

УДК 577.21

ОСНОВАННЫЕ НА МОЛЕКУЛЯРНОМ ПУТИ EGFR ГЕННЫЕ ПОДПИСИ КЛИНИЧЕСКИ ЗНАЧИМЫХ МУТАЦИЙ ПРИ РАКЕ МОЛОЧНОЙ ЖЕЛЕЗЫ, ЛЕГКИХ, ЩИТОВИДНОЙ ЖЕЛЕЗЫ И ПРИ МЕЛАНОМЕ

© 2021 М. Раевский^{1,2}, М. Сорокин^{3,4}, У. Владимирова³, М. Сунцова³, В. Ефимов²,
А. Гаража², А. Дробышев³, А. Моисеев³, П. Румянцев⁵, З. Ли⁶, А. Буздин^{1,2,3,7*}

¹ Omicsway Corp., 340 S Lemon Ave, 6040, Walnut, 91789 CA, USA; E-mail: buzdin@oncobox.com

² Московский физико-технический институт, 141701 Долгопрудный, Московская область, Россия

³ Первый Московский государственный медицинский университет имени И.М. Сеченова, 119991 Москва, Россия

⁴ Онкобокс Ltd., 121205 Москва, Россия

⁵ Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр эндокринологии» Министерства здравоохранения Российской Федерации, 117312 Москва, Россия

⁶ Кафедра патологии и лабораторной медицины Калифорнийского университета в Лос-Анджелесе, 90095 Лос-Анджелес, Калифорния, США

⁷ Институт биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова, 117997 Москва, Россия

Поступила в редакцию 30.07.2021

После доработки 04.10.2021

Принята к публикации 04.10.2021

Гены *EGFR*, *BRAF*, *PIK3CA* и *KRAS* играют важную роль в молекулярном пути EGFR и при раке часто несут активирующие диагностические мутации, предсказывающие ответ на многие таргетные терапевтические препараты. Тем не менее связь между этими мутациями и особенностями генной экспрессии систематически не изучалась. Мы исследовали такие связи тремя способами. Во-первых, исследовали экспрессию этих генов в опухолях в зависимости от наличия мутаций с использованием базы данных профилей РНК-секвенирования проекта The Cancer Genome Atlas ($n = 3660$). Во-вторых, сравнили мутационный статус с уровнем активации молекулярного пути EGFR. В-третьих, сопоставили мутационный статус с расчетным уровнем экспрессии транскрипционной подписи, полученной из дифференциально экспрессируемых генов пути EGFR между опухолями, мутантными и референсными по данным генам. Активация пути EGFR оказалась связана с мутациями генов *BRAF* (рак щитовидной железы, меланома) и *PIK3CA* (рак молочной железы). Было показано, что уровни экспрессии генных подписей связаны с мутациями в *BRAF* (рак щитовидной железы, меланома), *EGFR* (плоскоклеточный рак легкого), *KRAS* (колоректальный рак), *PIK3CA* (рак молочной железы), однако приемлемый биомаркерный потенциал со значением AUC > 0,7 (0,809) мы наблюдали лишь для подписи гена *BRAF* при раке щитовидной железы. Эта подпись была затем проверена на независимом наборе литературных данных ($n = 127$, свежемороженые образцы, AUC 0,912), а также на экспериментальном наборе данных, полученном в Эндокринологическом научном центре г. Москвы ($n = 42$, образцы парафинизированной ткани, AUC 0,822). Полученные данные показывают, что профили генной экспрессии могут использоваться для надежной детекции мутации *BRAF* V600E при папиллярном подтипе рака щитовидной железы. Результаты демонстрируют применимость количественного анализа генной экспрессии для четкого определения статуса образца по драйверным онкогенным мутациям.

КЛЮЧЕВЫЕ СЛОВА: мутации, EGFR, *BRAF* V600E, транскриптомный биомаркер, РНК-секвенирование, папиллярный рак щитовидной железы.

DOI: 10.31857/S0320972521110099

ВВЕДЕНИЕ

Гены *EGFR*, *BRAF*, *PIK3CA* и *KRAS* играют важную роль в сигнальном пути EGFR [1], в котором рецепторная тирозинкиназа EGFR явля-

ется первичным получателем сигнала, а продукты остальных генов — его нижестоящими передатчиками и регуляторами. Активирующие мутации в этих генах связаны с ответом на таргетные препараты гефитиниб/панитумумаб/эрло-

Принятые сокращения: АЛ — аденокарцинома легких; КРР — колоректальный рак; МКРЛ — мелкоклеточный рак легкого; РМЖ — рак молочной железы; РЩЖ — рак щитовидной железы; WT — дикий тип.

* Адресат для корреспонденции.

