DOI: 10.11134/btp.4.2026.5

UDC 616.348-006:575.224:577.21

MICROSATELLITE INSTABILITY PROFILING WITH MSISENSOR2 IN WHOLE-EXOME SEQUENCING DATA FROM HEALTHY CONTROLS AND COLORECTAL CANCER PATIENTS

Akbota Aitkulova*, Nurai Tynyshtykbayeva, Asset Daniyarov, Ainur Akilzhanova, Dos Sarbassov

Center for Life sciences, National Laboratory Astana, Nazarbayev University, 010000, Astana, Kazakhstan
Corresponding Author: akbota.aitkulova@nu.edu.kz

ABSTRACT

Microsatellite instability (MSI) is a molecular phenotype linked to defects in DNA mismatch repair (MMR) and serves as a biomarker in colorectal cancer (CRC). Computational MSI assessment using next-generation sequencing data enables characterisation of microsatellite instability without a dedicated MSI assay. This study evaluated MSIsensor2 scores in blood-derived whole-exome sequencing (WES) data from healthy individuals and two CRC patients.

WES data from peripheral blood samples of 235 healthy Kazakh individuals were obtained from the DNA Biobank of the Laboratory of Genomic and Personalised Medicine, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan. Blood-derived WES data from two patients with CRC, designated BR110 and BR114, were also analysed. Sequencing libraries were prepared using Illumina workflows and sequenced on the NovaSeq 6000 platform. MSIsensor2 was used to calculate MSI scores and quantify unstable microsatellite loci. Relationships among MSI scores, locus coverage, and sequencing data volume were examined descriptively.

Among the 235 healthy individuals, 205 samples (87.2%) had MSIsensor2 scores below 10%, whereas 30 samples (12.8%) had scores between 10% and 40%. No sample had a score above 40%. These intervals were used for descriptive stratification and did not represent clinically validated MSI categories. The CRC-associated blood samples BR110 and BR114 had scores of 5.43% and 7.81%, respectively. BR110 had 10 unstable microsatellite sites among 184 covered sites, whereas BR114 had 26 among 333 covered sites. The higher score in BR114 was accompanied by a greater number of covered and evaluable loci. The scores of both CRC samples were below the 20% computational threshold referenced in the MSIsensor2 documentation; however, this finding does not independently establish their tumour MSI status.

MSIsensor2 generated MSI-associated scores from peripheral blood WES data obtained from healthy individuals and two patients with CRC. The 10% and 40% boundaries were used solely for descriptive analysis, whereas the 20% threshold served as an algorithm-derived reference rather than a clinically validated diagnostic cut-off for the samples analysed. The clinical significance of the observed scores remains undetermined. Further investigation using tumour-derived material, established MSI reference methods, and larger clinical cohorts is required to determine the relationship between blood-derived MSIsensor2 scores and tumour MSI status.

Keywords: Microsatellite instability; MSIsensor2; whole-exome sequencing; colorectal cancer; mismatch repair.

INTRODUCTION

Microsatellite instability (MSI) is a molecular phenotype resulting from impaired DNA mismatch repair (MMR), which leads to the accumulation of insertion and deletion mutations in repetitive DNA sequences [1]. MSI serves as a well-established biomarker in colorectal cancer (CRC) and various other malignancies. Tumours often have high (MSI-H) tumour mutational burden and may demonstrate responsiveness to immune checkpoint inhibitors, underscoring the clinical importance of accurate MSI assessment [2,3].

Traditional MSI assessment primarily utilises polymerase chain reaction (PCR)-based analysis of microsatellite markers and immunohistochemistry (IHC) for mismatch repair proteins. The advent of next-generation sequencing (NGS) has facilitated computational MSI assessment across a broader range of microsatellite loci, including methods based on whole-exome sequencing (WES) data [2,4]. However, analytical methodology, sequencing quality, coverage of informative loci, and the biological source of the analysed DNA influence the interpretation of computational MSI scores.

MSIsensor2 is a computational tool for MSI detection from sequencing data that does not require a matched normal sample. It uses a machine learning approach to assess microsatellite instability in individual sequencing samples and supports various sequencing strategies, including WES [5]. The developers recommend an MSI score threshold of 20% to identify MSI-high samples, classifying scores below this threshold as microsatellite-stable under the tool's binary scheme [5]. However, computational classifications require interpretation within the context of sample type, sequencing coverage, and independent clinical validation.

In this study, we analysed MSI-associated profiles in whole-exome sequencing data from 235 healthy Kazakh individuals and two patients with CRC using MSIsensor2. We characterised the distribution of MSI scores, quantified the numbers of covered and unstable microsatellite sites, and examined the relationships between sequencing characteristics and MSI-associated measurements. Given the exploratory nature of the study and the limited number of CRC samples, our objective was to describe patterns in computational MSI scores and their technical context rather than establish clinically confirmed MSI status.

