

РАЗРАБОТКА И ВАЛИДАЦИЯ ПРЕДИКТИВНОЙ МОДЕЛИ ДЛЯ ПРЕДСКАЗАНИЯ ВЫЯВЛЕНИЯ ТАРГЕТИРУЕМЫХ АЛЬТЕРАЦИЙ У ПАЦИЕНТОК С РАСПРОСТРАНЕННЫМИ ФОРМАМИ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

П. С. Шило^{1,2}, М. Л. Макаркина³, А. А. Захаренко¹

¹ ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова» Министерства здравоохранения Российской Федерации, Санкт-Петербург

² ООО клиника «Лакhta»

³ ГБУЗ «Санкт-Петербургский клинический научно-практический центр специализированных видов медицинской помощи (онкологический) имени Н. П. Напалкова»

Цель. Разработать и валидировать предиктивную модель для прогнозирования вероятности выявления таргетируемых альтераций методом комплексного геномного профилирования (КГП) у больных метастатическим раком молочной железы (РМЖ).

Материалы и методы. Включены данные о 28 больных метастатическим РМЖ, которым проведено КГП методом таргетного секвенирования нового поколения (NGS) и сопоставимая когорта без выполненного КГП — 140 пациенток. Разработка предиктивной модели осуществлялась с использованием метода дерева решений с интеграцией априорных данных. Модель валидировалась с помощью *k-fold* перекрестной проверки, оценка качества осуществлялась с использованием AUC ROC. Для подтверждения стабильности результатов применялся бутстреп-анализ.

Результаты. Потенциально таргетируемые альтерации выявлены у 67,9 % пациенток, из которых 35,7 % соответствовали уровням ESCAT I/II. Модель продемонстрировала высокую точность предсказаний (AUC ROC = 0,78). Ключевыми предикторами выявления альтераций оказались фенотип РМЖ, количество пересмотров биоматериала и тип материала (операционный или биопсийный). Результаты бутстреп-анализа подтвердили стабильность модели.

Заключение. Разработанная предиктивная модель позволяет с высокой точностью прогнозировать вероятность выявления таргетируемых мутаций у больных метастатическим РМЖ. Использование модели способствует оптимизации молекулярного тестирования и повышению эффективности персонализированной терапии. Дальнейшие исследования с внешней валидацией необходимы для подтверждения результатов.

Ключевые слова: метастатический рак молочной железы, комплексное геномное профилирование, таргетируемые мутации, предиктивная модель, секвенирование нового поколения

DEVELOPMENT AND VALIDATION OF A PREDICTIVE MODEL FOR THE PREDICTION OF THE DETECTION OF TARGETABLE ALTERATIONS IN PATIENTS WITH ADVANCED TYPES OF BREAST CANCER

P. S. Shilo^{1,2}, M. L. Makarkina³, A. A. Zakharenko¹

¹ Federal State Budgetary Educational Institution of Higher Education "Academician I. P. Pavlov First Saint-Petersburg State Medical University" of the Ministry of Healthcare of the Russian Federation

² Limited Liability Company (LLC) Clinic "Lakhta"

³ State Budgetary Healthcare Institution "Saint — Petersburg Clinical Scientific Practical Center of Specialized Types of Medical Care (Oncological) named after N. P. Napalkov"

Objective of the study is to develop and validate a predictive model for the prognosis of the probability of detection of targetable alterations using comprehensive genomic profiling (CGP) in patients with metastatic breast cancer (BC).

Materials and Methods. Data of 28 patients with metastatic breast cancer, who had undergone comprehensive genomic profiling (CGP) using targeted next generation sequencing (NGS) and of a comparison cohort without comprehensive genomic profiling (CGP) — 140 patients were included in the model. The predictive model was developed using decision tree approach with integration of a-priori data. The model was validated using k-fold cross-validation, and the quality assessment was carried out using AUC ROC. Bootstrap analysis was performed to confirm the stability of the results.

Results. Potentially targetable alterations were revealed in 67,9% of patients, of whom 35,7% corresponded to ESCAT I/II levels. The model demonstrated high prediction accuracy (AUC ROC = 0,78). The phenotype of breast cancer, number of revisions of biomaterial and the type of material (operative or biopsy) were the key predictors for the detection of alterations. The results of bootstrap analysis confirmed the stability of the model.

