHC data were available for 133 samples, including 107/118 MSS samples and 26/28 MSI samples in MSI-PCR.

All of these 26 MSI samples (100%) displayed loss of expression of at least one of the MMR proteins (MLH1, MSH2, MSH6 or PMS2). Among the 107 MSS samples, 104 showed no loss of expression of MMR proteins in MMR-IHC (96.3%; Table 4). As the results obtained with MEM were entirely consistent with those of MSI-PCR, the same comparison was observed with our algorithm, with three samples showing MSS status with MEM and loss of expression of at least one MMR protein in MMR-IHC. For the three samples, the percentage of cancer cells was greater than 25%, all the microsatellites analysed were stable, and the loss of expression in MMR-IHC affected the MLH1 and PMS2 proteins, a situation typically observed in the case of sporadic dMMR.

**Table 4.** Concordance between MEM results for MSI status and MMR-IHC.

		MMR-IHC		Cohen's Kappa
		Loss of Expression of $\geq 1$ MMR Protein	No Loss of Expression	
MEM	MSI	26	0	0.931
	MSS	3	104	

## 4. Discussion

MEM is the first MSI-NGS algorithm using only five microsatellites validated by international guidelines for colorectal cancer management (Bethesda revised, ESMO). MEM is compatible with small NGS panels and attempts to faithfully reproduce the principles of MSI-PCR, both in the establishment of microsatellite length distributions and their interpretation. Because of the closeness between the two methods, MEM produces results that are fully consistent with MSI-PCR for the analysis of colorectal cancer samples at both MSI status and individual microsatellite levels.

Several MSI-NGS algorithms have been developed to assess MSI status from NGS data (mSINGS/MSIplus [16,17], MANTIS [18], MSIseq [19], MSIsensor [20], MSI-ColonCore [21],

ELMSI [24], MSICare [25], mSILICO [26], NovoPM-MSI [27], USCI-MSI [28], Cortes-Ciriano et al. [22], Lu et al. [23]). Similar to MEM, these algorithms roughly operate according to two main steps: firstly, they determine the length distribution of microsatellite sequences, and secondly, they compare the observed distribution to a reference distribution to determine the stable or unstable status of the microsatellite in a standardised and reproducible manner. These algorithms differ essentially in the way in which both steps are performed.

Most algorithms identify and determine the length of microsatellite sequences by recognising the sequences flanking the microsatellite, on aligned reads in BAM format (mSINGS/MSIplus [16,17], MANTIS [18], ELMSI [24], Cortes-Ciriano et al. [22]), while some algorithms are based on the analysis of indel variants identified in microsatellite sequences (MSIseq [19], Lu et al. [23]) or determine the length distribution of microsatellite sequences by alignment on all possible length variants of the microsatellite (MSIsensor [20], MSI-ColonCore [21]). The MEM algorithm determines the length distribution of the microsatellites by recognising the flanking sequences directly on unaligned, unfiltered, paired-end reads in FASTQ format. This method seeks to replicate as far as possible the MSI-PCR approach in which the amplicons generated are selected by primer choice, regardless of the internal sequence of the amplified locus. In this way, MEM maximises the number of sequences obtained for each microsatellite and essentially limits the introduction of bias in the length distribution related to the exclusion of some reads with low sequencing quality (as is often the case for homopolymeric sequences) or where marked instability would interfere with mapping by rendering their sequence too different from the reference sequence.

MSI-NGS algorithms also differ in terms of the methods used to interpret the length distribution of microsatellite sequences. Some algorithms use statistical methods to compare the distribution, such as the Kolmogorov–Smirnov test (Cortes-Ciriano et al. [22]), Chi-squared test (MSIsensor [20]) or scores approaching it (MANTIS [18]). Other algorithms analyse parameters related to distribution dispersion, such as the number of different lengths observed at a significant proportion in microsatellite sequences (mSINGS/MSIplus [16,17]) or the proportion of microsatellite sequences whose length is within a reference interval (MSI-ColonCore [21]), or parameters related to distribution skewness (mSILICO [26]). Finally, some algorithms use interpretation methods based on machine learning algorithms (MSIseq [19]). For its part, the MEM algorithm replicates the MSI-PCR interpretation method by a human operator, evaluating whether the observed length distribution of a microsatellite is deemed a mixture model, using an Expectation-Maximisation algorithm and empirical reference length distributions obtained on a cohort of MSS samples. ELMSI, a recently published algorithm, also uses an Expectation-Maximisation algorithm to determine the stability of microsatellites over large panels but in our experience, the normal distribution used to characterise the reference distribution of each microsatellite does not match the distributions actually observed [24].

