



Published in final edited form as:

Methods Mol Biol. 2020 ; 2055: 119–132. doi:10.1007/978-1-4939-9773-2_5.

Detection of microsatellite instability biomarkers via next-generation sequencing

Russell Bonneville^{1,2,*}, Melanie A. Krook^{1,*}, Hui-Zi Chen^{1,3,*}, Amy Smith¹, Eric Samorodnitsky¹, Michele R. Wing¹, Julie W. Reeser¹, Sameek Roychowdhury¹

¹Department of Internal Medicine, Division of Medical Oncology, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA

²Biomedical Sciences Graduate Program, The Ohio State University, Columbus, Ohio, USA

³Department of Internal Medicine, Hematology and Oncology Fellowship Program, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA

Abstract

A high level of microsatellite instability (MSI-H+) is an emerging predictive and prognostic biomarker for immunotherapy response in cancer. Recently, MSI-H+ has been detected within a variety of cancer types, in addition to the classical cancers associated with Lynch Syndrome. Clinical testing for MSI-H+ is currently performed primarily through traditional polymerase chain reaction (PCR) or immunohistochemistry (IHC) assays. However, next-generation sequencing (NGS) based approaches have been developed which have multiple advantages over traditional assays. For instance, NGS has the ability to interrogate thousands of microsatellite loci compared with just 5–7 loci that are detected by PCR. In this chapter, we detail the biochemical and computational steps to detect MSI-H+ from analysis of paired tumor and normal samples through NGS. We begin with DNA extraction, describe sequencing library preparation and quality control (QC), and outline the bioinformatics steps necessary for sequence alignment, pre-processing and MSI-H+ detection using the software tool MANTIS. This workflow is intended to facilitate more widespread usage and adaptation of NGS-powered MSI detection, which can be eventually standardized for routine clinical testing.

Keywords

Microsatellite instability; next-generation sequencing; biomarker; bioinformatics; clinical trials; tissue-agnostic; PD-1 inhibition

1. Introduction

In May 2017, the U.S. Food and Drug Administration (FDA) granted accelerated approval to pembrolizumab (KEYTRUDA®), a humanized antibody against the programmed death receptor-1 (PD-1), for treatment of patients with any advanced solid cancer harboring a high tumor mutation burden as measured by the presence of microsatellite instability (MSI-H+).

*These authors contributed equally to this work

This was the FDA's first cancer type-agnostic approval and resonated with the emerging theme of biomarker-driven therapy for cancer patients.

1.1. MSI as a biomarker for PD-1 inhibition

In normal cells, microsatellites are 10–60 base pair (bp) regions in the genome containing a defined number of repetitive DNA sequences ranging from 1–5 bp (Fig. 1). Germline (inherited) or somatic (acquired) mutations in DNA mismatch repair (MMR) proteins characteristically lead to unorganized expansion and/or contraction of these repetitive DNA sequences termed MSI-H+ (high levels of microsatellite instability). Defective mismatch repair results in high tumor mutation burden and abundant neo-antigen formation, which can be recognized by the host immune system. Cancer cells with high expression of PD-L1, a ligand of PD-1, evade detection and killing by the body's adaptive immune system. Blocking the PD-1/PD-L1 interaction re-primed a patient's immune system to attack previously 'invisible' cancer cells (1). Multiple studies have reported positive correlations between MSI-H+ status and increased PD-1/PD-L1 expression, as well as between MSI-H+ status and overall tumor mutational burden (2). For example, in MSI-positive colorectal cancer, elevated PD-L1 expression was detected on tumor-infiltrating immune (lymphoid and myeloid) cells (3). For these reasons, the use of pembrolizumab and nivolumab in patients with MSI-H cancers has proved to be efficacious. The observation of deep, durable responses in these patients has demonstrated the utility and validity of MSI as a predictive marker for response to immunotherapy with PD-1 blockade (1, 2).

