

Для цитирования: Семина Е.В., Кубасова К.А., Какоткин В.В., Родимов С.В., Агапов М.А. Геномный ландшафт колоректального рака. Пилотное исследование в Российской Федерации. Сибирский онкологический журнал. 2025; 24(3): 65–75. – doi: 10.21294/1814-4861-2025-24-3-65-75

For citation: Semina E. V., Kubasova K. A., Kakotkin V. V., Rodimov S. V., Agapov M. A. Genetic landscape of colorectal cancer. A pilot study in the Russian Federation. Siberian Journal of Oncology. 2025; 24(3): 65–75. – doi: 10.21294/1814-4861-2025-24-3-65-75

ГЕНОМНЫЙ ЛАНДШАФТ КОЛОРЕКТАЛЬНОГО РАКА. ПИЛОТНОЕ ИССЛЕДОВАНИЕ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Е.В. Семина^{1,2}, К.А. Кубасова³, В.В. Какоткин¹, С.В. Родимов², М.А. Агапов^{1,2}

¹ФГАОУ ВО «Балтийский федеральный университет им. И. Канта»
Россия, 236041, г. Калининград, ул. Александра Невского, 14

²Московский государственный университет имени М.В. Ломоносова
Россия, 119991, г. Москва, ул. Ленинские Горы, 1

³ООО «Биотек кампус»
Россия, 117437, г. Москва, ул. Миклухо-Маклая, 16/10, корп. 16

Аннотация

Цель исследования – комплексный анализ молекулярно-генетических изменений в опухолевых и нормальных тканях больных колоректальным раком с использованием полногеномного секвенирования для выявления мутаций *de novo*, паттернов микросателлитной нестабильности и мутационных сигнатур, связанных с возникновением опухоли. **Материал и методы.** Исследование являлось одномоментным поперечным, проводилось на территории Калининградской области. Проведен анализ образцов колоректального рака с использованием полногеномного секвенирования нового поколения (WGS). Для картирования прочтений был использован референсный геном человека GRCh38.p14. Биоинформатический анализ сигнатур проводился отдельно по каждому пациенту. Была секвенирована 71 пара образцов: полные геномы биопсии опухоли и ДНК, выделенная из крови, для 71 пациента. Средний возраст пациентов – 69 лет, медианный – 71 год. **Результаты.** Выявлены генетические изменения, связанные с развитием и прогрессированием опухоли, а также потенциальные терапевтические мишени. Используя базу данных OncoPrint, была обнаружена 151 драйверная мутация у 62 пациентов. С использованием базы данных CIVIC выявлено 70 маркерных мутаций у 45 пациентов. Обнаружено, что самые распространенные драйверные мутации в исследуемой когорте – мутации в гене KRAS (KRAS:G35T, KRAS:G35A, KRAS:G38A), ассоциированные с устойчивостью опухоли к препаратам Панитумумаб и Цетуксимаб. Частота наиболее распространенных драйверных мутаций в исследуемой когорте низкая, что говорит о высокой гетерогенности колоректального рака. При анализе мутационной нагрузки средний показатель TMB был равен 5,5, что свидетельствует о высоком потенциале использования иммунотерапии при колоректальном раке. В исследуемой популяции преобладают мутационные сигнатуры SBS1, SBS5 и SBS39. **Заключение.** Выявленное многообразие изменений генома может быть связано с ошибками в репликации ДНК, воздействием внешних факторов и генетической предрасположенностью. Профиль мутационных сигнатур очень схож в исследуемых образцах всех пациентов, что предполагает действие одного/единого мутагенного фактора в развитии колоректального рака у всех пациентов.

Ключевые слова: колоректальный рак, полногеномное секвенирование, секвенирование нового поколения, Next-Generation Sequencing, NGS.

GENETIC LANDSCAPE OF COLORECTAL CANCER. A PILOT STUDY IN THE RUSSIAN FEDERATION

E.V. Semina^{1,2}, K.A. Kubasova³, V.V. Kakotkin¹, S.V. Rodimov², M.A. Agapov^{1,2}

¹Immanuel Kant Baltic Federal University

4, Alexander Nevsky St., Kaliningrad, 236041, Russia

²Lomonosov Moscow State University

1, Leninskie Gory St., Moscow, 119991, Russia

³Biotek Campus

build. 16, 16/10, Mikluho-Maklaya St., Moscow, 117437, Russia

Abstract

The purpose of the study was a comprehensive analysis of molecular genetic changes in tumor and normal tissues of patients with colorectal cancer using whole-genome sequencing to identify *de novo* mutations, microsatellite instability patterns, and mutational signatures associated with tumor development. **Material and Methods.** A single-time cross-sectional study was conducted in the Kaliningrad region. Colorectal cancer samples were analyzed using whole-genome next-generation sequencing (WGS). The human reference genome GRCh38.p14 was used for read mapping. Bioinformatic analysis of signatures was conducted separately for each patient. 71 pairs of samples were sequenced: complete genomes of tumor biopsies and DNA isolated from blood samples for 71 patients. The average age among the analyzed patients was 69 years, with a median of 71 years. **Results.** Genetic alterations associated with tumor development and progression, as well as potential therapeutic targets were identified. Using the Oncovar database, 151 driver mutations were detected in 62 patients. Using the CIVIC database, 70 marker mutations were found in 45 patients. It was found that the most common driver mutations in the study cohort were mutations in the *KRAS* gene (*KRAS*:G35T, *KRAS*:G35A, *KRAS*:G38A), associated with tumor resistance to Panitumumab and Cetuximab. The frequency of the most common driver mutations in the study cohort was low, indicating high heterogeneity of colorectal cancer. Analysis of the tumor mutational burden (TMB) revealed a high average value of 5.5, suggesting a significant potential of immunotherapy in colorectal cancer treatment. In the studied population, the mutational signatures SBS1, SBS5, and SBS39 were the most common. **Conclusion.** The identified diversity of genomic changes may be associated with DNA replication errors, exposure to external factors, and genetic predisposition. The pattern of mutational signatures was very similar among all patients in the studied samples, suggesting a single mutagenic factor is involved in the development of colorectal cancer in all patients.

Key words: colorectal cancer, Whole Genome Sequencing, Next-Generation Sequencing, NGS.

Введение

Колоректальный рак (КРР) является одной из наиболее распространенных и агрессивных форм злокачественных новообразований, он входит в число основных причин смертности от онкологических заболеваний по всему миру. Понимание молекулярной природы этого заболевания имеет критическое значение для разработки методов диагностики, прогноза и терапии. В последние годы значительно расширились знания о генетических и эпигенетических изменениях, ассоциированных с развитием и прогрессированием КРР. Одним из ключевых методов исследования геномных изменений в образцах КРР является секвенирование нового поколения (Next-Generation Sequencing, NGS), которое позволяет проводить глубокий анализ геномных и транскриптомных изменений с высокой степенью чувствительности и точности.