Although most MSI-NGS algorithms can determine the MSI status of a cancer sample with reported good sensitivity and specificity (Table 5) [16,18–24], they may be unreliable for determining the stable or unstable status of each microsatellite individually. These methods manage to compensate for this shortcoming by simultaneously analysing several tens to several thousands of microsatellites, combined with comparison to a matched non-tumour sample for certain algorithms, thus reducing the risk and impact of microsatellite misclassification. However, analysing this number of microsatellites requires the use of large NGS panels, culminating in whole-exome sequencing, which is problematic since it increases analytical costs and diminishes multiplexing capacity. On the other hand, the use of microsatellite panels not validated by international recommendations may be a limiting factor in terms of therapeutic indications for immunotherapy.



**Table 5.** Reported diagnostic performances for the main MSI-NGS algorithms for the determination of MSI status, considering MSI-PCR as gold-standard.

MSI-NGS Algorithm	Normal Sample Required	Microsatellite Count	Validation Cohort	Sensitivity	Specificity
mSINGS/ MSIplus	No	N/A (TGS)	Zhao [27]: <i>n</i> = 113	88.9%	99.0%
		15 to 2957	Salipante [16]: <i>n</i> = 108	96.4–100%	97.2–100%
		2539	Kautto [18]: <i>n</i> = 275	76.1%	99.7%
		3154	Lee [26]: <i>n</i> = 117	92.5%	100%
		230	Lee [26]: <i>n</i> = 117	95.0%	100%
		23	Lee [26]: <i>n</i> = 117	95.0%	93.7%
		11	Hempelmann [17]: <i>n</i> = 81	97.1%	100%
MSIsensor	Yes	N/A (WES)	Niu [20]: <i>n</i> = 242	98.6%	98.2%
		2539	Kautto [18]: <i>n</i> = 275	96.5%	98.7%
		N/A (WES)	Ratovmanana [25]: <i>n</i> = 333	85.5–90.2%	95.5–100%
		441	Ratovmanana [25]: <i>n</i> = 152	97.1%	73.3%
MANTIS	Yes	2539	Kautto [18]: <i>n</i> = 275	97.2%	99.7%
		N/A (TGS)	Zhao [27]: <i>n</i> = 113	88.9%	86.5%
MSI-ColonCore	Yes	90	Zhu [21]: <i>n</i> = 91	97.9%	100%
ELMSI	No	20 to 100	Wang [24]: simulated data	70.0–82.1%	N/A
MSICare	Yes	N/A (WES)	Ratovmanana [25]: <i>n</i> = 333	96.1–100%	97.0–100%
		441	Ratovmanana [25]: <i>n</i> = 152	99.3%	100%
mSILICO	No	3154	Lee [26]: <i>n</i> = 117	100%	100%
		230	Lee [26]: <i>n</i> = 117	100%	77.2%
		23	Lee [26]: <i>n</i> = 117	95.0%	100%
NovoPM-MSI	Yes	19	Zhao [27]: <i>n</i> = 113	88.9%	97.1%
USCI-MSI	Yes	9	Zheng [28]: <i>n</i> = 64	100%	100%
MEM	No	5	(present study): <i>n</i> = 146	100%	100%

N/A: Not Available; TGS: Targeted-Genome Sequencing; WES: Whole-Exome Sequencing.

The MEM algorithm analyses each microsatellite individually in a standardised manner and is highly consistent with MSI-PCR results. Indeed, in this study, we have shown perfect agreement between MSI-PCR results and MEM results for BAT-25, BAT-26, NR-21 and NR-24, while NR-27, which was not evaluated in MSI-PCR, demonstrated excellent agreement with the MSI status of samples, with performance levels similar to those previously described for this panel of microsatellites in MSI-PCR (sensitivity and specificity: 93.9% and 99.0%, respectively, in MSI-PCR according to Goel et al., vs. 92.9% and 98.3%, respectively, with MEM, in our study [7]). Given this level of reliability for each individual microsatellite, MEM can be used to determine the MSI status of colorectal cancer samples by analysing only the microsatellites validated by international recommendations. In addition, the reduced number of microsatellites analysed and the non-requirement of a matched non-tumour sample considerably limits the size of the NGS panel to be used, limits the run-time (less than 10 s per sample for the whole process) and maintains a large multiplexing capacity, compatible with the number of samples to be analysed assuming a site-agnostic indication for immunotherapy.

