

ОПРЕДЕЛЕНИЕ ПРОГНОСТИЧЕСКИ ЗНАЧИМОЙ СИГНАТУРЫ ДНК-МЕТИЛИРОВАНИЯ У ПАЦИЕНТОК С РАЗЛИЧНЫМИ ТИПАМИ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

А. И. Калинин¹✉, В. О. Сигин¹, М. В. Немцова^{1,2}, В. В. Стрельников^{1,3}

¹ Медико-генетический научный центр имени Н. П. Бочкова, Москва, Россия

² Первый Московский государственный медицинский университет имени И. М. Сеченова (Сеченовский университет), Москва, Россия

³ Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва, Россия

Рак молочной железы (РМЖ) — наиболее часто диагностируемое онкологическое заболевание и одна из ведущих причин смертности среди женского населения. Разработка прогностических моделей с использованием мультиомиксных данных является главной целью прецизионной онкологии. Аберрантное метилирование ДНК в РМЖ представляет собой информативный маркер канцерогенеза. Несмотря на существующие факторы прогноза РМЖ, введение маркеров метилирования позволит получать более точную прогностическую оценку. Целью работы было изучить сигнатуры метилирования ДНК в различных подтипах РМЖ для клинических конечных точек и клинико-патологических характеристик пациенток. Данные об уровнях метилирования CpG-динуклеотидов (зондов) и клинические характеристики образцов РМЖ были получены из базы данных The Cancer Genome Atlas Breast Cancer. С помощью метода одномерной регрессии Кокса были выбраны CpG-динуклеотиды, ассоциированные с выбранными конечными точками. Методом LASSO осуществляли поиск стабильных зондов, а дальнейшее построение сигнатур и независимость клинических характеристик выполняли с помощью многофакторной регрессии Кокса. Диагностический и прогностический потенциал сигнатур оценивали с помощью метода ROC-анализа и кривых Каплан–Майера. Показано, что сигнатуры отобранных зондов обладают значимым диагностическим (AUC от 0,76 до 1) и прогностическим ($p < 0,05$) потенциалом. С помощью данного подхода удалось идентифицировать 47 генов, связанных с хорошим и плохим прогнозом, из которых пять уже были описаны ранее. При наличии результатов широкогеномного анализа ДНК примененный исследовательский подход можно использовать для изучения не только молекулярного патогенеза РМЖ, но и для других заболеваний.

Ключевые слова: рак молочной железы, молекулярные подтипы, выживаемость, метилирование ДНК, прогностические маркеры

Финансирование: работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Федеральной научно-технической программы развития генетических технологий на 2019–2027 годы (соглашение № 075-15-2021-1073).

Вклад авторов: А. И. Калинин — дизайн исследования, сбор, анализ и интерпретация данных, написание статьи; В. О. Сигин — написание статьи; М. В. Немцова — концепция и дизайн исследования; В. В. Стрельников — концепция и дизайн исследования, научное редактирование.

✉ **Для корреспонденции:** Алексей Игоревич Калинин
ул. Москворечье, д. 1, г. Москва, 115522; alexeika2@yandex.ru

Статья получена: 18.10.2022 **Статья принята к печати:** 11.11.2022 **Опубликована онлайн:** 25.11.2022

DOI: 10.24075/vrgmu.2022.056

IDENTIFICATION OF PROGNOSTICALLY SIGNIFICANT DNA METHYLATION SIGNATURES IN PATIENTS WITH VARIOUS BREAST CANCER TYPES

Kalinkin AI¹✉, Sigin VO¹, Nemtsova MV^{1,2}, Strelnikov VV^{1,3}

¹ Research Centre of Medical Genetics, Moscow, Russia

² Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

³ Pirogov Russian National Research Medical University, Moscow, Russia

Breast cancer (BC) is the most frequently diagnosed cancer and one of the major causes of female mortality. The development of prognostic models based on multiomics data is the main goal of precision oncology. Aberrant DNA methylation in BC is a diagnostic marker of carcinogenesis. Despite the existing factors of BC prognosis, introduction of methylation markers would make it possible to obtain more accurate prognostic scores. The study was aimed to assess DNA methylation signatures in various BC subtypes for clinical endpoints and patients' clinicopathological characteristics. The data on methylation of CpG dinucleotides (probes) and clinical characteristics of BC samples were obtained from The Cancer Genome Atlas Breast Cancer database. CpG dinucleotides associated with the selected endpoints were chosen by univariate Cox regression method. The LASSO method was used to search for stable probes, while further signature construction and testing of the clinical characteristics independence were performed using multivariate Cox regression. The diagnostic and prognostic potential of the signatures was assessed using ROC analysis and Kaplan–Meier curves. It has been shown that the signatures of selected probes have a significant diagnostic (AUC 0.76–1) and prognostic ($p < 0.05$) potential. This approach has made it possible to identify 47 genes associated with good and poor prognosis, among these five genes have been described earlier. If the genome-wide DNA analysis results are available, the research approach applied can be used to study molecular pathogenesis of BC and other disorders.

Keywords: breast cancer, molecular subtypes, survival analysis, DNA methylation, prognostic markers

Funding: the study was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the Federal Scientific and Technical Program for the Development of Genetic Technologies in 2019–2027 (agreement № 075-15-2021-1073).

Author contribution: Kalinkin AI — study design, data acquisition, analysis and interpretation, manuscript writing; Sigin VO — manuscript writing; Nemtsova MV — study concept and design; Strelnikov VV — study concept and design, scientific editing.

✉ **Correspondence should be addressed:** Alexey I. Kalinkin
Moskvorechye, 1, Moscow, 115522; alexeika2@yandex.ru

Received: 18.10.2022 **Accepted:** 11.11.2022 **Published online:** 25.11.2022

DOI: 10.24075/brsmu.2022.056

По данным организации мониторинга онкологических заболеваний GLOBOCAN, в 2020 г. зарегистрировано около 2,3 млн новых случаев и 684 996 смертей от рака

молочной железы (РМЖ), который занимает первое место по частоте среди онкологических заболеваний во всем мире [1] и представляет собой высокогетерогенное

заболевание с различными молекулярными и клиническими характеристиками [2].

В настоящее время в широкой клинической практике подтип РМЖ определяют в опухолевой ткани иммуногистохимическим (ИГХ) методом [3] в том числе по экспрессии в опухоли белков-рецепторов эстрогена, прогестерона, HER2, а также скорости деления опухолевых клеток. Разработка методов анализа экспрессии генов с помощью ДНК-микрочипов сыграла важную роль в определении молекулярных подтипов РМЖ. С помощью использования классификатора на основе экспрессии 50 генов PAM50 удается четко выделить люминальный A (LumA), люминальный B (LumB), HER2-обогащенный (HER2+) молекулярные подтипы и базальноподобный или тройной негативный РМЖ (ТН РМЖ) [4]. ТН РМЖ включает в себя 15–20% всех случаев РМЖ и характеризуется агрессивным течением, высоким уровнем метастазирования, частым возникновением рецидивов и низкой выживаемостью по сравнению с другими подтипами РМЖ [5]. Мультигенные микрочиповые диагностические системы позволяют получить важную прогностическую информацию для онкологических больных, особенно в случае двусмысленного прогноза по результатам клинических характеристик и иммуногистохимических маркеров. К таким системам относят MammaPrint/Blueprint и Prosigna/PAM50, которые в дополнении к их прогностической и предикторной ценности предоставляют возможность разбиения на молекулярные подтипы [6]. Данные системы можно использовать для определения оценки высокого и низкого рисков рецидива для пациенток с РМЖ, однако для ТН РМЖ и HER2+-молекулярного подтипа такая возможность пока отсутствует ввиду недостатка клинических исследований.