1.2. New assays and methods for MSI detection

Up until approximately two years ago, the detection of MSI from tumor specimens has relied on the immunohistochemical detection (MSI-IHC) of the following four MMR proteins: MLH1, MSH2, MSH6 and PMS2. One main disadvantage of MSI-IHC is the inability to detect MSI caused by point mutations or small insertion/deletion mutations in MMR proteins that can still produce a positive IHC result. A second standard method for MSI detection has been a PCR-based assay of five (Bethesda panel) or eight (Promega® panel) standardized microsatellite loci (MSI-PCR). The main disadvantage of MSI-PCR is the severely limited number (5–8) of MSI loci that are evaluated by the assay. Both of these assays were originally designed for the diagnosis of inherited mismatch repair deficiency, or Lynch syndrome. Since Lynch syndrome predominantly includes colorectal and uterine cancers, these assays may be biased for these cancer types.

To overcome the limitations of traditional methods of MSI-IHC and MSI-PCR, integrated next generation sequencing (NGS) assays and novel computational methods have been developed to detect MSI-H+ (4). Some examples of NGS gene panels for MSI-H+ detection include MSIPlus (5) and ColoSeq (6). MSIPlus is an assay that has been optimized for colorectal cancer and evaluates 16 microsatellite loci along with mutational hotspots in oncogenes (*KRAS*, *NRAS*, and *BRAF*). ColoSeq is an alternative NGS assay that is designed to detect mutations, deletions or complex structural rearrangements in seven DNA repair genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, *APC*, and *MUTYH*) that can lead to MSI.

New bioinformatics approaches developed for MSI detection include mSINGS (7), MSISensor (8), MOSAIC (9), MANTIS (10, 11) and others (Table 1). Most of these computational algorithms compare the length distribution of a selection of microsatellite loci by determining the read counts of all alleles. Their development has enabled the assessment of thousands of microsatellite loci located predominantly in the protein-coding region of the genome (compared with just 5–8 loci using MSI-PCR) across many cancer types through analysis of publically available and previously annotated datasets. For example, utilizing the program MOSAIC, Hause *et al.* evaluated MSI-H+ status in n=5,930 cases spanning 18 cancer types from the The Cancer Genome Atlas (TCGA) (9). Adding to this knowledge, Bonneville *et al.* assessed MSI-H+ status with the program MANTIS in n=11,139 cases spanning 39 distinct cancer types from the TCGA and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (10). In a third study, Middha *et al.* evaluated MSI-H+ with MSISensor in n=12,288 advanced solid cancers profiled with the NGS assay, Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) (8). Finally, methods that assess MSI-H+ based on mutation burden in microsatellites are available (12–14). An example of this is MSIseq Index (12), which is the only MSI-H+ detection method that utilizes RNA sequencing data to determine proportion of insertion/deletions in microsatellites relative to all insertion/deletions in RNA transcripts.

In summary, numerous studies that have applied novel computational approaches have revealed an unexpectedly high incidence of MSI-H+ in a diverse range of human cancers. Importantly, these studies identify patients with non-Lynch cancer types affected by MMR deficiencies leading to MSI-H+ who may benefit from immunotherapy. Given the validity of MSI-H+ as a predictive biomarker of response to PD-1 inhibition, it is likely that standardized clinical MSI-H+ testing will become incorporated into the routine care of cancer patients in the near future.

In the following Methods section of this Chapter, we provide detailed protocols of DNA extraction from tissue, sequencing library generation, targeted hybridization/capture and bioinformatics methods (i.e. MANTIS) for computational MSI detection. It is important to note that the target region for hybridization and capture is dependent on the end user's needs and resources, and therefore the size may vary accordingly. Our laboratory targets 99 top performing microsatellite loci for determination of MSI status. Due to the small amount of genomic space occupied by these loci, we have chosen to use this design in combination with a larger panel (~1 megabase) for the detection of single nucleotide variants (SNV) and copy number variation (CNV). The methodologies described below are applicable across a variety of capture region sizes, however some optimization may be required.

1.3. Concluding remarks

Microsatellite instability has proven to be a clinically important biomarker for predicting response to immunotherapy. MSI has been observed across a wide variety of cancer types, and this requires a pan-cancer scope of testing. Next-generation sequencing and new analytical software have permitted expanded testing for MSI-H+ detection. NGS-based methods demonstrate superior performance to previous technologies, and MSI-H+ testing

can be easily integrated into other sequencing assays for more comprehensive genomic analysis.