Conclusion. The developed predictive model allows for highly accurate prediction of the probability of detecting targetable mutations in patients with metastatic breast cancer. The use of the model contributed to the optimization of molecular testing and enhances the effectiveness of personalized therapy. Further studies with external validation are required to confirm the results.

Keywords: metastatic breast cancer, comprehensive genomic profiling (CGP), targetable mutations, predictive model, next generation sequencing

Введение

Прецизионная онкология стремительно развивается, предоставляя новые возможности для персонализированного лечения онкологических заболеваний. Одним из ключевых аспектов этой области является использование комплексного геномного профилирования (КГП) опухолей, которое позволяет выявлять биологические события, ассоциированные с эффективностью специфических лекарственных препаратов [1, 2].

КГП, основанное на секвенировании нового поколения (NGS), позволяет выявлять широкий спектр генетических альтераций, включая мутации, амплификации и транслокации, что способствует оптимизации выбора таргетной терапии [3, 4]. Среди наиболее часто встречающихся генетических изменений при раке молочной железы выделяют мутации в генах *PIK3CA*, *ESR1*, *ERBB2*, *BRCA1/2*, которые играют ключевую роль в патогенезе заболевания и определяют чувствительность к терапии [5, 6].

Несмотря на прогресс, внедрение молекулярного профилирования в клиническую практику сопряжено с рядом вызовов, включая ограниченную доступность методов NGS, высокую стоимость тестирования и сложность интерпретации данных [7, 8]. Для оптимизации использования этих технологий необходимы предиктивные модели, которые позволяют с высокой точностью прогнозировать вероятность выявления таргетируемых альтераций на основании клинических и биологических параметров [9].

Целью данной работы является разработка и валидация предиктивной модели, которая позволяет предсказывать вероятность выявления дополнительных таргетируемых альтераций методом КГП больным метастатическим РМЖ.

Пациенты и методы

Дизайн исследования

В одноцентровое ретроспективное когортное исследование были включены клинические данные о 28 больных метастатическим раком молочной железы, которым было выполнено КГП на базе клиники «Лахта» (ранее — «Луч») и сопоставимая когорта без выполненного КГП — 140 пациенток. Критерии включения в исследование: подтвержденный диагноз метастатического РМЖ, наличие адекватного биоматериала для выполнения секвенирования для группы КГП, а также согласие пациенток на использование данных в научных целях. Решение вопроса о назначении молекулярно-направленной терапии (МНТ) проводилось в рамках междисциплинарного онкологического консилиума.

Биоматериал был представлен как операционными образцами, так и биопсийным материалом. Количество пересмотров биоматериала варьировалось от 0 до 3, что также учитывалось в качестве предиктора в модели.

Комплексное геномное профилирование

Секвенирование нового поколения (NGS) выполнялось с использованием коммерческих панелей не менее 300 генов. Объем тестирования

приведен в рисунках. Успешность выполнения секвенирования оценивалась по наличию информативных данных, а также по количеству целевых участков, покрытых на глубине > 500х. Интерпретация обнаруженных биологических событий проводилась с применением шкалы ESCAT.

Построение и валидация модели

Модель предикции в данной работе были построена с использованием метода дерева решений, который позволяет учитывать нелинейные зависимости между переменными и позволяет интегрировать априорные вероятности.

Для минимизации переобучения применялась перекрестная проверка (k-fold cross-validation). Валидация модели проводилась с использованием метрик AUC ROC (площадь под кривой «приёмник — оператор»).

Учитывая небольшой объем выборки, для оценки стабильности модели использовался бутстреп-анализ. Этот метод включал многократную случайную генерацию подвыборок из исходного набора данных с последующим пересчетом метрик модели. Это позволило оценить надежность результатов и уменьшить влияние случайных факторов на финальные выводы.

Источники априорных вероятностей

Для расчета априорных вероятностей наличия мутаций использовались данные опубликованных исследований и баз данных (TCGA), включающих крупные когорты больных раком молочной железы. Основными источниками служили работы, описывающие частоту мутаций в генах *PIK3CA*, *ESR1*, *ERBB2*, *BRCA1/2*, *TP53* и других в зависимости от подтипов РМЖ и клинических характеристик пациентов [3, 5, 7].