Эпигенетические изменения модулируют использование генома с помощью модификаций гистонов, вариантного состава гистонов, ремоделинга хроматина, метилирования ДНК, расположения нуклеосом и некодирующих РНК (экспрессия специфичных микроРНК). Для проявления своего эффекта вышеперечисленные эпигенетические изменения действуют совместно. Метилирование ДНК является одним из наиболее известных факторов регуляции экспрессии генов. Оно происходит за счет ковалентной модификации цитозина путем присоединения метильной группы в контексте CpG-динуклеотида к 5'-углероду пиримидинового кольца [7]. CpG-динуклеотиды обычно концентрируются в CG-богатых участках ДНК под названием CpG-островки, значительная часть которых сосредоточена в промоторных участках гена и в областях с длинными повторами, например в ретротранспозонных элементах или центромерных повторах. Реакцию метилирования цитозина опосредует класс ферментов под названием метилтрансферазы (DNMT) [7]. Всего у млекопитающих идентифицировано пять членов семейства DNMT: DNMT1, DNMT2, DNMT3a, DNMT3b и DNMT3L. DNMT3a и DNMT3b являются метилтрансферазами *de novo*, которые взаимодействуют с неметилированными CpG-динуклеотидами. DNMT1 выполняет функцию поддержания метилирования при репликации в S-фазе. Было показано, что DNMT3L стимулирует метилирование *de novo*, которое происходит с помощью DNMT3a и опосредует репрессию транскрипции с помощью гистондеацетилазы 1 (HDAC1) [7]. Аберрантное метилирование ДНК связано с широким спектром заболеваний и наиболее выражено в злокачественных опухолях [8]. Исследования последних лет продемонстрировали, что каждая эпителиальная

опухоль содержит приблизительно 10–15 генов, инактивированных структурными изменениями генома, и несколько сотен, инактивированных гиперметилированием ДНК, что демонстрирует значение этой модификации в развитии опухоли. Другой особенностью опухолевых геномов является их тотальное гипометилирование. Это полногеномное гипометилирование, происходящее главным образом из-за потери метилирования повторяющихся элементов, ведет к геномной нестабильности и хромосомным перестройкам [8]. Важную роль в патогенезе РМЖ играет повышенное промоторное метилирование в генах-супрессорах опухолевого роста, которые подавляют различные механизмы опухолевого прогрессирования, приводящее к их эпигенетическому «умолчанию» и обратимой инактивации [8]. Идентификация опухолеспецифичных паттернов аберрантного метилирования ДНК может быть полезным для ранней диагностики рака, дифференциальной диагностики злокачественных новообразований, в качестве прогностических и предикторных маркеров [9]. Изучение специфичных паттернов метилирования ДНК, выявляемых с использованием широкогеномных методов анализа, вносит важный вклад в понимание молекулярного патогенеза рака молочной железы [10]. Как отмечалось выше, каждый тип рака подразделяется на подтипы, и существуют геномные паттерны, в том числе эпигенетические, характерные для этих подтипов. Таким образом, необходимо проведение специфического широкогеномного профилирования метилирования ДНК при онкологических заболеваниях, наряду с традиционными исследованиями точечных событий гиперметилирования промоторов отдельных генов [11].

Прогнозирование подразумевает под собой предсказание вероятного течения и исхода онкологического заболевания. Анализ выживаемости основан на математическом подходе к прогнозированию онкологического заболевания и позволяет предсказать вероятность жизни после определенного времени. Маркеры метилирования ДНК благодаря их биологической важности и стабильности являются эффективным прогностическим фактором [12]. В одной из работ на основе данных результатов широкогеномного анализа метилирования ДНК в образцах РМЖ в составе базы The Cancer Genome Atlas Breast Cancer (TCGA-BRCA) была построена модель из семи CpG-динуклеотидов, в которой хорошо различают опухоли молочной железы всех подтипов и нормальные ткани, а также идентифицировано шесть сайтов метилирования, которые высоко коррелировали с общей выживаемостью (ОВ) [13]. В ходе анализа данных метилирования из открытых источников с помощью LASSO-регрессии и бустинга были обнаружены соответственно 29 и 11 CpG-динуклеотидов, связанных с ОВ [14]. Исследование данных из открытого источника TCGA-BRCA тоже позволило выявить три гена (TDRD10, PRAC2 и TMEM132C), состояние метилирования которых имеет прогностическую ценность, но преимущественно для эстроген-положительных опухолей молочной железы [15]. Для ТН РМЖ была разработана прогностическая модель на основании данных из источника TCGA-BRCA, состоящая из пяти генов (*TGFBR2*, *EIF4EBP1*, *FOSB*, *BCL2A1*, *ADRB2*), которая одинаково хорошо прогнозирует ОВ и безрецидивную выживаемость (БРВ) [16].

Необходимость в проведении исследования обусловлена отсутствием таких сигнатур для HER2-обогащенного подтипа и довольно ограниченным количеством для остальных

Таблица 1. Клинико-патологические характеристики и данные о статусе клинических конечных точек для пациенток с LumAB, TH PMЖ и HER2-обогащенным молекулярными подтипами PMЖ из открытого источника TCGA-BRCA

Характеристики	LumAB	TH PMЖ	HER2-обогащенный
Число образцов (%)	555	134	46
Возраст (медиана), лет	59	54	58
T (%)			
T1	148 (26.49)	26 (19.4)	12 (26.09)
T2	310 (55.86)	87 (64.93)	28 (60.87)
T3	81 (14.59)	16 (11.94)	3 (6.52)
T4	14 (2.52)	4 (2.99)	3 (6.52)
Нет информации	2 (0.36)	1 (0.75)	-
N (%)			
N0	233 (41.98)	78 (58.21)	16 (34.78)
N1	197 (35.5)	41 (30.6)	17 (36.96)
N2	75 (13.51)	11 (8.21)	6 (13.04)
N3	42 (7.57)	4 (2.99)	4 (8.7)
Нет информации	8 (1.44)	-	3 (6.52)
M (%)			
M0	431 (77.66)	110 (82.09)	37 (80.43)
M1	6 (1.08)	3 (2.24)	1 (2.17)
Нет информации	118 (21.26)	21 (15.67)	8 (17.39)
Стадия (%)			
I	98 (17.66)	16 (11.94)	4 (8.7)
II	291 (52.43)	94 (70.15)	28 (60.87)
III	154 (27.75)	19 (14.18)	12 (26.09)
IV	5 (0.9)	2 (1.49)	1 (2.17)
Нет информации	7 (1.26)	3 (2.24)	1 (2.17)
Тип выживаемости			
Общая выживаемость (%)			
Без события	490 (88.29)	113 (84.33)	37 (80.43)
Смерть	65 (11.71)	21 (15.67)	9 (19.57)
Безрецидивная выживаемость (%)			
Без рецидива	445 (80.18)	102 (76.12)	35 (76.09)
Наступление рецидива	39 (7.03)	20 (14.93)	4 (8.7)
Нет информации	71 (12.79)	12 (8.96)	7 (15.22)
Выживаемость без признаков прогрессирования заболевания (%)			
Без прогрессирования	486 (87.57)	109 (81.34)	38 (82.61)
Наступление прогрессирования	69 (12.43)	25 (18.66)	8 (17.39)

молекулярных подтипов PMЖ. Диагностический потенциал существующих моделей для прогнозирования выживаемости тоже остается неясным, поэтому мы использовали модифицированный алгоритм для поиска CpG-динуклеотидов, ассоциированных со всеми доступными клиническими конечными точками из базы данных TCGA-BRCA.

