



Возможность биоинформатического определения микросателлитной нестабильности при полнотранскриптомном анализе образцов опухолей прямой и толстой кишки

А.В. Каныгина¹, Т.Н. Гарманова², Ю.В. Долудин³, Л.О. Скородумова¹, Е.И. Шарова¹

¹ Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства, Москва, Россия;

² Московский государственный университет имени М.В. Ломоносова, Москва, Россия;

³ Национальный медицинский исследовательский центр терапии и профилактической медицины, Москва, Россия

АННОТАЦИЯ

Обоснование. Статус микросателлитной нестабильности опухоли (microsatellite instability/microsatellite stability, MSI/MSS) является важным параметром, определяющим как прогноз её развития, так и потенциал ответа на иммунотерапию. Пациенты направляются на MSI-типирование в соответствии с показаниями, однако только небольшая часть пациентов имеет эти показания. Поэтому и большинство наборов данных секвенирования, доступных в репозиториях и собираемых локально, не будут по умолчанию содержать информацию об MSI в метаданных. Особенности финансирования проектов и количество собираемого биоматериала также не всегда позволяют выделить ДНК и использовать её для типирования на MSI методами полимеразной цепной реакции. Результаты секвенирования ПНК опухолевой ткани также могут использоваться для определения MSI-статуса, но для этого необходимо включение в протокол полнотранскриптомного анализа опухолей биоинформатического инструмента для определения MSI.

Цель. Оценить наличие и применимость биоинформатических инструментов по определению статуса MSI на локальной выборке полных транскриптомов опухолей прямой и толстой кишки.

Методы. Для анализа отбирали находящиеся в свободном доступе биоинформатические инструменты, предназначенные для оценки MSI-статуса по данным секвенирования ПНК, которые затем протестировали на локальной выборке полных транскриптомов опухолей толстой и прямой кишки от 13 пациентов после резекции первичной опухоли. Оценивали число соматических мутаций как суррогатный маркер статуса MSI/MSS, статуса MSI/MSS по результатам тестирования биоинформатических инструментов, а также их корреляцию.

Результаты. Нами опробованы биоинформатические инструменты PreMSIm и MIRACLE, предназначенные для определения MSI-статуса по транскриптомным данным. При использовании MIRACLE нестабильность повторов выявлена у трёх из 13 образцов. Статус MSI, определённый MIRACLE, коррелирует с мутационной нагрузкой (в среднем 2163 мутации в образцах с MSI и 122,9 — с MSS) и позволяет определить образцы как нестабильные. PreMSIm выявил MSI также в 3 образцах, однако его результаты показали ограниченную согласованность с мутационной нагрузкой. Для двух из трёх образцов с высокой мутационной нагрузкой выявлены известные патогенные и возможно патогенные варианты в генах *MSH2* и *MLH1*, вызывающие синдром Линча и подтверждающие MSI-статус. Для одного из образцов мы предположили спорадическую природу MSI из-за гиперметилирования гена *MLH1*.

Заключение. Инструменты PreMSIm и MIRACLE определяют MSI-статус с разной чувствительностью и специфичностью относительно мутационной нагрузки как суррогатного маркера MSI/MSS в аденокарциноме толстой и прямой кишки. Инструмент MIRACLE технически легко встраивается в протокол полнотранскриптомного анализа опухоли, позволяет получать биологически правдоподобную оценку статуса MSI/MSS, коррелирующую с мутационной нагрузкой опухоли по данным транскриптомного анализа.

Ключевые слова: микросателлитная нестабильность; мутационная нагрузка; иммунотерапия опухолей; опухоль толстой и прямой кишки; колоректальный рак; полный транскриптом; ПНК-секвенирование.

Как цитировать:

Каныгина А.В., Гарманова Т.Н., Долудин Ю.В., Скородумова Л.О., Шарова Е.И. Возможность биоинформатического определения микросателлитной нестабильности при полнотранскриптомном анализе образцов опухолей прямой и толстой кишки // Гены и клетки. 2025. Т. 20, № 3. С. 241–250. DOI: 10.17816/gc642339 EDN: ROTPJB

DOI: <https://doi.org/10.17816/gc642339>

EDN: ROTPJB

Bioinformatic Detection of Microsatellite Instability Using Whole Transcriptome Analysis of Colorectal Cancer Samples

Alexandra V. Kanygina¹, Tatiana N. Garmanova², Yuriy V. Doludin³, Liubov O. Skorodumova¹, Elena I. Sharova¹

¹ Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia;

² Lomonosov Moscow State University, Moscow, Russia;

³ National Medical Research Center for Therapy and Preventive Medicine, Moscow, Russia

ABSTRACT

BACKGROUND: Tumor microsatellite instability/microsatellite stability (MSI/MSS) status is a crucial parameter determining both disease prognosis and potential response to immunotherapy. Patients are referred for MSI typing based on clinical indications; however, only a small proportion of patients meet these criteria. Consequently, most sequencing datasets available in repositories and collected locally do not inherently contain MSI information in their metadata. Project funding limitations and insufficient biomaterial quantities often preclude DNA isolation and subsequent MSI typing using polymerase chain reaction methods. RNA sequencing results from tumor tissue may also be used to determine MSI status, but this requires incorporating a bioinformatics tool for MSI assessment into the tumor whole transcriptome analysis protocol.

AIM: To evaluate the availability and applicability of bioinformatics tools for determining MSI status in locally derived datasets of whole transcriptomes of colorectal cancer samples.

METHODS: Publicly available bioinformatics tools designed for MSI status assessment using RNA sequencing data were selected for analysis. These tools were tested on a locally derived dataset of whole tumor transcriptomes from 13 patients following primary colorectal tumor resection. The number of somatic mutations was assessed as a surrogate marker of MSI/MSS status, along with MSI/MSS status based on the results of bioinformatics tool testing, and their correlation.

