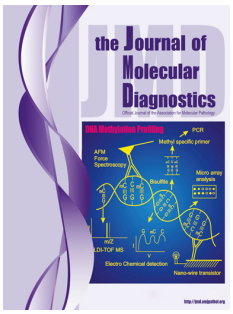
Accepted Manuscript



A Novel and Reliable Method to Detect Microsatellite Instability in Colorectal Cancer by Next-Generation Sequencing

Lizhen Zhu, Yanqin Huang, Xuefeng Fang, Chenglin Liu, Wanglong Deng, Chenhan Zhong, Jinghong Xu, Dong Xu, Ying Yuan

PII: S1525-1578(17)30369-0

DOI: 10.1016/j.jmoldx.2017.11.007

Reference: JMDI 657

To appear in: *The Journal of Molecular Diagnostics*

Accepted Date: 22 November 2017

Please cite this article as: 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, *The Journal of Molecular Diagnostics* (2018), doi: 10.1016/j.jmoldx.2017.11.007.

A novel and reliable method to detect microsatellite instability in colorectal cancer by next-generation sequencing

Lizhen Zhu,*† Yanqin Huang,† Xuefeng Fang,*† Chenglin Liu,‡ Wanglong Deng,§ Chenhan Zhong,*† Jinghong Xu,¶ Dong Xu,‖ and Ying Yuan*†

From the Department of Medical Oncology,* The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou; the Cancer Institute (Key Laboratory of Cancer Prevention and Intervention, Chinese National Ministry of Education; Key Laboratory of Molecular Biology in Medical Sciences), and the Departments of Pathology¶ and Surgical Oncology,‖ The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou; and the Departments of Bioinformatics,‡ and Research and Development,§ Burning Rock Biotech, Guangzhou, China

Short title: A novel and reliable method to detect MSI in CRC by NGS

Footnote: L.Z. and Y.H. contributed equally.

Funding: Supported by Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (No. 2014BAI09B07; Y.Y.) and the Natural Science Foundation of Zhejiang Province (LY16H160027; X.F.).

Corresponding author

Ying Yuan, Email: yuanying1999@zju.edu.cn

Disclosures: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Two types of molecular tests have been established to assess the deficiency of DNA mismatch repair (MMR) system: microsatellite instability (MSI) and immunohistochemical (IHC) analysis. We have developed a reliable method to analyze the MSI status by next-generation sequencing (NGS) based on read count distribution. A total of 91 patients with primary colorectal cancer were recruited. These patients included 54 cases with the loss of expression of any MMR protein in IHC suggesting deficient MMR (dMMR), and 37 cases of colorectal cancer with staining of all four MMR proteins in IHC, suggesting proficient MMR (pMMR) in postoperative sample. DNA was extracted from paired tumor-normal tissue for MSI detection by both ColonCore NGS panel and PCR. The sequencing data from NGS panel was processed using various MSI detection pipelines for a comparison with the ColonCore panel. Using MSI-PCR test as the gold standard, MSI-ColonCore achieved 97.9% sensitivity (47/48) and 100% specificity (37/37) for the detection of MSI status. MSI-ColonCore also showed more efficient and robust performance compared with other NGS-based MSI detection algorithms. The concordance rate was 92.3% between MSI-ColonCore and IHC testing, and 93.4% between MSI-PCR and IHC testing. These results suggest that MSI-ColonCore is a

reliable and robust method for MSI status detection by NGS-based on read count distribution.

Introduction

Microsatellites are tandem DNA repeats with one to six bases in coding and non-coding regions throughout the genome. The polymerase slippage during DNA synthesis leads to accumulation of mutations in microsatellites, and the two main types of errors are base–base mismatches and insertion–deletion. These errors are usually detected and corrected by the DNA mismatch repair (MMR) system. Deficient MMR (dMMR) activity caused by germline mutations or hypermethylation of MMR genes can lead to a hypermutable phenotype at the genomic level, named microsatellite instability (MSI).[1] Therefore, the MMR function can be detected by MSI analysis or immunohistochemical (IHC) loss of expression of any MMR proteins.

