

УДК 618.19-006.6:577.217

<https://doi.org/10.20538/1682-0363-2026-2-91-100>

Возможность использования метода «случайного леса» для поиска взаимосвязи микроРНК с тканевыми и молекулярно-генетическими маркерами рака молочной железы

Студеникина А.А.^{1,2}, Михайловский М.А.³, Тимофеев В.С.³, Архипов С.А.^{1,2},
Аутеншлюс А.И.^{1,2}

¹ Новосибирский государственный медицинский университет (НГМУ)
Россия, 630091, г. Новосибирск, Красный пр., 52

² Научно-исследовательский институт молекулярной биологии и биофизики (НИИМББ), Федеральный исследовательский центр фундаментальной и трансляционной медицины (ФИЦ ФТМ)
Россия, 630060, г. Новосибирск, ул. Тимакова, 2

³ Новосибирский государственный технический университет (НГТУ)
Россия, 630073, г. Новосибирск, пр. Карла Маркса, 20

РЕЗЮМЕ

Цель. Поиск возможности выявления с помощью метода «случайного леса» (МСЛ) взаимосвязи между уровнем микроРНК miR-181a, miR-25 и экспрессией маркеров: Е-кадгерина (CDH1), коллагена II (CII), интегрина бета-1 (ITGB1), циклина D1 (CCND1) и белка-костимулятора Т-лимфоцитов (CD86), а также их ассоциации с клинически значимыми молекулярно-генетическими маркерами (ER, PR, HER2, Ki-67) у пациенток с раком молочной железы (РМЖ).

Материалы и методы. У 44 пациенток с инвазивной протоковой карциномой молочной железы (без неоадъювантной терапии) определяли экспрессию микроРНК miR-181a и miR-25 в сыворотке крови, используя цифровую капельную полимеразную цепную реакцию (ПЦР). Экспрессию CDH1, CII, ITGB1, CCND1, CD86, ER, PR, HER2 и Ki-67 анализировали с помощью иммуногистохимии в образцах опухолевой ткани. Для анализа взаимосвязей использовали корреляционный анализ Спирмена и алгоритм МСЛ.

Результаты. Корреляционный анализ не выявил статистически значимых линейных взаимосвязей между исследуемыми параметрами ($p > 0,05$). Метод «случайного леса» показал наличие нелинейных ассоциаций: для miR-181a оптимальная прогностическая модель включала CDH1 и CCND1 (MAE = 4,267), для miR-25 – ITGB1 и CCND1 (MAE = 16,255). Из клинико-патологических характеристик статистически значимые положительные корреляции выявлены для CII с ER ($r = 0,498$; $p = 0,001$) и PR ($r = 0,354$; $p = 0,018$). Модели МСЛ продемонстрировали наибольшую точность предсказания HER2-статуса (82,6%) при использовании комбинации ITGB1, miR-181a и CDH1.

Заключение. Алгоритм МСЛ выявил сложные нелинейные взаимосвязи между miR-181a, miR-25 и тканевыми маркерами опухолевой прогрессии, которые ранее не обнаружены стандартными статистическими методами. Полученные результаты носят поисковый характер и формируют основу для дальнейших исследований по валидации данных биомаркеров и их роли в патогенезе различных подтипов РМЖ.

Ключевые слова: рак молочной железы, HER2, микроРНК, машинное обучение, метод «случайного леса», молекулярные маркеры, эпителиально-мезенхимальный переход

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Источник финансирования. Исследование выполнено в рамках темы государственного задания ФГБОУ ВО «Новосибирский государственный медицинский университет» Министерства здравоохранения Российской Федерации.

✉ Студеникина Анастасия Александровна, aastudenikina@frcftm.ru

ской Федерации (регистрационный номер 115060410035) и в рамках темы НИР ФГБНУ «Федеральный исследовательский центр фундаментальной и трансляционной медицины», Научно-исследовательский институт молекулярной биологии и биофизики (регистрационный номер 125031203556-7).

Соответствие принципам этики. Все пациенты подписали информированное согласие на участие в исследовании. Исследование одобрено этическим комитетом НИИМББ, НИЦ АТМ (протокол № 28 от 27.09.2023).

Для цитирования: Студеникина А.А., Михайловский М.А., Тимофеев В.С., Архипов С.А., Аутеншлюс А.И. Возможность использования метода «случайного леса» для поиска взаимосвязи микроРНК с тканевыми и молекулярно-генетическими маркерами рака молочной железы. *Бюллетень сибирской медицины*. 2026;25(2):91–100. <https://doi.org/10.20538/1682-0363-2026-2-91-100>.

The potential application of the random forest method to explore the association between microRNAs and tissue and molecular genetic biomarkers of breast cancer

Studenikina A.A.^{1,2}, Mihailovsky M.A.³, Timofeev V.S.³, Arkhipov S.A.^{1,2}, Autenshlyus A.I.^{1,2}

¹ Novosibirsk State Medical University

52 Krasny Ave., 630091 Novosibirsk, Russian Federation

² Institute of Molecular Biology and Biophysics, Federal Research Center of Fundamental and Translational Medicine

2 Timakova St., 630060 Novosibirsk, Russian Federation

³ Novosibirsk State Technical University

20 Karl Marx Ave., 630073 Novosibirsk, Russian Federation

ABSTRACT

Aim. Exploring the potential of using a random forest method (RFM) to identify associations between serum levels of microRNAs miR-181a and miR-25 and the expression of the following markers: E-cadherin (CDH1), type II collagen (CII), integrin beta-1 (ITGB1), cyclin D1 (CCND1), and T-lymphocyte costimulator (CD86), as well as their relationship with clinically relevant molecular genetic markers (ER, PR, HER2, Ki-67) in patients with breast cancer (BC).

Materials and methods. We recruited 44 BC patients with invasive ductal carcinoma who had not received neoadjuvant treatment. Serum levels of miR-181a and miR-25 were measured using digital droplet PCR, and the expression of CDH1, CII, ITGB1, CCND1, CD86, ER, PR, HER2, and Ki-67 in tumor tissue samples was assessed via immunohistochemistry. Spearman's correlation analysis and the RFM were applied to assess associations between variables.