2. Materials

2.1. DNA extraction

1. QIAamp DNA Blood Mini Kit (for DNA extraction from blood)
2. QIAamp DNA FFPE Tissue Kit (for DNA extraction from FFPE tissue)
3. DNase/RNase-Free 1.5 mL centrifuge tubes
4. Qiagen collection tubes
5. Qiagen RNase A
6. Pipettes (0.5–10 uL, 2–20 uL, 20–200 uL, 200–1000 uL) and plastic pipette tips
7. Ethanol: 200 proof
8. Centrifuge, thermomixer, and vortexer

2.2. Nucleic acid and library quality control

1. TapeStation 2200 (Agilent Technologies) with:
 - a. Strip tubes
 - b. Strip caps
 - c. Loading tips
 - d. Genomic DNA ScreenTape (4°C)
 - e. Genomic DNA Reagents (4°C)
 - f. D1000 ScreenTape (4°C)
 - g. D1000 Reagents (4°C)
2. Nanodrop 2000 Spectrophotometer with:
 - a. Blanking solution: Qiagen AE Buffer, Nuclease-free water, or Qiagen ATE Buffer
3. Qubit 2.0 Fluorometer with:
 - a. Qubit dsDNA High Sensitivity or Broad Range Assay Kit
 - b. Qubit Standards
 - c. Qubit Assay Tubes

2.3. Library preparation, hybridization, and sequencing

1. Pipettes (0.5–10 uL, 2–20 uL, 20–200 uL, 200–1000 uL) and plastic pipette tips
2. Centrifuge

3. Thermomixer
4. Vortexer
5. Heatblock
6. Magnet strip
7. Thermal cycler
8. Microcentrifuge tubes
9. PCR strip tubes
10. Ice bucket
11. Ethanol: 200 proof, room temperature
12. NaOH: 2N, -20°C
13. AMPure XP beads, 4°C
14. 2X KAPA HiFi HotStart ReadyMix (-20°C , included in kit)
15. Nuclease-free water
16. KAPA Library Amplification Primer Mix (-20°C , included in kit)
17. KAPA End Repair & A-Tailing Buffer (-20°C , included in kit)
18. KAPA End Repair & A-Tailing Enzyme (-20°C , included in kit)
19. KAPA Ligation Buffer (-20°C , included in kit)
20. KAPA DNA Ligase (-20°C , included in kit)
21. IDT Single Index Duplex Adapters (-20°C)
22. Cot-1 DNA: 1 ug/ul, -20°C
23. xGen Universal Blockers – TS Mix (-20°C)
24. xGen Hybridization and Wash Kit (-20°C)
25. DynaBeads® M-270 Streptavidin Beads (4°C)
26. xGen Library Amplification Primer – TS Mix (-20°C)
27. IDTE pH 8.0 (1X TE Solution) (room temperature)

2.4. Bioinformatics

1. Linux workstation, with at least 32 GB of memory and 8 CPU cores
2. Required software:

Software	Website	Citation
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	(17)
samtools	http://www.htslib.org/	(18)

Software	Website	Citation
bedtools	https://bedtools.readthedocs.io/en/latest/	(19)
bwa	http://bio-bwa.sourceforge.net/	(20)
GATK version 4.0 or later	https://software.broadinstitute.org/gatk/download/	(21)
MANTIS	https://github.com/OSU-SRLab/MANTIS	(11)

3. Required data files:

File	Website	Citation
hg19 reference FASTA	http://hgdownload.cse.ucsc.edu/downloads.html	(22)
dbSNP VCF	https://www.ncbi.nlm.nih.gov/variation/docs/human_variation_vcf/	(23)

3. Methods

For general notes pertaining to the entire workflow (see Notes 1–5).

3.1. GenExtraction of DNA from blood

Genomic DNA is extracted from whole blood using the **Qiagen QIAamp DNA Blood Mini Kit**, following the manufacturer's protocol. For points to be considered with respect to the protocol (see Notes 6–10).