тиниб/цетуксимаб, алпелисиб, дабрафениб/вемурафениб (табл. 1). Так, активирующие мутации в гене *EGFR* являются предпосылкой для эффективного применения некоторых ингибиторов *EGFR* [2]. В свою очередь, активирующие мутации в генах нижестоящих передатчиков сигнала свидетельствуют о потенциально низкой эффективности *EGFR*-направленной терапии [3]. Например, для гена *BRAF* наиболее частой активирующей мутацией является *BRAF*^{V600E} [4], которая часто встречается при меланоме (~ 35–50% случаев [5]) и при папиллярном раке щитовидной железы (~ 51% [6]). Помимо блокаторов *EGFR*, были разработаны также специфические ингибиторы мутантной формы серин-треониновой киназы *BRAF* [7] и фосфоинозитид-3-киназы *PIK3CA* [8]. На графе молекулярного пути *EGFR* белки *PIK3CA* и *BRAF* расположены как нижележащие передатчики сигнала относительно представителей семейства *RAS* [9] и, следовательно, они могут быть эффективны даже в опухолях с мутированными *RAS*-белками.

С другой стороны, профили экспрессии различных молекул РНК оказались эффективными маркерами чувствительности опухоли к таргетным терапевтическим препаратам, согласно данным микрочипового профилирования [19] и РНК-секвенирования (RNA-seq) [20]. Профили РНК-секвенирования могут служить альтернативой иммуногистохимии в отдельных опухолях и могут служить маркерами ответа на таргетные препараты [21]. Кроме того, в качестве биомаркеров нового поколения могут быть использованы относительные уровни активации молекулярных путей, рассчитанные с использованием широкомасштабных данных генной экспрессии [22] для алгоритмической оценки эффективности противораковых препаратов [23]. Агрегация отдельных генов в пути или подписи приводит к более надежным результатам экспрессионного профилирования, что было предсказано теоретически [24] и доказано на экспериментальном материале опухолевых тканей [25].

Упомянутые выше результаты позволяют рассматривать анализ данных профилирования

Таблица 1. Одобренные для клинического использования таргетные противоопухолевые препараты, для которых известны диагностические активирующие мутации в генах *EGFR*, *BRAF*, *PIK3CA* и *KRAS*

Официальное название	Одобрены для видов рака	Сопутствующие биомаркеры	Ссылки
Цетуксимаб	колоректальный рак, рак головы и шеи	дикий тип <i>KRAS</i> , <i>NRAS</i> , <i>BRAF</i> (для КРР)	[10, 11]
Панитумумаб	колоректальный рак	дикий тип <i>KRAS</i> , <i>NRAS</i> , <i>BRAF</i>	[10]
Афатиниб	рак легких, рак молочной железы	активирующие мутации в <i>EGFR</i> (для РЛ): делеции экзона 19 или мутации экзона 21, L858R (для РЛ)	[12]
Гефитиниб	рак легких	активирующие мутации в <i>EGFR</i> : делеции экзона 19 или мутации экзона 21, L858R	[13]
Эрлотиниб	рак легких, панкреатический рак	активирующие мутации в <i>EGFR</i> : делеции экзона 19 или мутации экзона 21, L858R (для АЛ)	[12]
Дакомитиниб	рак легких	активирующие мутации в <i>EGFR</i> : делеции экзона 19 или мутации экзона 21, L858R	[12]
Осимертиниб	рак легких	активирующие мутации в <i>EGFR</i> : делеции экзона 19 или мутации экзона 21, L858R, также мутация T790M; но нет мутации экзона 20 C797S	[14]
Дабрафениб	рак легких, рак щитовидной железы	активирующая мутация в <i>BRAF</i> : V600E	[15]
Вемурафениб	меланома	активирующая мутация в <i>BRAF</i> : V600E	[16]
Алпелисиб + фулвестрант	рак щитовидной железы, папиллярный подтип	активирующая мутация в <i>BRAF</i> : V600E	[17]
Алпелисиб	рак молочной железы	активирующие мутации в <i>PIK3CA</i> : экзон 7 C420R; экзон 9 E542K, E545A, E545D (только 1635G>T), E545G, E545K, Q546E, Q546R; экзон 20 H1047L, H1047R, H1047Y	[18]

Примечание. АЛ – аденокарцинома легкого, РК – рак легкого, КРР – колоректальный рак.

РНК как новый перспективный подход в клинической онкологии, который в будущем может улучшить или заменить некоторые из нынешних диагностических методов [26]. Поэтому представляется важным также получить возможность идентифицировать клинически значимые мутации на основании данных генной экспрессии. Такие экспрессионные подписи действительно были реконструированы для некоторых мутаций, например, для амплификации гена транскрипционного фактора MYCN в нейробластоме [27], однако же эта задача остается в значительной степени нерешенной для активирующих мутаций генов *BRAF*, *PIK3CA*, *EGFR* и *KRAS*.