Results. Spearman's correlation analysis did not reveal any significant linear relationships between the studied parameters ($p > 0.05$). However, the RFM identified nonlinear associations: for miR-181a, the optimal predictive model included CDH1 and CCND1 (MAE = 4.267); for miR-25, ITGB1 and CCND1 (MAE = 16.255). Among clinical and pathological characteristics, statistically significant positive correlations were found between CII and ER ($r = 0.498$; $p = 0.001$) and PR ($r = 0.354$; $p = 0.018$). The RFM models demonstrated the highest accuracy in predicting the HER2 status (82.6%) when using a combination of ITGB1, miR-181a, and CDH1.

Conclusion. The RFM revealed complex nonlinear associations between microRNAs and tissue markers of tumor progression that were not detected by conventional statistical methods. These findings are exploratory in nature and provide a foundation for further research aimed at validating these biomarkers and elucidating their role in the pathogenesis of various breast cancer subtypes.

Keywords: breast cancer, HER2, microRNA, machine learning, random forest, molecular markers, epithelial – mesenchymal transition

Conflict of interest. The authors declare the absence of obvious or potential conflict of interest related to the publication of this article.

Source of financing. The study was conducted within the state assignment of Novosibirsk State Medical University, Ministry of Health of the Russian Federation (Registration No. 115060410035), and the research project of the Federal Research Center for Fundamental and Translational Medicine, Research Institute of Molecular Biology and Biophysics (Registration No. 125031203556-7).

Conformity with the principles of ethics. All patients signed an informed voluntary consent to participate in the study. The study was approved by the Ethics Committee at the Research Institute of Molecular Biology and Biophysics of the Federal Research Center for Fundamental and Translational Medicine (Minutes No. 28 dated September 27, 2023).

For citation: Studenikina A.A., Mihailovsky M.A., Timofeev V.S., Arkhipov S.A., Autenshlyus A.I. The potential application of the random forest method to explore the association between microRNAs and tissue and molecular genetic biomarkers of breast cancer. *Bulletin of Siberian Medicine*. 2026;25(2):91–100. <https://doi.org/10.20538/1682-0363-2026-2-91-100>.

ВВЕДЕНИЕ

Рак молочной железы (РМЖ) ежегодно уносит жизни тысяч женщин, что подчеркивает необходимость углубленного изучения молекулярных механизмов патогенеза этого заболевания. Особое внимание следует уделить анализу связи потенциальных маркеров с классическими молекулярно-генетическими показателями: рецепторами эстрогена (ER), прогестерона (PR), рецептором эпидермального фактора роста человека 2-го типа (HER2) и индексом пролиферации Ki-67.

МикроРНК (miR) – небольшие некодирующие РНК, регулирующие экспрессию генов на посттранскрипционном уровне [1]. Они функционируют как онкогены и как супрессоры опухолевого роста, влияя на стабильность и трансляцию мРНК. Нарушение регуляции микроРНК связано с различными видами рака [2]. В частности, miR-181a и miR-25 вовлечены в процессы пролиферации, инвазии и модуляции иммунного ответа [3, 4]. Известно, что miR-181a может снижать экспрессию ER в клеточной культуре РМЖ, а ее высокий уровень коррелирует с низкой экспрессией PR, гиперэкспрессией HER2 и высоким показателем Ki-67 [5–7]. MiR-25, в свою очередь, участвует в модуляции сигнальных путей ER и PR [5].

Исследование J. Feng и соавт. подтверждает, что аномальная экспрессия микроРНК сопровождается развитием клеточной трансформации в направлении канцерогенеза, что также связано со способностью микроРНК способствовать эпителиально-мезенхимальному переходу (ЭМП) [8]. Этот процесс характеризуется снижением экспрессии эпителиальных маркеров, таких как E-кадгерин (CDH1), и активацией мезенхимальных маркеров, включая интегрин бета-1 (ITGB1) [9]. В ходе этого процесса опухолевые клетки теряют межклеточные связи и приобретают повышенную подвижность и инвазивность, что сопряжено с более высокой степенью злокаче-

ственности и метастазированием в лимфатические узлы [10]. Согласно данным А.Н. Али и соавт., низкая экспрессия CDH1 была характерна для пациентов с негативным статусом ER и HER2 [11].

Ранее нами было показано, что у пациенток с РМЖ повышены уровни miR-181a, miR-25, CD86, ITGB1 и циклина D1 (CCND1) наряду со сниженной экспрессией CDH1 по сравнению с пациентами с доброкачественными заболеваниями молочной железы [12–14]. Однако статистически значимых взаимосвязей между этими биомаркерами не выявлено, что указывает на необходимость использования более сложного аналитического инструментария. Следует подчеркнуть, что каждый из этих биомаркеров играет важную роль в патогенезе РМЖ. Циклин D1 является ключевым регулятором клеточного цикла и способен повышать транскрипционную активность ER, способствуя формированию устойчивости к эндокринной терапии [15]. CD86 представляет собой костимуляторную молекулу, участвующую в активации Т-лимфоцитов и формировании противоопухолевого иммунного ответа [16], а коллаген II типа (CII), являясь основным компонентом хрящевого матрикса, при РМЖ может участвовать в ремоделировании внеклеточного матрикса и создании метастатических ниш [17]. Несмотря на то, что CCND1, CD86 и CII не являются классическими индукторами ЭМП, их влияние на ключевые сигнальные пути, такие как TGF- β , Wnt/ β -катенин, и интегрин-зависимую передачу сигналов позволяет предположить их косвенную роль в прогрессировании ЭМП [17, 18].

Алгоритмы машинного обучения предлагают объективные и воспроизводимые средства для выявления скрытых закономерностей и взаимосвязей биомаркеров, которые могут быть упущены из виду при традиционном анализе, что является перспективным направлением изучения молекулярных механизмов развития и прогрессирования онкологических

заболеваний [19]. МСЛ представляет собой алгоритм машинного обучения, основанный на построении множества деревьев решений, каждое из которых строится на своем уникальном подмножестве данных и признаков (бутстрэп-выборка), а затем – усреднении их предсказаний. Это позволяет значительно снизить дисперсию прогноза и склонность к переобучению по сравнению с одиночными деревьями [20].