3.2. Extraction of DNA from FFPE tissues

Genomic DNA is extracted from FFPE tissue using the **Qiagen QIAamp DNA FFPE Tissue Kit**, following the manufacturer's protocol. For points to be considered with respect to the protocol (see Notes 11–15).

3.3. Nucleic acid quality control (QC)

Extracted DNA needs to be quantified and its quality should be assessed. Three methods that are readily available and recommended for use include Agilent TapeStation, Thermo Scientific Nanodrop, and Thermo Scientific Qubit. After determining the quantity and quality of DNA extracted with one of the recommended methods, one can proceed to the library preparation, hybridization, and sequencing of samples for MSI detection.

TapeStation—Genomic DNA should be run on the TapeStation using Genomic DNA Reagents and ScreenTapes according to the manufacturer's protocol. After the run is complete, examine the electropherogram traces and note the quality of genomic DNA. General guidelines are that if more than 75% of the DNA sample is 10,000 bp or larger the sample is considered “intact”. If between 50–75% of the DNA sample is 10,000 bp or larger the sample is considered “partially degraded”. If less than 50% of the DNA sample is 10,000 bp the sample is considered “degraded”. Adjustments to these guidelines can be made in individual circumstances.

NanoDrop—Genomic DNA should be assessed by the NanoDrop according to the manufacturer's protocol. Instrument blanking should be performed with the buffer used in genomic DNA elution. To ensure proper blanking, the blanking solution should also be measured and this value recorded. Additionally, the A260/280 and A260/230 values and provided concentration should be recorded (although this is not the concentration that will be used for calculating input volume for subsequent steps).

Qubit—Genomic DNA should be quantified using the double stranded DNA (dsDNA) quantification assay (either high sensitivity (HS) or broad range (BR) based on sample concentrations) by the Qubit. This assay should be used according to the manufacturer's protocol, including use of the provided standards. The dsDNA concentration reported by the Qubit will be used for calculating the input volume for subsequent steps.

3.4. Fragmentation of DNA

Fragmentation of DNA is completed using a Covaris S220 Focused-ultrasonicator. The following protocol is recommended:

1. Prior to fragmentation of DNA, fill Covaris S220 water tank with deionized water to lower marker 15. Turn on unit and pump, and open sonification software. Allow one hour for system to initialize.
2. Dilute 200 ng of each gDNA sample from above steps (Extraction of DNA from Blood and Extraction of DNA from FFPE Tissues) in 10mM Tris-HCl pH 8.0 to a total volume of 55 uL. Ensure that the same input is used for all samples that will be combined into the same hybridization reaction.
3. Transfer 55 uL of diluted gDNA to a Covaris microTUBE, place in tube holder, and fragment the DNA using the following settings which are optimized for creating library inserts of 200–400 bp:

DUTY FACTOR	10%
CYCLES PER BURST	200
DURATION	120 SECONDS
PEAK POWER	175.0
MODE	FREQUENCY SWEEPING
TEMPERATURE	7°C

4. Label 0.2 mL strip tubes for each sample and transfer 50 uL fragmented dsDNA sample from microTUBE to corresponding strip tube.

3.5. Library preparation and QC

The **KAPA Hyper Prep Kit** is used for the rapid preparation of libraries from fragmented, double-stranded DNA for sequencing on the Illumina platform. The KAPA Hyper Prep Kit contains all of the necessary reagents for the three steps to generate libraries: end repair and A-tailing, adapter ligation, and library amplification. For points to be considered with respect to the protocol (see Notes 16–19).

3.6. Hybridization and final library QC

Hybridization and capture of the above DNA libraries is completed using **IDT xGEN® Lockdown Probes and Reagents**. The protocol should be followed according to manufacturer's recommendations. For recommendations to best optimize the protocol (see Notes 20–26).

3.7. Sequencing

Prior to sequencing on the MiSeq, final capture libraries must be denatured and diluted to an appropriate concentration for optimal cluster generation. Additionally, commercially available PhiX library should be denatured and diluted for use as a sequencing control. This should be done according to the MiSeq System Denature and Dilute Libraries Guide from Illumina. For recommendations to best optimize the protocol (see Notes 27–28).