MATERIALS AND METHODS

Sample collection

This exploratory study analysed blood-derived whole-exome sequencing (WES) data from two groups: 235 healthy individuals (175 female, 60 male) and 2 colorectal cancer

DOI: 10.11134/btp.4.2026.5

(CRC) patients. Blood samples from healthy individuals were obtained from the DNA Biobank of the Laboratory of Genomic and Personalised Medicine, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan. Blood samples from CRC patients were collected at KazIOR, Almaty, Kazakhstan. All participants provided written informed consent prior to inclusion in the study.

DNA LIBRARY PREPARATION AND SEQUENCING

Genomic DNA was extracted from peripheral blood samples using the Illustra Blood GenomicPrep Mini Spin Kit (Cytiva, USA) following the manufacturer's instructions. DNA libraries were prepared with the Illumina DNA Exome 2.5 Panel according to the manufacturer's protocol. Sequencing was conducted on the NovaSeq 6000 platform (Illumina, USA).

Computational MSI assessment

Microsatellite instability (MSI) scores were evaluated using MSIsensor2 (version 0.1) with whole-exome sequencing (WES) data. For each sample, the analysis documented the total number of microsatellite sites assessed, the number of sites with sufficient coverage, the number of sites classified as unstable, and the resulting MSI score.

The MSI score was reported as a percentage, calculated by dividing the number of unstable microsatellite sites by the number of microsatellite sites with sufficient coverage and multiplying by 100.

According to MSIsensor2 documentation, 20% is the recommended computational threshold for MSI-H classification in this sequencing context. In the present study, the 20% threshold served as an algorithm-derived reference rather than a clinically validated diagnostic cut-off for the analysed samples. MSI scores were also examined descriptively using study-specific intervals to characterise their distribution within the cohort. The 10% and 40% boundaries were applied solely for descriptive stratification and were not intended to define MSI-L or MSI-H categories. Conventional MSI-L and MSI-H categories were not assigned because the study did not include independent validation against an established MSI reference method.

Assessment of sequencing characteristics

The relationship between the number of covered microsatellite loci and the number of detected unstable sites was examined to assess the potential influence of locus coverage on MSI measurements. MSI score distributions were summarised for the healthy individual cohort and the two colorectal cancer (CRC) patient samples.

The relationship between FASTQ file size and average sequencing coverage was examined as an exploratory measure of sequencing data volume. These descriptive analyses aimed to identify technical factors that may influence the interpretation of computational MSI scores.

Statistical analysis

Descriptive statistics and graphical summaries were used to characterise MSI score distributions and sequencing characteristics. Due to the limited number of CRC patient samples, no formal inferential comparisons between the CRC and healthy-individual cohorts were performed. Correlation coefficients and significance tests were not calculated unless explicitly specified. Results were interpreted as exploratory, considering differences in sequencing coverage and the limited number of microsatellite sites with available coverage.