NGS предоставляет возможности для выявления точечных мутаций, копийных вариаций, а также изменений в структуре генома, что особенно важно для изучения гетерогенности опухолей и

индивидуальных особенностей патогенеза КРР у различных пациентов. В частности, значительное внимание уделяется детекции соматических мутаций, т. к. они могут служить потенциальными биомаркерами для ранней диагностики, прогноза и предсказания ответа на терапию. Генетические изменения в таких ключевых генах, как *APC*, *KRAS*, *TP53*, *PIK3CA*, а также в участках микросателлитной нестабильности (MSI) могут существенно повлиять на прогноз заболевания и ответ на лечение. Современные NGS-технологии, такие как панели таргетного секвенирования, полногеномное и экзомное секвенирование, открывают возможность для создания индивидуализированных подходов к лечению пациентов с КРР.

Обработка данных NGS представляет собой многосоставной процесс, включающий этапы выравнивания последовательностей на референсный геном, выявления и аннотации вариантов, фильтрации артефактов и ложноположительных результатов, а также интерпретации биологически значимых изменений. Точность анализа NGS

данных играет решающую роль в получении достоверных результатов, которые могут быть применены в клинической практике.

В отечественной литературе описаны пилотные исследования, направленные на идентификацию наследственных форм онкологических заболеваний у лиц с повышенным риском развития и использование образцов крови для этой цели [1]. Однако на российской выборке больных отсутствуют исследования, включающие анализ первичного опухолевого узла и образца крови как контроля мутаций и исключающие из анализа мутации, обнаруженные в крови, для выявления мутаций *de novo*, которые могут быть связаны с возникновением опухоли.

В настоящем исследовании впервые на российской популяции проведен анализ 71 образца KPP с использованием полногеномного NGS (WGS). В работе представлен алгоритм обработки данных NGS и анализа мутаций, выявленных в образцах колоректального рака. Описание этапов обработки данных включает подготовку первичного (сырого) материала, фильтрацию и аннотирование полученных данных, а также их интерпретацию.

Цель исследования — комплексный анализ молекулярно-генетических изменений в опухолевых и нормальных тканях больных колоректальным раком с использованием полногеномного секвенирования для выявления мутаций *de novo*, паттернов микросателлитной нестабильности и мутационных сигнатур, связанных с возникновением опухоли.

Материал и методы

Исследование являлось одномоментным поперечным, проводилось на территории Калининградской области в рамках проекта «Национальная генетическая инициатива 100000+Я». Критерии включения в исследование: возраст старше 18 лет; подписанное информированное согласие на участие в исследовании и использование биологического материала в целях исследования; обнаружение при колоноскопии эпителиального образования (типы 2B-3 по JNET, наличие ямочного рисунка Vi и Vn по S. Kudo) с последующей верификацией диагноза «аденокарцинома толстой кишки» или «аденома толстой кишки» с помощью морфологического исследования.

Критерии не включения: возраст до 18 лет; наличие носительства вирусных инфекций с парентеральным путем передачи (ВИЧ, гепатит С, гепатит В, сифилис); отказ от участия в исследовании или невозможность получить согласие из-за мнестикоинтеллектуальных нарушений или иных причин; невозможность получения достаточного объема биологического материала (фрагмента слизистой оболочки кишки) для проведения рутинного патологоанатомического исследования и выделения ДНК.

Критерии исключения: отсутствие опухолевого роста в полученных фрагментах слизистой по результатам морфологического исследования; отказ добровольца от обработки результатов секвенирования ДНК, полученной из биологического материала, на любом этапе.

Для участия в исследовании приглашались пациенты, обратившиеся для выполнения эндоскопического скрининга колоректального рака. Непосредственно на этапе подготовки к процедуре пациенты были информированы о цели, задачах исследования, возможности участия в исследовании в качестве добровольца в случае выявления по данным первичного эндоскопического скрининга новообразований ободочной кишки, характеризующихся наличием паттернов метаплазии, дисплазии или злокачественной трансформации. В случае подписания пациентом добровольного информированного согласия и обнаружения образования в толстой кишке производился забор фрагмента слизистой оболочки толстой кишки в наиболее измененной части, объемом 2–3 мм³, который помещался в пробирку с буферным раствором, а также 16 мл венозной крови.

В случае выявления критериев не включения/исключения на этапе эндоскопического или морфологического исследования врач-исследователь сообщал об этом добровольцам, а биологические образцы подлежали утилизации.

Геномная ДНК была выделена методом сорбции на магнитных частицах (набор MGIEasy Magnetic Beads Blood Genomic DNA Extraction Kit, MGI), оценка качества выделенной геномной ДНК проведена методами спектрофотометрии и флуориметрии, затем использована для приготовления геномных библиотек для секвенирования (набор для подготовки библиотек без ПЦР, с энзиматической фрагментацией ДНК MGIEasy FS PCR-Free Library Prep Set, 96 реакций (MIX), MGI). Полученные библиотеки анализировали на секвенаторе (DNBSEQ-T7), методом PE150.

Для тримминга и фильтрации прочтений был использован инструмент fastp [2, <https://github.com/OpenGene/fastp>] со следующими параметрами: -5 -3 -W 4 -M 20 -l 70 -q 20 -u 20. Тримминг производился с 5' и с 3' конца прочтения (-5, -3) плавающим окном длиной 4 нуклеотида (-W 4). Если среднее качество прочтений внутри плавающего окна было меньше 30 (-M 30), то этот фрагмент обрезался. Также отфильтровывались прочтения с длиной после тримминга меньше 70 (-l 70), а также прочтения, у которых после тримминга более 30 % нуклеотидов имели качество ниже Q30 (-q 30, -u 30). Последовательности адаптеров детектировались и удалялись автоматически.

Для картирования прочтений был использован референсный геном человека GRCh38.p14, скачанный из Ensembl (https://www.ensembl.org/Homo_sapiens/Info/Index). Картирование про-

чтений на индексированный геном выполнялось при помощи программы bwa mem со следующими параметрами: -M -v 3 [3]. Для маркирования дуплицированных прочтений была использована программа Picard MarkDuplicates с опцией -MAX_RECORDS_IN_RAM 50000000. Прочтения с маркированными дубликатами в формате bam были проиндексированы при помощи samtools index.

Для поиска герминальных мутаций (SNP, InDel) использована программа DeepVariant [4, <https://github.com/google/deepvariant>]. DeepVariant был запущен в контейнере Singularity (google/deepvariant:latest) с параметром --model_type=»WGS» и остальными настройками по умолчанию. Для поиска соматических мутаций (SNP, InDel) был использован инструмент Strelka2 со стандартными настройками [5, <https://github.com/Illumina/strelka>], поскольку этот инструмент продемонстрировал высокую точность при выявлении соматических мутаций в ряде сравнительных исследований [6]. Strelka2 работает гораздо быстрее, что важно при обработке большого количества тяжелых данных полногеномного секвенирования. Ансамбли из инструментов незначительно повышают чувствительность или специфичность, но значительно увеличивают вычислительную нагрузку на анализ образцов. Фильтр в виде частоты мутации был использован, т. к. показана низкая точность определения мутаций с частотой меньше 5 %, особенно при секвенировании с глубиной 100X.