One of the main constraints associated with this algorithm relates to the need for sufficient coverage of each microsatellite in order to obtain homogeneous length distribution. According to the quality parameters used in our study, a minimum of 689, 958, 707, 579 and 746 reads were required to allow the interpretation of BAT-25, BAT-26, NR-21, NR-24 and NR-27 length distributions, respectively. Optimisation of the NGS panel used may be necessary to guarantee sufficient coverage.

Given the proximity between MEM and MSI-PCR, both analyses also have similar limits and continue to complement MMR-IHC. Although MSI-NGS can be coupled to the sequencing of MMR genes to identify mutations associated with Lynch syndrome,

neither MSI-NGS nor MSI-PCR can identify the loss of expression of an MMR protein: MMR-IHC, possibly associated with the search for MLH1 promoter hypermethylation, remains essential in order to distinguish sporadic dMMR from Lynch syndrome [6]. In addition, both MSI-PCR and MSI-NGS can produce false-negative results that MMR-IHC can help to identify: in our study, we identified three potential cases in which loss of expression of MMR proteins was observed in MMR-IHC despite MSS status with both MSI analyses. Several factors can limit the sensitivity of MSI-PCR and, by extension, the sensitivity of MSI-NGS: on the one hand, the detection of an unstable microsatellite is conditioned by the proportion of unstable alleles in the sample and can therefore be limited for samples with a small proportion of cancer cells. However, this limitation was not encountered in our study, since almost all of the samples (the three potential false-negatives in particular) presented tumour cellularity greater than 25%, while MEM showed an ability to identify microsatellite instability at allelic frequencies as low as 5%. Conversely, the sensitivity of MSI-PCR and MSI-NGS is also determined by the degree of instability: the lower the variation in length between an unstable microsatellite and the wild-type allele, the harder the identification process. If this phenomenon does not interfere to any considerable extent with the analysis of colorectal cancer samples, which are generally unstable to a significant degree, it is more frequently described in other tumour types, such as endometrial cancers [4,30]. As our study assessed the performance levels of the MEM algorithm only in evaluating the MSI status of colorectal cancers, a new assessment seems necessary to determine its analytical performance in other tumour types. In addition, some authors suggest that the analysis of only five microsatellites could be insufficient to identify all of the MSI samples, particularly for non-colorectal cancers for which the Bethesda panel microsatellites are not validated [3,4]. If the Bethesda panel is extended to incorporate new microsatellites or if new microsatellite panels adapted for MSI status determination are defined in some non-colorectal cancers, MEM could be adapted by updating the NGS panel and defining reference distributions for these new microsatellites on MSS control samples.

Conversely, MSI-NGS and MSI-PCR can be used to identify certain false-negatives of MMR-IHC, in particular when dMMR is linked to a loss of function of one of the MMR proteins without loss of expression. Both MSI-NGS and MSI-PCR have excellent specificity, with no false positives provided that care is taken to recognise potential polymorphisms of microsatellites. Although parallel analysis of a non-tumour sample is not required for the assessment of colorectal cancer, it should be carried out whenever the presence of a microsatellite polymorphism is suspected and particularly when the status of this microsatellite conditions the MSI status of the sample [31].

The MEM algorithm has been used since April 2020 at Nantes University Hospital for microsatellite instability assessment using NGS, thereby replacing MSI-PCR. From April 2020 to March 2021, 292 colorectal cancer samples were analysed: 36/292 were MSI (12.3%) with MEM. MMR-IHC was reported in 256 samples, including 34/36 MEM-MSI samples: all of these MEM-MSI samples demonstrated loss of expression of at least one MMR protein in MMR-IHC. Of the remaining 222 MEM-MSS samples, 221/222 had no loss of MMR protein expression, and 1 had a loss of expression of MLH1 and PMS2. Results were returned in conjunction with KRAS/NRAS/BRAF status within a median of 6 working days after receiving the samples (Q1–Q3: 6–7 days).

## 5. Conclusions

The MEM algorithm allows systematic determination of MSI status, analysing only the five microsatellites validated by international guidelines. MEM can be combined with the test for therapeutic mutations on small NGS panels for all colorectal cancer samples. Further studies are needed to compare the MEM performance with other MSI-NGS algorithms and to evaluate it for other cancer types.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13164203/s1>, Table S1. Monomorphic microsatellites used by MEM for the determination of MSI status, genomic position of QIAseq AMP primers and MEM trimming parameters.