Целью данного исследования было получить различные сигнатуры на основании открытых данных ДНК-метилирования в PMЖ The Cancer Genome Atlas Breast Cancer для прогнозирования различных типов выживаемости пациенток (общей выживаемости, безрецидивной выживаемости и выживаемости без прогрессирования) в молекулярных подтипах PMЖ и проверить зависимость клинико-патологических характеристик от полученных сигнатур.

МАТЕРИАЛЫ И МЕТОДЫ

Публично доступные клинические параметры, а также данные широкогеномного профилирования метилирования ДНК, полученные с помощью гибридационных чипов HumanMethylation450 (HM450) (Illumina Inc.; CLIA) проекта The Cancer Genome Atlas Breast Cancer (TCGA-BRCA) (<https://portal.gdc.cancer.gov/projects/TCGA-BRCA>), получали и обрабатывали с помощью пакета программ TCGAbiolinks [17]. Критерии включения пациентов для дальнейшего отбора кандидатных CpG-пар были следующими: наличие соответствующего молекулярного подтипа PMЖ, наличие доступной клинико-патологической информации, наличие данных о профилировании метилирования ДНК с чипов Illumina HumanMethylation450. Критерии исключения: отсутствие данных о показателях времени для клинических

Таблица 2. Суммарное количество сигнатур и CpG-пары, полученные в результате LASSO регрессии Кокса для каждого типа выживаемости и молекулярного подтипа РМЖ. Для некоторых CpG-пар не удалось установить принадлежность к гену

Тип выживаемости + молекулярный подтип	Идентификаторы зондов HM450	Гены	Число полученных комбинаций
ОВ + LumAB	cg02287630; cg20417424 cg05828605 cg00297993 cg20471297 cg08133669 cg17323488 cg08241401 cg00815177 cg08442529 cg20746134 cg01821113 cg04523731 cg11140305 cg22067527	<i>SLC30A7; EXTL2; ST6GALNAC5; C15orf41; DYNC1H1; MIA3; NIPAL3; HEY2; HK1; DIRC3; TMEM41A; SH3BP5L; RFX2</i>	32752
БПВ + LumAB	cg22790777 cg23667405 cg01017355 cg09561458 cg08128789 cg27304144 cg08039281 cg13486627 cg04833210 cg27439396 cg24347894	<i>SLC25A39; BAT2; ZNF417; LRRC8B; HSPG2; PSMA6; RG9MTD3; RBM19; N6AMT2; ZNF827</i>	2036
ВБП + LumAB	cg13792075 cg08128789 cg13447284 cg00815177 cg10466124 cg13486627 cg15481636 cg00120948 cg05564086 cg23667405 cg17960080	<i>LRRC8B; HIST3H2A; ABCC5; BAT2; SPAG5; RERE; NIPAL3; HLA-DRB5; RG9MTD3</i>	2036
ОВ + TH РМЖ	cg03512997 cg07804617 cg12814969 cg14293027 cg15355719 cg17053075 cg19002462 cg26401512 cg02567719	<i>FAM136A; HNRPDL; ENOPH1; LIN54; DNAJB4; ZNF643; TAP1; RASGRP2; LDLRAD3</i>	502
БПВ + TH РМЖ	cg20154816 cg02927111 cg18701707 cg12484411 cg20222926 cg02338142 cg06667406 cg13420273 cg22512222 cg17804981 cg13745678	<i>FEZF1; PLIN5; KCNMB2; AASS; HDAC9; ZFAND1; TRHR; PKNOX1</i>	2036
ВБП + TH РМЖ	cg01652244 cg02927111 cg20154816 cg00355315 cg24083274 cg23390595 cg13420273 cg10170774 cg01323371	<i>SSU72; DPPA5; PEX5L; HDAC9; CADPS2; STC1; PKNOX1</i>	502
ОВ + HER2-обогащенный	cg19236995 cg01564068 cg07351262 cg23409370 cg26290926 cg22043168 cg19986472 cg01647795	<i>GSTM4; BDNF; SLC43A1; PATL2; DHX8</i>	247
БПВ + HER2-обогащенный	cg02327465 cg11261264 cg23302638 cg27252154 cg10660854 cg02796790 cg04407660 cg23183932	<i>BIRC5; EDARADD; TAPBPL; QTRT1; PTPRH; SNRPB</i>	247

конечных точек, возрасте пациента, классификации TNM и стадии. В дальнейшем из матрицы данных профилирования были исключены результаты, полученные с FFPE (formalin fixed paraffin-embedded) блоков пациенток, а также кросс-гибридизационные зонды.

Отбор CpG-пар, ассоциированных с ОВ, БПВ или выживаемостью без признаков прогрессирования заболевания (ВБП), производили с помощью одномерной регрессии Кокса (Univariate Cox regression) [18]. Из отобранных CpG-пар в дальнейший анализ были включены прошедшие поправку на множественное тестирование (скорректированный показатель $p < 0,05$, использовали тест Уальда) методом средней доли ложных отклонений гипотез (False Discovery Rate, FDR). В целях отбора наиболее стабильных CpG-пар использовали метод LASSO-регрессии Кокса (Cox LASSO-regression) [19]

пакета программ SurvHiDim [20]. Многомерную регрессию Кокса [21] применяли для расчета сигнатур из CpG-пар и проверки независимости клинических параметров пациенток на эти сигнатуры. С помощью логистической регрессии определяли способность к классификации различных исходов. Для стратификации на высокий и низкий уровень риска использовали медиану. Методом cvROC (cross-validated receiver operative curve) [22] определяли качество построенных моделей и строили ROC-кривые. Наилучшие показатели чувствительности и специфичности определяли с помощью индекса Юдена. Кривые Каплан–Майера были построены с помощью пакета survminer [23]. Для сравнения двух кривых выживаемости использовали критерий Мантеля–Кокса. Для всех этапов выбора маркеров и вычисления сигнатур использовали 10-кратную рандомизированную кросс-валидацию. Все

Таблица 3. Лучшая сигнатура, количество зондов в сигнатуре и показатели cvAUC (cross-validated area under curve; средняя площадь под кривой, полученная на всех этапах кросс-валидации), чувствительности, специфичности и точности для каждого типа выживаемости и молекулярного подтипа РМЖ