RESULTS

Analysis of MSIsensor2 scores derived from WES data

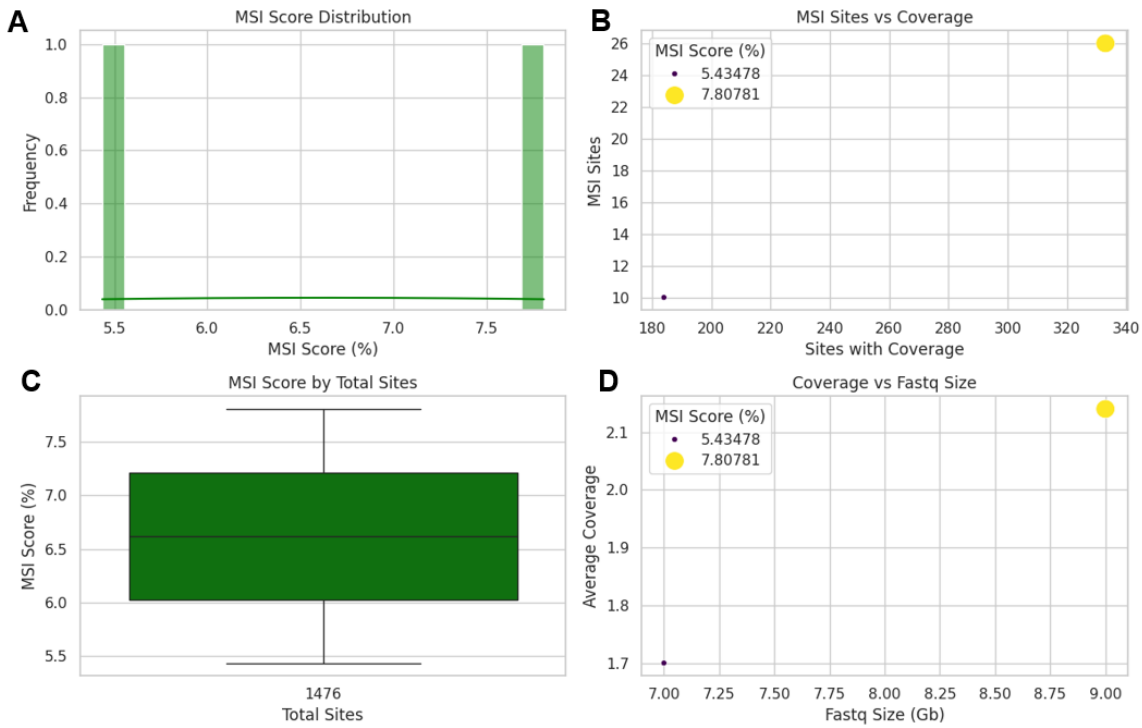


Figure 1. MSI profiling results in the healthy WGS cohort (n = 235): (A) MSI score distribution, (B) MSI sites vs. covered loci, (C) MSI score boxplot, (D) coverage vs. FASTQ size.

DOI: 10.11134/btp.4.2026.5

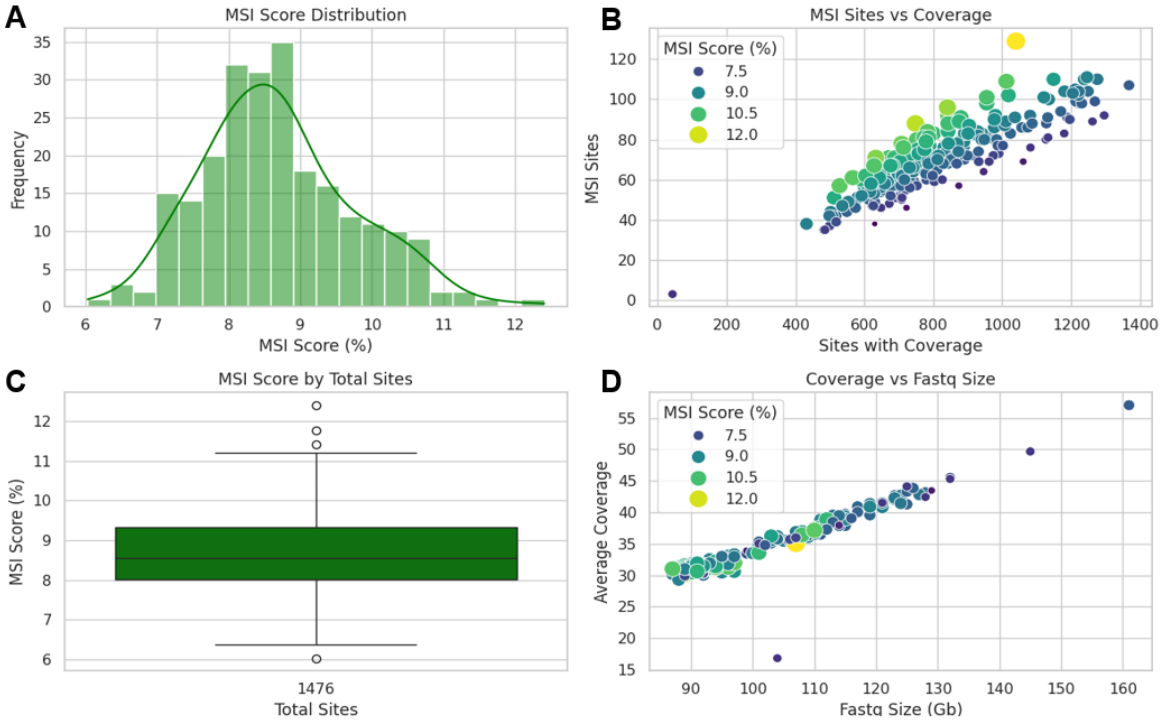


Figure 2. MSI profiling results in CRC WES samples (BR110 and BR114): (A) MSI score distribution, (B) MSI sites vs. covered loci, (C) MSI score boxplot, (D) coverage vs. FASTQ size.

from 235 healthy individuals indicated that 205 samples (87.2%) had scores below 10%, while 30 samples (12.8%) had scores betwescored 40%. No sample exhibited a score above the described intervals; intervals were selected to describe the observed score distribution and do not correspond to clinically validated MSI categories. Scores were predominantly concentrated in the 7–10% range, with most observations clustered around 8–9%. The boxplot demonstrated limited variability for most samples, with only a small subset exceeding 11% (Fig. 1C).

The number of unstable microsatellite sites detected varied with the number of loci covered by sequencing (Fig. 1B). Samples with more covered loci generally had higher counts of detected unstable sites. This trend was consistent across the cohort; however, we did not calculate a formal correlation coefficient or assess statistical significance. The number of covered loci therefore provides essential context for interpreting differences in the absolute number of detected unstable sites.

FASTQ data volume varied among samples and increased with higher average sequencing coverage (Fig. 1D). This relationship characterises the sequencing properties of the WES dataset and contextualises the observed differences in the number of evaluable microsatellite loci. FASTQ file size and average coverage were evaluated descriptively, without formal statistical analysis of their association.