Результаты

Описание когорты

Средний возраст 28 больных метастатическим РМЖ составил 54 года (диапазон: 38–68 лет). Распределение фенотипов следующее: люминальный HER2-негативный — 50 %, люминальный HER2-положительный — 21 %,

HER2-положительный HR-негативный — 14 %, тройной негативный — 14 %. Для сравнения результатов лечения в группе наличия или отсутствия выполненного КГП была подобрана когорта пациенток, сопоставимая по базовым клиническим характеристикам. Более подробно базовые характеристики пациенток описаны в таблице 1.

Результаты КГП

и молекулярно-направленной терапии

В исследуемой выборке потенциально таргетируемые альтерации обнаружены у 19 (67,9 %) пациенток, при этом у 10 из них (35,7 %) соответствовали уровню ESCAT I/II. Общее количество выявленных таргетируемых альтераций ESCAT I/II составило 12, при этом наиболее часто выявлялись биомаркеры ESCAT III/IV. Распределение выявленных альтераций по шкале ESCAT отображено в таблице 2.

В когорте рака молочной железы генами с наибольшей частотой альтераций оказались гены *TP53* ($n = 14$, 50,0 %), *PIK3CA* ($n = 8$, 28,6 %).

МНТ была назначена четырем пациенткам. Длительность ответа на МНТ описана в рисунке 1. В двух случаях в качестве максимального эффекта был зарегистрирован частичный ответ, в одном случае — стабилизация, и еще в одном случае — прогрессирование. Клинические характеристики пациенток, получивших МНТ, описаны в таблице 3.

BLM — ген, кодирующий белок из семейства RecQ-хеликаз. *PIK3CA* — ген, кодирующий альфа-субъединицу фосфатидилинозитид-3-киназы

Для оценки клинической пользы от КГП в качестве контроля была подобрана ретроспективная контрольная выборка пациенток с распространенными в соотношении 1:5. Подбор пациенток осуществлялся с сопоставлением по известным характеристикам, которые могут повлиять на результаты лечения:

- фенотип РМЖ;
- ECOG статус;
- возраст;
- количество предшествующих линий терапии;

Таблица 1

Базовые характеристики пациентов, включенных в исследование

		Группа КГП, n = 28	Группа без КГП, n = 140
Возраст, М (МКР)		52 (45–62)	54 (48–67)
Фенотип РМЖ, количество (%)	Люминальный HER2-негативный	14 (50 %)	70 (50 %)
	Люминальный HER2-позитивный	6 (21 %)	35 (25 %)
	Нелюминальный HER2-позитивный	4 (14 %)	21 (15 %)
	Трижды негативный	4 (14 %)	14 (10 %)
Количество линий терапии до выполнения комплексного геномного профилирования, количество (%)	0–2 линии терапии	12 (43 %)	64 (46 %)
	Три и более линии терапии	16 (57 %)	76 (54 %)
Длительность ответа на предшествующую линию терапии, М (МКР)		9 (6–13)	8 (5–12)
ECOG статус на момент выполнения комплексного геномного профилирования	0	10 (36 %)	48 (34 %)
	1	12 (43 %)	56 (40 %)
	2	4 (14 %)	20 (14 %)
	3	2 (7 %)	16 (11 %)
Название диагностического теста	Atlas Solo	16 (57 %)	–
	FoundationOne	12 (42 %)	–
Год выполнения диагностического теста	2020	5 (18 %)	–
	2021	6 (21 %)	–
	2022	7 (25 %)	–
	2023	10 (36 %)	–

КГП — комплексное геномное профилирование. МКР — межквартильный размах. РМЖ — рак молочной железы. HER2 — рецептор эпидермального фактора роста человека 2-го типа.

• длительность ответа на предыдущую линию терапии.

Соотношение ВБП2/ВБП1 > 1,3 в группе пациенток с выполненным КГП оказалось статистически значимо более высоким в сравнении с пациентками без выполненного КГП (25 % vs 15 %, $p = 0,039$).

Медианы ВБП в обеих группах составили девять недель, статистически значимых различий обнаружено не было ($p = 0,282$).