На первом этапе формируется обучающая выборка: $D = \{(X_i, y_i)\}_{i=1}^N$, где N – количество объектов, X_i – набор из k признаков для i -го наблюдения, y_i – значение целевой переменной. Затем производится поиск функции, вычисляющей итоговый прогноз значений интересующей переменной. При этом формируется ансамбль из T регрессионных деревьев $\{h_t\}_{t=1}^T$, где h_t – отдельно взятое дерево решений. Далее для каждого дерева h_t формируется бутстрэп-выборка, а именно из обучающей выборки D случайным отбором N наблюдений создается обучающий набор D_t . При этом некоторые наблюдения попадут в D_t несколько раз, а некоторые не попадут совсем. Построение регрессионного дерева решений на D_t начинается с корневого узла, содержащего всю выборку D_t . Далее для каждого узла случайно выбирается подмножество признаков из всех признаков, и среди выбранных признаков находится лучший признак и лучшее значение для разбиения (точка), чтобы максимизировать уменьшение дисперсии целевой переменной y в двух дочерних узлах. Критерием является минимизация среднеквадратичной ошибки (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

где $\hat{y}_i$ – полученный прогноз целевой переменной, n – количество наблюдений. Затем выборка разделяется на две по найденной точке и признакам. Процедура деления рекурсивно повторяется для каждого дочернего узла, пока не будет выполнено одно из условий остановки (достигнута максимальная глубина дерева, разбиение не уменьшает MSE достаточно сильно и т.д.). Когда достигнуто условие остановки, предсказанием в этом узел становится среднее значение целевой переменной y всех обучающих наблюдений, попавших в этот узел. Прогноз обученной модели для нового наблюдения x_{new} получается путем нахождения терминального узла для x_{new} в каждом дереве h_t ансамбля. Терминальный узел находится, следуя правилам деления, составленным во время обучения дерева решений. За прогноз модели считается предсказание терминального узла. Итоговый прогноз МСЛ находится путем усреднения прогнозов всех деревьев:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t.$$

МСЛ продемонстрировал многообещающие результаты в прогнозировании риска развития РМЖ, раннего выявления рецидивов и предсказания ответа на химиотерапию [21–23]. Настоящее исследование носит поисковый характер и направлено на определение возможности эффективного использования алгоритмов машинного обучения для анализа взаимосвязей между микроРНК и молекулярными маркерами опухоли в условиях ограниченной выборки.

Цель исследования – поиск возможности выявления с помощью алгоритма МСЛ взаимосвязи между уровнями miR-181a, miR-25 и экспрессией молекулярных маркеров CDH1, CIP1, ITGB1, CCND1 и CD86, а также их ассоциации с ключевыми клинико-патологическими характеристиками (ER, PR, HER2, Ki-67) у пациенток с РМЖ.

МАТЕРИАЛЫ И МЕТОДЫ

В исследовании приняли участие 44 пациентки с диагнозом «инвазивная протоковая карцинома молочной железы неспецифического типа» в возрасте от 23 до 72 лет (средний возраст 55 лет), получавших лечение в отделении опухолей молочной железы городской клинической больницы (ГКБ) № 1 (г. Новосибирск). Критерии включения: первичный РМЖ без неoadъювантной терапии; гистологически подтвержденная «инвазивная протоковая карцинома неспецифического типа»; степень гистологической дифференцировки G2. Критерии исключения: сопутствующие злокачественные новообразования; острые инфекционные заболевания или обострение хронических инфекций; отдаленные метастазы. Экспрессия ER, PR, HER2 и Ki-67 определялась в соответствии со стандартными клиническими протоколами ГКБ № 1 и предоставлена в виде карт пациентов. Подробная характеристика пациенток представлена в табл. 1.

Таблица 1

Клинико-морфологическая характеристика опухолей пациенток с РМЖ			
Параметр	Категория	Абс. число, n	Доля, %
Размер опухоли (Т)	T1 (<2 см)	13	29,5
	T2 (2–5 см)	31	70,5
Вовлеченность лимфоузлов (N)	N0	28	63,6
	N1	12	27,3
	N2	4	9,1
Эстрогеновые рецепторы (ER)	ER+ ($\geq 1\%$)	25	56,8
	ER– ($< 1\%$)	19	43,2
Прогестероновые рецепторы (PR)	PR+ ($\geq 1\%$)	22	50,0
	PR– ($< 1\%$)	22	50,0
Рецептор эпидермального фактора роста человека 2-го типа	3+	9	20,5
	0 или 1+	35	79,5

Окончание табл. 1

Параметр	Категория	Абс. число, <i>n</i>	Доля, %
Ki-67 пролиферативный индекс	<20%	18	40,9
	≥20%	26	59,1

МикроРНК были выделены из 300 мкл сыворотки крови с использованием набора NucleoSpin miRNA Plasma Kit (Macherey-Nagel, Германия). Для обратной транскрипции использовалась M-MuLV-RH (компания «Биолабмикс», Россия) и miR-специфичные праймеры. Подробные характеристики праймеров представлены в ранее опубликованной работе [16]. Количественную оценку экспрессии выполняли методом цифровой капельной ПЦР на системе QX200 AutoDG (Bio-Rad, США) с детекцией TaqMan. Малая ядерная РНК U6 служила внутренним контрольным образцом.

Иммуногистохимический анализ. Экспрессию исследуемых маркеров определяли на депарафинизированных срезах опухолевой ткани с применением моноклональных антител: анти-Е-кадгерин (Anti-E-Cadherin, Cat. No. 610181, BD Biosciences, США), анти-ITGB1 (Anti-CD29/Clone 18, Cat. No. 610467, BD Transduction Laboratories, США), анти-СП (COL2A1/M2139, Cat. No. sc-52658, Santa Cruz Biotechnology, Inc., США), анти-Cyclin D1 и анти-B7-2 (PAA585Hu01, Cloud-Clone Corp., Китай). Визуализацию осуществляли с помощью VECTASTAIN ABC Kit (Vector Laboratories, США) и анализировали с помощью программного обеспечения ImageJ (National Instruments Corporation, США). Уровень экспрессии маркеров рассчитывали как долю (%) позитивно окрашенных клеток, используя не менее 10 полей зрения при увеличении 400 для каждого пациента.