RESULTS: We tested two bioinformatics tools, PreMSIm and MIRACLE, designed to determine MSI status using transcriptomic data. When using MIRACLE, microsatellite instability was detected in 3 out of 13 samples. The MSI status determined by MIRACLE correlated with tumor mutational burden (TMB) (mean, 2163 mutations in MSI samples vs. 122.9 in MSS samples) and reliably identified unstable samples. PreMSIm also detected MSI in 3 samples, but its results showed limited concordance with TMB. For two of the three samples with high TMB, we identified known pathogenic and likely pathogenic variants in *MSH2* and *MLH1* genes associated with Lynch syndrome, confirming MSI status. For one sample, we proposed sporadic MSI etiology due to *MLH1* gene hypermethylation.

CONCLUSION: PreMSIm and MIRACLE demonstrate different sensitivity and specificity profiles for MSI status determination when using TMB as a surrogate MSI/MSS marker in colorectal adenocarcinoma. The MIRACLE tool can be easily integrated into whole-transcriptome tumor analysis protocols and provides biologically plausible MSI/MSS assessments that correlate with transcriptome-derived TMB.

Keywords: microsatellite instability; tumor mutational burden; cancer immunotherapy; colorectal tumor; colorectal cancer; whole transcriptome; RNA sequencing.

To cite this article:

Kanygina AV, Garmanova TN, Doludin YuV, Skorodumova LO, Sharova EI. Bioinformatic Detection of Microsatellite Instability Using Whole Transcriptome Analysis of Colorectal Cancer Samples. *Genes & cells*. 2025;20(3):241–250. DOI: 10.17816/gc642339 EDN: ROTPJB

Submitted: 27.11.2024

Accepted: 21.07.2025

Published online: 04.09.2025

ВВЕДЕНИЕ

Нарушения в генах репарации ошибочно спаренных нуклеотидов приводят к накоплению мутаций в опухолевых клетках. При этом сами нарушения могут быть вызваны как мутациями в генах систем репарации, так и их гиперметилированием. Дефектное функционирование системы репарации неспаренных нуклеотидов приводит к накоплению мутаций в микросателлитах (коротких tandemных повторах) [1], которые особенно подвержены ошибочному спариванию. Это состояние известно как микросателлитная нестабильность (*microsatellite instability*, MSI) в отличие от микросателлитной стабильности (*microsatellite stability*, MSS). Определение уровня MSI в опухоли имеет диагностическую ценность в связи с тем, что для опухолей с высоким уровнем MSI возможно применение ингибиторов контрольных точек иммунитета в качестве монотерапии. У пациентов с продвинутой стадией (метастатической и/или нерезектабельной) неизлечимой, прогрессирующей болезни или с непереносимостью стандартных методов лечения, с MSI и дефицитом систем репарации неспаренных оснований стойкое противоопухолевое действие оказывал пембролизумаб [2].

Для клинической диагностики MSI золотым стандартом является тестирование с помощью полимеразной цепной реакции (ПЦР) на панель мононуклеотидных и/или динуклеотидных микросателлитов, например BAT-25, BAT-26, D2S123, D5S346 и D17S250 [3–5]. Статус MSI также можно определять иммуногистохимически по наличию белков системы MMR (кодируемых генами *MLH1*, *MSH2*, *MSH6* и *PMS2*), однако в 5–10% случаев, несмотря на наличие белка, он не функционален [6]. Пациенты направляются на MSI-типирование в соответствии с показаниями, при этом показания имеет только небольшая часть пациентов. Однако для целей трансляционных исследований и молекулярной медицины зачастую есть данные анализа методами высокопроизводительного секвенирования (полные транскрипты, геномы или экзомы опухолевой ткани), но нет результатов непосредственно MSI-типирования ни с помощью ПЦР, ни иммуногистохимическим методом. Поэтому нам представляется обоснованным включение в протокол полнотранскриптомного анализа опухолей биоинформатического инструмента для определения MSI.

ЦЕЛЬ

Оценить наличие и применимость биоинформатических инструментов для определения статуса MSI на локальной выборке полных транскриптомов опухолей прямой и толстой кишки.

МЕТОДЫ

Включение пациентов в исследование и сбор образцов

Подбор пациентов и сбор образцов проходили на базе университетской клинической больницы № 2 Первого Московского государственного медицинского университета имени И.М. Сеченова в 2018–2020 гг. в рамках одобренного протокола (заседание Локального этического комитета от 13.09.2017 по протоколу «Изучение генома, протеома и биомаркёров болезней человека с использованием биологических образцов человека») на базе Биобанка Первого Московского государственного медицинского университета имени И.М. Сеченова. Все участники (либо их официальные представители) до включения в исследование добровольно подписали форму информированного согласия, утверждённую в составе протокола исследования этическим комитетом.

Критерии включения: возраст от 18 до 90 лет, дееспособность; письменное согласие субъекта исследования; диагноз, входящий в C18–20 по МКБ-10 для рака толстой и прямой кишки; наличие подозрения на аденокарциному.

Критерии исключения: сопутствующий диагноз гепатита вида В/С, носительство вируса иммунодефицита человека, сахарный диабет, сифилис, алкоголизм, наркомания, токсикомания и любые психические заболевания; невозможность или нежелание дать письменное информированное согласие на участие в исследовании; послеоперационное патоморфологическое заключение, не подтверждающее диагноз в рамках группы диагнозов «аденокарцинома толстой или прямой кишки».

Образцы получали в результате резекции первичной опухоли или резекции метастазов. В течение 30 мин после резекции образец опухоли размером не менее 7×7×10 мм отделяли, помещали в среду с антибиотиком и на холоде транспортировали в Федеральный научно-клинический центр физико-химической медицины имени Ю.М. Лопухина Федерального медико-биологического агентства России. Дополнительно перед операцией или интраоперационно у пациента забирали 30–40 мл крови в 3–4 вакуумные пробирки с фиолетовой крышкой с K2EDTA и в течение 30–60 мин на холоде транспортировали в Федеральный научно-клинический центр физико-химической медицины имени Ю.М. Лопухина Федерального медико-биологического агентства России. Все собранные образцы опухоли и клеток крови биобанкировались в соответствии с внутренними процедурами.