Для аннотации мутаций использована программа Ensembl VEP [7, <https://www.ensembl.org/info/docs/tools/vep/script/index.html>]. Для этой программы были использованы следующие параметры: «--sift b --polyphen b --protein --symbol --ccds --uniprot --per_gene --domains --check_existing --clin_sig_allele 1 --max_af--af_1kg --af_gnomade --pubmed --var_synonyms --failed 1». Также по каждой мутации был произведен поиск в базе данных CIVIC [<https://civicdb.org/releases/main>, база данных Evidence] и в базе данных OncoPrint [<https://oncoprint.org/>].

Для поиска амплификаций и делеций (вариантов числа копий – CNV) использована программа CNVkit [8, <https://cnvkit.readthedocs.io/en/stable/>]. Данная программа запущена в режиме batch (cnvkit.py batch), с опциями -m wgs, --annotate refFlat38.txt, --target-avg-size 1000. Для аннотации найденных CNV и определения количества копий каждого гена использована аннотация генома человека версии GRCh38, скачанная с сайта UCSC (<https://hgdownload.soe.ucsc.edu/downloads.htm>).

Для оценки качества сырых прочтений в формате fastq использована программа FASTQC [9, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>]. Графики распределения качества, GC-состава и неопознанных нуклеотидов (N) по позиции в прочтении после тримминга и фильтрации

получены при помощи программы fastp [2, <https://github.com/OpenGene/fastp>]. Анализ контаминации в образце выполнен при помощи программы FastQ_Screen [10, https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/]. Для анализа качества картирования прочтений на референсный геном использована программа Picard, функции CollectAlignmentSummaryMetrics и CollectInsertSizeMetrics [<https://broadinstitute.github.io/picard/>]. Получены график доли картированных на геном прочтений и график распределения размера вставки (insert size). Для анализа покрытия референсного генома использована программа Mosdepth [11, <https://github.com/brentp/mosdepth>].

Для анализа микросателлитной нестабильности (MSI) использован инструмент MSISensor-Pro [12, <https://github.com/xjtu-omics/msisensor-pro>]. Для запуска этой программы на первом этапе подготовлен файл с микросателлитами генома человека, используя команду msisensor-pro scan и референсный геном (GRCh38.p14) в формате fasta. Затем, используя этот файл и файлы с картированием на геном прочтений нормы и опухоли (.bam файлы), выполнен запуск программы msisensor-pro msi со стандартными настройками. В результате получен файл в формате tsv, который содержал следующую информацию: количество сайтов с микросателлитами в геноме человека, количество сайтов с нестабильными микросателлитами в образце и доля нестабильных микросателлитных сайтов (MSI) в образце.

Для анализа дефектов гомологичной рекомбинации (HRD) в образце использована программа HRDCNA, выдающая значения в диапазоне от 0 до 1 [13, 2023, <https://github.com/XSLiuLab/HRDCNA>]. Для запуска этой программы использован файл с количеством копий сегментов (*.call.cns), полученный при помощи программы CNVkit. В результате работы программы получен HRDCNAScore для образца (значение от 0 до 1). Чем выше значение HRDCNAScore, тем выше вероятность того, что в образце присутствуют дефекты гомологичной рекомбинации.

Биоинформатический анализ сигнатур проводился отдельно по каждому пациенту. Для анализа мутационных сигнатур использована программа SigProfilerAssignment [14, <https://github.com/ShixiangWang/sigminer/>]. Для запуска этой программы использован vcf файл с соматическими мутациями (SNP, InDel). На выходе получена таблица со значением вклада каждой мутационной сигнатуры (COSMIC v.3.4, <https://cancer.sanger.ac.uk/signatures/sbs/>) в формирование мутаций у пациента.

Результаты

В данной работе секвенирована 71 пара образцов: полные геномы биопсии опухоли и ДНК, выделенная из крови, для 71 пациента. Средний

Таблица/Table

Клинико-патоморфологические характеристики больных КРР
Clinical and pathological characteristics of patients with colorectal cancer (CRC)

Признак/Sing	Доля в выборке/Proportion
Локализация опухоли/Tumor localization	
Прямая кишка/Rectum	17 (23,94 %)
Ректосигмоидный переход/Rectosigmoid junction	8 (11,27 %)
Сигмовидная кишка/Sigmoid colon	18 (25,35 %)
Вышележащие отделы/Upper colon	28 (39,44 %)
Степень злокачественности/Tumor grade	
High grade	64 (90,14 %)
Low grade	7 (9,86 %)
Стадия опухолевого процесса/Tumor stage	
I	9 (12,68 %)
II	18 (25,36 %)
III	21 (29,57 %)
IV	23 (32,39 %)
pT	
pT2	11 (15,49 %)
pT3	41 (57,75 %)
pT4a	15 (21,13 %)
pT4b	4 (5,63 %)
pN	
pN0	28 (39,44 %)
pN1a	22 (30,99 %)
pN1b	5 (7,04 %)
pN1c	3 (4,23 %)
pN2a	5 (7,04 %)
pN2b	8 (11,26 %)
M	
cM0	48 (67,61 %)
pM1a	12 (16,90 %)
pM1b	11 (15,49 %)
Лимфатическая инвазия/Lymphovascular invasion	
L0	43 (60,56 %)
L1	28 (39,44 %)
Сосудистая инвазия/Vascular invasion	
V0	62 (87,32 %)
V1	9 (12,68 %)
Периневральная инвазия/Perineural invasion	
Pn0	60 (84,51 %)
Pn1	11 (15,49 %)
Опухолевый баддинг/Tumor budding	
Bd1	32 (45,07 %)
Bd2	26 (36,62 %)
Bd3	13 (18,31 %)

Примечание: таблица составлена авторами.

Note: created by the authors.

возраст пациентов – 69 (SD=11,1) лет, медианный – 71 (Q1/Q3=61/75) год (рис. 1). Распределение по полу: 35 мужчин, 36 женщин (рис. 1). Клинико-патоморфологические характеристики больных КРР представлены в таблице.

После фильтрации прочтений средний процент оснований с качеством не менее Q30 в образцах нормы равняется 92,3 %, в образцах опухоли – 91,9 %.