Уровни экспрессии рецепторов эпидермального фактора роста 2-го типа Her2/neu, рецепторов к эстрогену (ER) и прогестерону (PR), а также маркера пролиферации Ki-67 оценивали в образцах РМЖ по методикам [24], рекомендованным для определения молекулярно-генетических субтипов рака молочной железы (Gallen International Expert Consensus, 2011). Определение рецепторного статуса по ER и PR проводилось с использованием набора ER/PR pharmDx Kit (K4071, Dako). Для определения HER2/neu статуса использовали набор Ultra View Universal DAB Detection Kit (антитела 4B5). Белок Ki-67 (клон MIB-1) выявляли при использовании набора диагностических антител фирмы Dako. Все антитела использовались в разведении 1 : 100. Оценку уровней экспрессии ER и PR проводили по методу Allred [25]. Оценку уровней экспрессии HER2 проводили по общепринятой балльной системе (от 0 до 3+).

Обработку данных осуществляли с использованием языка программирования Python (версия 3.12). Для расчета коэффициентов ранговой корреляции Спирмена (*r*) и разработки модели МСЛ использовался пакет для Data Science и Machine Learning – Scikit-learn (версия 1.4.0). Проверку нормальности распределения проводили с помощью критерия Шапиро – Уилка. Для оценки различий в уровне экспрессии маркеров между группами пациентов использовался непараметрический *U*-критерий Манна – Уитни. Данные представлены в виде медианы, 25-го и 75-го процентилей *Me* (Q_{25} – Q_{75}). Различия считали статистически значимыми при $p < 0,05$. Качество регрессионных моделей МСЛ оценивали с помощью средней абсолютной ошибки (Mean Absolute Error, MAE), которая показывает среднее отклонение предсказанных значений от реальных: чем меньше MAE, тем точнее модель описывает данные [26]. MAE выражается в тех же единицах, что и целевая переменная.

Формула MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

где y_i – истинное значение, $\hat{y}_i$ – предсказанное значение, n – количество наблюдений. Для классификационных задач МСЛ использовалась метрика precision (точность), рассчитываемая как отношение правильно предсказанных положительных случаев к общему числу предсказанных положительных случаев. Учитывая ограниченный размер выборки ($n = 44$), который не позволял разделить данные на обучающую и тестовую подвыборки без потери статистической репрезентативности, обучение и валидация моделей МСЛ проводились на полном наборе данных. Относительно небольшое число предикторов – молекулярных маркеров – позволило найти наиболее подходящую для описания существующих взаимосвязей модель, путем полного перебора всех комбинаций предикторов. Для каждой из возможных комбинаций предикторов строилось 500 моделей с различными параметрами случайности, качество оценивалось путем усреднения MAE по всем моделям.

РЕЗУЛЬТАТЫ

Первичный анализ с использованием коэффициента ранговой корреляции Спирмена не выявил статистически значимых зависимостей между уровнем miR-181a, miR-25 и маркерами адгезии, пролиферации и иммунного ответа (табл. 2, $p > 0,05$ для всех сравнений). Это указывало на отсутствие простых линейных зависимостей и необходимость применения более сложных аналитических подходов.

Таблица 2

Корреляционные взаимосвязи между уровнем микроРНК и молекулярными маркерами				
Маркер	miR-181a		miR-25	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
CDH1	– 0,049	0,753	– 0,018	0,910
ITGB1	0,036	0,815	– 0,071	0,648
CCND1	0,006	0,970	– 0,124	0,421
СП	0,069	0,657	– 0,091	0,557
CD86	– 0,081	0,600	0,053	0,731

Для выявления более сложных нелинейных взаимосвязей был использован МСЛ. Оптимальная модель соответствовала комбинации предикторов, дающей наименьший MAE. Для miR-181a наилучшая прогностическая модель включала комбинацию маркеров CDH1 и CCND1 (MAE = 4,267). Для miR-25 оптимальный набор предикторов составил ITGB1 и CCND1 (MAE = 16,255). На рисунке представлено сравнение предсказанных и фактических значений для обеих моделей, демонстрирующее их способность выявлять скрытые закономерности в данных.

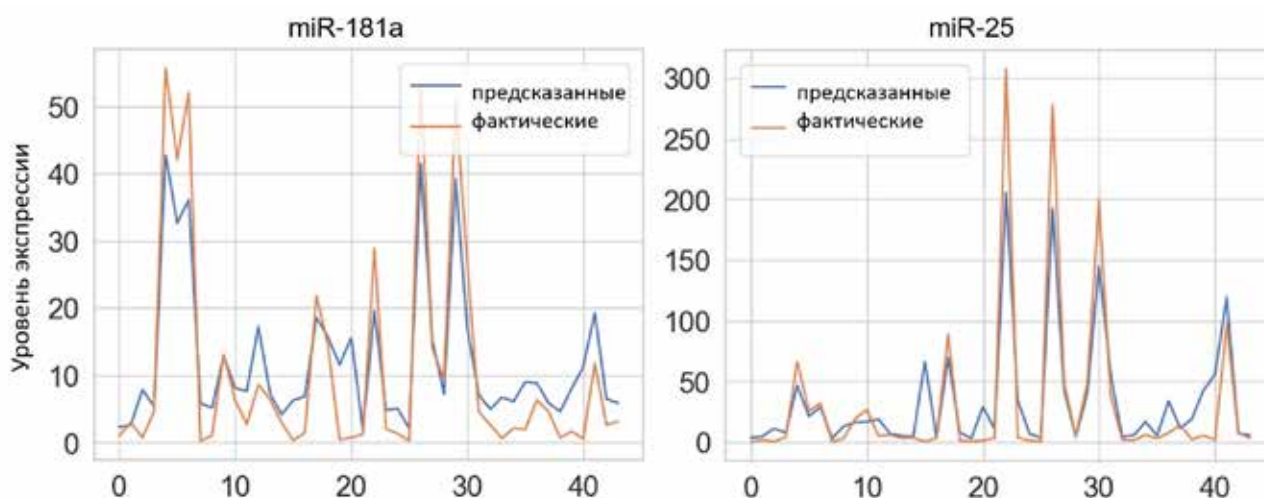


Рисунок. Сравнение предсказанных и фактических уровней miR-181a и miR-25: ось абсцисс – пациенты, $n = 44$; ось ординат – концентрация микроРНК