Подготовка и секвенирование РНК-библиотек

Фрагменты опухоли размером 3×3×5 мм помещали в раствор IntactRNA («Евроген», Россия), выдерживали

24 ч при температуре 4–8 °C, а затем хранили при температуре –80 °C до выделения РНК. Фрагменты ткани размораживали, отмывали фосфатно-солевым буфером («ПанЭко», Россия) от консерванта и гомогенизировали в лизирующем буфере металлическими шариками диаметром 3,5 мм на приборе Tissue Lyser (QIAGEN, США). РНК выделяли с помощью набора RNeasy Mini Kit (QIAGEN, США) по протоколу производителя. ДНК удаляли ДНКазой TURBO DNA-free Kit (Thermo Fisher Scientific, США). Рибодеплецию проводили с использованием набора NebNext rRNA Depletion Kit (Human/Mouse/Rat) (New England Biolabs, США). Цепь-специфичные библиотеки с одиночными баркодами конструировали с использованием наборов NEBNext Ultra II Directional Library Prep Kit for Illumina и Multiplex Oligos for Illumina (96 Indexes) — оба производства New England Biolabs (США). Секвенирование транскриптомных библиотек проводили в режиме парных прочтений 2×125 циклов на приборе HiSeq 2500 (Illumina, США).

Подготовка и секвенирование ДНК-библиотек

Из биобанкированных образцов крови с использованием набора Gentra Puregene Kit (QIAGEN, США) выделяли ДНК из клеток крови. Полноэкзомные библиотеки конструировали с использованием набора SureSelect V7 XT2 (Agilent, США). Секвенирование полноэкзомных библиотек проводили в режиме парных прочтений 2×125 циклов на приборе HiSeq 2500 (Illumina, США).

Биоинформатическая обработка результатов ДНК- и РНК-секвенирования

Первичный контроль качества как транскриптомных, так и экзомных прочтений проводили с использованием программы FastQC (Babraham Bioinformatics, Великобритания). Утилитой cutadapt [7] удаляли участки низкого качества чтения с 3'-концов прочтений, а также остатки адаптеров секвенирования. С помощью программы BWA-MEM проводили выравнивание экзомных прочтений, а с помощью программы STAR [8] — выравнивание транскриптомных прочтений на референсный геном человека версии GRCh38 (вариант GRCh38.d1.vd1) с использованием аннотации GENCODE v. 43. Маркировку дубликатов выполняли с помощью утилиты MarkDuplicates из программного пакета Picard (Broad Institute, США). Для транскриптома прочтения, отнесённые к участкам сплайсинга, разбивали на участки непрерывного выравнивания с помощью модуля SplitNCigarReads программного пакета GATK (Broad Institute, США) [9]. Рекалибровку значений качества чтения производили с помощью утилиты BaseRecalibrator (GATK). Для выявления соматических вариантов в образцах транскриптома опухоли применяли программу Mutect2 (GATK) в двух вариантах: с использованием парных образцов экзоса крови (режим tumor-normal, «парный») и без них (режим tumor only, «только опухоль»). В обоих случаях для дополнительной фильтрации ошибочно идентифицированных

(артефактных) вариантов применяли Panel of Normals (PoN) — данные секвенирования РНК 46 образцов нормальной ткани кишечника (кишечные органоиды) из открытых источников данных (датасет GSE165512 из базы данных Gene Expression Omnibus [10]). Аннотацию полученных вариантов производили с помощью пакета программ ANNOVAR [11]. Убирали варианты, не прошедшие внутренний фильтр программы Mutect2, а также варианты, поддержанные менее чем 4 прочтениями или в позициях, покрытых менее чем 7 прочтениями. Исключали также варианты, встречающиеся более чем в 2 образцах исследуемой выборки и при этом отсутствующие в базе данных соматических мутаций COSMIC, а также однонуклеотидные инсерции в гомополимерных участках. При анализе без использования парных образцов экзоса крови также исключали мутации, популяционная частота которых в геномной или экзомной выборке базы данных gnomAD v. 4.0 была выше 0,001. Для определения статуса MSI опухоли по данным секвенирования транскриптома использовали утилиты MIRACLE [12] и PreMSIm [13]. Визуализация результатов проводилась в среде R с использованием пакета ggplot2. Значимость различий между группами MSI и MSS определяли с помощью рангового теста Краскела–Уоллиса для групп разного размера.

РЕЗУЛЬТАТЫ

Нам удалось обнаружить 7 оригинальных исследований, предлагающих свой подход к вычислению статуса MSI/MSS на основании данных высокопроизводительного секвенирования (табл. 1 [12–18]). Три из них основаны на анализе данных секвенирования ДНК опухоли и потому не рассматривались в рамках нашего исследования. Среди оставшихся четырёх подходов, использующих данные секвенирования РНК, три основаны на анализе экспрессионных геновых сигнатур, однако готовое программное обеспечение предлагается только для алгоритма PreMSIm, который оценивает экспрессию 15 MSI-ассоциированных генов. Четвёртый подход — инструмент MIRACLE — предназначен для прямой оценки длин микросателлитных повторов в опухолевом образце и выявления статистически значимой вариативности их длин относительно референсного распределения, рассчитанного для 63 219 локусов на основе последовательностей мРНК из базы RefSeq. Анализируются только те локусы, для которых в образце возможна оценка длины повтора на основании не менее 5 прочтений. Для предсказания статуса MSI/MSS инструмент MIRACLE использует бинарный классификатор XGBoost.

Для анализа были использованы результаты секвенирования парных образцов [экзом крови–транскриптом опухоли] от пациентов с аденокарциномой толстой или прямой кишки, полученные нами в 2018–2020 гг. Всего было подобрано 17 пар образцов, из которых контроль качества по результатам секвенирования прошли

Таблица 1. Инструменты вычисления микросателлитной нестабильности по данным высокопроизводительного секвенирования
Table 1. Tools for calculating microsatellite instability using high-throughput sequencing data

Название программы/ способа	Год выпуска	На каком типе данных работает	Какой формат данных нужен	Применимость к данным секвенирования РНК	Ссылка
MSISensor	2014	Геном, экзом, таргетная панель	bam	Нет (только ДНК)	[14]
MANTIS	2016	Экзом	bam	Нет (только ДНК)	[15]
mSINGS	2014	Экзом, таргетная панель	bam	Нет (только ДНК)	[16]
PreMSIm	2020	Транскриптом	Таблица экспрессии генов	Да	[13]
25-генная сигнатура	2019	Транскриптом	Таблица экспрессии генов	Да, но авторы не предоставляют готовое к использованию про- граммное обеспечение	[17]
MSI predictor	2019	Транскриптом	Таблица экспрессии генов	Да, но авторы не предоставляют готовое к использованию про- граммное обеспечение	[18]
MIRACLE	2024	Транскриптом	bam	Да	[12]

13 транскриптомов и 12 экзомов. Метаданные пациентов указаны в табл. 2.