The DNA mismatch repair (MMR) function detection is applied to not only the initial molecular screening for Lynch syndrome, a major type of hereditary colorectal cancer (CRC) characterized by germline mutations in MMR genes, but also the selection of suitable patients for immunotherapy, since anti–programmed death-1 (anti–PD-1) therapies have achieved significant success in various MSI-H/dMMR cancers with the fact that Pembrolizumab (anti–PD-1 therapy) is recently approved by Food and Drug Administration for the treatment of

patients with unresectable or metastatic solid tumors who are referred to be MSI-H or dMMR. [2,3]

Before the era of massively parallel DNA sequencing, MSI is detected by PCR-based methods at specific microsatellite markers and CRCs can be classified into MSI-high (MSI-H), MSI-low (MSI-L), and microsatellite stable (MSS) according to the proportion of unstable markers. [4,5] As next-generation sequencing (NGS) is increasingly applied to detect tumor gene mutations, combining MSI status and mutation detection into the same sequencing process would highly decrease the demand of tissue samples and increase the efficiency of tests.

Currently, two types of methods have been proposed for the detection of MSI status by NGS. The first type tries to postulate MSI status from mutation burden, which is usually detected by whole-exome sequencing, and demonstrates a significant correlation between total mutation burden and MSI status. [6-8] The other type of methods directly measure the level of microsatellite instability by the read count distribution of a selected set of loci with different repeat lengths. Current approaches based on read count distribution include MSIsensor and mSINGS. [9,10] MSIsensor requires paired tumor and normal samples, and compares the histogram of read counts covering different repeat lengths of the loci using standard χ^2 test.

A locus is considered length-instable if the adjusted *P*-value is less than a pre-determined threshold. The percentage of length-instability loci are used to determine the MSI status. mSINGS does not require normal controls for MSI status detection. It determines the length-instable locus if the number of the types of repeat lengths is larger than

$$[\text{reference mean} + 3 \times \text{SD}].$$

The differences of length-instable loci percentage are statistically significant between MSI-H and MSS samples as to both methods.

Here, we have developed a new reliable algorithm to analyze MSI by NGS read count distribution, and compare the performance with MSIsensor and mSINGS. This algorithm, combined with the ColonCore panel, has also been validated against conventional PCR-MSI tests in a pool of samples with known IHC status of four major MMR proteins: MLH1, MSH2, MSH6, and PMS2.

Materials and Methods

With the approval of Ethics Committee of the Second Affiliated Hospital of Zhejiang

University School of Medicine and informed consent of all patients or their relatives, a total of 91 patients with primary CRC were recruited from January 2015 to January 2017., Among these, 54 cases were randomly selected from CRCs with IHC loss of expression of any of four MMR proteins (MLH1, MSH2, MSH6, and PMS2), and 37 cases were randomly recruited from those with intact expression of all four MMR proteins. They were not randomly selected from all the CRCs. For each tumor, formalin-fixed, paraffin-embedded tumor tissue was obtained postoperatively, as well as normal tissue from negative surgical margin, and the necrotic area ≤ 50%, and the percentage tumor cellularity was checked: 80.2% samples(73/91) ≤ 50%, 19.8% samples (18/91) range from 30% to 50% and none lower than 30%. DNA was extracted from paired tumor-normal tissue for MSI detection by NGS and PCR. Each group of researchers interpreting the MSI-ColonCore status, IHC results, and MSI-PCR results was blinded to the results of the other two tests.

MSI detection by ColonCore panel

ColonCore panel (Burning Rock, Guangzhou, China) is designed for simultaneous detection of MSI status and mutations in 36 CRC-related genes, including *KRAS*, *NRAS*, *BRAF*,

hereditary CRC genes, and other genes related to carcinogenesis and tumor development (Supplemental Table S1). The MSI phenotype detection method of MSI-ColonCore was a read-count–distribution-based approach. It utilized the coverage ratio of a specific set of repeat lengths as the main characteristic of each microsatellite locus, and categorized a locus as unstable if the coverage ratio was less than a given threshold. The MSI status of a sample was determined by the percentage of unstable loci in the given sample. The details of the method are described below. The raw reads were deposited in NCBI Sequence Read Archive (SRA) (https://trace.ncbi.nlm.nih.gov/Traces/sra/) accession number SRP119517.

Data preprocessing

To determine the MSI status, the microsatellite loci were firsst scanned from the reference genome, and the number of reads aligned to the loci of different repeat lengths was calculated in a training set of samples of known MSI status. The scanned loci were restricted to mononucleotide repeats, as those were reported as the most sensitive and specific for MSI detection. [10] Sequencing reads were aligned by BWA (v0.7.10) against the reference genome (hg19/GRCh37). Reads aligned to the loci at every possible repeat length were counted respectively using the same strategy as proposed by MSIsensor (9). To obtain the read counts

with highest specificity and also allow for sequencing errors, especially for loci with low coverage, two versions of read count statistics were generated by applying perfect match restriction and allowing 1bp-mismatch, respectively, in the alignment.