Анализ взаимосвязей между исследуемыми биомаркерами и ключевыми клинико-патологическими параметрами опухоли (ER, PR, HER2, Ki-67) показал статистически значимые корреляции только для коллагена II типа. Установлены умеренные положительные корреляции экспрессии СП со статусом рецепторов эстрогена ($r = 0,498$; $p = 0,001$) и прогестерона ($r = 0,354$; $p = 0,018$, табл. 3).

Таблица 3

Корреляция между уровнем микроРНК, молекулярными маркерами и молекулярно-генетическими показателями опухоли					
Маркер		ER	PR	HER2	Ki-67
miR-181a	<i>r</i>	– 0,156	– 0,036	0,150	0,176
	<i>p</i>	0,311	0,819	0,330	0,253
miR-25	<i>r</i>	– 0,166	– 0,007	0,057	0,300
	<i>p</i>	0,282	0,967	0,712	0,048
CDH1	<i>r</i>	0,089	0,047	0,024	– 0,051
	<i>p</i>	0,568	0,763	0,879	0,744
ITGB1	<i>r</i>	0,326	0,180	0,168	– 0,121
	<i>p</i>	0,031	0,243	0,275	0,433
СП	<i>r</i>	0,498*	0,354*	0,076	– 0,196
	<i>p</i>	0,001	0,018	0,623	0,202

Окончание табл. 3

Маркер		ER	PR	HER2	Ki-67
CCND1	<i>r</i>	– 0,016	0,118	0,076	– 0,164
	<i>p</i>	0,917	0,446	0,622	0,288
CD86	<i>r</i>	0,030	– 0,058	– 0,277	0,204
	<i>p</i>	0,846	0,708	0,069	0,183

* $p < 0,05$.

Дальнейший анализ различий в уровне маркеров между группами пациентов показал, что экспрессия СП статистически значимо выше в опухолях с положительным статусом гормональных рецепторов (табл. 4). В ER-положительных образцах *Me* уровня экспрессии составила 20,3% по сравнению с 6,9% в ER-отрицательных ($p = 0,010$). Аналогично для PR-статуса: 20,2% в PR-положительных против 9,8% в PR-отрицательных опухолях ($p = 0,016$). Для ITGB1 отмечена тенденция к более высокой экспрессии в ER-положительных опухолях (*Me* 18,8% против 11,5%), однако различие не достигло уровня статистической значимости ($p = 0,052$).

Таблица 4

Анализ экспрессии исследуемых маркеров в зависимости от молекулярно-генетических показателей опухоли, Me (Q_{25} – Q_{75})

Маркер	ER–, n = 19	ER+, n = 25	PR–, n = 22	PR+, n = 22	HER2–0 или 1+, n = 35	HER2+ 3+, n = 9	Ki-67 < 20%, n = 18	Ki-67 ≥ 20%, n = 26
miR-181a	2,9 (1,6–14,4)	3,0 (0,8–11,3)	3,0 (1,5–18,0)	2,9 (0,9–10,3)	2,9 (1,0–9,4)	3,1 (1,4–52,8)	2,6 (0,8–18,1)	4,5 (1,6–13,1)
	$p = 0,414$		$p = 0,438$		$p = 0,383$		$p = 0,557$	
miR-25	5,0 (3,1–26,7)	3,3 (1,3–25,8)	3,6 (2,2–28,1)	4,5 (1,3–24,2)	3,6 (1,8–25,5)	5,0 (0,8–49,2)	3,3 (1,0–16,6)	5,5 (3,1–38,6)
	$p = 0,337$		$p = 0,656$		$p = 0,977$		$p = 0,162$	
CDH1	81,5 (30,8–94,0)	90,3 (59,5–96,4)	84,0 (36,8–94,5)	89,4 (67,1–96,3)	88,5 (58,9–94,8)	74,1 (31,4–96,3)	88,5 (49,3–95,4)	86,4 (38,8–94,8)
	$p = 0,507$		$p = 0,664$		$p = 0,942$		$p = 0,925$	
ITGB1	11,5 (5,4–19,1)	18,8 (10,5–80,6)	12,0 (6,1–20,4)	18,5 (8,2–80,1)	13,5 (6,3–46,1)	16,5 (9,8–24,9)	14,3 (5,9–26,4)	16,4 (7,4–46,1)
	$p = 0,052$		$p = 0,236$		$p = 0,827$		$p = 0,690$	
CII	6,9 (6,3–15,7)	20,3 (12,7–30,0)	9,8 (6,4–15,9)	20,2 (11,2–27,3)	14,4 (6,4–26,5)	14,4 (9,0–18,5)	15,7 (8,4–27,2)	14,4 (6,5–20,5)
	$p = 0,010^*$		$p = 0,016^*$		$p = 0,942$		$p = 0,259$	
CCND1	93,9 (77,1–97,2)	93,9 (83,1–97,6)	93,9 (74,6–96,8)	93,9 (85,8–97,7)	93,9 (80,7–97,3)	93,9 (87,9–97,4)	93,9 (83,1–98,0)	93,9 (77,1–96,7)
	$p = 0,849$		$p = 0,451$		$p = 0,759$		$p = 0,207$	
CD86	21,1 (10,9–26,4)	19,9 (15,2–29,2)	21,3 (10,9–27,3)	19,0 (2,4–29,0)	20,7 (12,8–30,2)	19,0 (8,9–23,6)	20,7 (11,2–29,2)	19,9 (16,1–26,4)
	$p = 0,924$		$p = 0,673$		$p = 0,244$		$p = 0,991$	

Применение МСЛ для прогнозирования рецепторного статуса РМЖ показало различные уровни точности в зависимости от маркера: наилучший результат был достигнут для HER2-статуса (точность 82,6%) с использованием комбинации ITGB1, miR-181a и CDH1. Для ER-статуса точность составила 69,5% при использовании miR-25, CDH1 и CII, а для PR-статуса – 66,3% с предикторами miR-181a, ITGB1 и CCND1. В то же время прогнозирование уровня Ki-67 с помощью miR-25 и CCND1 оказалось наименее точным, на что указывает высокая средняя абсолютная ошибка (MAE = 30,411).