Тип выживаемости + молекулярный подтип	Число зондов в комбинации	Гены в комбинации	cvAUC	Чувствительность	Специфичность	Точность
ОВ + LumAB	12	<i>SLC30A7, EXTL2, C15orf41, MIA3, NIPAL3, HEY2, HK1, DIRC3, TMEM41A, SH3BP5L, RFX2</i>	0,797	0,829	0,629	0,805
БРВ + LumAB	6	<i>SLC25A39, BAT2, ZNF417, PSMA6, RG9MTD3, ZNF827</i>	0,831	0,838	0,716	0,828
ВБП + LumAB	9	<i>ABCC5, NIPAL3, HLA-DRB5, RG9MTD3, HIST3H2A, RERE, SPAG5, BAT2, cg13447284</i>	0,761	0,875	0,562	0,836
ОВ + ТН РМЖ	5	<i>cg03512997, LIN54, RASGRP2, LDLRAD3, ZNF643</i>	0,969	0,864	0,939	0,876
БРВ + ТН РМЖ	5	<i>PKNOX1, KCNMB2, ZFAND1, HDAC9, cg13745678</i>	0,834	0,902	0,673	0,865
ВБП + ТН РМЖ	6	<i>DPPA5, cg02927111, PKNOX1, SSU72, CADPS2, PEX5L</i>	0,844	0,952	0,674	0,900
ОВ + HER2-обогатенный	3	<i>GSTM4 (TSS200), GSTM4 (Body), cg26290926</i>	0,898	1	0,883	0,977
БРВ + HER2-обогатенный	2	<i>BIRC5, cg10660854</i>	1	1	1	1
ВБП + HER2-обогатенный	4	<i>SLC43A1, BOD1, cg00297843, KCNN1</i>	1	0,947	1	0,956

вышеуказанные вычисления выполняли с помощью языка статистического программирования R [24].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Исследуемый массив данных TCGA-BRCA состоял из профиля ДНК-метилирования, полученного с помощью чипов HM450 и клинико-патологических характеристик 735 образцов первичных опухолей РМЖ. После исключения образцов из парафиновых блоков суммарно осталось 555 образцов LumA+B подтипа (LumAB), 134 образца ТН-подтипа и 46 образцов HER2-обогатенного подтипа (табл. 1). Перед отбором признаков для дальнейшего анализа из матрицы метилирования исключили кросс-гибридизационные зонды, вследствие чего число зондов снизилось с 485 577 до 456 344 соответственно.

Дальнейшим этапом анализа было применение одномерной регрессии Кокса для поиска сайтов метилирования, которые коррелировали с продолжительностью ОВ, БРВ и ВБП в разных молекулярных подтипах РМЖ. Всего, после первоначального отбора, с учетом *p*-value с поправкой на множественное тестирование, были выделены:

– в подтипах LumAB — 10 433 зондов, ТН — 3214 зондов, HER2-обогатенном — 6471 зонд, ассоциированный с ОВ;

– в подтипах LumAB — 4419 зондов, ТН — 168 зондов, HER2-обогатенном — 483 зонда, ассоциированных с БРВ;

– в подтипах LumAB — 2345 зондов, ТН — 43 зонда, HER2-обогатенном — 3216 зондов, ассоциированных с ВБП.

Для каждого из этих наборов была применена LASSO-регрессия Кокса, что позволило отобрать наиболее важные для анализа CpG-динуклеотиды. На каждом этапе кросс-валидации выявляли разное число таких CpG-пар, и были отобраны такие CpG-пары, которые встречались больше чем в 50% кросс-валидационных разбиений (табл. 2).

Для выбора сочетаний CpG-динуклеотидов, обеспечивающих значимую связь для различных типов выживаемости, были оценены все возможные комбинации (сигнатуры) таких CpG-динуклеотидов в различных

молекулярных подтипах РМЖ. Для каждой клинической конечной точки и молекулярного подтипа РМЖ были получены результаты cvAUC (cross-validated area under curve, усредненная площадь под кривой на каждом этапе кросс-валидации), чувствительности, специфичности и точности для различных комбинаций. Первые 10 комбинаций с высокими показателями cvAUC были проверены на независимость от клинико-патологических характеристик, и диагностические характеристики этих комбинаций вместе с числом зондов и указанием принадлежности генов к зондам представлены в табл. 3.

Наибольшей установленной нами комбинацией является комбинация из двенадцати CpG-динуклеотидов для прогнозирования ОВ в LumAB подтипе, а меньшей — сочетание из двух CpG-динуклеотидов для прогнозирования БРВ в HER2-экспрессирующем подтипе. Для каждой сигнатуры были построены кривые cvROC (cross-validated receiver operative characteristics, на каждом этапе кросс-валидации строится ROC-кривая, в дальнейшем строится усредненная кривая) и кривые Каплан-Майера для отображения диагностического потенциала и оценки функции выживаемости. Комбинации для LumAB имели меньший показатель cvAUC (от 0,76 до 0,83), тогда как комбинации для ТН и HER2-экспрессирующего подтипа показывали высокий уровень cvAUC при меньших количествах сочетаний (от 0,83 до 1) (рис. 1).

Наши комбинации являются независимыми от клинических характеристик (табл. 4), что позволяет использовать показатели риска данных типов выживаемости для любых групп пациенток.

Анализ кривых Каплана-Майера показал статистически значимое (*p* < 0,05) снижение ОВ, БРВ и ВБП в группе пациенток с высоким риском летального исхода, рецидива и прогрессирования болезни по сравнению с группой пациенток с низким риском при всех молекулярных подтипах РМЖ для каждой из выбранных комбинаций (рис. 2).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В настоящем исследовании с помощью методов анализа выживаемости, а также с использованием