Loci characterization and baseline construction

Among the loci scanned from the reference genome, those showed high consistency between the stability of their repeat lengths and the MSI status of their corresponding sample were selected as marker microsatellite loci. In microsatellite stable samples, the length of homopolymers in the marker microsatellite loci were relatively stable. In other words, reads were aligned to only a few types of repeat-lengths. For each microsatellite locus, the specific repeat-length covered by the largest amount of reads was called peak length, and the read count was called peak count. Repeat lengths covered by no less than 75% peak count were recorded for each normal sample in the training set. This length set was called as reference length set which is then used for baseline construction. At the baseline construction stage, the ratio of read count covering the reference length set divided by the total read count covering all possible lengths of the locus was calculated for each normal control. The average coverage ratio (mean) and the standard deviation level (SD) of all normal samples in the training set

were calculated afterwards, and the threshold of

$$[\text{mean} - 3 \times \text{SD}]$$

was set as the lower limit for a length-stable locus.

A Locus with coverage ratio less than the threshold was determined as a length-instable one.

The final set of marker microsatellite loci were then selected using the training sample set with the following criteria: length-instable in more than 75% MSI samples, length-stable in more than 75% MSS samples, and has an average Spearman correlation higher than 0.8 between the ratio of reads covering each type of repeat length of the loci of each pair of normal samples.

MSI status determination for samples

After the selection of marker microsatellite loci and the establishment of the ratio reference, the MSI status of a tumor sample could be determined based on the percentage of length-instable loci, without a paired normal control. For each marker locus, the read count histogram was constructed and the coverage ratio of the reference length set was calculated and compared to the reference threshold. A locus with a coverage ratio less than

$$[\text{mean} - 3 \times \text{SD}]$$

of the reference ratio was determined as a length-instable locus. A tumor sample is considered MSI-H if more than 40% of the marker loci are length-instable, MSS if the percentage of length-instable loci were less than 15%, or MSI-L for if the percentage is between 15% to 40%.

Performance evaluation index for the NGS-based approaches

MSI status reported by PCR was set as the ground truth, with MSI-H (PCR) samples as positive and MSS(PCR) samples as negative. Four widely used measurements were adopted for performance evaluation, including sensitivity (SN), specificity (SP), accuracy (ACC) and Matthew's correlation coefficient (MCC), as illustrated in Eq. 1-4.

$$SN = \frac{TP}{TP + FN} \tag{1}$$

$$SP = \frac{TN}{TN + FP} \tag{2}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+FP)(\mathrm{TN}+\mathrm{FN})}} \tag{4}$$

where TP, TN, FP, and FN denoted the true positive, true negative, false positive, and false negative, respectively.

MSI detection by PCR

MSI-PCR testing was performed as the gold standard of MSI status. Genomic DNA extracted from all paired tumor-normal samples was tested by the Beijing Microread Genetics Co. Ltd. using the MSI detection kit (Microread Genetics Co. Ltd., Beijing, China, Patent No: ZL 201110152226.X) on the ABI 3730xl Genetic Analyzer (Applied Biosystems, Foster City, CA). The panel employed for the microsatellite instability analysis composed of nine markers, including six mononucleotide repeat sequences (NR-21, BAT-26, NR-27, BAT-25, NR-24, and MONO-27) , two pentanucleotide repeat sites (Penta C and Penta D), and a sex loci (Amelogeni). Penta C, Penta D, and Amelogeni were used for sample contamination control only. Data were collected and analyzed with the GeneMapper v4.0 software (Applied Biosystems, Foster City, CA). Microsatellite instability of any marker was defined when there were peaks in the fluorescence profile of the amplified microsatellite DNA from tumor tissue that were absent in a corresponding profile of the paired normal tissue. Samples were categorized into MSI-H (≥ 2 mononucleiotide markers instable), MSI-L (one mononucleiotide marker instable) and MSS (none of the mononucleiotide markers showed instability), and the cutoff value was 33.3%, within the range of 30% to 40%.