В настоящей работе мы исследовали транскрипционную проекцию мутаций *BRAF*, *PIK3CA*, *EGFR* и *KRAS* тремя способами. Во-первых, сравнивали экспрессию самих этих генов в опухолях, где они мутированы или имеют статус дикого типа. Сравнение проводили с использованием базы данных проекта The Cancer Gene Atlas (TCGA), $n = 3660$. Во-вторых, мутантный статус по каждому из генов сопоставляли с уровнем активации пути EGFR. В-третьих, мутации ассоциировали с генными подписями, полученными из наборов дифференциально экспрессирующихся генов пути EGFR между опухолями с мутантным статусом и статусом дикого типа.

Мы обнаружили, что мутации гена *EGFR* были связаны с повышенной экспрессией этого гена в аденокарциноме легкого (АЛ, $p = 0,021$), но не при плоскоклеточном раке легкого (ПКРЛ). Мутации *KRAS* были связаны с его транскрипционной активацией при колоректальном раке (КРР, $p = 4 \times 10^{-6}$), а мутация *BRAF* была связана с его повышенной экспрессией в меланоме ($p = 1 \times 10^{-7}$). Активация молекулярного пути EGFR была связана с мутациями *BRAF* при раке щитовидной железы (РЩЖ, $p = 1 \times 10^{-8}$), при меланоме ($p = 0,006$), а также с мутациями *PIK3CA* при раке молочной железы (РМЖ, $p = 2 \times 10^{-5}$). С другой стороны, генные экспрессионные подписи оказались статистически значимо связаны с мутациями *BRAF* при РЩЖ ($p < 1 \times 10^{-8}$) и при меланоме ($p < 1 \times 10^{-8}$), с мутациями *EGFR* для МКРЛ ($p = 2 \times 10^{-4}$), с мутациями *KRAS* при КРР ($p = 1 \times 10^{-8}$) и с мутациями *PIK3CA* при РМЖ ($p = 1 \times 10^{-9}$). Однако же большинство этих показателей имеют недостаточно сильный биомаркерный потенциал, что отражается в значениях ROC AUC (receiver operating characteristic, area under curve) $< 0,7$. Только лишь для подписи гена *BRAF* при РЩЖ был зафиксирован высокий биомаркерный потенциал со значением AUC, равным 0,809. Эту подпись мы проверили на независимом наборе литера-

турных данных из работы Yoo et al. [28] ($n = 127$, AUC 0,912, свежемороженые образцы) и наборе экспериментальных данных, полученных в Федеральном государственном бюджетном учреждении «Национальный медицинский исследовательский центр эндокринологии» Министерства здравоохранения Российской Федерации (Москва, РФ; далее в тексте – ЭНЦ) ($n = 42$, AUC 0,822, FFPE-образцы). Полученные нами данные свидетельствуют, что профили РНК-секвенирования могут быть использованы для надежной идентификации *BRAF^{V600E}* при раке щитовидной железы.

МАТЕРИАЛЫ И МЕТОДЫ

Экспериментальные клинические образцы биоматериала. Исследование проводилось в соответствии с этическими принципами Хельсинкской декларации. Ретроспективные биологические образцы были получены от пациентов с диагностированной первичной опухолью щитовидной железы, перенесших операцию в ЭНЦ. От всех пациентов или их законных представителей было получено письменное информированное согласие на участие в исследовании. Процедура получения письменного информированного согласия и дизайн исследования были одобрены локальным этическим комитетом ЭНЦ под председательством Е.Н. Андреевой. Для всех пациентов было получено согласие на публикацию информации о поле, возрасте, гистологическом типе опухоли, диагнозе и молекулярных данных, включая данные РНК-секвенирования, статусе мутации *BRAF*, но не о данных полногеномного или полноэкзомного секвенирования.

Биопрепараты представляли собой фиксированные в формалине парафинизированные (FFPE) блоки опухолевой ткани, которые были исследованы патологом для определения гистологического типа и оценки доли опухолевых клеток. В этом исследовании использовались только блоки FFPE, содержащие не менее 60% опухолевых клеток. В общей сложности для дальнейшего молекулярного скрининга было получено 42 образца папиллярного типа рака щитовидной железы, отвечающих вышеуказанным критериям. Клинические аннотации опухолевых образцов приведены в таблице в Приложении к статье.

Оценка статуса мутации *BRAF^{V600E}*. Статус мутации *BRAF^{V600E}* определяли с помощью аллель-специфической ПЦР с флуоресцентным зондом с двойной меткой согласно протоколу, опубликованному в работе [29]. Условия амплифика-

ции: начальная денатурация при 95 °С в течение 2 мин, затем 50 циклов по следующей программе – денатурация (94 °С, 10 с), отжиг и элонгация (60 °С, 15 с). Статус мутации считали положительным, если содержание мутантного аллеля составляло 3% и более от содержания аллеля дикого типа.