Мы применили инструменты MIRACLE и PreMSIm к группе из 13 транскриптомов. В документации PreMSIm не указан предпочтительный способ нормализации уровней экспрессии генов, поэтому нами были использованы лог-трансформированные значения RSEM, так как эта метрика применялась авторами при обучении модели. PreMSIm классифицировал как MSI три из 13 образцов (p2_1003, p2_1013, p2_1027). При использовании MIRACLE для двух из 13 образцов (p2_1004, p2_1013) вероятность

классификации как MSI оказалась равна единице, для одного (p2_1003) — 0,38, для оставшихся 10 образцов — нулю (табл. 3). Неоднозначность определения статуса для образца p2_1003 связана, вероятно, с тем, что из-за недостаточного количества прочтений, отнесённых к анализируемому локусам, оценка производилась на основании 78% локусов из референсного набора, в то время как для остальных образцов этот показатель составил в среднем 97%. Поскольку абсолютное количество значимо варьируемых повторов для p2_1003 составило 273 в сравнении со средним $17,4 \pm 7,6$ для 10 образцов,

Таблица 2. Клинические и демографические данные пациентов, включённых в анализ полных транскриптомов опухолей
Table 2. Clinical and demographic characteristics of patients included in the analysis of whole tumor transcriptomes

Участник	TNM	Возраст, лет	Пол	Локализация опухоли в кишечнике
p2_1003	T3NxM0	54	Мужской	Сигмовидная
p2_1013	T3N1M0	71	Женский	Слепая
p2_1019	T2N2M0	35	Женский	Прямая
p2_1024	T2N0M0	62	Женский	Сигмовидная
p2_1027	T3bN2bM0	76	Женский	Прямая
p2_1029	T2N1M0	45	Мужской	Прямая
p2_1031	T3bN1M0	54	Женский	Прямая
p2_1032	cT3N2M1a	74	Мужской	Прямая
p2_1034	T3N1M1	84	Мужской	Сигмовидная
p2_1036	mtT2N0M0	80	Мужской	Прямая
p2_1008	T2N0M0	80	Мужской	Сигмовидная
p2_1009	T3NxM0	81	Женский	Восходящая ободочная
p2_1004	T4bNxM0	55	Женский	Сигмовидная

Примечание: TNM: T — Tumor (размер и степень инвазии опухоли), N — Nodes (вовлечённость лимфоузлов), M — Metastasis (наличие отдалённых метастазов).
Note: TNM: T, tumor (size and invasion degree), N, nodes (lymph node involvement), M, metastasis (presence of distant metastases).

Таблица 3. MSI/MSS-статус опухолевых образцов по оценке инструмента MIRACLE

Table 3. MSI/MSS status of tumor samples according to MIRACLE tool assessment

Образец	MSI, вероятность	Количество значимо вариабельных повторов	Доля вошедших в анализ локусов относительно референсного набора
p2_1003	0,38	273	0,781
p2_1004	1,0	488	0,975
p2_1008	0	12	0,955
p2_1009	0	23	0,978
p2_1013	1,0	405	0,974
p2_1019	0	26	0,971
p2_1024	0	17	0,988
p2_1027	0	11	0,928
p2_1029	0	11	0,968
p2_1031	0	6	0,962
p2_1032	0	9	0,963
p2_1034	0	13	0,982
p2_1036	0	21	0,960

однозначно классифицированных как MSS, данный образец был отнесён нами к MSI.

Для подтверждения полученных результатов нами была проведена оценка числа соматических мутаций в транскриптомах двумя подходами: в режиме парного коллинга с участием PoN и только для опухолей с PoN с дополнительной фильтрацией вариантов по частотам gnomAD. Анализ результатов показал, что образцы, классифицированные инструментом MIRACLE как MSI, значительно отличаются от образцов со статусом MSS как по общему числу соматических мутаций, так и по числу несинонимичных соматических мутаций (рис. 1) ($p < 0,03$ во всех 4 вариантах сравнений). Интересно отметить, что не только количество, но и доля несинонимичных мутаций также значительно различается между образцами с разным статусом (рис. 2) ($p < 0,03$ для обоих режимов выявления соматических вариантов). Из трёх образцов, классифицированных PreMSIm как MSI, два совпали с предсказаниями MIRACLE (p2_1003, p2_1013); в одном случае наблюдалось расхождение (p2_1027, MSS по MIRACLE); один из образцов, определённый MIRACLE как MSI, PreMSIm классифицировал как MSS (p2_1004). При этом сравнение числа соматических мутаций для классификаций, полученных с помощью PreMSIm, показало значимые различия только в части сравнений. Если считать, что все образцы с высокой мутационной нагрузкой являются образцами с MSI, то инструмент MIRACLE показал 100% чувствительность и специфичность в сравнении с 66,7 и 92,3% инструмента PreMSIm соответственно.

Для объяснения нестабильного статуса нами был проведён анализ врождённых и соматических вариантов в генах системы репарации ДНК для образцов, классифицированных как MSI. В экзоме и в транскриптоме участника p2_1003 была обнаружена герминальная делеция

rs63749831 в рамке считывания гена *MSH2*. Эта мутация, согласно данным базы CLINVAR, является патогенной и ассоциирована с врождённым синдромом Линча. Сочетание высокой мутационной нагрузки и патогенного варианта в гене *MSH2* позволяет считать, что нестабильный статус для данного транскриптома определён инструментом MIRACLE верно. В образце p2_1004 были обнаружены две потенциально каузальные соматические однонуклеотидные замены в генах системы репарации ДНК. Мутация NM_001258281:c.447+1G>A (rs267607689) в регионе сплайсинга гена *MSH2* является патогенной согласно данным базы CLINVAR, а мутация NM_001167619:c.G187T в гене *MLH1* оценивается как вероятно патогенная. Это в сочетании с высокой мутационной нагрузкой позволяет считать результат инструмента MIRACLE корректным, а результат инструмента PreMSIm — менее корректным.