MMR Analysis by IHC Staining

Immunohistochemistry staining of CRCs was performed to examine the expression of four MMR proteins, MLH1, MSH2, MSH6, and PMS2 on formalin-fixed, paraffin-embedded tissue. Primary monoclonal antibodies against MLH1 (clone ES05, diluted 1:50 [DAKO, Carpinteria, CA]), MSH2 (clone FE11, diluted 1:50 [Oncogene Research Products, Boston, MA]), MSH6 (clone EP49, diluted 1: 150 [DAKO, Carpinteria, CA]), and PMS2 (clone EP51, diluted 1: 50 [DAKO, Carpinteria, CA]) were used with external controls. Deficient MMR (dMMR) was interpreted when any of these MMR proteins is totally absent in the nuclear staining of tumor tissue while present in nuclear staining of adjacent benign tissue, and any convincing nuclear staining of all of these four proteins was considered proficient MMR (pMMR). The IHC results were assessed by two specialized pathologists, and only concordant samples were included in the present study.

Statistical Analysis

Categorical variables were analyzed using χ^2 test, and continuous variables were analyzed using unpaired Student's test. Two-sided *P*-value of <0.05 was considered statistically significant.

Results

Marker microsatellites loci of ColonCore panel

In all target regions of ColonCore panel covering 36 CRC-related genes, 90 microsatellite loci with homopolymers no less than 8bp long were scanned out. The read count ratio covering the reference length set was calculated for each locus using the normal controls of the training sample set (20 tumor-normal sample pairs with previously determined MSI phenotype, and the lower limit for a stable locus was set as

[mean – 3 X SD].

Twenty-two marker microsatellite loci were selected as the final set according to criteria described in the *Materials and Methods*. The list of the marker loci and the baseline statistics are listed in Table 1.

Validation of MSI-ColonCore MSI method against conventional PCR-MSI

MSI-ColonCore achieved 97.9% sensitivity (47/48) and 100% specificity (37/37) for the detection of MSI when MSI-PCR testing performed as the gold standard, with one PCR

MSI-H sample labeled as MSI-L (Table 2).

Correlation between MSI and IHC status

According to MSI-ColonCore, all of 37 IHC pMMR cases were identified as MSS, whereas 47 of 54 IHC dMMR cases were MSI-H, with the remaining seven interpreted as MSS/MSI-L. In MSI-PCR testing, the results were almost the same as MSI-ColonCore, except that the case interpreted as MSI-L in MSI-ColonCore turned out to be MSI-H in MSI-PCR, and one considered as MSI-L in MSI-ColonCore turned to be MSS in MSI-PCR. Therefore, the concordance rate was 92.3% between MSI-ColonCore and IHC testing, and 93.4% between MSI-PCR and IHC testing. (Table 2)

Comparison of MSI status detection ability among different NGS-based methods

Here, we compared the performance of our approach to that of two previously published read-count–distribution-based methods: MSIsensor(v0.2) and mSINGS(v2.0). The MSI status of 79 samples was first determined by PCR method as the ground truth. One sample was reported as MSI-L by PCR method, and was excluded from the further performance

evaluation. The 90 microsatellite loci with the length of homopolymers no less than 8bp were canned out of the target region, and were utilized as marker loci for MSIsensor. The threshold for the percentage of length-instable loci was set at 30% for MSIsensor. It achieved the best performance based on this threshold. For mSINGS, loci that may cause artifacts were excluded from the baseline according to the recommendation of the software. Forty-four loci were retained as marker loci afterwards. The threshold for the percentage of length-instable loci was set at the default value of 10% recommended by mSINGS. The performance indexes of these three methods are shown in Table 3. ColonCore panel achieved best performance according to Table 3. In addition, the independency of paired normal sample of the algorithm makes it more practical in clinical applications.

ColonCore: a more robust MSI status detection method comparing to mSINGS

As the percentage of length-instable loci is the key index to distinguish between MSI statuses for a sample, the distribution of the percentage in MSS and MSI-H samples were compared between the three methods. The percentages were most distinguished by ColonCore panel, which demonstrated its robustness in MSI status detection (Figure 1).

Mutation burden comparison between MSI-H and MSS samples

The high correlation of MSI status and mutation burden has been demonstrated by recent reports benefiting from whole-exon sequencing. The mutation burden per Mb of MSI-H and MSS samples reported by MSI-ColonCore was presented as two violin plots (Figure 2). Although MSI-H samples tended to have higher mutation burden, the two types of samples were not as highly distinguished as reported.