На референсный геном было выровнено 100 % прочтений нормы и 99,4 % прочтений опухоли. Средний процент дублированных прочтений в образцах нормы равняется 4,4 %, в образцах опухоли – 13,2 %. Средний размер вставки в образцах нормы – 388,6, в образцах опухоли – 395,5. Средняя глубина прочтения образцов нормы – 26.9X, образцов опухоли – 100.0X.

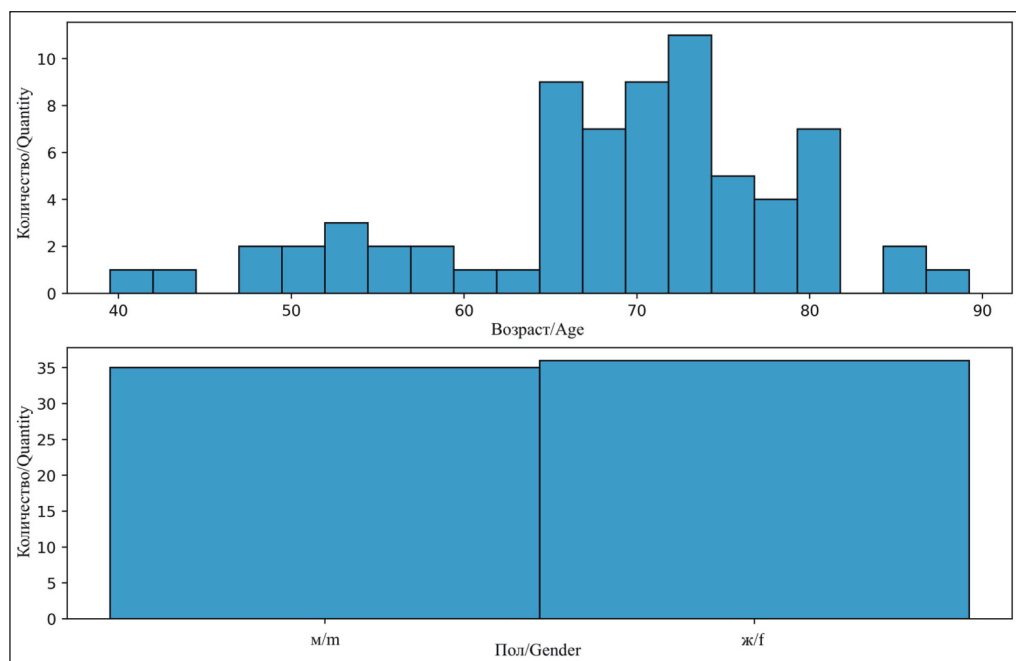


Рис. 1. Распределение пациентов по возрасту и полу. Горизонтальная ось – значение признака (возраст на верхнем графике и пол на нижнем графике), вертикальная ось – количество пациентов с данным значением признака; рисунок выполнен авторами
Fig. 1. Patients' distribution charts by age and gender. The horizontal axis is the characteristic value (age on the top graph and gender on the bottom graph), the vertical axis is the number of patients with given characteristic value; created by the authors

Среднее количество обнаруженных сырых соматических мутаций до фильтрации в одном образце – 1462120. После фильтрации соматических мутаций (были оставлены только мутации, влияющие на белок, с частотой мутантного аллеля $>5\%$ и с количеством прочтений, содержащих мутантный аллель, не менее 5) в среднем для каждого пациента осталось 328 мутаций. Всего после такой фильтрации в выборке обнаружено 22 275 соматических мутаций. Больше всего мутаций было обнаружено в генах APC (79 мутаций у 50 пациентов), TTN (63 мутации у 32 пациентов), TP53 (50 мутаций у 46 пациентов), MUC19 (42 мутации у 17 пациентов), MUC4 (36 мутаций у 24 пациентов), KRAS (32 мутации у 30 пациентов).

Наиболее представленными конкретными мутациями среди обнаруженных генов были следующие: ENST00000311936(KRAS):c.225C>A (у 8 пациентов), ENST00000311936(KRAS):c.228C>T (у 6 пациентов), ENST00000311936(KRAS):c.225C>T (у 6 пациентов) и ENST00000634670(KLF18):c.2224-2225A>TAT (у 7 пациентов). Таблица «Top_somatic_mutations» со всеми уникальными мутациями, сортированными по убыванию частоты встречаемости, приведена в приложении электронного варианта статьи.

Далее, используя базу данных Oncovar, нами обнаружена 151 драйверная мутация у 62 пациентов. Самые частые драйверные мутации – KRAS: G35T (8 пациентов), KRAS:G35A (6 пациентов), KRAS:G38A (6 пациентов). Таблица «Top_driver_mutations» со всеми уникальными онкогенными мутациями, сортированными по убыванию частоты

встречаемости, приведена в приложении электронного варианта статьи. Используя базу данных CIVIC, обнаружено 70 маркерных мутаций у 45 пациентов. Наиболее частые из них: KRAS:G35T (8 пациентов), KRAS:G35A (6 пациентов), KRAS:G38A (6 пациентов). Также у 4 пациентов идентифицирована маркерная мутация в гене BRAF V600E. Таблица «Top_marker_mutations» со всеми уникальными маркерными мутациями, сортированными по убыванию частоты встречаемости, приведена в приложении электронного варианта статьи.

Среднее количество обнаруженных герминальных мутаций в образце – 5 694 810. После фильтрации герминальных мутаций были оставлены только мутации, влияющие на белок в 80 генах (полный перечень приведен в Приложении), их оказалось 4 713. Больше всего мутаций было обнаружено в генах HLA-B (2 450 мутаций у 67 пациентов), EIF3E (227 мутаций у 60 пациентов), PMS2 (194 мутации у 66 пациентов), RNF43 (131 мутация у 65 пациентов), TP53 (129 мутаций у 66 пациентов). Таблица «Top_germline_mutations» со всеми уникальными герминальными мутациями в 80 генах, сортированными по убыванию частоты встречаемости, приведена в приложении электронного варианта статьи.