Ни в экзоме крови, ни в транскриптоме опухоли участника p2_1013 нам не удалось выявить патогенные и возможно патогенные варианты в генах системы репарации *MLH1*, *MSH2*, *MSH2*, *MSH6*, *PMS2*, однако была обнаружена мутация V600E в гене *BRAF*, при этом в двух других транскриптомах с MSI она не обнаружена. Известно, что данная мутация ассоциирована со спорадической MSI в колоректальном раке и не встречается при синдроме Линча [19].

ОБСУЖДЕНИЕ

Опухолевая мутационная нагрузка (tumor mutation burden, TMB) — это количество соматических мутаций в опухоли, выраженное в числе мутаций на 1 млн пар оснований генома. Показатель TMB является важной характеристикой опухоли, так как мутации приводят к образованию неоантигенов [20, 21]. Ранее многими группами была исследована корреляция между TMB и MSI.

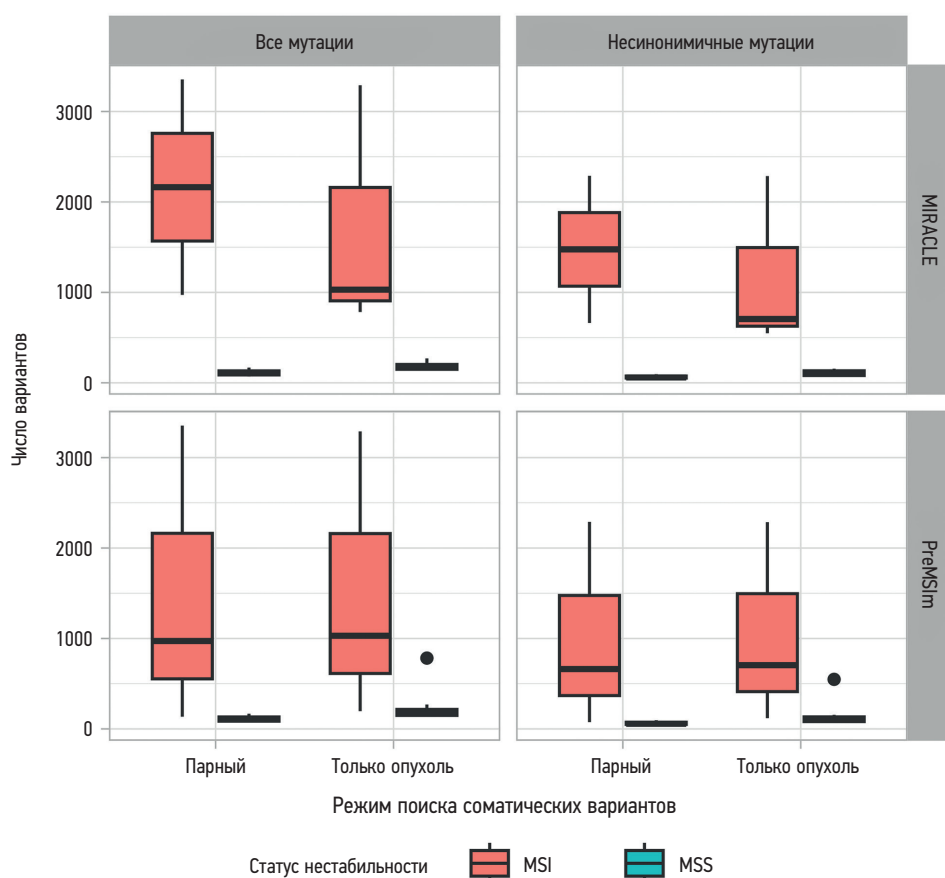


Рис. 1. Соотношение числа мутаций и статуса микросателлитной нестабильности, выявленной инструментами MIRACLE и PreMSIm, в образцах опухолей прямой и толстой кишки ($n=13$). MSI — микросателлитная нестабильность (microsatellite instability), MSS — (microsatellite stability).

Fig. 1. Association between the number of mutations and microsatellite instability status identified by MIRACLE and PreMSIm tools in colorectal cancer samples ($n = 13$). MSI — microsatellite instability, MSS — microsatellite stability.