Методом деревьев решений с интеграцией априорных данных была разработана предиктивная модель для предсказания вероятности выявления ESCAT I/II биомаркеров. В качестве предикторов в модель были включены характеристики биоматериала, которые могут влиять на вероятность успешного выполнения секвенирования, клинические характеристики пациенток, наследственный анамнез и информация о ранее выявленных альтерациях.

Таблица 2

Распределение обнаруженных альтераций по шкале ESCAT (ESCAT — шкала клинической значимости геномных альтераций)

		Количество	%
Распределение альтераций и препаратов по шкале ESCAT	I	7	14,9
	II	5	10,6
	III	16	34,0
	IV	19	40,4

Клиническая характеристика пациентов, получивших молекулярно-направленную терапию

№	Диагноз, группа	Краткая клиническая характеристика	Обнаруженная альтерация	Назначенный препарат	Максимальный эффект терапии
1	Рак молочной железы	Женщина, 39 лет Трижды негативный рак молочной железы (метапластическая карцинома), состояние после восьми линий лекарственной терапии	PIK3CA	Алпелисиб	Частичный регресс
2	Рак молочной железы	Женщина, 68 лет, люминальный HER2-негативный РМЖ, состояние после двух линий гормонотерапии (тамоксифен, анастрозол + рибоциклиб)	PIK3CA	Алпелисиб	Частичный регресс
3	Рак молочной железы	Женщина, 35 лет Трижды негативный РМЖ, состояние после пяти линий лекарственной терапии	BLM	Олапариб	Стабилизация
4	Рак молочной железы	Женщина, 27 лет Трижды негативный рак молочной железы, состояние после четырех линий лекарственной терапии	Высокая мутационная нагрузка (12 мут/мб)	Пембролизумаб	Прогрессирование

Длительность ответа на МНТ

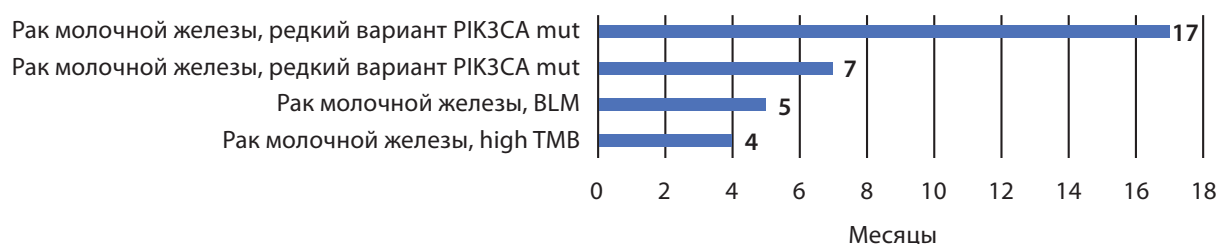


Рис. 1. Сравнение результатов лечения у пациенток с выполненным КГП и с отсутствием выполнения КГП. МНТ — молекулярно-направленная терапия. PIK3CA mut — мутация в гене PIK3CA. BLM — ген, кодирующий белок из семейства RecQ-хеликаз. High TMB — высокая мутационная нагрузка

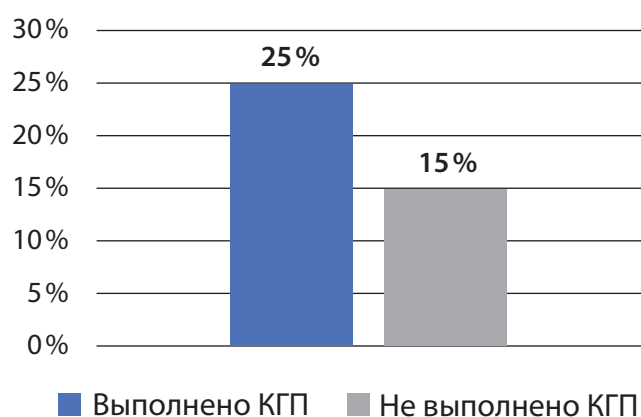


Рис. 2. Процент пациенток с ВБП2/ВБП1 > 1,3 в группе пациенток с выполненным и невыполненным КГП. КГП — комплексное геномное профилирование

В качестве предикторов использовались:

- фенотип РМЖ:
 - люминальный HER2-негативный,
 - люминальный HER2-положительный,
 - HER2-положительный HR-негативный,
 - тройной негативный;
- возраст пациенток (как количественная переменная);
- характеристика биоматериала:
 - количество пересмотров биоматериала (0, 1 ≥ 2),
 - тип биоматериала (операционный или биопсийный);
- выявленные ранее альтерации.