In the colorectal cancer cohort, BR110 and BR114 exhibited MSI scores of 5.43% and 7.81%, respectively (Fig. 2A). BR114 demonstrated 26 unstable microsatellite sites among 333 covered loci, whereas BR110 showed 10 unstable sites among 184 covered loci (Fig. 2B). The workflow evaluated a total of 1,476 microsatellite sites, with only 184 (12.5%) and 333 (22.6%) covered in BR110 and BR114, respectively. Sequencing metrics indicated higher average coverage and greater FASTQ data volume for BR114 than for BR110 (Fig. 2D).

The higher MSI score in BR114 coincided with greater microsatellite locus coverage and higher sequencing depth; therefore, the difference between the samples cannot be attributed solely to biological variation. The limited proportion of covered loci also restricts the interpretation of both scores. Although both values were below the study-specific threshold of 10%, they should not be interpreted as definitive evidence of microsatellite stability or used to assign clinical MSS, MSI-L or MSI-H status without independent validation using an established MSI assay. These findings are therefore preliminary and hypothesis-generating, not evidence of clinical diagnostic accuracy.

DISCUSSION

In this study, computational analysis of whole-exome sequencing (WES) data from 235 healthy individuals showed that 205 samples (87.2%) had MSI scores below 10%, whereas 30 samples (12.8%) fell within the 10–40% interval. No sample exceeded 40%. These findings describe the distribution of scores generated by the analytical approach used in this study and provide a reference for assessing score variation within the investigated cohort.

However, these operational categories should not be interpreted as clinically validated microsatellite-stable (MSS), MSI-low (MSI-L), or MSI-high (MSI-H) classifications. Computational MSI algorithms differ in their underlying methods, selected microsatellite loci, coverage requirements, and scoring procedures. Consequently, numerical scores and classification thresholds are not necessarily interchangeable across tools or sequencing platforms [6–10]. Previous studies have demonstrated that MSI assessment using sequencing data can be informative, but interpretation depends on the analytical method and the characteristics of the sequencing data [7,9,10].

DOI: 10.11134/btp.4.2026.5

The distribution observed in the healthy cohort is therefore best considered a study-specific reference rather than a universal baseline for MSI in healthy individuals. In particular, the presence of scores between 10% and 40% does not by itself indicate biological microsatellite instability or mismatch-repair deficiency. Establishing the clinical significance of these scores would require comparison with an independently validated MSI assay and, where appropriate, mismatch-repair protein immunohistochemistry (MMR IHC).

The number of microsatellite loci covered by sequencing is an important consideration when interpreting computational MSI scores. In the present cohort, the number of covered loci generally increased with the number of detected unstable sites (Figure 1B). This pattern is descriptively consistent with the expectation that the ability to detect unstable loci depends, at least in part, on how many loci can be evaluated. Nevertheless, because no formal correlation analysis was performed, this observation should not be described as a statistically established association.

Previous studies have evaluated computational MSI detection using WES and targeted next-generation sequencing data, demonstrating that sequencing-based approaches can identify MSI-associated patterns but that their performance depends on the data and analytical pipeline [7,9,10]. Tools such as MANTIS and MSIsensor use distinct computational approaches to evaluate instability, and results should therefore be interpreted within the context of the specific method employed [8,10]. More broadly, analyses across cancer types have demonstrated that microsatellite instability can vary in its genomic distribution, further supporting the importance of the loci interrogated and the analytical strategy used [9,12].

In the current study, the number of covered loci should consequently be considered alongside the MSI score rather than treated as a secondary technical detail. Samples with limited locus coverage may provide an incomplete representation of microsatellite instability across the targeted genomic regions. Conversely, a larger number of covered loci does not automatically establish that an MSI classification is clinically accurate. Reliable classification requires adequate data quality and validation against an appropriate reference method.

Figure 1D shows the relationship between FASTQ file size and mean sequencing coverage in the investigated samples. The observed pattern is descriptive and indicates that larger sequencing data volumes were accompanied by higher mean coverage in this dataset. However, FASTQ file size is not a direct measure of MSI analytical sensitivity: it can also reflect read length, compression, sequencing format, and other technical characteristics. Similarly, mean coverage alone does not indicate whether individual microsatellite loci meet the minimum requirements for reliable evaluation.

Computational MSI analysis using WES has been evaluated in colorectal cancer cohorts, while other studies have investigated MSI detection using targeted sequencing data [7,9,10]. These studies support the use of sequencing-based methods under appropriate analytical conditions, but they do not establish that every dataset with a given file size or mean coverage is suitable for reliable MSI classification.

The distinction is particularly important in the present study because the two colorectal cancer samples had low mean coverage and differed in the proportion of covered microsatellite loci. Therefore, differences in their MSI scores cannot be interpreted independently of differences in sequencing depth and locus coverage. A larger number of detected unstable sites may reflect a greater number of evaluable loci as well as underlying biological differences.