**Author Contributions:** Conceptualisation, G.H.; data curation, G.H.; formal analysis, G.H.; investigation, G.H., F.A. and S.B.; methodology, G.H.; resources, F.A., S.B., S.G. and H.B.; software, G.H.; supervision, M.G.D.; validation, G.H., F.A., M.G.D., E.P.-N., S.G., H.B., A.V., S.T. and S.B.; writing—original draft preparation, M.G.D. and G.H.; writing—review and editing, M.G.D. and G.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** According to French and European legislation, the use of anonymous data does not require ethics committee approval. This study has been registered at Nantes Hospital by the Local Data Protection Officer under reference TS005-BIO.2019\_4.

**Informed Consent Statement:** Not applicable. According to French and European legislation, the use of anonymous data does not require ethics committee approval.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors would like to thank Réjane Lapied, Jean-Jacques Paillusson and Florian Pecquet for their expert technical assistance.

**Conflicts of Interest:** The authors have no conflict of interest to declare.

## Appendix A MEM Algorithm

### Appendix A.1 Determination of the Length Distribution of Microsatellite Sequences

MEM is a bioinformatics algorithm built on CLC Genomics Workbench 20 (QIAGEN, Hilden, Germany). This algorithm directly analyses unaligned reads in FASTQ format (preventing mapping from triggering a bias in length distribution by excluding reads in which the microsatellite would differ substantially from the reference sequence, especially in cases of marked instability).

The first step of the MEM analysis consists in identifying, for each microsatellite, the 5' and 3' flanking sequences of the microsatellite among the unmapped and quality-unfiltered, paired-end reads, using Smith–Waterman alignment (See Supplementary Table S1 for the list of sequences used and the alignment parameters). If both 5' and 3' flanking sequences are identified within a read, then this read is trimmed from these sequences, and the resulting sequence is retained if it contains a homopolymeric sequence. Otherwise, if at least one of the flanking sequences is not identified or if the resulting trimming sequence does not contain any homopolymer, then the read is excluded from the analysis. The resulting trimmed sequences essentially correspond to the sequences of the microsatellite of interest: their length is measured, and the probability of each length is then determined by reporting the number of trimmed sequences presenting this length in relation to the total number of trimmed sequences obtained to establish the length distribution of the microsatellite.

$$p_i = \frac{N_i}{\sum_{i=0}^{L_{max}} N_i} \quad (A1)$$

$p_i$ : probability of occurrence of a microsatellite sequence of length  $i$

$N_i$ : number of trimmed sequences of length  $i$

$L_{max}$ : maximum length

The reference length distribution of each microsatellite was determined according to the method described, using unaligned reads from merged NGS data from 36 MSS samples with the stability of the five microsatellites in MSI-PCR and expressing MLH1, MSH2, MSH6 and PMS2 in MMR-IHC.

### Appendix A.2 Modelling of Length Distribution Using a Mixture Model

To determine whether a microsatellite is unstable, MEM determines a mixture model with three sub-distributions (assuming one sub-distribution for the stable allele and two for two unstable alleles) representing the observed length distribution with the maximum

likelihood. This mixing model is defined by a vector  $\theta$  of five independent parameters:  $m_1$ ,  $m_2$  and  $m_3$ , the expected value of each sub-distribution, and  $P_1$  and  $P_2$ , the proportions of sub-distributions one and two, respectively, within the mixture model (given that proportion  $P_3$  of the third sub-distribution is not an independent parameter, since  $P_1 + P_2 + P_3 = 1$ ).

To create this type of mixture model, a continuous density function has to be defined for each sub-distribution, the expected value  $m$  of which can be parameterised. To this end, the shape of each sub-distribution is considered similar to the reference distribution. Thus, a continuous density function  $f(x, m)$  is extrapolated from this empirical discrete reference distribution using Parzen–Rosenblatt kernel density estimation:

$$f(x, m) = \frac{1}{h} \sum_{i=0}^{L_{max}} \left[ K\left(\frac{S(x, m) - i}{h}\right) \times p_i \right] \quad (\text{A2})$$

$h$ : a smoothing parameter called “bandwidth”. MEM uses a bandwidth  $h = 1$

$K(x)$ : the kernel function. MEM uses the density function of the standard normal distribution as the kernel function, such as:

$$K(x) = \frac{1}{\sqrt{2}} e^{-\frac{1}{2}x^2} \quad (\text{A3})$$

$S(x, m)$ : a scaling function allowing a switch from a reference distribution with an expected value,  $m_{ref}$ , to distribution with an expected value,  $m$ , such as:

$$S(x, m) = \frac{x \times m_{ref}}{m} \quad (\text{A4})$$

The variance of this density function cannot be configured but varies depending on its expected value: the greater the distribution value expected (i.e., the average length of the microsatellite), the greater the variance of the distribution. This reliably reflects the behaviour of replication slippage errors, which increase in proportion to the length of the microsatellite sequence.