Подготовка транскриптомных библиотек и РНК-секвенирование. Библиотеки РНК были созданы, отсекуены и первично проанализированы согласно протоколам, ранее использованным для создания Атласа профилей РНК-секвенирования нормальных тканей человека [30]. Выделение РНК проводили с использованием набора RecoverAll™ Total Nucleic Acid Isolation Kit («Invitrogen», США), а длину выделенной РНК (RIN) измеряли с помощью биоанализатора Agilent 2100 («Agilent», США). Концентрацию РНК измеряли с помощью набора Qubit RNA Assay Kit («Thermo Fisher Scientific», США). Рибосомную РНК удаляли с использованием набора KAPA RNA Hyper Kit with RiboErase («Roche», Швейцария). Концентрации библиотек и распределение длин амплифицированных кДНК измеряли с помощью набора Qubit dsDNA HS Assay kit («Life Technologies», США) и Agilent TapeStation («Agilent») соответственно. Образцы секвенировали с использованием прибора Illumina HiSeq 3000 по протоколу для одноконцевых прочтений со средней длиной 50 п.н. с глубиной прочтения около 30 миллионов на образец. Первичную проверку качества данных отсекуенных библиотек проводили с помощью программы Illumina SAV. Демультиплексирование производили с использованием программы Illumina Bcl2fastq2 v2.17 согласно публикации [30].

Обработка данных РНК-секвенирования. Полученные прочтения РНК-секвенирования в формате файлов FASTQ обрабатывали с помощью программы STAR aligner [31] в режиме GeneCounts с использованием аннотации транскриптома человека Ensembl (сборка GRCh38, аннотация GRCh38.89). Идентификаторы генов Ensembl были преобразованы в названия генов согласно номенклатуре HGNC в соответствии с версией от 16 апреля 2021 (<https://www.genenames.org>). Уровни генной экспрессии были установлены для 36 596 аннотированных генов с HGNC-идентификаторами. Контроль качества полученных данных секвенирования проводили с использованием программы NCBI MAGIC [32]. Затем уровни прочтений индивидуальных генов были нормализованы с помощью метода квантильной нормализации [33]. Анализ дифференциальной экспрессии выполняли с использованием исходных

уровней генной экспрессии при помощи программы DESeq2, визуализацию данных проводили с помощью пакета EnhancedVolcano R.

Литературные транскриптомные данные с аннотацией мутационного статуса. Мы использовали транскриптомные данные, для которых также имелась аннотация мутационного статуса, полученные на открытом портале проекта TCGA (<https://portal.gdc.cancer.gov/>) [34]. Всего был проанализирован 491 образец РЩЖ, из которых 42,4% опухолей имели мутацию *BRAF^{V600E}*; 548 образцов меланомы, из которых 27,6% имели мутацию *BRAF^{V600E}*; 579 образцов КРР, из которых 37,1% имели мутацию *KRAS*; 1005 образцов РЛ, из которых 10,5% имели мутацию *EGFR*; 1037 образцов РМЖ, из которых 16,0% имели мутацию *PIK3CA*.

Для валидационных исследований использовали другой открытый набор данных РНК-секвенирования с аннотированным статусом мутации *BRAF^{V600E}* для образцов папиллярного РЩЖ, ранее опубликованный в работе Yoo et al. [28].

Аннотация молекулярного пути EGFR и оценка уровня его активации. В этом исследовании мы использовали две альтернативно аннотированные версии пути EGFR: «biocarta_egf_signaling_Main_Pathway» из базы данных Biocarta версии 1.2 [35] и «EGF_Pathway» из базы данных Qiagen Pathway Central [36]. Алгоритмически проиндексированные аннотации молекулярных функций для узлов пути были взяты из ранее опубликованных нами данных молекулярных путей [36]. Уровни активации молекулярных путей (PAL, Pathway activation levels) рассчитывали с помощью биоинформатической платформы Онкобокс (<https://oncobox.com/ru/>) [37].