Future analyses should incorporate prespecified quality-control criteria, including minimum coverage requirements at individual loci, the proportion of evaluable loci, and the number of loci contributing to the final score. Such measures would make comparisons between samples more transparent and help distinguish technical limitations from potentially meaningful biological variation.

The two colorectal cancer samples, BR110 and BR114, had MSI scores of 5.43% and 7.81%, respectively. BR110 had 184 covered loci out of 1,476 total loci, corresponding to approximately 12.5% coverage, whereas BR114 had 333 covered loci, corresponding to approximately 22.6%. The numbers of detected unstable sites were 10 and 26, respectively. Although BR114 had both a higher score and more detected unstable sites, it also had a greater number of covered loci and higher mean sequencing coverage. The observed difference therefore cannot be attributed solely to biological variation.

Computational MSI assessment has been investigated in colorectal cancer using WES and targeted sequencing data [7,10,11]. Nevertheless, MSI scores require careful interpretation when sequencing coverage is limited, and a score generated by one computational method should not be directly equated with an MSI classification obtained using another assay. Comparative clinical studies have shown that the agreement between sequencing-based MSI assessment, MSI-PCR, and MMR IHC depends on the assay and the criteria used for classification [15]. These findings reinforce the need to distinguish exploratory computational results from clinically validated MSI status.

An additional limitation concerns the source of the samples. BR110 and BR114 were analysed using DNA from peripheral blood. Consequently, their scores cannot be used to establish the MSI status of the corresponding colorectal tumours. In this setting, the values of 5.43% and 7.81% should be reported as computational scores obtained from the analysed blood-derived DNA, not as definitive evidence that the tumours were MSS.

This distinction is clinically important because MSI and mismatch-repair deficiency are established biomarkers relevant to treatment selection, including immune checkpoint inhibition in appropriately assessed tumours [14]. However, the clinical significance of MSI depends on reliable identification of the biomarker in the relevant tumour context. The present two-sample analysis does not provide sufficient evidence to infer tumour MSI status, predict immunotherapy response, or establish the diagnostic performance of the computational method.

Regarding score interpretation, the 20% threshold recommended in the MSIsensor2 documentation serves as an algorithm-derived reference for MSI-H classification within the intended sequencing context. Its clinical applicability to the peripheral blood DNA analysed in this study, however, has not been established. The 10% and 40% boundaries were used ex-

DOI: 10.11134/btp.4.2026.5

clusively to describe the distribution of scores within the study cohort and should not be interpreted as alternative diagnostic thresholds. Similarly, the descriptive interval of 10–40% does not correspond to the conventional MSI-low (MSI-L) category. Accordingly, these boundaries represent study-specific analytical choices rather than clinically validated criteria, and the clinical significance of the observed scores remains undetermined. Further validation against an established MSI reference method is required to assess their clinical relevance.

The observed differences in MSIsensor2 scores between the healthy cohort and the two CRC-associated blood samples should be interpreted cautiously. The number of evaluable microsatellite loci differed between BR110 and BR114 (184 and 333, respectively), potentially affecting the number of detected unstable loci and the resulting MSI scores. Although BR114 had a higher MSI score and a greater number of unstable loci, the available data do not establish whether this difference reflects biological variation or differences in locus coverage and evaluability. Moreover, the analysis of only two CRC samples does not allow robust conclusions regarding differences between the healthy cohort and patients with CRC. These findings should therefore be considered descriptive, and further analyses using adequately covered samples and a larger clinical cohort are required to investigate the influence of sequencing coverage and the number of evaluable loci on MSIsensor2 scores.

Limitations and future directions

Several limitations should be considered when interpreting these findings. First, the study-specific score intervals were used to describe the distribution of computational results and were not validated as clinical MSI categories. Second, although the healthy cohort included 235 individuals, the colorectal cancer analysis comprised only two samples and therefore cannot support robust conclusions about MSI patterns in colorectal cancer. Third, both cancer-associated samples were analysed using peripheral-blood DNA rather than tumour material, preventing direct determination of tumour MSI status from these results.

Fourth, sequencing coverage was limited in both colorectal cancer samples. The proportion of covered loci was approximately 12.5% for BR110 and 22.6% for BR114, which constrains interpretation of their scores and comparison with samples having more complete locus coverage. Fifth, the study did not include an independent validation of the reported scores using MSI-PCR or MMR IHC. Such validation is important because sequencing-based MSI classifications may differ according to assay design and classification thresholds [7,11,15].

Future studies should evaluate the approach in a larger and clinically characterised colorectal cancer cohort using tumour-derived DNA. Where feasible, computational scores should be compared with established MSI-PCR and MMR IHC results, and the analytical workflow should include predefined coverage and quality-control criteria. The clinical relevance of any proposed score thresholds should be established independently rather than inferred from the distribution in healthy individuals. Further validation would also help determine whether the method is suitable for exploratory genomic profiling or whether it can achieve sufficient analytical performance for clinical MSI assessment.