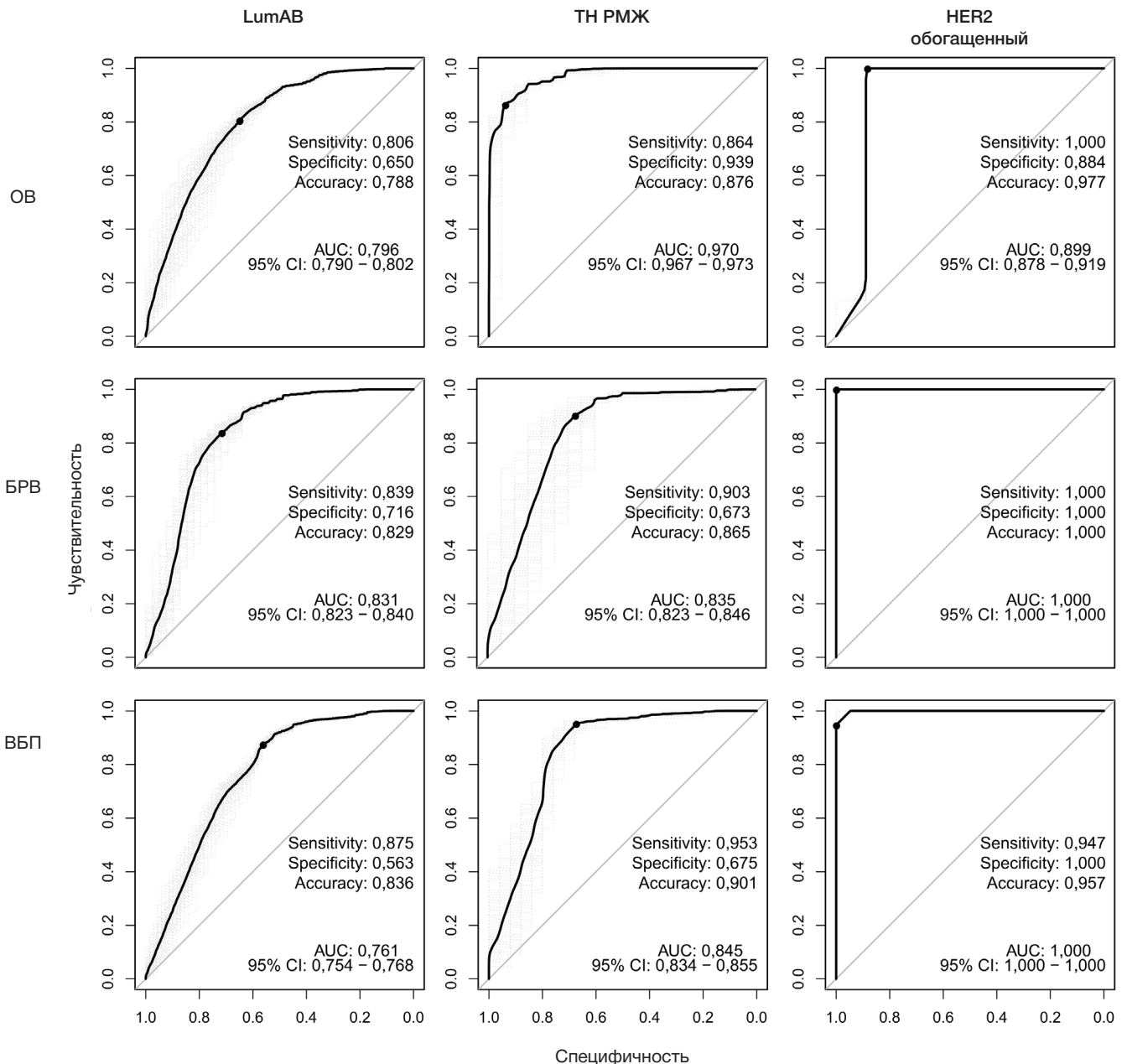


Рис. 1. Кривые cvROC (cross-validated receiver operative curve; на каждом этапе кросс-валидации строится ROC-кривая и в результате — результирующая средняя ROC-кривая) для лучших сигнатур. По оси абсцисс показана чувствительность (от 0 до 1), по оси ординат — специфичность (от 0 до 1), по строкам — типы выживаемости, по столбцам — молекулярные подтипы РМЖ

данных ДНК-метилирования была рассмотрена возможность идентификации сайтов дифференциального метилирования CpG-динуклеотидов для прогнозирования видов выживаемости в различных молекулярных подтипах РМЖ. Подход для расчета дифференциального метилирования с помощью одномерной регрессии Кокса широко используют в различных работах. Так, этот метод применяли для идентификации 249 810 и 249 811 зондов в данных ДНК-метилирования при раке яичников и при РМЖ соответственно [12] и для идентификации зондов в данных ДНК-метилирования при меланоме кожи [25].

Нами показано, что при использовании различных комбинаций (от 2 до 12 CpG-динуклеотидов) можно добиться приемлемого (cvAUC между 0,7 и 0,8), хорошего (от 0,8 до 0,9) и очень хорошего (от 0,9 до 1) качества классификации высокого и низкого риска летального исхода, рецидива и прогрессирования. В рамках данной работы были идентифицированы 47 зондов/генов

(*SLC30A7, EXTL2, C15orf41, MIA3, NIPAL3, HEY2, HK1, DIRC3, TMEM41A, SH3BP5L, RFX2, SLC25A39, BAT2, ZNF417, PSMA6, RG9MTD3, ZNF827, ABCC5, HLA-DRB5, HIST3H2A, RERE, SPAG5, cg13447284, cg03512997, LIN54, RASGRP2, LDLRAD3, ZNF643, PKNOX1, KCNN1, ZFAND1, HDAC9, cg13745678, DPPA5, cg02927111, PKNOX1, SSU72, CADPS2, PEX5L, GSTM4, cg26290926, BIRC5, cg10660854, SLC43A1, BOD1, cg00297843, KCNN1*), метилирование которых ассоциировано с ОБ, БРВ и ВБП, при этом для семи из них (*cg13447284, cg03512997, cg13745678, cg02927111, cg26290926, cg10660854, cg00297843*) не удалось установить принадлежность к генам, а пять из них (*BIRC5, PKNOX1, SPAG5, HDAC9, PSMA6*) ранее были описаны в научной литературе как молекулярные маркеры выживаемости больных с РМЖ на основе данных об количественной экспрессии генов [26–31].

Примечательно, что в рамках данного исследования мы обнаружили такое же число CpG-динуклеотидов

Таблица 4. Результаты многофакторной регрессии Кокса для лучших сигнатур и клинико-патологических характеристик. HR — hazard ratio (показатель относительного риска), P — отвечает за p-val

Переменная / тип выживаемости + молекулярный подтип	ОВ + LumAВ		БРВ + LumAВ		ВБП + LumAВ		ОВ + ТН РМЖ		БРВ + ТН РМЖ		ВБП + ТН РМЖ		ОВ + HER2-обогащенный		БРВ + HER2-обогащенный		ВБП + HER2-обогащенный	
	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P	HR (95% дов. интервал)	P
Показатель Риска	1 (1-1)	< 0.001	1.1 (1.05-1.10)	< 0.001	1.03 (1.02-1.04)	< 0.001	1.1 (1.004-1.2)	< 0.001	1.02 (1.01-1.03)	< 0.001	1.02 (1.001-1.03)	< 0.001	1.0 (1-1)	0.03	1.02 (1.009-1.03)	0.006	1.005 (1.001-1.000)	0.01
Возраст (Меньше медианы/больше медианы)	1,38 (0,84-2,30)	0,201	1,2 (0,58-2,40)	0,638	2 (0,9-3,30)	0,231	0,76 (0,28-2,00)	0,575	0,51 (0,160-1,600)	0,25	0,66 (0,270-1,600)	0,35	6,2 (0,619-62,900)	0,12	1,9 (0,878-2,344)	0,512	2,06 (0,27-15)	0,47
T (T1-2/3-4)	0,87 (0,48-1,60)	0,64	1,1 (0,48-2,60)	0,78	0,77 (0,38-1,60)	0,463	2,52 (0,75-8,40)	0,134	1,06 (0,310-3,600)	0,93	1,50 (0,430-5,200)	0,52	2,6 (0,169-39,900)	0,49	0,9 (0,625-1,932)	0,404	7,00 (0,50-98)	0,14
N (N0/N1-3)	1,23 (0,74-2,00)	0,42	2,1 (0,73-6,00)	0,16	1,1 (0,56-2,20)	0,097	2,95 (0,88-9,90)	0,08	0,96 (0,280-3,300)	0,95	1,14 (0,390-3,300)	0,81	1,1 (0,174-7,500)	0,89	0,34 (0,120-1,500)	0,463	1,09 (0,07-16)	0,94
M (M0/M1)	1,03 (0,51-2,10)	0,94	1,6 (0,66-3,80)	0,30	1,9 (0,90-2,20)	0,076	0,36 (0,09-1,30)	0,125	1,56 (0,360-6,700)	0,55	0,89 (0,290-2,800)	0,83	1,2 (0,089-15,300)	0,90	0,63 (0,105-2,000)	0,376	0,02 (0,00-1,05)	0,053
Стадия (I-III/III-IV)	1,20 (0,71-2,00)	0,49	2,3 (0,95-5,50)	0,06	1,54 (0,80-2,90)	0,19	1,48 (0,36-6,00)	0,58	4,19 (0,94-18,60)	0,06	3,76 (0,97-14,50)	0,055	1,10 (0,09-14,30)	0,91	2,6 (0,78-2,80)	0,16	1,83 (0,12-27,00)	0,66