Конструирование и оценка эффективности генных подписей. Генные подписи были сконструированы с использованием данных TCGA независимо для каждого типа рака на основе дифференциально экспрессируемых генов, относящихся к молекулярному пути EGFR в каждой из версий. Генные подписи оценивали по формуле:

$$GSS = \sum_{i=1}^n w_i g_i,$$

где *GSS* – это оценка значения генной подписи, *n* – общее число генов в подписи, *w_i* и *g_i* – двоичный логарифм отношения уровня экспрессии в образце и в контрольной группе и уровень экспрессии данного гена, включенного в подпись, соответственно. Таким образом, сконструированные генные подписи представляют собой

взвешенное среднее арифметическое экспрессии генов, включенных в соответствующую генную подпись.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Анализ ассоциаций между мутациями и особенностями транскриптома. Путь EGFR играет решающую роль в развитии различных солидных опухолей, включая НМРЛ, РМЖ, РЩЖ и некоторые другие [13]. В настоящем исследовании мы ставили своей целью обнаружение специфических транскриптомных особенностей, связанных с наличием онкогенных мутаций в генах *BRAF*, *PIK3CA*, *EGFR* и *KRAS* тремя способами. Сначала мы проанализировали экспрессию этих генов в опухолях, имеющих по ним мутантный статус и статус дикого типа. Затем статус мутации этих генов сопоставляли с уровнем активации EGFR-пути. Далее мутации ассоциировали с генными подписями, сформированными из генов пути EGFR, которые дифференциально экспрессировались между опухолями с

мутированным статусом и статусом дикого типа для каждого из анализируемых генов.

Для решения указанной задачи мы использовали профили РНК-секвенирования, аннотированные мутационным статусом опухолей, из базы данных проекта TCGA [34]. Для нескольких видов рака мы исследовали, как уровень активации пути EGFR связан с онкогенными мутациями (табл. 2, 3): *KRAS* при КРП, *EGFR* при РЛ, *BRAF* при РЩЖ и меланоме, *PIK3CA* при РМЖ.

Для расчета уровня активации молекулярного пути (PAL) мы использовали два альтернативных варианта пути EGFR: «Основной путь передачи сигналов EGFR biocarta» из базы данных Biocarta и «Путь EGFR» из базы данных Qiagen Pathway Central, которые были единообразно алгоритмически аннотированы в отношении функциональной роли каждого из узлов пути [35, 36, 43].

Метод PAL для оценки активации молекулярного пути был выбран как отличающийся большей точностью и надежностью обнаружения новых биомаркеров по сравнению с уровнем экспрессии отдельных генов, что было по-

Таблица 2. Результаты ассоциативного исследования

Тип мутации	Тип опухоли	Генная экспрессия	Версия пути EGFR		Значения генной подписи для пути EGFR		Валидация генной подписи	
			Biocarta	Qiagen	Biocarta	Qiagen	литературная (PRJEB11591) [28]	экспериментальная (GSE138042) [38]
<i>BRAF</i> V600E [40]	РЩЖ	0,69 (0,073)	0,564 (0,015)	0,651 (9,7×10 ⁻⁹)	–	0,809 (2,2×10 ⁻¹⁶)	0,912 (3,5×10 ⁻¹⁵)	0,822 (3,7×10 ⁻⁴)
	меланома	0,69 (1,1×10 ⁻⁷)	0,568 (0,014)	0,576 (0,006)	–	0,692 (4,0×10 ⁻¹³)	–	–
<i>EGFR</i> : L858R и делеции в 19 и 28 экзонах [40]	АЛ	0,567 (0,021)	0,518 (0,59)	0,536 (0,34)	–	0,531 (0,43)	–	–
	ПРЛ	0,562 (0,95)	0,509 (0,83)	0,507 (0,89)	–	0,675 (1,7×10 ⁻⁴)	–	–
<i>KRAS</i> : мутации в 12, 13, 59, 61, 117, 146 кодонах [41]	КРП	0,69 (3,8×10 ⁻⁶)	0,534 (0,19)	0,508 (0,74)	–	0,646 (1,1×10 ⁻⁸)	–	–
<i>PIK3CA</i> [42]*	РМЖ	0,541 (0,46)	0,513 (0,6)	0,606 (1,5×10 ⁻⁵)	–	0,648 (1,3×10 ⁻⁹)	–	–

Примечание. Каждая строка представляет результаты сравнения между группами с мутантным статусом и статусом дикого типа по соответствующему гену для выбранных клинически значимых мутаций, в каждой ячейке показано значение ROC AUC (первое значение) и значение *p* (дано в скобках) согласно статистическому тесту по Вилкоксоу. Значения ROC AUC, превышающие 0,8, выделены жирным шрифтом.

* Экзон 7 C420R; экзон 9 E542K, E545A, E545D [1635G>T], E545G, E545K, Q546E, Q546R и экзон 20 H1047L, H1047R, H1047Y. АЛ – аденокарцинома легкого, РЩЖ – рак щитовидной железы, ПРЛ – плоскоклеточный рак легкого, КРП – колоректальный рак.