Overall, the present study characterises the distribution of computational MSI scores in a cohort of healthy individuals and illustrates how locus coverage and sequencing depth affect the interpretation of results. The two colorectal cancer samples provide preliminary examples of the analytical challenges encountered with low-coverage WES data, but they do not establish the MSI status of the corresponding tumours. The findings should therefore be considered exploratory and hypothesis-generating.

CONCLUSION

This study reports MSIsensor2 scores derived from whole-exome sequencing data of peripheral blood DNA from 235 healthy individuals and two patients with colorectal cancer. The score distribution provides a descriptive overview of MSI-associated computational signals in the analysed cohort. The 10% and 40% boundaries were applied for exploratory data stratification, while the 20% threshold corresponds to the algorithm-derived reference recommended in the MSIsensor2 documentation. However, none of these thresholds has been clinically validated for the samples analysed in this study. Consequently, the observed scores cannot determine patients' MSI status or be classified into conventional MSI-L or MSI-H categories. The clinical relevance of these findings remains unclear and requires further validation using appropriate reference methods and tumour-derived material.

FUNDING

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grants No. AP22788205, AP23489913) awarded to AA.

AUTHOR CONTRIBUTIONS

Conceptualization, A.A.; methodology, A.A. and N.T.; bioinformatic analysis, A.D. and N.T.; supervision, Ai.A. and D.S.; writing-original draft preparation, A.A.; writing -review and editing, A.A., Ai.A. and D.S. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Boland CR, Thibodeau SN, Hamilton SR, et al. 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 Research. – 1998. – Vol. 58, No. 22. – P. 5248–5257.
2. Sepulveda AR, Hamilton SR, Allegra CJ, et al. Molecular biomarkers for the evaluation of colorectal cancer: guideline from the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology // Journal of Clinical Oncology. – 2017. – Vol. 35, No. 13. – P. 1453–1486. – DOI: 10.1200/JCO.2016.71.9807.
3. André T, Shiu KK, Kim TW, et al. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer // New England Journal of Medicine. – 2020. – Vol. 383. – P. 2207–2218. – DOI: 10.1056/NEJMoa2017699.
4. Takamatsu M, et al. Clinical predominance of

DOI: 10.11134/btp.4.2026.5

whole-exome sequencing to evaluate microsatellite instability status // Cancer Science. – 2023. – Vol. 114, No. 7. – P. 2848–2859. – DOI: 10.1111/cas.15813.
5. Jia P, Yang X, Guo L, et al. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability // Genomics, Proteomics & Bioinformatics. – 2020. – Vol. 18, No. 1. – P. 65–72. – DOI: 10.1016/j.gpb.2020.02.001.

6. Niu-lab. MSIsensor2: Microsatellite instability detection for tumour-only sequencing data [Internet] // GitHub. – Available at: https://github.com/niu-lab/msisensor2 (accessed 10 November 2026).
7. Johansen AF, Kassentoft CG, Knudsen M, et al. Validation of computational determination of microsatellite status using whole exome sequencing data from colorectal cancer patients // BMC Cancer. – 2019. – Vol. 19. – Art. 971. – DOI: 10.1186/s12885-019-6227-7.
8. Kautto EA, Bonneville R, Miya J, et al. Performance Evaluation for Rapid Detection of Pan-Cancer Microsatellite Instability with MANTIS // Oncotarget. – 2017. – Vol. 8, No. 5. – P. 7452–7463. – DOI: 10.18632/oncotarget.13918.
9. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and Characterization of Microsatellite Instability Across 18 Cancer Types // Nature Medicine. – 2016. – Vol. 22. – P. 1342–1350. – DOI: 10.1038/nm.4191.
10. Middha S, Zhang L, Nafa K, et al. Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data // JCO Precision Oncology. – 2017. – Vol. 1. – P. 1–17. – DOI: 10.1200/PO.17.00084.
11. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Detection of Mismatch Repair Deficiency and Microsatellite Instability in Colorectal Adenocarcinoma by Targeted Next-Generation Sequencing // The Journal of Molecular Diagnostics. – 2018. – Vol. 20, No. 1. – P. 1–11. – DOI: 10.1016/j.jmoldx.2017.09.004.
12. Bonneville R, Krook MA, Kautto EA, et al. Landscape of Microsatellite Instability Across 39 Cancer Types // JCO Precision Oncology. – 2017. – Vol. 1. – P. 1–15. – DOI: 10.1200/PO.17.00073.
13. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability // Journal of the National Cancer Institute. – 2004. – Vol. 96, No. 4. – P. 261–268. – DOI: 10.1093/jnci/djh034.
14. Le DT, Durham JN, Smith KN, et al. Mismatch Repair Deficiency Predicts Response of Solid Tumors to PD-1 Blockade // Science. – 2017. – Vol. 357, No. 6349. – P. 409–413. – DOI: 10.1126/science.aan6733.
15. Kang SY, Kim DG, Ahn S, Ha SY, Jang KT, Kim KM. Comparative Analysis of Microsatellite Instability by Next-Generation Sequencing, MSI PCR and MMR Immunohistochemistry in 1942 Solid Cancers // Pathology – Research and Practice. – 2022. – Vol. 233. – Art. 153874. – DOI: 10.1016/j.prp.2022.153874.