Коэффициенты предиктивной модели представлены в таблице 4.

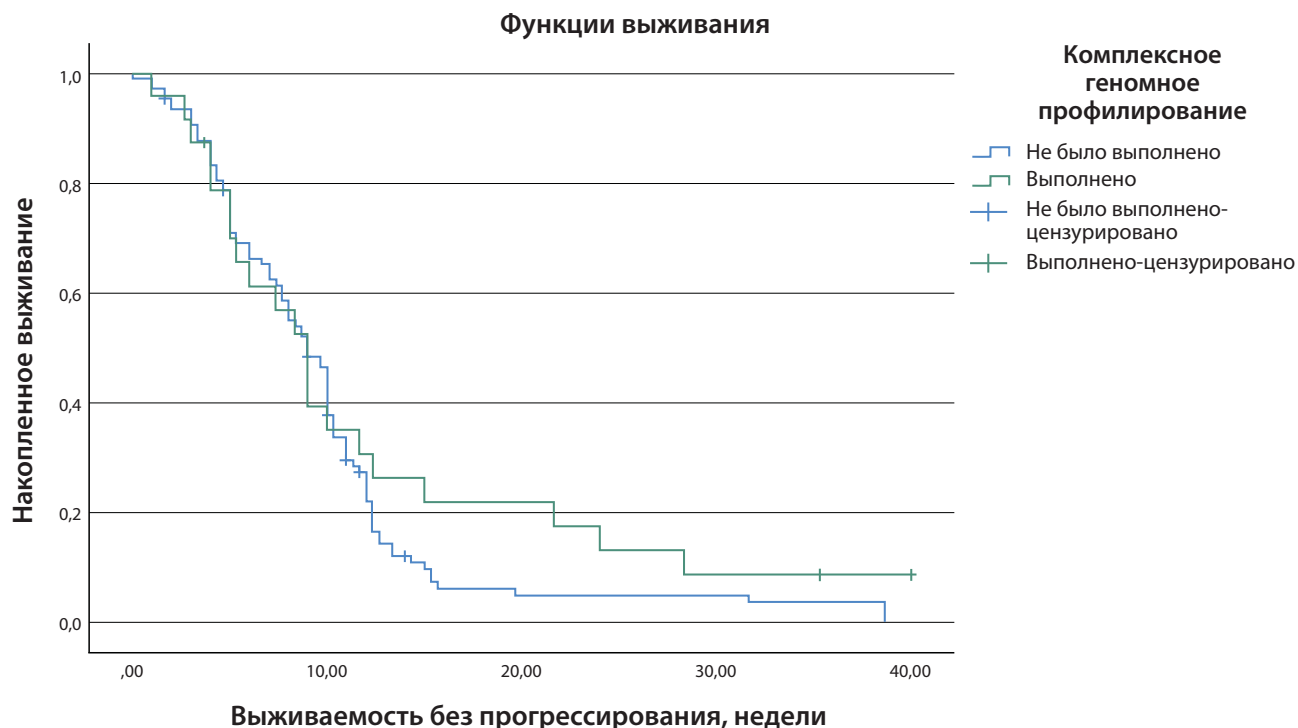


Рис. 3. Сравнительный анализ ВБП в группах пациенток с выполненным КГП и без выполнения КГП. Ось абсцисс отражает выживаемость без прогрессирования в неделях. Ось ординат отражает накопленное выживание в виде статистической величины. Предиктивная модель для предсказания вероятности выявления ESCAT I/II биомаркеров

Таблица 4

Коэффициенты дерева решений для итоговой предиктивной модели [9]