ОБСУЖДЕНИЕ

Полученные данные демонстрируют сложный характер взаимосвязей между экспрессией микроРНК miR-25, miR-181a и молекулярными маркерами при РМЖ. Способность алгоритма МСЛ выявлять нелинейные ассоциации, не детектируемые традиционными методами корреляционного анализа, подчеркивает потенциал машинного обучения в исследовании молекулярных механизмов канцерогенеза.

Выявленная ассоциация miR-181a с CDH1 и CCND1 имеет биологическое обоснование. MiR-181a может влиять на E-кадгерин через активацию TGF- β /Smad-зависимых механизмов, способствуя ЭМП [27]. Кроме того, miR-181a активирует сигнальный путь PI3K-Akt, усиливающий клеточный рост и пролиферацию, тем самым влияя на экспрессию CDH1 [28]. Участие как CCND1, так и CDH1 в регуляции сигнального пути Wnt/ β -катенина может объяснять их совместное включение в прогностическую модель для miR-181a [29]. В исследовании

множественной миеломы X. Yan о соавт. обозначены механизмы, лежащие в основе регуляции клеточного цикла miR-181a, путем воздействия на экспрессию CCND1 [30].

Связь miR-25 с ITGB1 и CCND1 также находит подтверждение на молекулярном уровне. MiR-25 участвует в активации пути Wnt/ β -катенина, способствуя накоплению β -катенина в ядре и усиливая транскрипцию генов-мишеней, включая CCND1 [31]. ITGB1 участвует в сигнальном пути ILK-Akt, который усиливает ЭМП, что может указывать на общие регуляторные механизмы с miR-25 [32].

Статистически значимые ассоциации коллагена II типа с гормональными рецепторами (ER и PR) представляют особый интерес, поскольку этот маркер мало изучен в контексте РМЖ. При раке молочной железы различные типы коллагена демонстрируют различные профили экспрессии [33]. Показано, что коллагены различных типов по-разному влияют на развитие раковых образований, в частности РМЖ, воздействуя на важнейшие клеточные процессы, включая апоптоз, пролиферацию и метастазирование [33, 34]. Например, показано, что при тройном негативном РМЖ снижение экспрессии COL2A1 (CII) связано с увеличением степени злокачественности опухоли, метастазированием, ухудшением прогноза и снижением выживаемости [33]. При этом уменьшение экспрессии COL1A1 может быть ассоциировано со снижением злокачественности РМЖ. Установлено, что экспрессия COL1A1 связана с экспрессией рецепторов эстрогена или прогестерона (ER/PR) и статусом метастазирования. У пациенток с ER-положительным РМЖ более высокая экспрес-

сия COL1A1 коррелирует с агрессивным поведением клеток и ухудшением прогноза [34]. Выявленная нами повышенная экспрессия СП в гормон-позитивных опухолях может отражать особенности ремоделирования внеклеточного матрикса при люминальных подтипах РМЖ, что требует дальнейшего изучения.

Высокая точность предсказания HER2-статуса (более 80%) с использованием комбинации ITGB1, miR-181a и CDH1 указывает на их потенциальную диагностическую значимость. Вероятно, эта комбинация комплексно отражает биологические особенности агрессивного HER2-позитивного подтипа РМЖ. ITGB1 посредством клеточного ремоделирования повышает агрессивность опухоли, а miR-181a, подавляя экспрессию CDH1, дополнительно усиливает пролиферацию и инвазивный потенциал. В то же время модели для статуса ER и PR показали умеренную точность. Скорее всего, это связано со значительной гетерогенностью люминальных подтипов РМЖ, которые сильно различаются как по молекулярным характеристикам, так и по клиническому поведению.

Следует подчеркнуть, что данное исследование носит поисковый характер и имеет ряд существенных ограничений, которые необходимо учитывать при интерпретации результатов. Основными ограничениями являются относительно небольшой размер выборки и ее гетерогенность по клинико-морфологическим и молекулярно-генетическим параметрам. Из-за малого числа наблюдений мы не разделяли данные на обучающую и тестовую выборки, что повышает риск переобучения модели и ограничивает возможность обобщения результатов. Поэтому полученные выводы следует рассматривать как предварительные, требующие валидации на более крупных и независимых когортах пациентов.

ЗАКЛЮЧЕНИЕ

Полученные данные демонстрируют, что применение метода машинного обучения «случайный лес» является эффективным подходом для выявления сложных, нелинейных взаимосвязей между циркулирующими микроРНК и маркерами в опухолевой ткани при раке молочной железы. Установленные ассоциации между miR-181a, CDH1 и CCND1, а также между miR-25, ITGB1 и CCND1 не обнаружены с помощью стандартного корреляционного анализа. Эти результаты имеют фундаментальное значение, поскольку позволяют выдвинуть гипотезу о координированной регуляции ключевых процессов опухолевой прогрессии. Потери клеточной адгезии (CDH1), усиление миграции (ITGB1) и нарушения клеточного цикла (CCND1). Хотя получен-

ные модели не предназначены для непосредственного клинического применения на данном этапе, они представляют собой важную отправную точку для дальнейших исследований в этом направлении. Результаты подчеркивают потенциал комбинаций биомаркеров для стратификации пациентов и предсказания молекулярных подтипов РМЖ, в особенности статуса HER2.