УДК 616.348-006:575.224:577.21

DOI: 10.11134/btp.4.2026.5

ПРОФИЛИРОВАНИЕ МИКРОСАТЕЛЛИТНОЙ НЕСТАБИЛЬНОСТИ С ПОМОЩЬЮ MSISENSOR2 В ДАННЫХ ПОЛНОГЕНОМНОГО ЭКЗОМНОГО СЕКВЕНИРОВАНИЯ ЗДОРОВЫХ ЛИЦ И ПАЦИЕНТОВ С КОЛОРЕКТАЛЬНЫМ РАКОМ

Акбота Айткулова[1],*, Нурай Тыныштыкбаева[1], Ассет Данияров[1], Айнур Акилжанова[1], Дос Сарбассов[1]

[1] Центр наук о жизни, ЧУ «National Laboratory Astana», Назарбаев Университет, 010000, Астана, Казахстан
Автор для корреспонденции: akbota.aitkulova@nu.edu.kz

АННОТАЦИЯ

Микросателлитная нестабильность (MSI) представляет собой молекулярный фенотип, связанный с нарушениями системы репарации неспаренных оснований ДНК (MMR), и служит биомаркером при колоректальном раке (КРР). Вычислительная оценка MSI на основе данных секвенирования нового поколения позволяет охарактеризовать микросателлитную нестабильность без проведения специализированного теста на MSI. В настоящем исследовании оценивали показатели MSIsensor2 в данных полногеномного экзомного секвенирования (WES), полученных из образцов крови здоровых лиц и двух пациентов с КРР.

В исследовании использованы данные полноэкзомного секвенирования (WES) ДНК периферической крови 235 здоровых лиц казахской популяции, полученные из Биобанка ДНК Лаборатории геномной и персонализированной медицины National Laboratory Astana, Назарбаев Университет (Астана, Казахстан). Дополнительно проанализированы данные WES ДНК периферической крови двух пациентов с колоректальным раком (КРР), обозначенных как BR110 и BR114. Библиотеки для секвенирования подготовлены с использованием протоколов Illumina, секвенирование выполнено на платформе NovaSeq 6000. Для вычислительной оценки показателей микросателлитной нестабильности (MSI) использовали MSIsensor2. Анализ включал определение количества микросателлитных локусов, доступных для оценки, числа локусов, классифицированных алгоритмом как нестабильные, и соответствующих показателей MSI. Взаимосвязи между показателями MSI, покрытием микросателлитных локусов и объёмом секвенирующих данных оценивали описательными методами.

Среди 235 здоровых лиц у 205 (87,2%) показатель MSIsensor2 был ниже 10%, а у 30 (12,8%) находился в диапазоне от 10 до 40%. Ни в одном образце показатель не превышал 40%. Указанные интервалы использовали исключительно для описания распределения значений в исследуемой когорте; они не являются клинически валидированными категориями MSI. В образцах крови пациентов с КРР BR110 и BR114 показатели MSIsensor2 составили 5,43 и 7,81% соответственно. В образце BR110 выявлено 10 нестабильных микросателлитных локусов из 184 локусов с достаточным покрытием, тогда как в образце BR114 - 26 из 333. Более высокий показатель в образце BR114 сопровождался большим числом локусов, доступных для оценки. Эти результаты демонстрируют вариабельность вычислительных показателей MSI в данных WES периферической крови, но не позволяют установить MSI-статус опухоли.

Таким образом, MSIsensor2 позволил охарактеризовать показатели, связанные с микросателлитной нестабильностью, в данных WES ДНК периферической крови здоровых лиц и двух пациентов с КРР. Порог 20%, указанный в документации MSIsensor2, рассматривался как алгоритмический ориентир, тогда как границы 10 и 40% использовались исключительно для описательной стратификации результатов. Ни один из этих порогов не был клинически валидирован в рамках настоящего исследования. Клиническая значимость полученных показателей и их связь с MSI-статусом опухоли остаются неустановленными. Для дальнейшей оценки клинической применимости необходимы исследования с использованием опухолевого материала, валидированных референсных методов определения MSI и более крупных независимых когорт.

Ключевые слова: микросателлитная нестабильность; MSIsensor2; полногеномное экзомное секвенирование; колоректальный рак; репарация неспаренных оснований ДНК.