The mixture model is therefore defined for a vector of parameters  $\theta$  by a density function  $g_\theta(x)$ , such as:

$$g_\theta(x) = f(x, m_1) \times P_1 + f(x, m_2) \times P_2 + f(x, m_3) \times (1 - P_1 - P_2) \quad (\text{A5})$$

### Appendix A.3 Expectation-Maximisation Algorithm

MEM uses the Expectation-Maximisation (EM) algorithm to determine which parameters, namely  $m_1$ ,  $m_2$ ,  $m_3$ ,  $P_1$  and  $P_2$ , are used by the mixture model to represent the observed distribution with maximum likelihood.

Briefly, an initial value is assigned to each parameter ( $m_1 = m_{ref} + 2$ ;  $m_2 = m_{ref} - 3$ ;  $m_3 = m_{ref} - 10$ ;  $P_1 = 33\%$ ;  $P_2 = 33\%$ ). The EM algorithm then operates according to two steps: at the expectation step, the algorithm determines, for each length  $i$  of the microsatellite, the extent to which the probability of occurrence  $p_i$  can be explained by each sub-distribution  $j$ , by calculating a weight  $w_{i,j}$  such that:

$$w_{i,j} = \frac{f(i, m_j) \times P_j}{g(i)} \quad (\text{A6})$$

At the maximisation step, the algorithm recalculates the value of the parameters from the observed data, weighted by  $w_{i,j}$ . Thus, the new value of the expected value  $m_j$  of the sub-distribution  $j$  is:

$$m_j = \sum_{i=0}^{L_{max}} i \times p_i \times w_{i,j} \quad (A7)$$

The new value of the proportion  $P_j$  of the sub-distribution  $j$  in the mixture model is:

$$P_j = \sum_{i=0}^{L_{max}} p_i \times w_{i,j} \quad (A8)$$

After the maximisation step, MEM calculates the log-likelihood  $L_g(\theta)$  of the mixture model defined by the vector  $\theta$  of the recalculated parameters  $m_1, m_2, m_3, P_1$  and  $P_2$ :

$$\log L_g(\theta) = \sum_{i=0}^{L_{max}} (\log g_\theta(i) \times N_i) \quad (A9)$$

Both steps are then repeated: with each repetition, the likelihood of the mixture model increases and tends asymptotically towards a maximum. The repetition of the EM algorithm is stopped when the log-likelihood stabilises or after 100 repetitions. The parameters  $m_1, m_2, m_3, P_1$  and  $P_2$ , as well as the log-likelihood value obtained at the last repetition, are retained and define the optimal mixture model.

#### Appendix A.4 Interpretation

If the expected value of a sub-distribution falls within an interval  $[m_{ref} \times 90\%; m_{ref} \times 110\%]$  (i.e., if the expected value of the sub-distribution is equal to the mean of the reference distribution,  $+/-10\%$ ), then this sub-distribution is deemed to be related to a stable allele. Otherwise, the sub-distribution is deemed related to a potentially unstable allele. If no sub-distribution is associated with a possibly unstable allele, or if the sub-distributions associated with instability represent less than 2% of the mixture model, then the microsatellite is considered stable.