3.8. Data file preparation

These steps only need to be performed once.

The reference genome FASTA file at /path/to/hg19.fa must be indexed with samtools:

```
samtools faidx /path/to/hg19.fa
```

A BED file of targeted microsatellite regions must be generated. First, in the tools directory of MANTIS, run:

```
make
```

to build the RepeatFinder utility. To extract microsatellite regions from the reference genome, run:

```
./RepeatFinder -i /path/to/hg19.fa -o hg19_microsatellites.bed
```

Using bedtools, filter this file for microsatellites within your genomic capture region or other desired regions to identify microsatellites:

```
bedtools intersect -a hg19_microsatellites.bed -b /path/to/targets.bed -wa >  
hg19_microsatellites_in_target.bed
```

For additional comments on the bioinformatics workflow outlined in this chapter (see Notes 29–32).

3.9. QC

Beginning with gzipped FASTQ files 123456_T_R1.fastq.gz and 123456_T_R2.fastq.gz from the tumor sample, initial QC is performed with fastqc as follows:


```
fastqc -o . 123456_T_R1.fastq.gz 123456_T_R2.fastq.gz
```

3.10. Alignment

Alignment will be performed by bwa using the MEM algorithm as follows:

```
bwa mem -M -t <number of threads> -R "@RG\tID:123456_T\tLB:123456_T\tSM:
123456_T\tPL:ILLUMINA" /path/to/hg19.fa 123456_T_R1.fastq.gz
123456_T_R2.fastq.gz | samtools view -s -b -o 123456_T.bam -
```

The resulting BAM file must now be sorted and indexed using samtools:

```
samtools sort -o 123456_T_sorted.bam 123456_T.bam
samtools index 123456_T_sorted.bam
```

3.11. Deduplication

PCR duplicates are removed using Picard tools:

```
gatk --java-options "-Xmx20g" MarkDuplicates --INPUT 123456_T_rmdup.bam --
OUTPUT 123456_T_rmdup.bam --METRICS_FILE 123456_T_rmdup_metrics.txt --
VALIDATION_STRINGENCY LENIENT --ASSUME_SORTED true --REMOVE_DUPLICATES true
```

This BAM file must also be indexed:

```
samtools index 123456_T_rmdup.bam
```

3.12. Base quality recalibration

Several post-alignment processing steps are recommended to improve alignment accuracy around indels and reduce quality score bias. These will be performed with the Genome Analysis Toolkit (GATK), using the dbSNP VCF listed in section 2.4:

```
gatk --java-options "-Xmx20g" BaseRecalibrator -R /path/to/hg19.fa -O
123456_T_recal.data.grp -I 123456_T_rmdup.bam --known-sites /path/to/
dbsnp.vcf
gatk --java-options "-Xmx20g" ApplyBQSR -R /path/to/hg19.fasta -O
123456_T_recal.bam -I 123456_T_rmdup.bam -bqsr 123456_T_recal.data.grp
samtools index 123456_T_recal.bam
```

Important: Repeat steps 3.9 to 3.12 for the matched normal sample, beginning with 123456_N_R1.fastq.gz and 123456_N_R2.fastq.gz and ending with 123456_N_recal.bam.

3.13. MSI calling

With the aligned and processed BAM files, along with the BED file of in-target microsatellite regions, we can now detect microsatellite instability in this sample. Run:

```
python /path/to/mantis.py -t 123456_T_recal.bam -n 123456_N_recal.bam -b  
hg19_microsatellites_in_target.bed -o 123456.mantis.txt --genome /path/to/  
hg19.fa
```

This will output several files, including 123456.mantis.txt.status. View this using cat:

```
cat 123456.mantis.txt.status
```

In the table, there will be a value listed for “Step-Wise Difference (DIF)”. This is the MSI score for this tumor. There will also be a status, either “Stable” for MSS or “Unstable” for MSI-H.