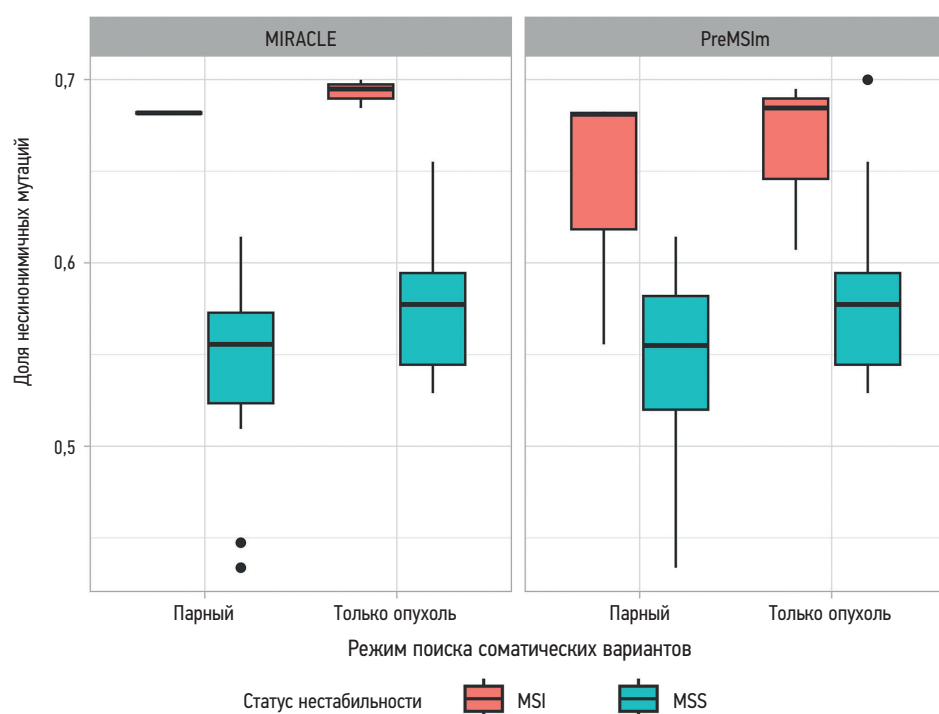


Рис. 2. Соотношение долей несинонимичных мутаций и статуса микросателлитной нестабильности, выявленной инструментами MIRACLE и PreMSIm, в образцах опухолей прямой и толстой кишки ($n=13$). MSI — микросателлитная нестабильность (microsatellite instability), MSS — (microsatellite stability).

Fig. 2. Association between the fraction of nonsynonymous mutations and microsatellite instability status identified by MIRACLE and PreMSIm tools in colorectal cancer samples ($n = 13$). MSI — microsatellite instability, MSS — microsatellite stability.

Например, в исследовании карцином желудка и кишечника при помощи секвенирования панели генов, ассоциированных с ответом на повреждение ДНК (DNA damage response, DDR), выявлена корреляция между опухолями с TMB-H (high, высокий) и/или имеющими повреждающие мутации в генах DDR и MSI-статусом: 87% таких опухолей

были также MSI-H [22]. В исследовании выборки образцов колоректального рака с помощью панели Foundation One (Genetico, Россия) практически все образцы (99,7%), которые были расценены как MSI-H, были также и TMB-H [23]. В другом исследовании среди образцов с MSI-H 77% были также TMB-H [24]. Хотя конкретные цифры зависят

от выставленных отсечек разделения на группы, для коло-ректального рака, по-видимому, существует высокая корреляция между наличием MSI и высоким уровнем TMB [25]. Это позволяет использовать TMB как суррогатный маркер статуса MSI/MSS при биоинформатическом исследовании транскриптома и проверке инструментов вычисления MSI. Однако не для всех типов рака опухоли TMB-H являются подгруппой опухолей MSI-H. В результате метаанализа 17 исследований было выявлено, что TMB-H являются подгруппой опухолей MSI-H также для аденокарциномы пищевода (27,7%) и рака эндометрия (31,0%) по сравнению с группой «все виды рака», меланомами и немелкоклеточным раком лёгкого ($p < 0,05$ для всех этих различий) [25]. Получение окончательных выводов по корреляции TMB-H и MSI-H также затруднено тем, что в разных исследованиях по-разному проводятся отсечки, разделяющие образцы с высокой или низкой TMB. В более ранних исследованиях в качестве уровня отсечки было использовано значение более 20 мутаций на 1 млн пар оснований генома, тогда как в исследовании, например, метастатического немелкоклеточного рака лёгкого наличие от 10 мутаций на 1 млн пар оснований ассоциировалось с увеличенной долей ответа и выживаемостью без прогрессии при комбинированной иммунотерапии [22, 26]. В общем случае TMB рассчитывается с учётом соматических замен и инсерций/делеций без исключения синонимичных замен. Считается, что учёт синонимичных замен снижает влияние шума выборки [27]. Хотя синонимичные мутации, скорее всего, непосредственно не участвуют в создании иммуногенности, их наличие является сигналом мутационных процессов, которые также привели к несинонимичным заменам в других частях генома. При этом чаще всего при дальнейшем анализе не производят разделения между долей синонимичных и несинонимичных замен в общем уровне TMB. В некоторых случаях оценивают только несинонимичные замены [28, 29]. С учётом этого обстоятельства при вычислении TMB нами для анализа было взято как число синонимичных, так и общее число соматических мутаций в образце транскриптома. Для транскриптома унифицированная оценка TMB затруднена из-за особенностей поведения таргетного региона. Однако при рассмотрении только одной нозологии нам кажется возможным использование не мутационной нагрузки (TMB на 1 млн п. о.), а абсолютного числа мутаций. В нашем исследовании и по общему числу соматических мутаций, и по числу несинонимичных мутаций наблюдалась очень резкая (на 1,5 порядка) разница между группами образцов, при этом группа образцов с более высокой мутационной нагрузкой предположительно и была образцами с MSI. Независимо от этого результата на данной же выборке образцов нами были использованы биоинформатические инструменты, позволяющие определить MSI из результатов секвенирования либо по экспрессионному профилю (PreMSIm), либо путём прямой оценки длин микросателлитов (MIRACLE). Алгоритм PreMSIm показал чувствительность и специфичность ниже на нашей

выборке, если мутационный профиль считать суррогатным маркером MSI/MSS. В результате для дальнейшей работы был выбран новый инструмент MIRACLE, который был опубликован в августе 2024 года (т. е. по нему нет накопленного опыта использования). Данный инструмент по сути является аналогом именно ПЦР, а не иммуногистохимической детекции MSI, так как учитывает разброс длин микросателлитов, а не экспрессию гена/белка. Это преимущество повышает результативность использования инструмента MIRACLE. В результате нами была выявлена чёткая зависимость между числом мутаций и статусом MSI. Дополнительно выявлено, что не только число, но и доля несинонимичных мутаций коррелирует с статусом MSI/MSS. Данное наблюдение нам не удалось подтвердить литературными данными, т. е., возможно, мы его обнаружили первые. Такой эффект стоит проверить на более крупной выборке, а также на образцах других нозологий. Кроме того, интересно отметить, что ещё в 2019 году А.М. Goodman и соавт. выяснили, что, возможно, именно мутационная нагрузка (а не MSI сама по себе) является лучшим предиктором выживаемости на фоне иммунотерапии [26]. Это ещё больше поднимает вопрос об одновременном анализе двух показателей в больших наборах данных секвенирования, среди которых транскриптомные данные занимают не последнее место.

Дополнительным ограничением нашей работы является отсутствие прямых валидационных данных по MSI для проанализированных образцов.