(зондов HM450) для прогнозирования ОВ и БРВ (по пяти зондам), что и в другой работе [16], но зонды различались по принадлежности к генам. Согласно нашим данным, сигнатура для прогнозирования ВБП состоит из шести зондов; в работе [16] сигнатуру ВБП не вычисляли. Другие исследователи предлагают использовать для прогноза ОВ и БРВ больных с ЭР + РМЖ индивидуальные маркеры промоторного гиперметилирования семи генов (*RASSF1*, *BRCA1*, *PITX2*, *RARB*, *PGR*, *CDH1* и *PCDH10*), а также рассматривают использование панели из трех генов (*GSTP1*, *RASSF1* и *RARB*) для предсказания ОВ, основываясь на анализе литературных данных (по результатам систематического обзора публикаций) [26], в то время как в нашем исследовании применена стратегия формирования панелей из шести, девяти и 12 маркеров метилирования с учетом диагностического потенциала маркеров, установленного статистическим анализом массива экспериментальных данных.

Среди генов, вошедших в полученные нами комбинации, внимание привлекает ген *BIRC5* (кодирует белок — бакуловирусный ингибитор мотива апоптозных повторов 5), который гиперэкспрессируется в большинстве опухолей, в том числе при РМЖ, связан с худшим

прогнозом общей, безрецидивной и метастатической выживаемости. Показано, что использование химиотерапевтических препаратов из группы таксанов может увеличивать экспрессию данного гена [27]. Ген *PKNOX1* (кодирует одноименный белок, расположен на коротком плече 21-й хромосомы, играет важную роль в эмбриональном развитии) является онкосупрессором, а его повышенная экспрессия ассоциирована с худшей выживаемостью [28]. Прогностическим фактором является также повышенная экспрессия гена *SPAG5* (кодирует белок, связанный с аппаратом митотического веретена), ассоциированная с худшим прогнозом ОВ, БРВ и выживаемости без метастазирования только для эстроген-положительных (ЭР+) опухолей молочной железы [29], что также подтверждается нашим исследованием. Результаты исследования, проведенного на образцах ЭР+ РМЖ, показывают, что повышенная экспрессия гена эпигенетического фермента HDAC9 (кодирует белок фермента гистондеацетилазы 9) в опухолях связана с худшим прогнозом БРВ [30]. В нашей работе показана ассоциация аномального метилирования этого гена с выживаемостью для больных с эстроген-отрицательными (ЭР-) опухолями, точнее с ТН РМЖ. Сниженная БРВ

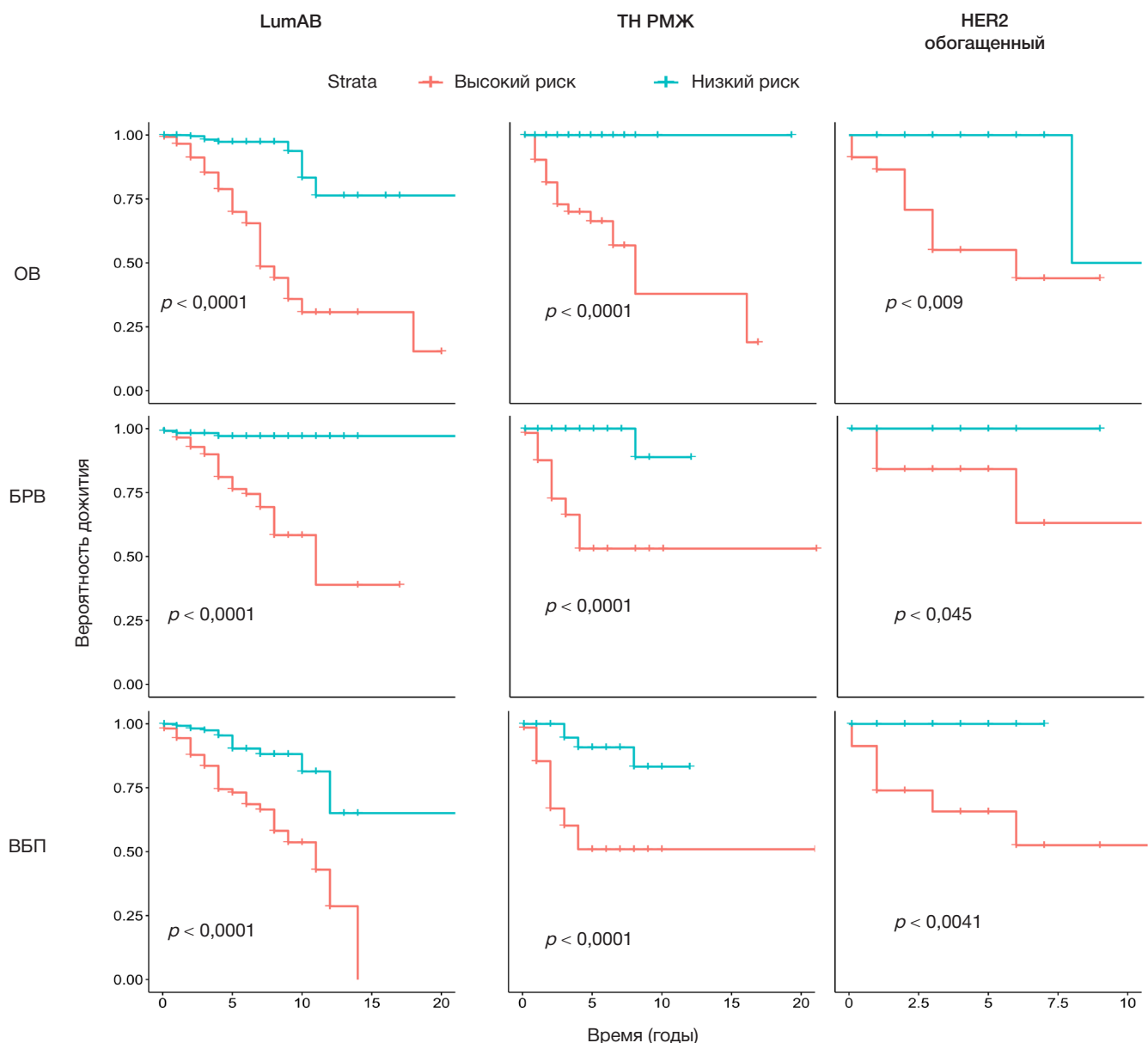


Рис. 2. Кривые Каплан–Мейера для лучших сигнатур. По оси ординат показано время в годах, по оси абсцисс — вероятность жизни (от 0 до 1). Красным цветом указан высокий риск летального исхода, прогрессирования или рецидива, бирюзовым — низкий риск летального исхода, прогрессирования или рецидива. По строкам обозначены типы выживаемости, по столбцам — молекулярные подтипы РМЖ

больных с ЭР+ РМЖ была показана при повышенной экспрессии гена PSMA6 (кодирует белок протеасомной субъединицы альфа-типа-6) [31], что также подтверждают результаты нашего исследования.