4. Notes

4.1. General notes

1. Laboratory personnel must wear gloves and laboratory coat at all times during extraction of DNA from blood samples.
2. Pipettes should be calibrated every 6 months and decontaminated with Sporidicin® Disinfectant Solution daily prior to and after each use.
3. Should blood not be available to be used as a normal control, DNA from buccal swabs or adjacent normal tissue from FFPE block can also be used.
4. This protocol can also be completed on fresh frozen tissues. To extract DNA from these tissues or buccal swabs use the QIAamp DNA Mini Kit.
5. In regard to the three options indicated for quantity and quality control of DNA, Nanodrop will quantify single and double stranded DNA, whereas Qubit only quantifies dsDNA. It is recommended to use the Qubit concentration for input calculations. However, the more degraded the sample, the bigger of a difference will be observed between Nanodrop and Qubit concentration values. Nanodrop also provides 260/280 and 260/230 ratios which are helpful to assess quality and can help determine protein or RNA contamination. Additionally, the Agilent Bioanalyzer can be used instead of the TapeStation.

4.2. Extraction of DNA from Blood

6. DNA must be extracted from whole blood samples stored at 4°C within one week (7 days). Whole blood samples can be stored at –80°C for months to years and then can subsequently be used for DNA extraction.

7. Prior to starting the protocol, allow samples to equilibrate to room temperature (15–25°C).
8. It is recommended to add RNase A stock solution (100mg/mL) to the sample prior to the addition of Buffer AL in order to yield RNA-free genomic DNA as the presence of RNA may inhibit downstream enzymatic reactions.
9. A second elution step with Buffer AE or distilled water may increase yields up to 15%.
10. DNA can be stored at 4°C for short-term storage but should be stored at –80°C for long-term storage.

4.3. Extraction of DNA from FFPE Tissues

11. DNA should be extracted from FFPE tissue as soon as possible upon receipt.
12. Depending on tissue size, 1–8 sections should be cut at a 10µm thickness and placed on a slide. If the tissue block is large enough to be macrodissected, scrape off the area of interest (enriched with tumor cells) using a scalpel and place in a 1.5 mL eppendorf tube, avoiding excess paraffin in the process. Alternatively, if using the entire tissue from the block, scrolls can be cut and placed directly into 1.5 mL eppendorf tubes.
13. If using only one heating block, leave the sample at room temperature after the 56°C incubation until the heating block has reached 90°C.
14. Incubating the QIAamp MinElute column with Buffer ATE on center of the membrane for 1–5 minutes prior to centrifugation can increase yields.
15. DNA can be stored at 4°C for short-term storage but should be stored at –80°C for long-term storage.

4.4. Library preparation and QC

16. Ensure that the same input is used for all samples that will be combined into the same hybridization reaction.
17. Adapters must be purchased separately and are available from several commercial companies. The adapter stock concentration needs to be adjusted depending on the amount of DNA input used for library preparation. Recommended adapter concentrations for DNA inputs ranging from 1 ng – 1 µg are found in the Technical Data Sheet for the KAPA Hyper Prep Kit.
18. The number of PCR cycles used for library amplification needs to be modified depending on the amount of DNA input used for library preparation. It is recommended to use 5 PCR cycles for an input of 200 ng. Recommended PCR cycle numbers to generate either 100 ng or 1 µg of amplified library from 1 ng – 1 µg of input DNA are found in the Technical Data Sheet for the KAPA Hyper Prep Kit.

19. After generation of the libraries, the Qubit and TapeStation (D1000) should be used to check library concentration and size.