УДК 616.348-006:575.224:577.21

ДЕНІ САУ АДАМДАР МЕН КОЛОРЕКТАЛЬДЫ ОБЫРМЕН АУЫРАТЫН ПАЦИЕНТТЕРДІҢ ТОЛЫҚ ЭКЗОМДЫҚ СЕКВЕНИРЛЕУ ДЕРЕКТЕРІ НЕГІЗІНДЕ MSISENSOR2 КӨМЕГІМЕН МИКРОСАТЕЛЛИТТІК ТҰРАҚСЫЗДЫҚТЫ ПРОФИЛЬДЕУ

Акбота Айткулова[1],*, Нурай Тыныштыкбаева[1], Ассет Данияров[1], Айнур Акилжанова[1], Дос Сарбассов[1]

[1] Өмір туралы ғылымдар орталығы, National Laboratory Astana, Назарбаев Университет, 010000, Астана, Қазақстан
Корреспондент-автор: akbota.aitkulova@nu.edu.kz

АҢДАТПА

Микросателлиттік тұрақсыздық (MSI) - ДНҚ-дағы қате жұптасқан нуклеотидтерді түзету жүйесінің (MMR) бұзылыстарымен байланысты молекулалық фенотип және колоректальды обырда (КРО) биомаркер ретінде қолданылады. Жаңа буын секвенирлеу деректеріне негізделген MSI-ді есептеу арқылы бағалау арнайы MSI тестін жүргізбей-ақ микросателлиттік тұрақсыздықты сипаттауға мүмкіндік береді. Бұл зерттеуде дені сау адамдар мен КРО-

DOI: 10.11134/btp.4.2026.5

мен ауыратын екі пациенттің қан үлгілерінен алынған толық экзомдық секвенирлеу (WES) деректеріндегі MSIsensor2 көрсеткіштері бағаланды.

Қазақстандық популяцияның дені сау 235 өкілінің шеткері қан үлгілерінен алынған толық экзомдық секвенирлеу (WES) деректері Астана қаласындағы Назарбаев Университетінің National Laboratory Astana Геномдық және дербестендірілген медицина зертханасының ДНҚ биобанкінен алынды. Сонымен қатар BR110 және BR114 деп белгіленген колоректалдық қатерлі ісігі (КҚІ) бар екі пациенттің шеткері қан үлгілерінің WES деректері талданды. Секвенирлеу кітапханалары Illumina хаттамаларына сәйкес дайындалып, секвенирлеу NovaSeq 6000 платформасында жүргізілді. Микросателлиттік тұрақсыздыққа (MSI) байланысты көрсеткіштерді есептеу және тұрақсыз деп жіктелген микросателлиттік локустар санын анықтау үшін MSIsensor2 бағдарламасы қолданылды. MSI көрсеткіштері, микросателлиттік локустардың қамтылуы және секвенирлеу деректерінің көлемі арасындағы байланыстар сипаттамалық әдістермен бағаланды.

Дені сау 235 адамның 205-інде (87,2%) MSIsensor2 көрсеткіші 10%-дан төмен болды, ал 30 адамда (12,8%) ол 10–40% аралығында болды. Бірде-бір үлгіде көрсеткіш 40%-дан асқан жоқ. Бұл аралықтар зерттелген когортадағы көрсеткіштердің таралуын сипаттау үшін ғана пайдаланылды және клиникалық тұрғыдан валидацияланған MSI санаттары болып табылмайды. КҚІ бар BR110 және BR114 пациенттерінің қан үлгілеріндегі MSIsensor2 көрсеткіштері тиісінше 5,43% және 7,81% құрады. BR110 үлгісінде жеткілікті қамтылған 184 микросателлиттік локустың 10-ы, ал BR114 үлгісінде 333 локустың 26-сы тұрақсыз деп жіктелді. BR114 үлгісіндегі жоғары көрсеткіш бағалауға қолжетімді локустар санының көп болуымен қатар байқалды. Бұл нәтижелер шеткері қанның WES деректерінен есептелген MSI көрсеткіштерінің өзгермелілігін көрсетеді, алайда ісіктің MSI мәртебесін анықтауға мүмкіндік бермейді.

Осылайша, MSIsensor2 дені сау адамдар мен КҚІ бар екі пациенттің шеткері қан үлгілерінің WES деректері негізінде MSI-ге байланысты есептеу көрсеткіштерін сипаттауға мүмкіндік берді. MSIsensor2 құжаттамасында ұсынылған 20% шекті мән алгоритмдік бағдар ретінде қарастырылды, ал 10% және 40% шекаралары нәтижелерді сипаттамалық жіктеу үшін ғана пайдаланылды. Осы зерттеу аясында аталған шекті мәндердің ешқайсысы клиникалық тұрғыдан валидацияланған жоқ. Алынған көрсеткіштердің клиникалық маңызы және олардың ісіктің MSI мәртебесімен байланысы әзірге анықталмаған. Бұл көрсеткіштердің клиникалық қолданылу мүмкіндігін бағалау үшін ісік тінін, MSI-ді анықтаудың валидацияланған референстік әдістерін және көлемі үлкенірек тәуелсіз когорталарды қамтитын қосымша зерттеулер қажет.

Түйін сөздер: микросателлиттік тұрақсыздық; MSIsensor2; толық экзомдық секвенирлеу; колоректальды обыр; ДНҚ-дағы қате жұптасқан нуклеотидтерді түзету жүйесі.