Таблица 3. Соотношение образцов с мутантным статусом и статусом дикого типа для каждого из проведенных сравнений

Ген	Рассматриваемые мутации	Тип опухоли	Источник данных	Количество образцов, статус дикого типа**	Количество образцов, мутированный статус**
<i>BRAF</i>	V600E [44]	РЩЖ	TCGA THCA	283 (57,6%)	208 (42,4%)
			PRJEB11591 [28]	75 (59,1%)	52 (40,9%)
			GSE138042 [38]	27 (64,3%)	15 (35,7%)
<i>EGFR</i>	L858R; делеции 19 и 28 экзона [40]	меланома	TCGA SKCM	397 (72,4%)	151 (27,6%)
			АЛ	TCGA LUAD	444 (86,9%)
<i>KRAS</i>	мутации в 12, 13, 59, 61, 117, 146 кодонах [41]	ПРЛ	TCGA LUSC	455 (92,1%)	39 (7,9%)
			КРР	TCGA COAD READ	364 (62,9%)
<i>PIK3CA</i>	[42]*	РМЖ	TCGA BRCA	871 (84,0%)	166 (16,0%)

Примечание. * Экзон 7 C420R; экзон 9 E542K, E545A, E545D (только 1635G>T), E545G, E545K, Q546E, Q546R и экзон 20 H1047L, H1047R, H1047Y.

** В скобках указан процент от общего числа образцов.

казано ранее [22]. Он также обладает большей стабильностью при независимом воспроизведении результатов [24]. Положительное значение PAL означает активацию молекулярного пути по сравнению с контрольной группой образцов, а отрицательное значение указывает на подавление пути, тогда как нулевое значение говорит об отсутствии изменений в активации пути [24]. В свою очередь, абсолютная величина PAL отражает степень изменения активации или ингибирования пути [22]. Таким образом, более высокое значение PAL отражает большую активацию пути, и наоборот [24].

Исследованные в данной работе генные подписи были сконструированы для каждой из двух версий пути EGFR с использованием следующих критериев: участвующие гены должны быть (i) включены в путь EGFR и (ii) дифференциально экспрессироваться ($p_{adj} - value < 0,05$ и $|\log_2 FoldChange| > 1$) в мутированных образцах по сравнению с образцами дикого типа данного вида опухоли в отношении рассматриваемого онкогена. Значения генной подписи рассчитывали следующим образом:

$$GSS = \sum_{i=1}^n w_i g_i / n,$$

где GSS – это значение генной подписи, n – число генов в подписи, w_i и g_i – это двоичный логарифм отношения уровня экспрессии в образце к

контрольной группе и нормализованное число генных прочтений для гена i из данной подписи соответственно.

Связь с мутационным статусом оценивали с использованием двух различных статистических подходов. Во-первых, оценивали значимость различий между двумя сравниваемыми группами с использованием критерия Вилкоксона. Во-вторых, на основании значений ROC AUC определяли, можно ли использовать рассматриваемый показатель в качестве высококачественного биомаркера наличия данной мутации в образце. Площадь под ROC-кривой (AUC) – это показатель устойчивости биомаркера, который зависит от его чувствительности и специфичности [45]. Он изменяется в диапазоне от 0,5 до 1, а порог дискриминации обычно устанавливается на уровне 0,7, где элементы с более высоким значением AUC считаются биомаркерами хорошего качества [46]. Оценка AUC широко используется для аннотации молекулярных маркеров в онкологии [38]. Таким образом, оценка значения p согласно критерию Вилкоксона и ROC AUC помогают ответить на два вопроса: (1) регулируется ли рассматриваемый параметр по-разному среди двух сравниваемых групп и (2) может ли он служить для прогнозирования наличия мутации в конкретной опухоли. В нашем исследовании был обнаружен ряд ассоциаций для шести изучаемых видов опухолей.

Рак щитовидной железы. Папиллярный подтип РЩЖ был исследован в отношении мута-

пии *BRAF^{V600E}*. Мы обнаружили, что экспрессия *BRAF* существенно не различалась между мутированными и образцами РЦЖ дикого типа по этому гену (Wilcoxon test p -value = 0,073, ROC AUC 0,69; рис. S1 в Приложении). Тем не менее уровни PAL *EGFR* были значительно повышены в опухолях с мутацией *BRAF* по сравнению с опухолями дикого типа ($p = 0,015$ и $9,7 \times 10^{-9}$ для версий пути *EGFR* Biocarta и Qiagen соответственно; рис. S2 в Приложении). При этом, однако, значения PAL не могли служить качественными биомаркерами, позволяющими отличать индивидуальные опухоли с мутацией *BRAF* от опухолей дикого типа – они показали относительно низкие значения ROC AUC (0,564 и 0,651 для версий Biocarta и Qiagen соответственно; рис. S2 в Приложении).