Discussion

The deficiency of DNA mismatch repair system can be assessed through approaches at two different levels: genomic level (MSI analysis, PCR or NGS-based) and protein level (IHC tests of MMR proteins). It has been reported that the MSI analysis and IHC testing are highly related, with a concordance rate ranging from 84.5% to 98.6%. [2,12-14] Here, the concordance rate was 92.3% between MSI-ColonCore and IHC testing, and 93.4% between MSI-PCR and IHC testing.

At the genomic level, PCR-based approaches have been the gold standard for MSI analysis.

With the development of NGS technology, NGS-based MSI analysis has been increasingly adopted for two major advantages. First, NGS sequencing panels, when properly designed, can capture the mutation spectrum and MSI status in CRC patients simultaneously, reducing the amount of tissue sample required and simplifying the testing process. Second, after NGS-based MSI analysis properly constructing baseline reference set, like in the ColonCore panel, it eliminates the need of normal control samples, which will benefit patients without surgery, especially for most metastatic cancers.

As described earlier, NGS-based MSI analyses fall into two categories: the mutation burden approach and the read count distribution approach. The application of the mutation burden approach is limited in clinical practice, for the need of large and costly sequencing panels of hundreds of genes or even whole exome analysis, as mutation burden calculated from small panels tends to deviate from the real value. For an example, in the present study, mutation burden calculated from our panel of 36 hotspot genes is obviously over-estimated than those from large panels, and due to the small number of genes, even adjusted mutation burdens will be biased. Another technical challenge for mutation-burden–based MSI analysis is that cutoff-values between MSI-H, MSI-L, and MSS samples have to be defined for each specific

sequencing panel.

In contrast, read-count–distribution-based MSI analysis not only shows high consistency with the gold standard, but also suits clinical applications for its compatibility with smaller, cheaper, and more efficient sequencing panels like the ColonCore panel in this study. It is also more versatile as the cutoff-values are easy to define for any given panel using a similar logic to that in PCR-based approaches: the percentage of instable microsatellite loci. Although all read-count–distribution-based NGS methods achieved similar performance in this experiment, the MSI-ColonCore showed the most robustness compared to MSIsensor and mSINGS, as the percentages of length-instable loci were most distinguished between the MSI-H and MSS samples. After NGS-based MSI analysis properly constructing baseline reference set, it also eliminated the need for normal controls of MSIsensor.

Although MSI-PCR and IHC, two methods to detect mismatch repair system function, are well established and relatively inexpensive, such methods have limited capability to multiplex. In contrary, NGS allows for large-scale parallel sequencing and has proved to be a cost-effective and accurate tool for the parallel profiling of different forms of genetic abnormalities including mutations, fusions, and amplifications across a large number of genes,

which could not be provided by MSI-PCR or IHC but are very important in clinical practice. Besides, previous studies had discussed the power of NGS in clinical practice [22, 23]. Furthermore, the performance of MSI-ColonCore is comparable with golden standard PCR-MSI. Therefore, the NGS-based ColonCore panel is cost-effective and promising in clinical practice.

Our study is to some extent limited by the relatively small number of cases, due to the only 15% CRCs driven by MMR deficiency. [24] More cases will be recruited to further validate our findings, and the capability of MSI-ColonCore will also be tested in other types of cancers with high MSI, such as endometrial cancer and gastric cancer.

In summary, MSI-ColonCore can detect MSI accurately and more robustly compared with current NGS methods based on read count distribution.

Acknowledgments

We thank Haiyan Xu and Han Han-Zhang for medical editorial assistance and Zhou Zhang for data analysis.

References

1. Duval A, Hamelin R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. Cancer Res 2002, 62(9):2447–54.

2. Moreira L, Balaguer F, Lindor N, de la Chapelle A, Hampel H, Aaltonen LA, Hopper JL, Le Marchand L, Gallinger S, Newcomb PA, Haile R, Thibodeau SN, Gunawardena S, Jenkins MA, Buchanan DD, Potter JD, Baron JA, Ahnen DJ, Moreno V, Andreu M, Ponz de Leon M, Rustgi AK, Castells A; EPICOLON Consortium. Identification of Lynch syndrome among patients with colorectal cancer. JAMA 2012, 308:1555- 1565.

3. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhaijee F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardoll DM, Papadopoulos N, Kinzler KW, Zhou S, Cornish TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B, Diaz LA Jr. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015, 372:2509-20.

4. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ,

Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 1998, 58:5248–5257.

5. 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 mono-nucleotide repeats and pentaplex PCR. Gastroenterology. 2002, 123:1804–1811.

6. Huang MN, McPherson JR, Cutcutache I, Teh BT, Tan P, Rozen SG. MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. Scientific reports. 2015, 5:13321.

7. Nowak JA, Yurgelun MB, Bruce JL, Rojas-Rudilla V, Hall DL, Shivdasani P, Garcia EP, Agoston AT, Srivastava A, Ogino S, Kuo FC, Lindeman NI, Dong F. Detection of Mismatch Repair Deficiency and Microsatellite Instability in Colorectal Adenocarcinoma by Targeted Next-Generation Sequencing. The Journal of molecular diagnostics : JMD. 2017, 19(1):84-91.

8. Stadler ZK, Battaglin F, Middha S, Hechtman JF, Tran C, Cercek A, Yaeger R, Segal NH, Varghese AM, Reidy-Lagunes DL, Kemeny NE, Salo-Mullen EE, Ashraf A, Weiser MR, Garcia-Aguilar J, Robson ME, Offit K, Arcila ME, Berger MF, Shia J, Solit DB, Saltz LB. Reliable Detection of Mismatch Repair Deficiency in Colorectal Cancers Using Mutational Load in Next-Generation Sequencing Panels. J Clin Oncol 2016, 34:2141-2147.

9. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics (Oxford, England) 2014, 30(7):1015-6.

10. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. Clinical chemistry 2014, 60(9):1192-9.

11. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. Journal of the National Cancer Institute 2004, 96(4):261-8.

12. Yan WY, Hu J, Xie L, Cheng L, Yang M, Li L, Shi J, Liu BR, Qian XP. Prediction of biological behavior and prognosis of colorectal cancer patients by tumor Msi/MMr in the chinese population. OncoTargets and Therapy 2016, 9:7415–7424.

13. Yuan L, Chi Y, Chen W, Chen X, Wei P, Sheng W, Zhou X, Shi D. Immunohistochemistry and microsatellite instability analysis in molecular subtyping of colorectal carcinoma based on mismatch repair competency. Int J Clin Exp Med 2015, 8(11):20988-21000.

14. Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Clendenning M, Sotamaa K, Prior T, Westman JA, Panescu J, Fix D, Lockman J, LaJeunesse J, Comeras I, de la Chapelle A. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol 2008, 26:5783–8.

15. Southey MC, Jenkins MA, Mead L, Whitty J, Trivett M, Tesoriero AA, Smith LD, Jennings K, Grubb G, Royce SG, Walsh MD, Barker MA, Young JP, Jass JR, St John DJ, Macrae FA, Giles GG, Hopper JL. Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. J Clin Oncol 2005, 23:6524–6532.

16. Umar A, Risinger JI, Glaab WE, Tindall KR, Barrett JC, Kunkel TA. Functional overlap in mismatch repair by human MSH3 and MSH6. Genetics 1998, 148:1637–1646.

17. Yamada NA, Castro A, Farber RA. Variation in the extent of microsatellite instability in human cell lines with defects in different mismatch repair genes. Mutagenesis 2003, 18:277–282.

18. Edelmann W, Yang K, Umar A, Heyer J, Lau K, Fan K, Liedtke W, Cohen PE, Kane MF, Lipford JR, Yu N, Crouse GF, Pollard JW, Kunkel T, Lipkin M, Kolodner R, Kucherlapati R. Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell 1997, 91:467–477.

19. Jass JR. Hereditary non-polyposis colorectal cancer: the rise and fall of a confusing term. World J Gastroenterol 2006, 12:4943–4950.

20. Malkhosyan SR, Yamamoto H, Piao Z, Perucho M. Late onset and high incidence of colon cancer of the mutator phenotype with hypermethylated hMLH1 gene in women. Gastroenterology 2000, 119:598.

21. Mekenkamp LJ, Heesterbeek KJ, Koopman M, Tol J, Teerenstra S, Venderbosch S, Punt CJ, Nagtegaal ID. Mucinous adenocarcinomas: Poor prognosis in metastatic colorectal cancer. European Journal of Cancer 2012, 48, 501–509.