Параметр	ОШ (95 % ДИ)	p-value	Коэффициент для итоговой модели
Первый узел — характеристики биоматериала (фиксированные значения для всех нозологий)			
Доступность только биопсийного материала	0,74 (0,58–0,94)	p = 0,012	0,821
> 2 предшествующих линии терапии	0,63 (0,47–0,85)	p = 0,003	0,774
> 2 дополнительных пересмотров биоматериала	0,81 (0,66–0,98)	p = 0,028	0,867
Второй узел — клинические параметры пациента			
Более молодой возраст	2,45 (1,78–3,38)	p < 0,001	1,205
Третий узел — морфологическая форма рака молочной железы			
Трижды негативный фенотип	1,89 (1,32–2,71)	p = 0,002	1,132
Люминальный HER2-негативный фенотип	2,08 (1,50–2,89)	p = 0,001	1,198
Четвертый узел — ранее обнаруженные альтерации			
Патогенный вариант BRCA1/2	0,56 (0,39–0,80)	p = 0,002	0,632
Другие альтерации в генах гомологичной рекомбинации	0,69 (0,52–0,91)	p = 0,009	0,713
Патогенный вариант PIK3CA	0,84 (0,70–0,99)	p = 0,043	0,945

HER2 — рецептор эпидермального фактора роста человека 2-го типа. TKI — ингибиторы тирозинкиназы. BRCA1/2 — гены, связанные с репарацией ДНК и подавлением опухолевого роста. PIK3CA — ген, кодирующий альфа-субъединицу фосфатидилинозитид-3-киназы

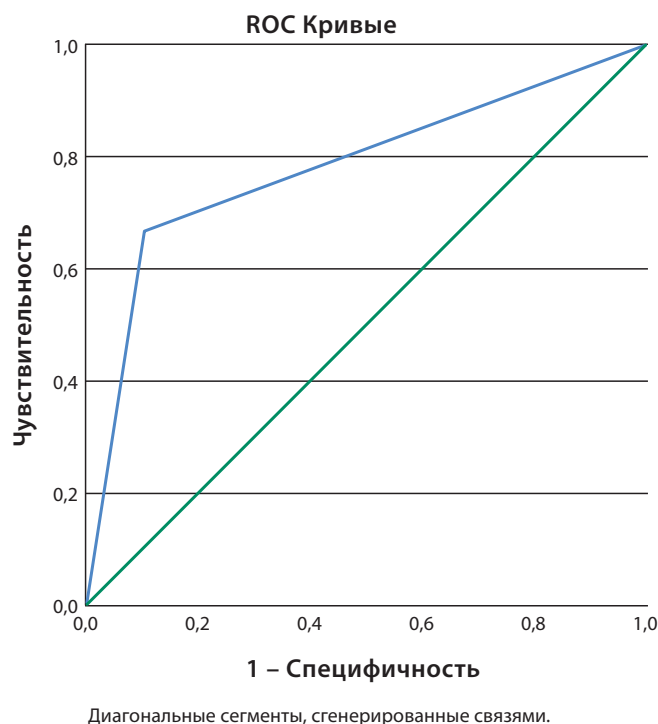


Рис. 4. AUC ROC анализ бинарного классификатора для выявления альтераций ESCAT I/II

Оценка качества модели

Точность бинарной классификации, измеренная с помощью AUC ROC, составила 0,78 (95 % ДИ 0,59–0,99), $p = 0,018$, что указывает на высокую способность различать пациенток с наличием и отсутствием таргетируемых мутаций. R2 Нейджелкерка для полученной модели оказался равен 0,68.

Заключение

Разработанная предиктивная модель, основанная на методе дерева решений, продемонстрировала высокую точность в прогнозировании наличия таргетируемых мутаций у больных метастатическим раком молочной железы.

Использование априорных данных о частоте мутаций и включение клинических предикторов позволило добиться AUC ROC = 0,89 для общей модели. Это подчеркивает значимость модели в контексте оптимизации молекулярного тестирования, особенно в условиях ограниченных ресурсов.

Ограничениями данного исследования являются ретроспективный наблюдательный дизайн исследования, небольшой объем выборки, применение только внутренней валидации предиктивной модели. Дальнейшие крупные исследования с валидацией на внешних выборках позволят дополнительно оценить потенциал применения данной модели.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Tredan O., Corset V., Wang Q. Routine molecular screening of advanced refractory cancer patients: An analysis of the first 2490 patients of the ProFiLER study. // *Journal of Clinical Oncology*. — 2017. — Vol. 35, N 18. DOI: 10.1200/JCO.2017.35.15_suppl.LBA100
2. Tsimberidou A. M. Initiative for Molecular Profiling and Advanced Cancer Therapy and challenges in the implementation of precision medicine // *Current Problems in Cancer*. — 2017. — Vol. 41, N 3. — P. 176–181. DOI: 10.1016/j.crrprobcan.2017.02.002
3. Tsimberidou A. M., Hong D. S., Wheeler J. J. Precision medicine: Clinical outcomes including long-term survival according to the pathway targeted and treatment period — The IMPACT study. // *Journal of Clinical Oncology*. — 2018. — Vol. 36, N 18. DOI: 10.1200/jco.2018.36.18_suppl.lba2553