If there are one or more unstable sub-distributions representing more than 2% of the mixture model, then MEM defines a second mixture model  $q(x)$ , including only stable sub-distributions:

$$q_\theta(x) = \frac{f(x, m_1) \times P_1 \times E_1 + f(x, m_2) \times P_2 \times E_2 + f(x, m_3) \times (1 - P_1 - P_2) \times E_3}{P_1 \times E_1 + P_2 \times E_2 + (1 - P_1 - P_2) \times E_3} \quad (A10)$$

whereby  $E_j$  is an indicator function equal to one if the sub-distribution  $j$  is considered stable, and equal to zero if the sub-distribution  $j$  is associated with instability. The log-likelihood  $L_q(\theta)$  of this new mixture model is then calculated:

$$\log L_q(\theta) = \sum_{i=0}^{L_{max}} (\log q_\theta(i) \times N_i) \quad (A11)$$

MEM finally evaluates the ability of the complete mixture model to represent the observed length distribution, compared to this new mixture model  $q(x)$ , including only the stable sub-distributions, using the log-likelihood ratio test. The statistic  $\lambda$  is:

$$\lambda = -2[\log L_q(\theta) - \log L_g(\theta)] \quad (A12)$$

A cut-off value of 4 was defined in a development cohort of 36 MSS patients with 5 stable microsatellites and 15 MSI patients with 5 unstable microsatellites in MSI-PCR. If the log-likelihood ratio is greater than four, then the complete mixture model represents the observed length distribution to a significantly better extent than the mixture model,



including only the stable sub-distributions: the microsatellite is deemed unstable, and MEM can estimate the allele frequency and the length of unstable alleles from the parameters of the mixture model.

Otherwise, it is necessary to determine whether the number of sequences obtained for the microsatellite is sufficient to detect an unstable allele. The statistical power of the log-likelihood ratio test according to the number of microsatellite sequences was determined in advance for each microsatellite, using simulated data. In the case of a negative log-likelihood ratio test, the microsatellite analysis is considered to be contributory only if the number of sequences is sufficient to obtain a statistical power greater than 80% to detect an unstable allele present at a frequency of 10%, and with a length equal to  $m_{ref} \times 90\% - 1$ . If so, the microsatellite is deemed stable; otherwise, the analysis of this microsatellite is considered non-contributory.

Similar to MSI-PCR, a sample is considered as MSI if at least two microsatellites out of five are unstable. The sample is considered MSS if at least four microsatellites out of five are stable. The sample analysis is deemed non-contributory if at least two microsatellites are non-contributory or if at least one microsatellite is non-contributory and another is unstable.

MEM source code and ready-to-use files for CLC Genomics Workbench are available at <https://github.com/MGPC-Nantes/MEM> (accessed on 19 August 2021).