В среднем у каждого пациента обнаружены амплификации 2 554 генов и делеции 2 619 генов. Среди онкогенов по версии базы данных Oncovar наиболее часто встречаются амплификации в следующих генах (в скобках указано число образцов, имеющих мутацию): GNAS (41), KCNQ2 (41), ACTB (28), COL4A2 (29), SOSTDC1 (27), FERD3L (27),

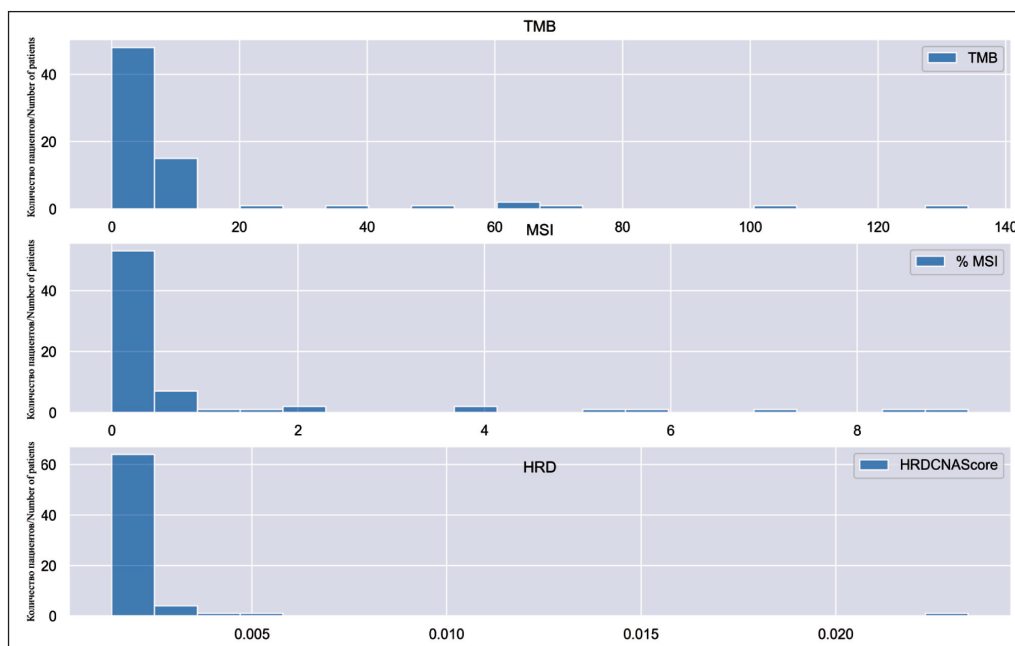


Рис. 2. Распределение TMB (мутационная нагрузка), MSI (доля сайтов с микросателлитной нестабильностью) и HRD (вероятность наличия дефектов в системе репарации путем гомологичной рекомбинации). Горизонтальная ось – значение параметра (TMB на верхнем графике, MSI на среднем и HRD на нижнем), вертикальная ось – количество пациентов с данным значением параметра; рисунок выполнен авторами

Fig. 2. Distribution charts of TMB (tumor mutational burden), MSI (proportion of sites with microsatellite instability) and HRD (defects' probability in the repair system by homologous recombination). The horizontal axis represents the parameter value (TMB on the top chart, MSI on the middle chart and HRD on the bottom chart), the vertical axis represents the number of patients with a given parameter value; created by the authors

BRAF (23), *ADCY8* (24), *TRPS1* (23). Делеции наиболее часто встречаются в генах *HDX* (38), *NKAP* (38), *THOC2* (38), *AFF2* (38), *RS1* (36), *TMX3* (32), *SMAD4* (32), *LIPG* (31), *TP53* (23). Таблица «Top_amplifications_deletions» (страница «Амплификации») со всеми амплификациями, сортированными по убыванию частоты встречаемости, приведена в приложении электронного варианта статьи. Таблица «Top_amplifications_deletions» (страница «Делеции») со всеми делециями, сортированными по убыванию частоты встречаемости, приведена в приложении электронного варианта статьи.

Согласно руководству FDA [15], высокий уровень TMB определяется по пороговому значению больше 10. Среднее значение TMB (количество мутаций на 1 млн пар оснований в геноме) составило 12,8 (SD=23,61, коэффициент вариации – 185,14 %), медианное значение – 5,5 (Q1/Q3=4,16/8,32) (рис. 2). Среднее значение MSI (процент сайтов с микросателлитной нестабильностью) составило 0,73 (SD=1,75, коэффициент вариации – 240,85 %), медианное значение – 0,07 (Q1/Q3=0,0/0,41) (рис. 2). Среднее значение HRD (значение, отображающее вероятность наличия дефектов в системе репарации путем гомологичной рекомбинации) составило 0,0019 (SD=0,0027, коэффициент вариации – 145,42 %), медианное значение – 0,0014 (Q1/Q3=0,0013/0,0014) (рис. 2). Максимальное значение HRD – 0,0234. Такие низкие значения HRD могут свидетельствовать о том, что в исследуемой когор-

те больных отсутствовали пациенты со значимыми дефектами в системе репарации двуцепочечных разрывов ДНК путем гомологичной рекомбинации. Распределение всех трех показателей (TMB, MSI, HRD) значительно отличалось от нормального, характеризовалось правосторонней асимметрией с выраженным эксцессом, что делает предпочтительным использование медианы и квартилей для оценки параметров распределения представленных величин в исследуемой популяции.

Мутационные сигнатуры были определены по всем мутациям, в том числе в некодирующей части генома. Несмотря на то, что отдельные мутации у пациентов сильно различаются, мутационные сигнатуры (паттерны нуклеотидных замен) у пациентов весьма схожи. Так, в исследуемой группе пациентов обнаружена активность 15 мутационных сигнатур (рис. 3). При этом значительный вклад в формирование мутаций (от 5 %) вносят только три из них – SBS1 (5 %), SBS5 (33 %) и SBS39 (57 %). Сигнатуры SBS1 и SBS5 являются clock-like сигнатурами, их накопление связано с возрастом. Этиология сигнатуры SBS39 не определена.

Обсуждение

В представленном исследовании проанализированы данные полногеномного секвенирования 71 образца колоректального рака и соответствующие им образцы клеток крови в качестве контроля. После фильтрации мутаций (SNP, Indel) было получе-

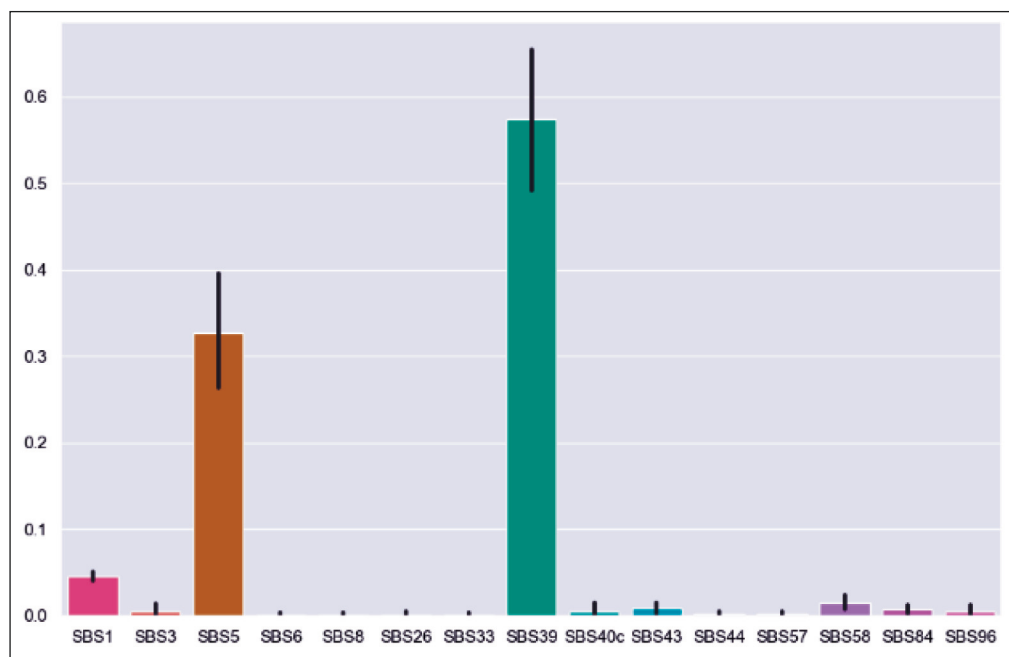


Рис. 3. Доля мутаций, относящихся к каждой мутационной сигнатуре с ненулевым вкладом. Горизонтальная ось – мутационные сигнатуры, вертикальная ось – средняя доля мутаций, которая относится к данной мутационной сигнатуре, по всем пациентам.