СПИСОК ИСТОЧНИКОВ

1. Rani V., Sengar R.S. Biogenesis and mechanisms of micro RNA-mediated gene regulation. *Biotechnol. Bioeng.* 2022;119(3):685–692. DOI: 10.1002/bit.28029.
2. Kumar S., Ranga A. Role of miRNAs in breast cancer development and progression: Current research. *BioFactors.* 2025;51(1):e2146. DOI: 10.1002/biof.2146.
3. Zhai Z., Mu T., Zhao L., Li Y., Zhu D., Pan Y. MiR-181a-5p facilitates proliferation, invasion, and glycolysis of breast cancer through NDRG2-mediated activation of PTEN/AKT pathway. *Bioengineered.* 2022;13(1):83–95. DOI: 10.1080/21655979.2021.2006974.
4. Zhao T., Meng W., Chin Y., Gao L., Yang X., Sun S. et al. Identification of miR-25-3p as a tumor biomarker: Regulation of cellular functions via TOB1 in breast cancer. *Mol. Med. Rep.* 2021; 23(6):406. DOI: 10.3892/mmr.2021.12045.
5. Kalinina T.S., Kononchuk V.V., Alekseenok E.Y., Obukhova D., Sidorov S.V., Strunkin D. et al. Expression of estrogen receptor- and progesterone receptor-regulating micro RNAs in breast cancer. *Genes.* 2021;12(4):582. DOI: 10.3390/genes12040582.
6. Andreeva O.E., Sorokin D.V., Mikhaevich E.I., Bure I.V., Shchegolev Y.Y., Nemtsova M.V. et al. Towards unravelling the role of ERα-targeting miRNAs in the exosome-mediated transferring of the hormone resistance. *Molecules.* 2021;26(21):6661. DOI: 10.3390/molecules26216661.
7. Gilam A., Shai A., Ashkenazi I., Sarid L.A., Drobot A., Bickel A. et al. MicroRNA regulation of progesterone receptor in breast cancer. *Oncotarget.* 2017;8(16):25963–25976. DOI: 10.18632/oncotarget.15657.
8. Feng J., Hu S., Liu K., Sun G., Zhang Y. The role of microRNA in the regulation of tumor epithelial-mesenchymal transition. *Cells.* 2022;11(13):1981. DOI: 10.3390/cells11131981.
9. Mendez M.J., Hoffman M.J., Cherry E.M., Lemmon C.A., Weinberg S.H. Cell fate forecasting: a data-assimilation approach to predict epithelial-mesenchymal transition. *Biophys. J.* 2020;118(7):1749–1768. DOI: 10.1016/j.bpj.2020.02.011.
10. Manfioletti G., Fedele M. Epithelial-mesenchymal transition (EMT). *Int. J. Mol. Sci.* 2023;24(14):11386. DOI: 10.3390/ijms241411386.
11. Ali A.H., Haider G., Ahmed K., Hashmi M., Kerio P., Irshad R. Correlation between e-cadherin and hormone receptor status among breast cancer patients. *Journal of the College of Physicians and Surgeons Pakistan.* 2020;30(10):1030–1034. DOI: 10.29271/jcpsp.2020.10.1030.
12. Autenshlyus A.I., Perepechaeva M.L., Studenikina A.A., Grishanova A.Y., Lyakhovich V.V. Serum miR-181a and miR-25 levels in patients with breast cancer or benign breast