22. Kurian AW, Mills MA, Kingham KE, McPherson L, Whittemore AS, McGuire V, Ladabaum U, Kobayashi Y, Lincoln SE, Cargill M, Ford JM. Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. Journal of Clinical Oncology, 2014 Jul 1;32(19):2001-9.

23. Shen T, Pajaro-Van de Dtadt SH, Yeat NC, Lin LC. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Frontiers in Genetics, 2015 Jun 17;6:215.

24. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008, 135(4):1079–99.

Figure legends

Fig. 1 The percentages of length-instability loci in microsatellite stable (MSS) and microsatellite instability (MSI)-high (MSI-H) samples for ColonCore panel, MSIsensor, and mSINGS. The MSS and MSI-H samples are colored as red and green. Each dot represents one sample. The percentages are most distinguished in ColonCore panel comparing to those from MSIsensor and mSINGS.

Fig.2 Mutation burden of microsatellite stable (MSS) and microsatellite instability (MSI)-high (MSI-H) samples reported by PCR method. The MSS and MSI-H samples are colored as red and green. Each dot represents one sample. The mutation burden is not highly distinguished between these two types of samples.

Table 1. Colorectal cancer–specific marker microsatellite loci and baseline statistics.

Loci Identity	chr	position	Homo-polymer		Left-mer†	Right-mer‡	Mean ratio	sd§	Mis-match
MS-BR1	1	161332091	14	[T]	ATTCC	GCTTT	0.661	0.039	0
MS-BR2	2	47635523	13	[T]	TGTAC	AAGGA	0.913	0.019	0
MS-BR3*	2	47641559	27	[A]	CAGGT	GGGTT	0.738	0.032	1
MS-BR4	2	48032740	13	[T]	TGTGA	AAGGT	0.974	0.011	0
MS-BR5	2	48033890	18	[T]	AAAAC	AATTT	0.873	0.055	0
MS-BR6*	2	95849361	23	[T]	TCCTA	GTGAG	0.747	0.058	0
MS-BR7*	4	55598211	25	[T]	TTTGA	GAGAA	0.440	0.037	0
MS-BR8	7	6037057	17	[A]	AACTG	TTCAC	0.895	0.042	0
MS-BR9	7	116381121	16	[T]	TGGTG	GGTTT	0.794	0.052	0
MS-BR10	7	116409675	15	[T]	CAACC	CCTTT	0.875	0.034	0
MS-BR11	11	108114661	15	[T]	AATAA	AAGAA	0.750	0.043	0
MS-BR12	11	108121410	15	[T]	TATCC	AGGCT	0.784	0.064	0
MS-BR13	11	108141955	15	[T]	TGAAC	ACCAC	0.638	0.031	0

MS-BR14	11	108188266	13	[T]	CTTGA	GCCTC	0.853	0.057	0
MS-BR15	11	108195976	19	[T]	CATAG	CATTT	0.733	0.072	0
MS-BR16*	11	125490765	21	[T]	GAAGA	AATAT	0.832	0.046	0
MS-BR17	12	133237753	14	[A]	ACCTG	GGCAA	0.724	0.038	0
MS-BR18	13	32905219	12	[T]	TTTGA	GAGGT	0.913	0.021	0
MS-BR19	13	32907535	11	[T]	CTGTC	GTAAA	0.913	0.019	0
MS-BR20*	14	23652346	21	[A]	TTGCT	GGCCA	0.792	0.089	1
MS-BR21	15	91303325	12	[T]	AAGAC	CCCTC	0.816	0.031	0
MS-BR22	18	48584855	16	[T]	GGCTA	GGTAG	0.776	0.059	1

*These loci were also used in the PCR method.

† The left side of the homopolymer.

‡ The right side of the homopolymer.

§standard deviation.

The homopolymer was described using the repeat length and the repeat unit. For example, the MS-BR1 is the loci in chromosome 1, with a 14 Ts homopolymer. The flanking sequences on the left and right side of the homopolymer are ATTCC and GCTTT, respectively. Mismatch described the maximum number of mismatch allowed when counting the reads aligned to the loci of different repeat length. Three marker loci allowed one mismatch during the alignment.

Table 2. Correlation of MSI-ColonCore, MSI-PCR, and IHC.

MSI / IHC	MSI-ColonCore		MSI-PCR	
	MSI-H	**MSI-L / MSS**	**MSI-H**	**MSI-L/MSS**
dMMR	47	7	48	6
pMMR	0	37	0	37

Table 3. Performance of MSI status detection for 91 samples.