Вертикальные линии обозначают 95 % ДИ; рисунок выполнен авторами

Fig. 3. Chart depicting the proportion of mutations belonging to each non-zero contribution mutation signature. The horizontal axis represents mutational signatures, the vertical axis represents the average proportion of mutations associated with each signature across all patients. Vertical lines indicate 95 % CI; created by the authors

но 22 275 соматических мутаций, 151 драйверная соматическая мутация, 70 маркерных соматических мутаций и 4 713 герминальных мутаций. Из них уникальными оказались 21 857, 111, 38, 405 мутаций соответственно. Выяснилось, что большая часть соматических мутаций являются уникальными, что свидетельствует о высокой мутационной гетерогенности исследуемой когорты.

В то же время мутации *KRAS* выявлены у 30 пациентов, что согласуется с литературными данными, указывающими на его роль как одного из наиболее часто мутирующих генов при развитии КРР [16]. Мутации *BRAF* V600E, хотя и менее частые, были обнаружены у четырех пациентов, что также свидетельствует о возможной важности этого маркера для прогноза заболевания и выбора терапии, особенно с учетом того, что *BRAF* V600E ассоциируется с более агрессивными формами некоторых злокачественных опухолей [17, 18].

При анализе CNV обнаружено, что в среднем примерно 20 % всех генов имеют измененное количество копий (амплификации и делеции примерно в равной пропорции). Среди CNV присутствуют амплификации важных онкогенов (например, *BRAF*) и делеции онкосупрессоров (например, *TP53*). Тот факт, что амплификации и делеции происходят в равной пропорции, позволяет предположить, что такие изменения могут играть роль не только в активации онкогенов, но и в инактивации генов-супрессоров, что представляет собой важный аспект патогенеза КРР [19].

Мутационная нагрузка (ТМВ) в исследуемой когорте достаточно высокая (медианное значение ТМВ 5,5). Наблюдаемая высокая мутационная нагрузка может быть важным предсказателем клинического ответа на иммунотерапию, особенно с учетом того, что ТМВ выше 10 связано с улучшением прогноза при применении пембролизумаба [20, 21]. Однако, несмотря на высокую ТМВ, значения MSI и дефектов в системе репарации путем гомологичной рекомбинации были существенно низкими. Это может указывать на то, что в исследуемой когорте отсутствуют пациенты, которые могли бы получить пользу от терапии, направленной на дефекты репарации ДНК, такие как ингибиторы поли(АДФ-рибоза) полимеразы (PARP) [22, 23].

Профиль мутационных сигнатур во всех образцах очень схожий. Анализ мутационных сигнатур показал, что в исследуемой популяции преобладают мутационные сигнатуры SBS1, SBS5 и SBS39, что отражает возрастные изменения в клеточной ДНК, а также указывает на активность других, пока не уточненных, механизмов, связанных с сигнатурой SBS39 [24, 25]. Это открывает новые перспективы для дальнейших исследований в области молекулярной онкологии и подчеркивает важность детального анализа мутационных паттернов для выявления новых мишеней для терапии.

Заключение

Полученные результаты позволяют предположить, что мутационный процесс колоректального

рака имеет многофакторный характер. Результаты проведенного исследования демонстрируют молекулярное разнообразие и генетическую гетерогенность опухолей у пациентов с колоректальным раком, а также подтверждают значимость ключевых генов и молекулярных изменений, таких как мутации в *KRAS*, *BRAF*, *TP53*, а также вариации числа копий генов в патогенезе заболевания. Вы-

явленное многообразие изменений генома может быть связано с ошибками в репликации ДНК, воздействием внешних факторов и генетической предрасположенностью. Дальнейшие исследования помогут лучше понять эти механизмы и разработать новые подходы к диагностике, лечению и профилактике колоректального рака.