## References

1. Strand, M.; Prolla, T.A.; Liskay, R.M.; Petes, T.D. Destabilization of Tracts of Simple Repetitive DNA in Yeast by Mutations Affecting DNA Mismatch Repair. *Nature* **1993**, *365*, 274–276. [\[CrossRef\]](#)
2. Schlötterer, C. Genome Evolution: Are Microsatellites Really Simple Sequences? *Curr. Biol.* **1998**, *8*, R132–R134. [\[CrossRef\]](#)
3. Bonneville, R.; Krook, M.A.; Kautto, E.A.; Miya, J.; Wing, M.R.; Chen, H.-Z.; Reeser, J.W.; Yu, L.; Roychowdhury, S. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis. Oncol.* **2017**, *2017*, 1–15. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Hause, R.J.; Pritchard, C.C.; Shendure, J.; Salipante, S.J. Classification and Characterization of Microsatellite Instability across 18 Cancer Types. *Nat. Med.* **2016**, *22*, 1342–1350. [\[CrossRef\]](#)
5. Umar, A.; Boland, C.R.; Terdiman, J.P.; Syngal, S.; de la Chapelle, A.; Rüschoff, J.; Fishel, R.; Lindor, N.M.; Burgart, L.J.; Hamelin, R.; et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *J. Natl. Cancer Inst.* **2004**, *96*, 261–268. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Luchini, C.; Bibeau, F.; Ligtenberg, M.J.L.; Singh, N.; Nottegar, A.; Bosse, T.; Miller, R.; Riaz, N.; Douillard, J.-Y.; Andre, F.; et al. ESMO Recommendations on Microsatellite Instability Testing for Immunotherapy in Cancer, and Its Relationship with PD-1/PD-L1 Expression and Tumour Mutational Burden: A Systematic Review-Based Approach. *Ann. Oncol.* **2019**, *30*, 1232–1243. [\[CrossRef\]](#)
7. Goel, A.; Nagasaka, T.; Hamelin, R.; Boland, C.R. An Optimized Pentaplex PCR for Detecting DNA Mismatch Repair-Deficient Colorectal Cancers. *PLoS ONE* **2010**, *5*, e9393. [\[CrossRef\]](#)
8. Le, D.T.; Uram, J.N.; Wang, H.; Bartlett, B.R.; Kemberling, H.; Eyring, A.D.; Skora, A.D.; Luber, B.S.; Azad, N.S.; Laheru, D.; et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **2015**, *372*, 2509–2520. [\[CrossRef\]](#)
9. André, T.; Shiu, K.-K.; Kim, T.W.; Jensen, B.V.; Jensen, L.H.; Punt, C.; Smith, D.; Garcia-Carbonero, R.; Benavides, M.; Gibbs, P.; et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N. Engl. J. Med.* **2020**, *383*, 2207–2218. [\[CrossRef\]](#)
10. Marabelle, A.; Le, D.T.; Ascierto, P.A.; Di Giacomo, A.M.; De Jesus-Acosta, A.; Delord, J.-P.; Geva, R.; Gottfried, M.; Penel, N.; Hansen, A.R.; et al. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *J. Clin. Oncol.* **2020**, *38*, 1–10. [\[CrossRef\]](#)
11. Marcus, L.; Lemery, S.J.; Keegan, P.; Pazdur, R. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. *Clin. Cancer Res.* **2019**, *25*, 3753–3758. [\[CrossRef\]](#)
12. Overman, M.J.; McDermott, R.; Leach, J.L.; Lonardi, S.; Lenz, H.-J.; Morse, M.A.; Desai, J.; Hill, A.; Axelson, M.; Moss, R.A.; et al. Nivolumab in Patients with Metastatic DNA Mismatch Repair-Deficient or Microsatellite Instability-High Colorectal Cancer (CheckMate 142): An Open-Label, Multicentre, Phase 2 Study. *Lancet Oncol.* **2017**, *18*, 1182–1191. [\[CrossRef\]](#)
13. Overman, M.J.; Lonardi, S.; Wong, K.Y.M.; Lenz, H.-J.; Gelsomino, F.; Aglietta, M.; Morse, M.A.; Van Cutsem, E.; McDermott, R.; Hill, A.; et al. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *J. Clin. Oncol.* **2018**, *36*, 773–779. [\[CrossRef\]](#)
14. Abdullaev, S.; André, T.; Lei, M.; Lenz, H.-J.; Novotny, J.; Paulson, A.S.; Tejpar, S.; Yamazaki, K.; Ledezne, J.-M. A Phase III Study of Nivolumab (NIVO), NIVO + Ipilimumab (IPI), or Chemotherapy (CT) for Microsatellite Instability-High (MSI-H)/Mismatch Repair-Deficient (dMMR) Metastatic Colorectal Cancer (MCRC): Checkmate 8HW. *JCO* **2020**, *38*, TPS266. [\[CrossRef\]](#)