Method	Accuracy	Sensitivity	Precision	MCC
ColonCore panel	98.90%	97.92%	100%	0.978
MSIsensor(v0.2)	96.70%	97.92%	95.35%	0.934
mSINGS(v2.0)	97.80%	95.83%	100%	0.957

Figure 1

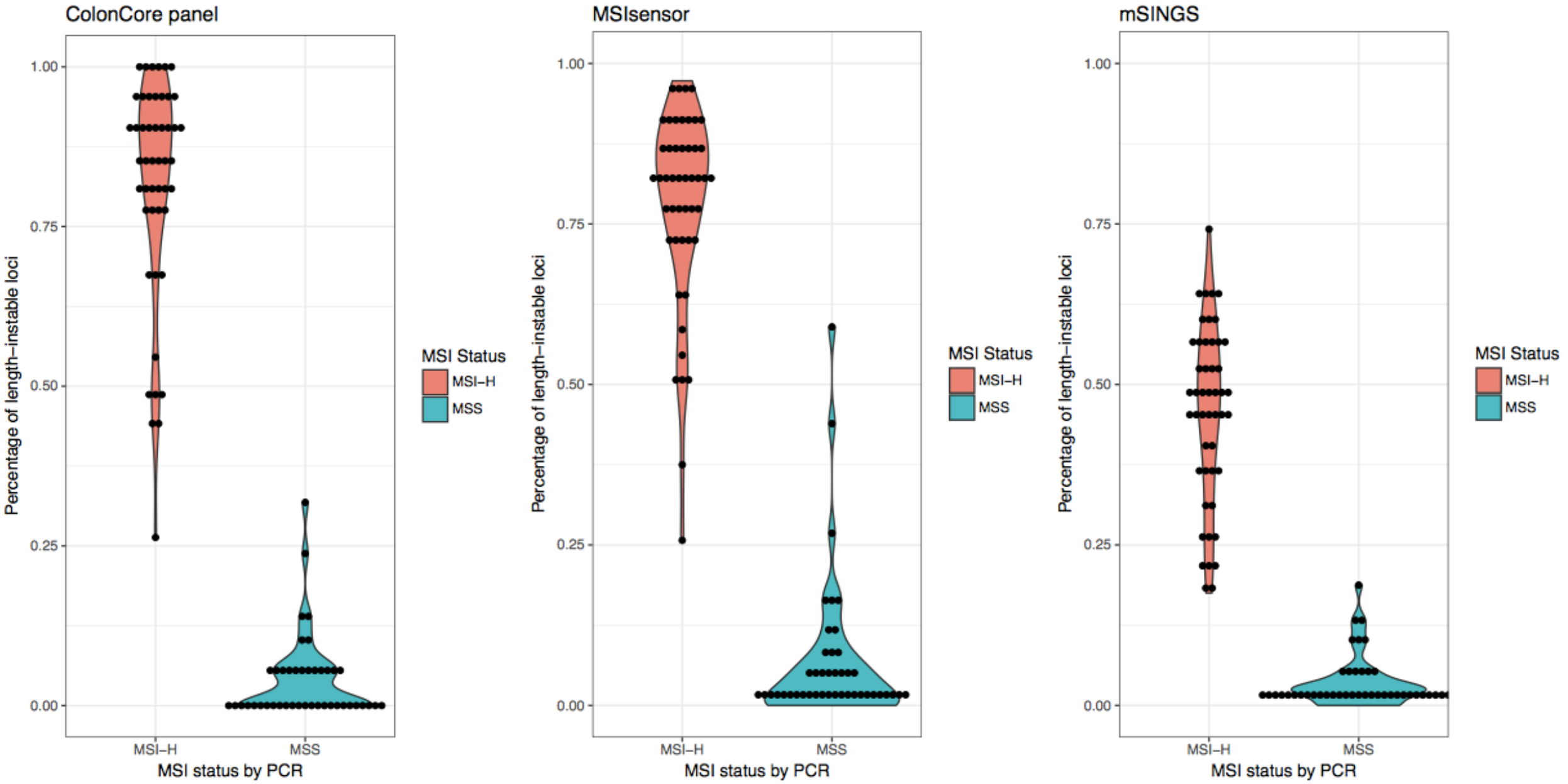


Figure 2