ЛИТЕРАТУРА/REFERENCES

1. Semenova A.B., Byakhova M.M., Makarova M.V., Galkin V.N., Nemtsova M.V., Chernevskiy D.K., Danishevich A.M., Shatalov V.G., Babkina A.V., Popova N.G., Gadzhieva S.M. Struktura patogennykh germinal'nykh variantov pri kolorektal'nom rake v vyborke patsientov Moskvyy [The structure of pathogenic germline variants in colorectal cancer in Moscow patients]. *Arkh Patol.* 2023; 85(6): 16–25. Russian. doi: 10.17116/patol20238506116.
2. Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018; 34(17): 884–90. doi: 10.1093/bioinformatics/bty560.
3. Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010; 26(5): 589–95. doi: 10.1093/bioinformatics/btp698.
4. Poplin R., Chang P.C., Alexander D., Schwartz S., Colthurst T., Ku A., Newburger D., Dijamco J., Nguyen N., Afshar P.T., Gross S.S., Dorfman L., McLean C.Y., DePristo M.A. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol.* 2018; 36(10): 983–87. doi: 10.1038/nbt.4235.
5. Kim S., Scheffler K., Halpern A.L., Bekritsky M.A., Noh E., Källberg M., Chen X., Kim Y., Beyter D., Krusche P., Saunders C.T. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods.* 2018; 15(8): 591–94. doi: 10.1038/s41592-018-0051-x.
6. Liu S.T.H., Lin H.M., Baine I., Wajnberg A., Gumprecht J.P., Rahman F., Rodriguez D., Tandon P., Bassily-Marcus A., Bander J., Sanky C., Dupper A., Zheng A., Nguyen F.T., Amanat F., Stadlbauer D., Altman D.R., Chen B.K., Krammer F., Mendu D.R., Firpo-Betancourt A., Levin M.A., Bagliella E., Casadevall A., Cordon-Cardo C., Jhang J.S., Arinsburg S.A., Reich D.L., Aberg J.A., Bouvier N.M. Convalescent plasma treatment of severe COVID-19: a propensity score-matched control study. *Nat Med.* 2020; 26(11): 1708–13. doi: 10.1038/s41591-020-1088-9.
7. McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., Thormann A., Flicek P., Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016; 17(1): 122. doi: 10.1186/s13059-016-0974-4.
8. Talevich E., Shain A.H., Botton D., Bastian B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol.* 2016; 12(4): e1004873. doi: 10.1371/journal.pcbi.1004873.
9. Andrews S. Babraham bioinformatics—FastQC a quality control tool for high throughput sequence data. [Internet]. [cited 18.12.2024]. URL: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
10. Wingett S.W., Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* 2018; 7: 1338. doi: 10.12688/f1000research.15931.2.
11. Pedersen B.S., Quinlan A.R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics.* 2018; 34(5): 867–68. doi: 10.1093/bioinformatics/btx699.
12. Jia P., Yang X., Guo L., Liu B., Lin J., Liang H., Sun J., Zhang C., Ye K. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. *Genomics Proteomics Bioinformatics.* 2020; 18(1): 65–71. doi: 10.1016/j.gpb.2020.02.001.
13. Yao H., Li H., Wang J., Wu T., Ning W., Diao K., Wu C., Wang G., Tao Z., Zhao X., Chen J., Sun X., Liu X.S. Copy number alteration features in pan-cancer homologous recombination deficiency prediction and biology. *Commun Biol.* 2023; 6(1): 527. doi: 10.1038/s42003-023-04901-3.
14. Díaz-Gay M., Vangara R., Barnes M., Wang X., Islam S.M.A., Vermes I., Narasimhan N.B., Yang T., Jiang Z., Moody S., Senkin S., Brennan P., Stratton M.R., Alexandrov L.B. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *bioRxiv: the preprint server for biology.* 2023 Jul 11: 2023.07.10.548264. doi: 10.1101/2023.07.10.548264.
15. FDA approves pembrolizumab for adults and children with TMB-H solid tumors. Food and Drug Administration. 2020. [Internet]. [cited 01.06.2025]. URL: <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-adults-and-children-tmb-h-solid-tumors>.
16. Yang Y., Zhang H., Huang S., Chu Q. KRAS Mutations in Solid Tumors: Characteristics, Current Therapeutic Strategy, and Potential Treatment Exploration. *J Clin Med.* 2023; 12(2): 709. doi: 10.3390/jcm12020709.
17. Lai H.F., Hang J.F., Kuo P.C., Kuo C.S., Yao S.F., Chen J.Y., Lee C.H. BRAF V600E Mutation Lacks Association with Poorer Clinical Prognosis in Papillary Thyroid Carcinoma. *Ann Surg Oncol.* 2024; 31(5): 3495–501. doi: 10.1245/s10434-024-14935-4.
18. Javed S., Benoist S., Devos P., Truant S., Guimbaud R., Lièvre A., Sefrioui D., Cohen R., Artru P., Dupré A., Bachet J.B., de la Fouchardière C., Ploquin A., Turpin A. Prognostic factors of BRAF V600E colorectal cancer with liver metastases: a retrospective multicentric study. *World J Surg Oncol.* 2022; 20(1): 131. doi: 10.1186/s12957-022-02594-2.
19. Li Q., Geng S., Luo H., Wang W., Mo Y.Q., Luo Q., Wang L., Song G.B., Sheng J.P., Xu B. Signaling pathways involved in colorectal cancer: pathogenesis and targeted therapy. *Signal Transduct Target Ther.* 2024; 9(1): 266. doi: 10.1038/s41392-024-01953-7.
20. Palmeri M., Mehnert J., Silk A.W., Jabbour S.K., Ganesan S., Popli P., Riedlinger G., Stephenson R., de Meritens A.B., Leiser A., Mayer T., Chan N., Spencer K., Girda E., Malhotra J., Chan T., Subbiah V., Groisberg R. Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. *ESMO Open.* 2022; 7(1): 100336. doi: 10.1016/j.esmoop.2021.100336.
21. Marcus L., Fashoyin-Aje L.A., Donoghue M., Yuan M., Rodriguez L., Gallagher P.S., Philip R., Ghosh S., Theoret M.R., Beaver J.A., Pazdur R., Lemery S.J. FDA Approval Summary: Pembrolizumab for the Treatment of Tumor Mutational Burden-High Solid Tumors. *Clin Cancer Res.* 2021; 27(17): 4685–89. doi: 10.1158/1078-0432.CCR-21-0327.
22. Wu F.H., Wei H.Z., Deng H.Y., Xiao G.H., Zhang Y.C. PARP in colorectal cancer: Molecular mechanisms, immunity, clinical trials, and drug combinations. *Neoplasma.* 2023; 70(1): 1–14. doi: 10.4149/neo.2022.220724N745.
23. Demuth P., Thibol L., Lemsch A., Potlitz F., Schulig L., Grathwol C., Manolikakes G., Schade D., Roukos V., Link A., Fahrer J. Targeting PARP-1 and DNA Damage Response Defects in Colorectal Cancer Chemotherapy with Established and Novel PARP Inhibitors. *Cancers (Basel).* 2024; 16(20): 3441. doi: 10.3390/cancers16203441.
24. Alexandrov L.B., Kim J., Haradhvala N.J., Huang M.N., Tian Ng A.W., Wu Y., Boot A., Covington K.R., Gordenin D.A., Bergstrom E.N., Islam S.M.A., Lopez-Bigas N., Klimczak L.J., McPherson J.R., Morganella S., Sabarinathan R., Wheeler D.A., Mustonen V.; PCAWG Mutational Signatures Working Group; Getz G., Rozen S.G., Stratton M.R.; PCAWG Consortium. The repertoire of mutational signatures in human cancer. *Nature.* 2020; 578(7793): 94–101. doi: 10.1038/s41586-020-1943-3. Erratum in: *Nature.* 2023; 614(7948): E41. doi: 10.1038/s41586-022-05600-5.
25. Spisak N., de Manuel M., Milligan W., Sella G., Przeworski M. The clock-like accumulation of germline and somatic mutations can arise from the interplay of DNA damage and repair. *PLoS Biol.* 2024; 22(6): e3002678. doi: 10.1371/journal.pbio.3002678.

Поступила/Received 20.12.2024

Одобрена после рецензирования/Revised 10.06.2025

Принята к публикации/Accepted 24.06.2025

СВЕДЕНИЯ ОБ АВТОРАХ

Семина Екатерина Владимировна, доктор биологических наук, профессор Образовательно-научного кластера, Институт медицины и наук о жизни (МЕДБИО), ФГАОУ ВО «Балтийский федеральный университет им. И. Канта» (г. Калининград, Россия); ведущий научный сотрудник лаборатории морфогенеза и репарации тканей университетской клиники, Московский

государственный университет имени М.В. Ломоносова (г. Москва, Россия). SPIN-код: 4586-4001. Researcher ID (WOS): AAB-4817-2020. Author ID (Scopus): 35081127300. ORCID: 0000-0002-3927-9286.