ЗАКЛЮЧЕНИЕ

В работе оценены два наиболее подходящих и технически доступных инструмента для определения MSI-статуса по данным полного транскриптома. Инструмент MIRACLE легко интегрируется в протоколы транскриптомного анализа и даёт биологически правдоподобные оценки MSI/MSS, хорошо коррелирующие с мутационной нагрузкой в опухолях толстой и прямой кишки. Инструмент PreMSIm показал недостаточную чувствительность и более низкую специфичность по сравнению с MIRACLE. Результаты демонстрируют, что MIRACLE может быть эффективно использован для оценки MSI-статуса в задачах персонализированной онкодиагностики с использованием полнотранскриптомных данных.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. А.В. Каныгина — осуществление научно-исследовательского процесса, включая выполнение вычислительных экспериментов, подготовка и написание текста статьи; Т.Н. Гарманова, Ю.В. Долудин — осуществление научно-исследовательского процесса, включая сбор данных, редактирование текста статьи; Л.О. Скородумова — осуществление научно-исследовательского процесса, включая выполнение экспериментов, редактирование статьи; Е.И. Шарова — административное управление планированием и проведением исследования, осуществление научно-исследовательского процесса, включая подготовку данных, написание текста и редактирование статьи.

Все авторы одобрили рукопись (версию для публикации), а также согласились нести ответственность за все аспекты работы, гарантируя надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой её части.

Благодарности. Неприменимо.

Этическая экспертиза. Исследование проведено в рамках одобренного протокола (заседание Локального этического комитета от 13.09.2017 по протоколу «Изучение генома, протеома и биомаркёров болезней человека с использованием биологических образцов человека») на базе Биобанка Первого Московского государственного медицинского университета имени И.М. Сеченова.

Согласие на публикацию. Неприменимо.

Источники финансирования. Данная публикация выполнена в рамках государственного задания «Персональный онкопрепарат», номер государственного учёта НИОКТР 124031200002-3. Часть результатов получена с использованием научного оборудования Центра коллективного пользования «Геномика, протеомика, метаболомика» Федерального научно-клинического центра физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства (<http://rcpcm.org/?p=2806>). Финансирующие организации не устанавливали ограничений на использование данных и распространение результатов исследования.

Раскрытие интересов. Авторы заявляют об отсутствии отношений, деятельности и интересов за последние три года, связанных с третьими лицами (коммерческими и некоммерческими), интересы которых могут быть затронуты содержанием статьи.

Доступ к данным. Доступ к данным, полученным в настоящем исследовании, за исключением приведённых в тексте рукописи, закрыт по причине конфиденциальности (наличия в базе данных сведений, на основании которых могут быть идентифицированы участники исследования, и отсутствия их согласия на распространение этих сведений).

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения (текст, иллюстрации, данные).

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовали.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовали три внешних рецензента, один член редакционной коллегии и научный редактор издания.

ADDITIONAL INFORMATION

Author contributions: A.V. Kanygina: investigation, formal analysis, writing—original draft; T.N. Garmanova, Yu.V. Doludin: investigation, data curation, writing—review & editing; L.O. Skorodumova: investigation, formal analysis, writing—review & editing; E.I. Sharova: project administration, investigation, data curation, writing—original draft, writing—review & editing. All authors approved the manuscript (version for publication) and agreed to take responsibility for all aspects of the work, ensuring the proper consideration and resolution of any issues related to the accuracy and integrity of any part of the study.

Acknowledgments: Not applicable.

Ethics approval: The study was conducted under an approved protocol (meeting of the Local Ethics Committee dated September 13, 2017, following the “Study of the genome, proteome, and biomarkers of human diseases using human biological samples” protocol) at the I.M. Sechenov First Moscow State Medical University Biobank.

Consent for publication: Not applicable.

Funding sources: This study was conducted as part of the state assignment “Personalized Anticancer Drug” (State Registration No. 1240312000-02-3). Part of the results was obtained using the equipment of the Shared Research Facility “Genomics, Proteomics, Metabolomics” at the Yu.M. Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of the Federal Medical-Biological Agency (<http://rcpcm.org/?p=2806>). The funding sources did not impose any restrictions on the use or dissemination of the study findings.

Disclosure of interests: The authors have no relationships, activities, or interests for the last three years related to for-profit or not-for-profit third parties whose interests may be affected by the content of the article.

Statement of originality: No previously published material (text, images, or data) was used in this work.

Data availability statement: Access to data obtained in this study, except for those presented in the manuscript text, is restricted due to confidentiality reasons (presence of potentially identifiable participant information in the database and lack of consent for its dissemination).

Generative AI: No generative artificial intelligence technologies were used to prepare this article.

Provenance and peer review: This paper was submitted unsolicited and reviewed following the standard procedure. The review process involved three external reviewers, one member of the editorial board, and the in-house scientific editor.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Vaish M, Mittal B. DNA mismatch repair, microsatellite instability and cancer. *Indian J Exp Biol*. 2002;40(9):989–994.
2. O'Malley DM, Bariani GM, Cassier PA, et al. Pembrolizumab in patients with microsatellite instability-high advanced endometrial cancer: results from the KEYNOTE-158 Study. *J Clin Oncol*. 2022;40(7):752–761. doi: 10.1200/JCO.21.01874 EDN: XBODRZ
3. Li K, Luo H, Huang L, et al. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int*. 2020;20:16. doi: 10.1186/s12935-019-1091-8 EDN: XXNZRC
4. Suraweera N, Duval A, Reperant M, et al. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology*. 2002;123(6):1804–1811. doi: 10.1053/gast.2002.37070
5. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96(4):261–268. doi: 10.1093/jnci/djh034 EDN: ISMQOL
6. Funkhouser WK Jr, Lubin IM, Monzon FA, et al. Relevance, pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas: a report of the association for molecular pathology. *J Mol Diagn*. 2012;14(2):91–103. doi: 10.1016/j.jmoldx.2011.11.001
7. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;17(1):10–12. doi: 10.14806/ej.17.1.200
8. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. doi: 10.1093/bioinformatics/bts635
9. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303. doi: 10.1101/gr.107524.110 EDN: NZKCFJ
10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–210. doi: 10.1093/nar/30.1.207 EDN: ITZHYB
11. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi: 10.1093/nar/gkq603 EDN: MZERFB
12. Choi JW, Lee JO, Lee S. Detecting microsatellite instability by length comparison of microsatellites in the 3' untranslated region with RNA-seq. *Brief Bioinform*. 2024;25(5):bbae423. doi: 10.1093/bib/bbae423 EDN: BGTUKJ
13. Li L, Feng Q, Wang X. PreMSIm: An R package for predicting microsatellite instability from the expression profiling of a gene panel in cancer. *Comput Struct Biotechnol J*. 2020;18:668–675. doi: 10.1016/j.csbj.2020.03.007 EDN: CARSRO