Дальнейший анализ пути *EGFR* выявил дифференциально экспрессируемые гены только для версии Qiagen: экспрессия генов *FOS* и *RHOV* была подавлена, тогда как экспрессия *EPS8*, *RND3*, *CAV2*, *CAV1* и *MUC1* – активирована в образцах РЦЖ с мутацией *BRAF*. Для дифференциально экспрессируемых генов была сконструирована генная подпись, для которой провели оценку ее биомаркерного потенциала.

Значения генной подписи статистически значимо различались между образцами РЦЖ с мутантным статусом и статусом дикого типа ($p = 2,2 \times 10^{-16}$), при этом значение ROC AUC составило 0,809 (рис. 1). Это свидетельствует в пользу того, что такая подпись может служить качественным биомаркером наличия мутации *BRAF^{V600E}* в образце РЦЖ.

Меланома. Были исследованы образцы меланомы кожи на наличие связи с мутацией *BRAF^{V600E}*. Мы обнаружили, что экспрессия *BRAF* была значительно выше в мутировавших опухолях ($p = 1,1 \times 10^{-7}$; рис. S3 в Приложении), но при этом уровень экспрессии этого гена не мог служить достаточно качественным биомаркером (ROC AUC = 0,69). Уровни активации путей *EGFR* Biocarta и Qiagen были также значительно повышены в мутированных меланомах ($p = 0,014$ и $0,006$ соответственно; рис. S4, *a* в Приложении), но при этом также не были расценены как значимые биомаркеры (AUC = 0,568 и 0,576 соответственно; рис. S4, *b* в Приложении).

Генные подписи содержали только один ген с повышенной экспрессией (*STAT1*) для версии пути *EGFR* Biocarta; девять генов с повышенной экспрессией (*STAT1*, *JAK2*, *EGFR*, *DOK2*, *VAV1*, *SHC3*, *RHOH*, *PRKCG*, *PIK3R5*) и восемь генов с пониженной экспрессией (*VAV2*, *MYC*, *MAP2K2*, *JUN*, *PIK3R2*, *RRAS2*, *RHOD*, *FOS*) для версии пути *EGFR* Qiagen (рис. S5 в Приложении). Обе

версии генных подписей продемонстрировали существенную разницу в значениях между мутированными и немутированными образцами ($p = 0,0081$ и $4,0 \times 10^{-13}$ соответственно), но показали невысокую прогностическую способность со значениями AUC, равными 0,586 и 0,692 соответственно (рис. S6 в Приложении).

Рак легкого. В раке легкого активирующие мутации в *EGFR* (*EGFR^{L858R}* и делеции в 19 и 28 экзонах) рассматривали для двух типов: аденокарцинома легкого (АЛ) и плоскоклеточный рак легких (ПРЛ).

Ген *EGFR* был статистически значимо сверхэкспрессирован в мутированных опухолях при АЛ, но не при ПРЛ ($p = 0,021$ и $0,95$ соответственно; рис. S7 в Приложении), однако экспрессия *EGFR* была слабым биомаркером даже для АЛ (AUC 0,567; рис. S7, *b* в Приложении). В то же время уровни активации обеих версий пути *EGFR* не были дифференциальными и оказались плохими биомаркерами как для АЛ, так и для ПРЛ (рис. S8 в Приложении).

Для АЛ удалось сгенерировать только одну генную подпись, которая содержала один ген с повышенной экспрессией (*EGFR*) и два – с пониженной (*VAV3* и *PIK3C2G*) (рис. S9, S10, *a* в Приложении). При этом указанная подпись не давала значимо различающихся значений между группами (рис. S11, *a* в Приложении) и не могла служить значимым биомаркером (рис. S11, *b* в Приложении).

Для ПРЛ также была выделена только одна генная подпись, включающая два гена с пониженной экспрессией (*HRAS* и *RELB*) из *EGFR* Qiagen пути (рис. S10, *b* в Приложении). Указанная подпись являлась дифференциальной ($p = 1,7 \times 10^{-4}$; рис. S12, *a* в Приложении), но не могла служить надежным биомаркером (AUC 0,675; рис. S12, *b* в Приложении).

Колоректальный рак. Для колоректального рака (КРР) оценивали ассоциации с мутациями гена *KRAS* (кодоны 12, 13, 59, 61, 117 и 146). Экспрессия *KRAS* была значительно выше в мутированных образцах ($p = 3,8 \times 10^{-6}$; рис. S13, *a* в Приложении), но при этом она являлась недостаточно надежным биомаркером (AUC 0,69, рис. S12, *b* в Приложении). Значения PAL для пути *EGFR* не были дифференциальными и служили плохими биомаркерами (рис. S14 в Приложении).

Единственную подпись удалось сконструировать для пути *EGFR* Qiagen с двумя активированными генами (*SHC2* и *PRKCG*; рис. S15 в Приложении). Она была дифференциальной ($p = 1,1 \times 10^{-8}$), но имела плохую прогностическую значимость (AUC 0,646; рис. S16 в Приложении).