Кубасова Кристина Андреевна, магистр биологических наук, научный сотрудник, ООО «Биотек кампус» (г. Москва, Россия). ORCID: 0009-0009-3811-2618.

Какоткин Виктор Викторович, ассистент кафедры хирургических дисциплин, ФГАОУ ВО «Балтийский федеральный университет им. И. Канта» (г. Калининград, Россия). SPIN-код: 4162-0600. Researcher ID (WOS): AAJ-1657-2021. Author ID (Scopus): 57202677166. ORCID: 0000-0003-0352-2317.

Родимов Сергей Викторович, кандидат медицинских наук, онколог отделения абдоминальной и торакальной онкологии университетской клиники, Московский государственный университет имени М.В. Ломоносова (г. Москва, Россия). SPIN-код: 8908-5592. Researcher ID (WOS): LVS-1489-2024. Author ID (Scopus): 57195214771. ORCID: 0000-0002-1771-9252.

Агапов Михаил Андреевич, доктор медицинских наук, профессор кафедры хирургических дисциплин, ФГАОУ ВО «Балтийский федеральный университет им. И. Канта» (г. Калининград, Россия); главный научный сотрудник отдела хирургии и онкологии университетской клиники, Московский государственный университет имени М.В. Ломоносова (г. Москва, Россия). SPIN-код: 2887-6341. Researcher ID (WOS): AAJ-2250-2021. Author ID (Scopus): 25931366400. ORCID: 0000-0002-6569-7078.

ВКЛАД АВТОРОВ

Семина Екатерина Владимировна: концепция и дизайн исследования, обработка данных, написание текста.

Кубасова Кристина Андреевна: сбор и обработка данных, утверждение окончательного варианта рукописи.

Какоткин Виктор Викторович: сбор и обработка данных, написание текста статьи.

Родимов Сергей Викторович: написание текста статьи, редактирование.

Агапов Михаил Андреевич: концепция и дизайн исследования, редактирование текста статьи.

Все авторы одобрили финальную версию статьи перед публикацией, выразили согласие нести ответственность за все аспекты работы, подразумевающую надлежащее изучение и решение вопросов, связанных с точностью и добросовестностью любой части работы.

Финансирование

Исследование было поддержано из средств программы стратегического академического лидерства «Приоритет 2030» БФУ им. И. Канта.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Соответствие принципам этики

Проведенное исследование соответствует стандартам Хельсинкской декларации, одобрено независимым этическим комитетом Балтийского федерального университета им. И. Канта (Россия, 236041, г. Калининград, ул. Александра Невского, 14), протокол № 43 от 18.10.23.

Информированное согласие

Все пациенты подписали письменное информированное согласие на публикацию данных в медицинском журнале, включая его электронную версию.

ABOUT THE AUTHORS

Ekaterina V. Semina, DSc, Professor, Educational and Scientific Cluster, Institute of Medicine and Life Sciences (MEDBIO), Immanuel Kant Baltic Federal University (Kaliningrad, Russia); Leading Researcher, Laboratory of Tissue Morphogenesis and Reparation, University Clinic, Lomonosov Moscow State University (Moscow, Russia). Researcher ID (WOS): AAB-4817-2020. Author ID (Scopus): 35081127300. ORCID: 0000-0002-3927-9286.

Kristina A. Kubasova, Master of Biological Sciences, Researcher, Biotek Campus (Moscow, Russia). ORCID: 0009-0009-3811-2618.

Viktor V. Kakotkin, Assistant, Department of Surgical Disciplines, Immanuel Kant Baltic Federal University (Kaliningrad, Russia). Researcher ID (WOS): AAJ-1657-2021. Author ID (Scopus): 57202677166. ORCID: 0000-0003-0352-2317.

Sergey V. Rodimov, MD, PhD, Oncologist, Department of Abdominal and Thoracic Oncology, University Clinic, Lomonosov Moscow State University (Moscow, Russia). Researcher ID (WOS): LVS-1489-2024. Author ID (Scopus): 57195214771. ORCID: 0000-0002-1771-9252.

Mikhail A. Agapov, MD, DSc, Professor, Department of Surgical Disciplines, Immanuel Kant Baltic Federal University (Kaliningrad, Russia); Chief Researcher, Department of Surgery and Oncology, University Clinic, Lomonosov Moscow State University (Moscow, Russia). Researcher ID (WOS): AAJ-2250-2021. Author ID (Scopus): 25931366400. ORCID: 0000-0002-6569-7078.

AUTHOR CONTRIBUTIONS

Ekaterina V. Semina: concept and design of the study, data processing, writing the text.

Kristina A. Kubasova: data collection and processing, approval of the final version of the manuscript.

Viktor V. Kakotkin: data collection and processing, writing the text.

Sergey V. Rodimov: writing the text, editing.

Mikhail A. Agapov: concept and design of the study, editing the text.

All authors approved the final version of the manuscript prior to publication and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Funding

The reported study was funded by Immanuel Kant Baltic Federal University according to the Strategic Academic Leadership Program "Priority 2030".

Conflict of interests

The authors declare that they have no conflict of interest.

Compliance with Ethical Standards

The study was conducted in accordance with ethical principles outlined in the Declaration of Helsinki approved by Ethics Committee of Immanuel Kant Baltic Federal University (4, Alexander Nevsky St., Kaliningrad, 236041, Russia), protocol No. 43 dated October 18, 2023.

Voluntary informed consent

Written informed voluntaries consents were obtained from the patients for the publication of data in medical journal.

Приложение/Supplement

Список проанализированных онкогенов/

The list of oncogenes analyzed

ACVR1B, ACVR2A, AKT1, AMER1, APC, ARHGAP5, ARID1A, ATM, AXIN1, AXIN2, B2M, BAX, BCL9, BCL9L, BCOR, BRAF, CASP8, CD58, CDC27, CDKN1B, CHD2, CTNNB1, CUX1, EIF3E, ELF3, EP300, ERBB2, ERBB3, ERBB4, FBXW7, GRIN2A, HIF1A, HLA-B, KRAS, MAP2K1, MAP2K4, MAX, MCF2, MDM2, MLH1, MSH2, MSH6, MUTYH, NBEA, NRAS, PCBP1, PIK3CA, PIK3R1, PMS1, PMS2, POLE, PRDM2, PTPRK, PTPRT, QKI, RAD21, RBM10, RNF43, ROBO2, RPL22, RSPO2, RSPO3, SALL4, SFRP4, SMAD2, SMAD3, SMAD4, SOX9, SRC, STAG1, TBL1XR1, TBX3, TCF7L2, TGFB2, TP53, UBR5, ZBTB20, ZC3H13, ZNF14, ZNRF3.