Выводы

С помощью методов анализа выживаемости, сочетания различных сайтов метилирования и оценки диагностических показателей были обнаружены молекулярные эпигенетические

сигнатуры для различных подтипов РМЖ. Данную методику можно рекомендовать для поиска сигнатур, характерных для РМЖ, а также для других опухолевых заболеваний. В дальнейшем обнаруженные нами эпигенетические сигнатуры можно использовать для разработки диагностических тест-систем на основе метилчувствительной количественной ПЦР. Такие тест-системы при прохождении клинических испытаний могут стать более дешевой и практичной альтернативой экспрессионным микрочипам без ущерба для диагностической эффективности.

Литература

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71 (3): 209–49. DOI: 10.3322/caac.21660
2. Bernhardt SM, Dasari P, Walsh D, Townsend AR, Price TJ, Ingman WW. Hormonal Modulation of Breast Cancer Gene Expression: Implications for Intrinsic Subtyping in Premenopausal Women. *Front Oncol.* 2016; 6: 241. DOI: 10.3389/fonc.2016.00241.
3. Zaha DC. Significance of immunohistochemistry in breast cancer.

- World J Clin Oncol. 2014; 5 (3): 382–92. DOI: 10.5306/wjco.v5.i3.382. PMID: 25114853; PMCID: PMC4127609.
4. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406 (6797): 747–52. DOI: 10.1038/35021093.
 5. Echeverria GV, Ge Z, Seth S, Zhang X, Jeter-Jones S, Zhou X, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci Transl Med*. 2019; 11 (488): eaav0936. DOI: 10.1126/scitranslmed.aav0936.
 6. Blanchette P, Sivajohanathan D, Bartlett J, Eisen A, Feilotter H, Pezo R, et al. Clinical Utility of Multigene Profiling Assays in Early-Stage Invasive Breast Cancer: An Ontario Health (Cancer Care Ontario) Clinical Practice Guideline. *Curr Oncol*. 2022; 29 (4): 2599–615. DOI: 10.3390/curroncol29040213.
 7. Edwards JR, Yarychivska O, Boulard M, Bestor TH. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin*. 2017; 10: 23. DOI: 10.1186/s13072-017-0130-8.
 8. Vietri MT, D'Elia G, Benincasa G, Ferraro G, Caliendo G, Nicoletti GF, et al. DNA methylation and breast cancer: A way forward (Review). *Int J Oncol*. 2021; 59 (5): 98. DOI: 10.3892/ijo.2021.5278.
 9. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*. 2007; 8 (4): 286–98. DOI: 10.1038/nrg2005.
 10. Karami Fath M, Azarگونjahromi A, Kiani A, Jalalifar F, Osati P, Akbari Oryani M, et al. The role of epigenetic modifications in drug resistance and treatment of breast cancer. *Cell Mol Biol Lett*. 2022; 27 (1): 52. DOI: 10.1186/s11658-022-00344-6.
 11. Lee G, Bang L, Kim SY, Kim D, Sohn KA. Identifying subtype-specific associations between gene expression and DNA methylation profiles in breast cancer. *BMC Med Genomics*. 2017; 10 (Suppl 1): 28. DOI: 10.1186/s12920-017-0268-z.
 12. Hu WL, Zhou XH. Identification of prognostic signature in cancer based on DNA methylation interaction network. *BMC Med Genomics*. 2017; 10 (Suppl 4): 63. DOI: 10.1186/s12920-017-0307-9.
 13. Zhang M, Wang Y, Wang Y, Jiang L, Li X, Gao H, et al. Integrative Analysis of DNA Methylation and Gene Expression to Determine Specific Diagnostic Biomarkers and Prognostic Biomarkers of Breast Cancer. *Front Cell Dev Biol*. 2020; 8: 529386. DOI: 10.3389/fcell.2020.529386.
 14. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A*. 2017; 114 (28): 7414–9. DOI: 10.1073/pnas.1703577114.
 15. de Almeida BP, Apolônio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. *BMC Cancer*. 2019; 19 (1): 219. DOI: 10.1186/s12885-019-5403-0.
 16. Gao Y, Wang X, Li S, Zhang Z, Li X, Lin F. Identification of a DNA Methylation-Based Prognostic Signature for Patients with Triple-Negative Breast Cancer. *Med Sci Monit*. 2021; 27: e930025. DOI: 10.12659/MSM.930025.
 17. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44 (8): e71. DOI: 10.1093/nar/gkv1507.
 18. Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxid Med Cell Longev*. 2021; 2021: 1302811. DOI: 10.1155/2021/1302811.
 19. Utazirubanda JC, Leon T, Ngom P. Variable selection with Group LASSO approach: Application to Cox regression with frailty model. *Commun Stat Simul Comput*. 2021; 50 (3): 881–901. DOI: 10.1080/03610918.2019.1571605.
 20. Bhattacharjee A, Pawar A. SurvHiDim: high dimensional survival data analysis. R package version 0.1.1. 2021. Available from: <https://CRAN.R-project.org/package=SurvHiDim>.
 21. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis — an introduction to concepts and methods. *Br J Cancer*. 2003; 89 (3): 431–6. DOI: 10.1038/sj.bjc.6601119.
 22. LeDell E, Petersen M, van der Laan M. cvAUC: Cross-Validated Area Under the ROC Curve Confidence Intervals. R package version 1.1-4. 2022 Available from: <http://CRAN.R-project.org/package=cvAUC>.
 23. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.6. Available from: <https://CRAN.R-project.org/package=survminer>.
 24. Team RCR Foundation for Statistical Computing; Vienna, Austria: 2015.
 25. Guo W, Zhu L, Zhu R, Chen Q, Wang Q, Chen JQ. A four-DNA methylation biomarker is a superior predictor of survival of patients with cutaneous melanoma. *Elife*. 2019; 8: e44310. DOI: 10.7554/eLife.44310.
 26. de Ruijter TC, van der Heide F, Smits KM, Aarts MJ, van Engeland M, Heijnen VCG. Prognostic DNA methylation markers for hormone receptor breast cancer: a systematic review. *Breast Cancer Res*. 2020; 22 (1): 13. DOI: 10.1186/s13058-020-1250-9.
 27. Dai JB, Zhu B, Lin WJ, Gao HY, Dai H, Zheng L, et al. Identification of prognostic significance of BIRC5 in breast cancer using integrative bioinformatics analysis. *Biosci Rep*. 2020; 40 (2): BSR20193678. DOI: 10.1042/BSR20193678.
 28. Jiang S, Bu X, Tang D, Yan C, Huang Y, Fang K. A tumor suppressor gene-based prognostic classifier predicts prognosis, tumor immune infiltration, and small molecule compounds in breast cancer. *Front Genet*. 2022; 12: 783026. DOI: 10.3389/fgene.2021.783026.
 29. Mohamadizadeh-Hanjani Z, Shahbazi S, Geranpayeh L. Investigation of the SPAG5 gene expression and amplification related to the NuMA mRNA levels in breast ductal carcinoma. *World J Surg Oncol*. 2020; 18 (1): 225. DOI: 10.1186/s12957-020-02001-8.
 30. Linares A, Assou S, Lapiere M, Thouennon E, Duraffourd C, Fromaget C, et al. Increased expression of the HDAC9 gene is associated with antiestrogen resistance of breast cancers. *Mol Oncol*. 2019; 13 (7): 1534–47. DOI: 10.1002/1878-0261.
 31. Li Y, Huang J, Sun J, Xiang S, Yang D, Ying X, et al. The transcription levels and prognostic values of seven proteasome alpha subunits in human cancers. *Oncotarget*. 2017 Jan 17; 8 (3): 4501–19. DOI: 10.18632/oncotarget.13885.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021; 71 (3): 209–49. DOI: 10.3322/caac.21660
2. Bernhardt SM, Dasari P, Walsh D, Townsend AR, Price TJ, Ingman WW. Hormonal Modulation of Breast Cancer Gene Expression: Implications for Intrinsic Subtyping in Premenopausal Women. *Front Oncol*. 2016; 6: 241. DOI: 10.3389/fonc.2016.00241.
3. Zaha DC. Significance of immunohistochemistry in breast cancer. *World J Clin Oncol*. 2014; 5 (3): 382–92. DOI: 10.5306/wjco.v5.i3.382. PMID: 25114853; PMCID: PMC4127609.
4. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406 (6797): 747–52. DOI: 10.1038/35021093.
5. Echeverria GV, Ge Z, Seth S, Zhang X, Jeter-Jones S, Zhou X, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci Transl Med*. 2019; 11 (488): eaav0936. DOI: 10.1126/scitranslmed.aav0936.
6. Blanchette P, Sivajohanathan D, Bartlett J, Eisen A, Feilotter H, Pezo R, et al. Clinical Utility of Multigene Profiling Assays in Early-Stage Invasive Breast Cancer: An Ontario Health (Cancer Care Ontario) Clinical Practice Guideline. *Curr Oncol*. 2022; 29 (4):