4.5. Hybridization and final library QC

20. Inspect the tube of 2X Hybridization Buffer for crystallization of salts. If crystals are present, heat the tube at 65°C, shaking intermittently, until the buffer is completely solubilized (may require heating for several hours).
21. The recommended protocol uses a 4-hour incubation for the hybridization, however the protocol can be modified for an overnight hybridization (16–18 hours).
22. DynaBeads® M-270 Streptavidin Beads should be prepared immediately before use. Do not allow beads to dry out.
23. Hybridization/capture reactions have been optimized to include a total of 4 samples (matched normal and tumor libraries from two patients) based on the size of our capture region and sequencing capacity on the MiSeq instrument. To bias sequencing in favor of the tumor samples, we input a greater quantity of tumor library (200 ng) compared to normal library (50 ng). Based on these inputs we are able to achieve sequencing depths of ~150x for normal samples and 500x for tumor samples.
24. It is recommended to use 12 cycles of PCR for the post-capture amplification step. The number of PCR cycles should be optimized based on the panel size and total amount of library in the hybridization reaction to ensure sufficient material for sequencing. Recommended PCR cycles based on these parameters can be found in the xGen® Hybridization Capture of DNA Libraries for NGS Target Enrichment protocol handbook available through IDT's website.
25. After hybridization, Qubit and TapeStation (D1000) should be used to check final library concentration and size. Alternatively, library quantification can be carried out using a qPCR kit.
26. To calculate the final nM concentration of your library, use the formula listed below:

$$Final\ Library\ Conc. \left(\frac{ng}{\mu l} \right) \times \frac{10^6 \mu l}{1L} \times \frac{1\ nmol}{660\ ng} \times \frac{1}{Fragment\ Size\ in\ bp} = Final\ Library\ Conc. (nM)$$

4.6. Sequencing

27. Prepare a fresh dilution of NaOH for each run and use within 12 hours.
28. Denatured 20pM PhiX library can be stored at –15°C to –25°C for up to 3 weeks.

4.7. Bioinformatics

29. This workflow can be adapted to hg38 (or future human genome builds) by substituting the appropriate genome FASTA file and a targeted capture BED in hg38 coordinates.
30. RepeatFinder has several options to fine tune the microsatellites considered for analysis. Of particular interest is -L to determine the maximum length of the microsatellite repeat unit (k-mer), along with -m and -M to control the minimum and maximum length of microsatellites respectively. The default settings are appropriate in most use cases.
31. If either your target BED or microsatellites BED files are very large, bedtools may fail due to excessive memory usage. In this case, position-sort both BED files by bedtools sort -i in.bed > out.bed , then run bedtools with the -sorted flag (19).
32. The flag “-Xmx20g” allocate 20 gigabytes of memory for Java. If you encounter an out of memory error, consider increasing this threshold.

5. References

1. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science*. 2018;359(6382):1350–5. [PubMed: 29567705]
2. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin Cancer Res*. 2016;22(4):813–20. [PubMed: 26880610]
3. 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 Discov*. 2015;5(1):43–51. [PubMed: 25358689]
4. Baudrin LG, Deleuze JF, How-Kit A. Molecular and Computational Methods for the Detection of Microsatellite Instability in Cancer. *Front Oncol*. 2018;8:621. [PubMed: 30631754]
5. Hempelmann JA, Scroggins SM, Pritchard CC, Salipante SJ. MSIplus for integrated colorectal cancer molecular testing by next-generation sequencing. *J Mol Diagn*. 2015;17(6):705–14. [PubMed: 26322950]
6. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn*. 2012;14(4):357–66. [PubMed: 22658618]
7. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem*. 2014;60(9):1192–9. [PubMed: 24987110]
8. Middha S, Zhang L, Nafa K, Jayakumaran G, Wong D, Kim HR, et al. Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precision Oncology*. 2017(1):1–17.
9. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med*. 2016;22(11):1342–50. [PubMed: 27694933]
10. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen H-Z, et al. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precision Oncology*. 2017(1):1–15.
11. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget*. 2017;8(5):7452–63. [PubMed: 27980218]
12. Lu Y, Soong TD, Elemento O. A novel approach for characterizing microsatellite instability in cancer cells. *PLoS One*. 2013;8(5):e63056. [PubMed: 23671654]