4. Vogelstein B., Papadopoulos N., Velculescu V. E. et al. Cancer genome landscapes // *Science*. — 2013. — Vol. 340, N 6127. — P. 1546–1558. DOI: 10.1126/science.1235122
5. Kumar-Sinha C., Chinnaiyan Arul. M Precision oncology in the age of integrative genomics // *Physiology & behavior*. — 2018. — Vol. 176, N 5. — P. 139–148. DOI: 10.1038/nbt.4017
6. Reitsma M., Fox J., Borre P. Vanden Effect of a Collaboration Between a Health Plan, Oncology Practice, and Comprehensive Genomic Profiling Company from the Payer Perspective // *Journal of Managed Care & Specialty Pharmacy*. — 2019. — Vol. 25, N 5. DOI: 10.18553/jmcp.2019.18309
7. Normanno N., Luca A. De, Abate R. E. et al. Current practice of genomic profiling of patients with advanced solid tumours in Italy: the Italian Register of Actionable Mutations (RATIONAL) study // *European Journal of Cancer*. — 2023. — Vol. 187. — P. 174–184. DOI: 10.1016/j.ejca.2023.03.027
8. Xue L., Guo C., Zhang K. et al. Comprehensive molecular profiling of extrahepatic cholangiocarcinoma in Chinese population and potential targets for clinical practice // *Hepatobiliary Surgery and Nutrition*. — 2019. — Vol. 8, N 6. — P. 615–622. DOI: 10.21037/hbsn.2019.08.05
9. Mosele M. F., Westphalen C. B., Stenzinger A. et al. Recommendations for the use of next-generation sequencing (NGS) for patients with advanced cancer in 2024: a report from the ESMO Precision Medicine Working Group // *Annals of Oncology*. — 2024. — Vol. 35, N 7. — P. 588–606. DOI: 10.1016/j.annonc.2024.04.005.

СВЕДЕНИЯ ОБ АВТОРАХ

Шило Полина Сергеевна, врач-онколог, Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, 197022 Россия, Санкт-Петербург; клиника «Лахта», 197183 Россия, Санкт-Петербург, ул. Савушкина, д. 73/50, e-mail: polinashilo0@gmail.com

Shilo Polina S., *Oncologist*, Pavlov First Saint Petersburg State Medical University, 6–8 Lva Tolstogo St., 197022, Saint Petersburg, Russia; Lahta Clinic, 73/50 Savushkina St., 197183, Saint Petersburg, Russia, e-mail: polinashilo0@gmail.com

Макаркина Мария Леонидовна, кандидат медицинских наук, врач-онколог, Санкт-Петербургский клинический научно-практический центр специализированных видов медицинской помощи (онкологический) имени Н. П. Напалкова, пос. Песочный, Ленинградская ул., д. 68А, лит. А, 197758 Россия, Санкт-Петербург, e-mail: stepanova100992@mail.ru

Makarkina Maria L., Ph.D. in Medical Sciences, Oncologist, Saint Petersburg Clinical Scientific and Practical Center for Specialized Types of Medical Care (Oncological) named after N. P. Napalkov, 68A Leningradskaya St., pos. Pesochny, 197758, Saint Petersburg, Russia, e-mail: stepanova100992@mail.ru

Захаренко Александр Анатольевич, доктор медицинских наук, профессор, заведующий кафедрой онкологии факультета последипломного обучения, Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, 197022 Россия, Санкт-Петербург, e-mail: 9516183@mail.ru

Zakharenko Alexander A., M.D., Ph.D. in Medical Sciences, Professor, Head of the Department of Oncology, Faculty of Postgraduate Studies, Pavlov First Saint Petersburg State Medical University, 6–8 Lva Tolstogo St., 197022, Saint Petersburg, Russia, e-mail: 9516183@mail.ru