- disease. *Dokl. Biochem. Biophys.* 2023;512(1):279–283. DOI: 10.1134/S1607672923700436.
13. Студеникина А.А., Архипов С.А., Варакин Н.А., Проскура А.В., Аутеншлюс А.И. Биомаркеры молочной железы в выявлении малигнизации при доброкачественных заболеваниях у женщин. *Профилактическая медицина.* 2024;27(1):63–66. DOI: 10.17116/profmed20242701163.
 14. Студеникина А.А., Архипов С.А., Жураковский И.П., Проскура А.В., Аутеншлюс А.И. Экспрессия маркеров эпителиально-мезенхимального перехода при раке молочной железы. *Архив патологии.* 2023;85(4):12–17. DOI: 10.17116/patol20238504112.
 15. Jeffreys S.A., Becker T.M., Khan S., Soon P., Neubauer H., de Souza P. et al. Prognostic and predictive value of CCND1/Cyclin D1 Amplification in Breast Cancer with a Focus on Postmenopausal Patients: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology.* 2022;13:895729. DOI: 10.3389/fendo.2022.895729.
 16. Wennhold K., Thelen M., Lehmann J., Schran S., Preugsatz E., Garcia-Marquez M. et al. CD86+ antigen-presenting B cells are increased in cancer, localize in tertiary lymphoid structures, and induce specific T-cell responses. *Cancer Immunol. Res.* 2021;9(9):1098–1108. DOI: 10.1158/2326-6066.CIR-20-0949.
 17. Yang H.L., Tsai Y.C., Korivi M., Chang C.T., Hseu Y.C. Lucidone Promotes the cutaneous wound healing process via activation of the PI3K/AKT, Wnt/ β -catenin and NF- κ B signaling pathways. *Biochimica et Biophysica Acta – Molecular Cell Research.* 2017;1864(1):151–168. DOI: 10.1016/j.bbamcr.2016.10.021.
 18. Li X., Jin Y., Xue J. Unveiling collagen's role in breast cancer: insights into expression patterns, functions and clinical implications. *International Journal of General Medicine.* 2024;17:1773–1787. DOI: 10.2147/IJGM.S463649.
 19. Khan M.J., Singh A.K., Sultana R., Singh P.P., Khan A., Saxena S. Breast cancer: A comparative review for breast cancer detection using machine learning techniques. *Cell Biochem. Funct.* 2023;41(8):996–1007. DOI: 10.1002/cbf.3868.
 20. Breiman L. Random forests. *Mach. Learn.* 2001;45:5–32. DOI: 10.1023/A:1010933404324.
 21. Zuo D., Yang L., Jin Y., Qi H., Liu Y., Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med. Inform. Decis. Mak.* 2023;23(1):276. DOI: 10.1186/s12911-023-02377-z.
 22. Jin Y., Lan A., Dai Y., Jiang L., Liu S. Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy. *Eur. J. Med. Res.* 2023;28(1):394. DOI: 10.1186/s40001-023-01361-7.
 23. Arika R.N., Mindila A., Cheruiyo W. Machine learning algorithms for breast cancer diagnosis: challenges, prospects and future research directions. *J. Oncol.* 2023;5(1):1–31. DOI: 10.30564/jor.v5i1.4977.
 24. Kondov B., Milenković Z., Kondov G., Petrushevska G., Basheska N., Bogdanovska-Todorovska M. et al. Presentation of the molecular subtypes of breast cancer detected by immunohistochemistry in surgically treated patients. *J. Med. Sci.* 2018;6(6): 961–967. DOI: 10.3889/oamjms.2018.231.
 25. Yumpu [Internet]. ER/PR PharmDx Interpretation Manual. – Dako. URL: <https://www.yumpu.com/en/document/read/18500132/er-pr-pharmdxm-interpretation-manual-dako>. Available at: 09.20.2013.
 26. Willmott C.J., Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* 2005;30(1):79–82. DOI: 10.3354/cr030079.
 27. Li L., Xu Q.H., Dong Y.H., Li G.X., Yang L., Wang L.W. et al. MiR-181a upregulation is associated with epithelial-to-mesenchymal transition (EMT) and multidrug resistance (MDR) of ovarian cancer cells. *Eur. Rev. Med. Pharmacol. Sci.* 2016;20(10):2004–2010.
 28. Strotbek M., Schmid S., Sánchez-González I., Boerries M., Busch H., Olayioye M.A. miR-181 elevates Akt signaling by co-targeting PHLPP2 and INPP4B phosphatases in luminal breast cancer. *Int. J. Cancer.* 2017;140(10):2310–2320. DOI: 10.1002/ijc.30661.
 29. Govatati S., Singamsetty G.K., Nallabelli N., Malempati S., Rao P.S., Madamchetty V.K.K. et al. Contribution of cyclin D1 (CCND1) and E-cadherin (CDH1) alterations to colorectal cancer susceptibility: a case-control study. *Tumor Biol.* 2014;35(12):12059–12067. DOI: 10.1007/S13277-014-2505-9.
 30. Yan X., Gao M., Zhang P., Ouyang G., Mu Q., Xu K. MiR-181a functions as an oncogene by regulating CCND1 in multiple myeloma. *Oncol. Lett.* 2020;20(1):758–764. DOI: 10.3892/OL.2020.11579.
 31. Yao J., Li G., Liu M., Yang S., Su H., Ye C. Inc-MICAL2 1 sponges miR 25 to regulate DKK3 expression and inhibits activation of the Wnt/ β catenin signaling pathway in breast cancer. *Int. J. Mol. Med.* 2022;49(2):23. DOI: 10.3892/ijmm.2021.5078.
 32. Jiang X., Wang J., Zhang K., Tang S., Ren C., Chen Y. The role of ITGB1-ILK-Akt signaling-mediated epithelial-mesenchymal transition of liver epithelial cells and chemoresistance and radioresistance in hepatocellular carcinoma cells. *Med. Oncol.* 2015;32(5):141. DOI: 10.1007/S12032-015-0595-X.
 33. Vora H., Gavali M., Patel N. A Study of Col2A1 expression in triple-negative breast cancer patients. *Glob. J. Clinic. Medici. Medica. Res.* 2023;1:1–5.
 34. Li X., Jin Y., Xue J. Unveiling collagen's role in breast cancer: insights into expression patterns, functions and clinical implications. *Int. J. Gen. Med.* 2024;17:1773–1787. DOI: 10.2147/IJGM.S463649.

Вклад авторов

Студеникина А.А. – проведение экспериментов, обсуждение результатов исследования, написание текста. Михайловский М.А. – проведение экспериментов, обсуждение результатов исследования. Тимофеев В.С., Аутеншлюс А.И. – концепция и руководство работой, обсуждение результатов исследования, редактирование текста статьи. Архипов С.А. – проведение экспериментов, обсуждение результатов исследования, редактирование текста статьи.

Информация об авторах

Студеникина Анастасия Александровна – канд. мед. наук, ст. науч. сотрудник, Центральная научно-исследовательская лаборатория (ЦНИЛ), НГМУ, г. Новосибирск; науч. сотрудник, НИИМББ, ФИЦ ФТМ, г. Новосибирск, aastudenikina@frcftm.ru, <https://orcid.org/0000-0003-3936-1316>

Михайловский Михаил Артемович – студент, НГТУ, г. Новосибирск, maiailovsky18@mail.ru, <https://orcid.org/0009-0009-2272-6295>

Тимофеев Владимир Семенович – д-р техн. наук, профессор кафедры теоретической и прикладной информатики, декан факультета прикладной математики и информатики, НГТУ, г. Новосибирск, v.timofeev@corp.nstu.ru, <https://orcid.org/0000-0002-6352-3218>

Архипов Сергей Алексеевич – д-р биол. наук, вед. науч. сотрудник, ЦНИЛ, НГМУ, г. Новосибирск; ст. науч. сотрудник, НИИМББ, ФИЦ ФТМ, г. Новосибирск, arhipowsergei@yandex.ru, <https://orcid.org/0000-0002-1390-4426>

Аутеншлюс Александр Исаевич – д-р биол. наук, профессор, зав. ЦНИЛ, НГМУ, г. Новосибирск; гл. науч. сотрудник, НИИМББ, ФИЦ ФТМ, г. Новосибирск, lpaiir@211.ru, <https://orcid.org/0000-0001-7180-010X>

(✉) Студеникина Анастасия Александровна, aastudenikina@frcftm.ru

Поступила в редакцию 23.06.2025;
одобрена после рецензирования 22.09.2025;
принята к публикации 30.10.2025