13. Huang MN, McPherson JR, Cutcutache I, Teh BT, Tan P, Rozen SG. MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci Rep*. 2015;5:13321. [PubMed: 26306458]
14. Nowak JA, Yurgelun MB, Bruce JL, Rojas-Rudilla V, Hall DL, Shivdasani P, et al. Detection of Mismatch Repair Deficiency and Microsatellite Instability in Colorectal Adenocarcinoma by Targeted Next-Generation Sequencing. *J Mol Diagn*. 2017;19(1):84–91. [PubMed: 27863258]
15. Zhu L, Huang Y, Fang X, Liu C, Deng W, Zhong C, et al. A Novel and Reliable Method to Detect Microsatellite Instability in Colorectal Cancer by Next-Generation Sequencing. *J Mol Diagn*. 2018;20(2):225–31. [PubMed: 29277635]
16. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun*. 2017;8:15180. [PubMed: 28585546]
17. Andrews S, Babraham Institute;Pages <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. [PubMed: 19505943]
19. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. [PubMed: 20110278]
20. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. [PubMed: 19451168]
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010;20:1297–303. [PubMed: 20644199]
22. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921. [PubMed: 11237011]
23. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine dbSNP accession:rs1799977. (dbSNP Build ID: 141). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.

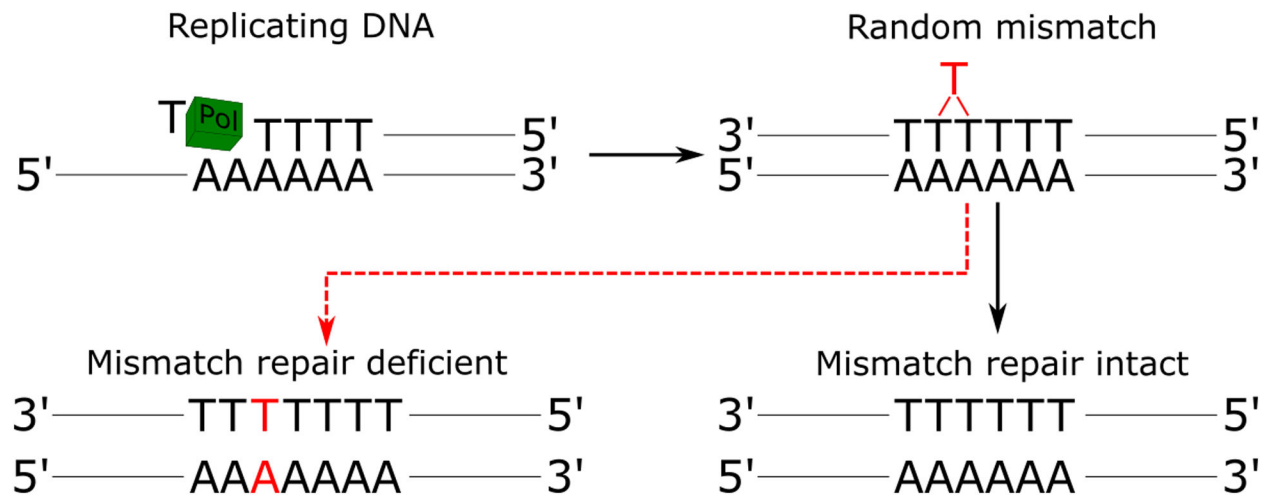


Figure 1:
Schematic of microsatellite instability due to deficient mismatch repair (MMR). Loss of MMR prevents a cell from repairing random mismatches between antiparallel strands that arise at repetitive regions during DNA replication, leading to propagation of random insertions/deletions (indels) in microsatellites.

Table 1.

Examples of Computational Methods for MSI Detection from NGS data (DNA)

Computational Method	Samples analyzed	MSI calling method
mSINGS	Tumor v. baseline normal	Binary MSI/MSS classifier
		MSI threshold: >20% unstable loci
MSIsensor	Tumor v. paired normal	Binary MSI/MSS classifier
		MSI threshold: >3.5% unstable loci
MANTIS	Tumor v. paired normal	Binary MSI/MSS classifier
		MSI threshold: average aggregate MSI score >0.4
MSI-ColonCore (15)	Tumor v. baseline normal	MSI-H/MSI-L/MSS classifier
		MSI-H threshold: >40% unstable loci
Cortes-Ciriano method (16)	Tumor v. paired normal	Binary MSI/MSS classifier
		Random forest based