15. Hampel, H.; Pearlman, R.; Beightol, M.; Zhao, W.; Jones, D.; Frankel, W.L.; Goodfellow, P.J.; Yilmaz, A.; Miller, K.; Bacher, J.; et al. Assessment of Tumor Sequencing as a Replacement for Lynch Syndrome Screening and Current Molecular Tests for Patients With Colorectal Cancer. *JAMA Oncol.* **2018**, *4*, 1–8. [[CrossRef](#)]
16. Salipante, S.J.; Scroggins, S.M.; Hampel, H.L.; Turner, E.H.; Pritchard, C.C. Microsatellite Instability Detection by next Generation Sequencing. *Clin. Chem.* **2014**, *60*, 1192–1199. [[CrossRef](#)] [[PubMed](#)]
17. Hempelmann, J.A.; Scroggins, S.M.; Pritchard, C.C.; Salipante, S.J. MSIplus for Integrated Colorectal Cancer Molecular Testing by Next-Generation Sequencing. *J. Mol. Diagn.* **2015**, *17*, 705–714. [[CrossRef](#)] [[PubMed](#)]
18. Kautto, E.A.; Bonneville, R.; Miya, J.; Yu, L.; Krook, M.A.; Reeser, J.W.; Roychowdhury, S. Performance Evaluation for Rapid Detection of Pan-Cancer Microsatellite Instability with MANTIS. *Oncotarget* **2017**, *8*, 7452–7463. [[CrossRef](#)]
19. Huang, M.N.; McPherson, J.R.; Cutcutache, I.; Teh, B.T.; Tan, P.; Rozen, S.G. MSIsq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci. Rep.* **2015**, *5*, 13321. [[CrossRef](#)] [[PubMed](#)]
20. Niu, B.; Ye, K.; Zhang, Q.; Lu, C.; Xie, M.; McLellan, M.D.; Wendl, M.C.; Ding, L. MSIsensor: Microsatellite Instability Detection Using Paired Tumor-Normal Sequence Data. *Bioinformatics* **2014**, *30*, 1015–1016. [[CrossRef](#)]
21. Zhu, L.; Huang, Y.; Fang, X.; Liu, C.; Deng, W.; Zhong, C.; Xu, J.; Xu, D.; Yuan, Y. A Novel and Reliable Method to Detect Microsatellite Instability in Colorectal Cancer by Next-Generation Sequencing. *J. Mol. Diagn.* **2018**, *20*, 225–231. [[CrossRef](#)]
22. Cortes-Ciriano, I.; Lee, S.; Park, W.-Y.; Kim, T.-M.; Park, P.J. A Molecular Portrait of Microsatellite Instability across Multiple Cancers. *Nat. Commun.* **2017**, *8*, 15180. [[CrossRef](#)] [[PubMed](#)]
23. Lu, Y.; Soong, T.D.; Elemento, O. A Novel Approach for Characterizing Microsatellite Instability in Cancer Cells. *PLoS ONE* **2013**, *8*, e63056. [[CrossRef](#)] [[PubMed](#)]
24. Wang, Y.; Zhang, X.; Xiao, X.; Zhang, F.-R.; Yan, X.; Feng, X.; Zhao, Z.; Guan, Y.; Wang, J. Accurately Estimating the Length Distributions of Genomic Micro-Satellites by Tumor Purity Deconvolution. *BMC Bioinform.* **2020**, *21*, 82. [[CrossRef](#)]
25. Ratovomanana, T.; Cohen, R.; Svrcek, M.; Renaud, F.; Cervera, P.; Siret, A.; Letourneur, Q.; Buhard, O.; Bourgoin, P.; Guillermin, E.; et al. Performance of Next Generation Sequencing for the Detection of Microsatellite Instability in Colorectal Cancer with Deficient DNA Mismatch Repair. *Gastroenterology* **2021**. [[CrossRef](#)] [[PubMed](#)]
26. Lee, Y.; Lee, J.A.; Park, H.E.; Han, H.; Kim, Y.; Bae, J.M.; Kim, J.H.; Cho, N.-Y.; Kim, H.-P.; Kim, T.-Y.; et al. Targeted Next-Generation Sequencing-Based Detection of Microsatellite Instability in Colorectal Carcinomas. *PLoS ONE* **2021**, *16*, e0246356. [[CrossRef](#)]
27. Zhao, L.; Shan, G.; Li, L.; Yu, Y.; Cheng, G.; Zheng, X. A Robust Method for the Rapid Detection of Microsatellite Instability in Colorectal Cancer. *Oncol. Lett.* **2020**, *20*, 1982–1988. [[CrossRef](#)]
28. Zheng, K.; Wan, H.; Zhang, J.; Shan, G.; Chai, N.; Li, D.; Fang, N.; Liu, L.; Zhang, J.; Du, R.; et al. A Novel NGS-Based Microsatellite Instability (MSI) Status Classifier with 9 Loci for Colorectal Cancer Patients. *J. Transl. Med.* **2020**, *18*, 215. [[CrossRef](#)]
29. Suraweera, N.; Duval, A.; Reperant, M.; Vaury, C.; Furlan, D.; Leroy, K.; Seruca, R.; Iacopetta, B.; Hamelin, R. Evaluation of Tumor Microsatellite Instability Using Five Quasimonomorphic Mononucleotide Repeats and Pentaplex PCR. *Gastroenterology* **2002**, *123*, 1804–1811. [[CrossRef](#)]
30. Dedeurwaerdere, F.; Claes, K.B.; Van Dorpe, J.; Rottiers, I.; Van der Meulen, J.; Breyne, J.; Swaerts, K.; Martens, G. Comparison of Microsatellite Instability Detection by Immunohistochemistry and Molecular Techniques in Colorectal and Endometrial Cancer. *Sci. Rep.* **2021**, *11*, 12880. [[CrossRef](#)]
31. Saul, M.; Poorman, K.; Tae, H.; Vanderwalde, A.; Stafford, P.; Spetzler, D.; Korn, W.M.; Gatalica, Z.; Swensen, J. Population Bias in Somatic Measurement of Microsatellite Instability Status. *Cancer Med.* **2020**, *9*, 6452–6460. [[CrossRef](#)] [[PubMed](#)]