14. Niu B, Ye K, Zhang Q, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*. 2014;30(7):1015–1016. doi: 10.1093/bioinformatics/btt755 EDN: SPHXRJ
15. Kautto EA, Bonneville R, Miya J, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget*. 2017;8(5):7452–7463. doi: 10.18632/oncotarget.13918 EDN: YFDYTZ
16. Salipante SJ, Scroggins SM, Hampel HL, et al. Microsatellite instability detection by next generation sequencing. *Clin Chem*. 2014;60(9):1192–1199. doi: 10.1373/clinchem.2014.223677
17. Pačínková A, Popovici V. Cross-platform data analysis reveals a generic gene expression signature for microsatellite instability in colorectal cancer. *Biomed Res Int*. 2019;2019:6763596. doi: 10.1155/2019/6763596
18. Danaher P, Warren S, Ong S, et al. A gene expression assay for simultaneous measurement of microsatellite instability and anti-tumor immune activity. *J Immunother Cancer*. 2019;7(1):15. doi: 10.1186/s40425-018-0472-1 EDN: OIZGNB
19. Deng G, Bell I, Crawley S, et al. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res*. 2004;10(1 Pt 1):191–195. doi: 10.1158/1078-0432.ccr-1118-3
20. Tran E, Ahmadzadeh M, Lu YC, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science*. 2015;350(6266):1387–1390. doi: 10.1126/science.aad1253
21. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348(6230):124–128. doi: 10.1126/science.aaa1348 EDN: URBFX
22. Parikh AR, He Y, Hong TS, et al. Analysis of DNA damage response gene alterations and tumor mutational burden across 17,486 tubular gastrointestinal carcinomas: implications for therapy. *Oncologist*. 2019;24(10):1340–1347. doi: 10.1634/theoncologist.2019-0034
23. Fabrizio DA, George TJ Jr, Dunne RF, et al. Beyond microsatellite testing: assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition. *J Gastrointest Oncol*. 2018;9(4):610–617. doi: 10.21037/jgo.2018.05.06
24. Innocenti F, Ou FS, Qu X, et al. Mutational analysis of patients with colorectal cancer in CALGB/SWOG 80405 identifies new roles of microsatellite instability and tumor mutational burden for patient outcome. *J Clin Oncol*. 2019;37(14):1217–1227. doi: 10.1200/JCO.18.01798
25. Luchini C, Bibeau F, Ligtenberg MJL, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol*. 2019;30(8):1232–1243. doi: 10.1093/annonc/mdz116 EDN: YROWKS
26. Goodman AM, Sokol ES, Frampton GM, et al. Microsatellite-stable tumors with high mutational burden benefit from immunotherapy. *Cancer Immunol Res*. 2019;7(10):1570–1573. doi: 10.1158/2326-6066.CIR-19-0149
27. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9(1):34. doi: 10.1186/s13073-017-0424-2 EDN: RJYZOM
28. Vanderwalde A, Spetzler D, Xiao N, et al. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Med*. 2018;7(3):746–756. doi: 10.1002/cam4.1372
29. Cheng DT, Mitchell TN, Zehir A, et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn*. 2015;17(3):251–264. doi: 10.1016/j.jmoldx.2014.12.006

ОБ АВТОРАХ

* Шарова Елена Ивановна;

адрес: Россия, 119435, Москва, ул. Малая Пироговская, д. 1а;

ORCID: 0000-0003-3208-9719;

eLibrary SPIN: 5226-3942;

e-mail: sharova78@gmail.com

Каныгина Александра Васильевна;

ORCID: 0000-0003-4993-9492;

eLibrary SPIN: 1100-0839;

e-mail: kanygina@rcpcm.org

Гарманова Татьяна Николаевна, канд. мед. наук;

ORCID: 0000-0003-2330-4229;

eLibrary SPIN: 7627-0586;

e-mail: tatianagarmanova@gmail.com

Долудин Юрий Валерьевич;

ORCID: 0000-0002-0554-9911;

eLibrary SPIN: 3496-7005;

e-mail: doludin@gmail.com

Скородумова Любовь Олеговна, канд. биол. наук;

ORCID: 0000-0002-7747-2730;

eLibrary SPIN: 5060-4554;

e-mail: lo.skorodumova@gmail.com

AUTHORS' INFO

* Elena I. Sharova;

address: 1a Malaya Pirogovskaya st, Moscow, Russia, 119435;

ORCID: 0000-0003-3208-9719;

eLibrary SPIN: 5226-3942;

e-mail: sharova78@gmail.com

Alexandra V. Kanygina;

ORCID: 0000-0003-4993-9492;

eLibrary SPIN: 1100-0839;

e-mail: kanygina@rcpcm.org

Tatiana N. Garmanova, MD, Cand. Sci. (Medicine);

ORCID: 0000-0003-2330-4229;

eLibrary SPIN: 7627-0586;

e-mail: tatianagarmanova@gmail.com

Yuriy V. Doludin;

ORCID: 0000-0002-0554-9911;

eLibrary SPIN: 3496-7005;

e-mail: doludin@gmail.com

Liubov O. Skorodumova, Cand. Sci. (Biology);

ORCID: 0000-0002-7747-2730;

eLibrary SPIN: 5060-4554;

e-mail: lo.skorodumova@gmail.com

* Автор, ответственный за переписку / Corresponding author