- 2599–615. DOI: 10.3390/curroncol29040213.
7. Edwards JR, Yarychivska O, Boulard M, Bestor TH. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin*. 2017; 10: 23. DOI: 10.1186/s13072-017-0130-8.
 8. Vietri MT, D'Elia G, Benincasa G, Ferraro G, Caliendo G, Nicoletti GF, et al. DNA methylation and breast cancer: A way forward (Review). *Int J Oncol*. 2021; 59 (5): 98. DOI: 10.3892/ijo.2021.5278.
 9. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*. 2007; 8 (4): 286–98. DOI: 10.1038/nrg2005.
 10. Karami Fath M, Azargoonjahromi A, Kiani A, Jalalifar F, Osati P, Akbari Oryani M, et al. The role of epigenetic modifications in drug resistance and treatment of breast cancer. *Cell Mol Biol Lett*. 2022; 27 (1): 52. DOI: 10.1186/s11658-022-00344-6.
 11. Lee G, Bang L, Kim SY, Kim D, Sohn KA. Identifying subtype-specific associations between gene expression and DNA methylation profiles in breast cancer. *BMC Med Genomics*. 2017; 10 (Suppl 1): 28. DOI: 10.1186/s12920-017-0268-z.
 12. Hu WL, Zhou XH. Identification of prognostic signature in cancer based on DNA methylation interaction network. *BMC Med Genomics*. 2017; 10 (Suppl 4): 63. DOI: 10.1186/s12920-017-0307-9.
 13. Zhang M, Wang Y, Wang Y, Jiang L, Li X, Gao H, et al. Integrative Analysis of DNA Methylation and Gene Expression to Determine Specific Diagnostic Biomarkers and Prognostic Biomarkers of Breast Cancer. *Front Cell Dev Biol*. 2020; 8: 529386. DOI: 10.3389/fcell.2020.529386.
 14. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A*. 2017; 114 (28): 7414–9. DOI: 10.1073/pnas.1703577114.
 15. de Almeida BP, Apolônio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. *BMC Cancer*. 2019; 19 (1): 219. DOI: 10.1186/s12885-019-5403-0.
 16. Gao Y, Wang X, Li S, Zhang Z, Li X, Lin F. Identification of a DNA Methylation-Based Prognostic Signature for Patients with Triple-Negative Breast Cancer. *Med Sci Monit*. 2021; 27: e930025. DOI: 10.12659/MSM.930025.
 17. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44 (8): e71. DOI: 10.1093/nar/gkv1507.
 18. Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxid Med Cell Longev*. 2021; 2021: 1302811. DOI: 10.1155/2021/1302811.
 19. Utazirubanda JC, Leon T, Ngom P. Variable selection with Group LASSO approach: Application to Cox regression with frailty model. *Commun Stat Simul Comput*. 2021; 50 (3): 881–901. DOI: 10.1080/03610918.2019.1571605.
 20. Bhattacharjee A, Pawar A. *SurvHiDim*: high dimensional survival data analysis. R package version 0.1.1. 2021. Available from: <https://CRAN.R-project.org/package=SurvHiDim>.
 21. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis — an introduction to concepts and methods. *Br J Cancer*. 2003; 89 (3): 431–6. DOI: 10.1038/sj.bjc.6601119.
 22. LeDell E, Petersen M, van der Laan M. *cvAUC*: Cross-Validated Area Under the ROC Curve Confidence Intervals. R package version 1.1-4. 2022 Available from: <http://CRAN.R-project.org/package=cvAUC>.
 23. Kassambara A, Kosinski M, Bieчек P. *survminer*: Drawing Survival Curves using 'ggplot2'. R package version 0.4.6. Available from: <https://CRAN.R-project.org/package=survminer>.
 24. Team RCR Foundation for Statistical Computing; Vienna, Austria: 2015.
 25. Guo W, Zhu L, Zhu R, Chen Q, Wang Q, Chen JQ. A four-DNA methylation biomarker is a superior predictor of survival of patients with cutaneous melanoma. *Elife*. 2019; 8: e44310. DOI: 10.7554/eLife.44310.
 26. de Ruijter TC, van der Heide F, Smits KM, Aarts MJ, van Engeland M, Heijnen VCG. Prognostic DNA methylation markers for hormone receptor breast cancer: a systematic review. *Breast Cancer Res*. 2020; 22 (1): 13. DOI: 10.1186/s13058-020-1250-9.
 27. Dai JB, Zhu B, Lin WJ, Gao HY, Dai H, Zheng L, et al. Identification of prognostic significance of BIRC5 in breast cancer using integrative bioinformatics analysis. *Biosci Rep*. 2020; 40 (2): BSR20193678. DOI: 10.1042/BSR20193678.
 28. Jiang S, Bu X, Tang D, Yan C, Huang Y, Fang K. A tumor suppressor gene-based prognostic classifier predicts prognosis, tumor immune infiltration, and small molecule compounds in breast cancer. *Front Genet*. 2022; 12: 783026. DOI: 10.3389/fgene.2021.783026.
 29. Mohamadizadeh-Hanjani Z, Shahbazi S, Geranpayeh L. Investigation of the SPAG5 gene expression and amplification related to the NuMA mRNA levels in breast ductal carcinoma. *World J Surg Oncol*. 2020; 18 (1): 225. DOI: 10.1186/s12957-020-02001-8.
 30. Linares A, Assou S, Lapierre M, Thouennon E, Duraffourd C, Fromaget C, et al. Increased expression of the HDAC9 gene is associated with antiestrogen resistance of breast cancers. *Mol Oncol*. 2019; 13 (7): 1534–47. DOI: 10.1002/1878-0261.
 31. Li Y, Huang J, Sun J, Xiang S, Yang D, Ying X, et al. The transcription levels and prognostic values of seven proteasome alpha subunits in human cancers. *Oncotarget*. 2017 Jan 17; 8 (3): 4501–19. DOI: 10.18632/oncotarget.13885.