Рак молочной железы. При РМЖ мы исследовали статус мутации гена *PIK3CA*. Уровень экспрессии *PIK3CA* не различался между мутированными опухолями и опухолями дикого типа (рис. S17 в Приложении). В то же время нами была обнаружена статистически значимая активация пути EGFR Qiagen (но не по версии Biocarta) в опухолях с мутацией *PIK3CA* ($p = 1,5 \times 10^{-5}$; рис. S18, *a* в Приложении), однако прогностическая значимость ее была низкой (рис. S18, *b* в Приложении).

Для пути EGFR Qiagen удалось сгенерировать подпись, включающую гены с повышенной регуляцией (*FOS* и *MUC1*) и гены с пониженной регуляцией (*PRKCG* и *ERBB2*; рис. S19 в Приложении). Значения для этой генной подписи оказались дифференциальными ($p = 1,3 \times 10^{-8}$), но имели недостаточную прогностическую значимость (AUC 0,648; рис. S19, *b* в Приложении).

Результаты описанных исследований обобщены в табл. 2. В целом, среди всех проверенных вариантов только один одновременно отличался достаточной статистической значимостью по обоим исследованным критериям: дифференциальности ($p = 2,2 \times 10^{-16}$) и значимости как биомаркера (AUC 0,809): 7-генная подпись для мутации *BRAF^{V600E}* при РЩЖ (рис. 1). Проверка этой подписи была в дальнейшем подтверждена на двух независимых наборах экспрессионных профилей РЩЖ.

Независимая проверка 7-генной подписи *BRAF^{V600E}* РЩЖ. Для 7-генной подписи оценивали биомаркерную значимость с использованием двух независимых наборов данных с профилями РНК-секвенирования для образцов папиллярного подтипа РЩЖ. Один набор данных ранее был опубликован в работе [28] и включал в себя 75 образцов дикого типа и 52 мутированных образца (*BRAF^{V600E}*). Другой набор данных включал образцы 27 опухолей дикого типа и 15 мутированных опухолей и был получен нами экспериментально в настоящем исследовании.

Для проверки подписи были собраны 42 ретроспективных гистологически и клинически охарактеризованных образца РЩЖ (таблица в Приложении) с установленным для них статусом мутации *BRAF^{V600E}*, и для них были получены профили РНК-секвенирования. Биологические образцы представляли собой парафинизированные блоки опухолевой ткани, содержащие не менее 60% опухолевых клеток. В исследование были включены 8 пациентов мужского пола и 34 – женского, возраст которых варьировал от 11 до 71 года (средний возраст – 45 лет; таблица в Приложении).

Семигенная подпись продемонстрировала хорошую биомаркерную значимость для обоих

наборов данных: AUC 0,912, значение $p = 3,5 \times 10^{-15}$ для литературного набора данных [28] (рис. 2, *a, b*) и AUC 0,822, $p = 3,7 \times 10^{-4}$ – для экспериментального набора (рис. 2, *c, d*).

Таким образом, на основе анализа трех независимых наборов данных, один из которых был экспериментальным, мы пришли к выводу, что 7-генная подпись, включающая гены *FOS*, *RHOB* с пониженной экспрессией и гены *PS8*, *RND3*, *CAV2*, *CAV1*, *MUC1* с повышенной экспрессией в мутированных опухолях, является надежным предсказательным маркером статуса мутации *BRAF^{V600E}* с использованием данных РНК-секвенирования.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В настоящем исследовании мы подтвердили гипотезу о том, что при некоторых видах опухолей активирующие мутации в онкогенах могут быть статистически достоверно связаны с повышенной экспрессией соответствующих генов. Например, мутации *EGFR* были связаны с активацией этого гена при аденокарциноме легкого ($p = 0,021$), мутации *KRAS* – при колоректальном раке ($p = 4 \times 10^{-6}$), а мутации *BRAF* – при меланоме ($p = 1 \times 10^{-7}$).

Тем не менее подобных зависимостей не наблюдалось для тех же классов мутаций при других опухолях: для *EGFR* – при плоскоклеточном раке легкого, для *BRAF* – при раке щитовидной железы и для *PIK3CA* – при раке молочной железы (табл. 2).

Следует отметить, что продукты генов *EGFR*, *BRAF*, *KRAS* и *PIK3CA* образуют основные узлы молекулярного пути EGFR (рис. 2, *a, b*), поэтому мы поставили задачу впервые сравнить активацию пути EGFR с соответствующими статусами онкогенных мутаций. Было обнаружено, что активация пути EGFR связана с мутацией *BRAF^{V600E}* при раке щитовидной железы ($p = 1 \times 10^{-8}$) и при меланоме ($p = 0,006$), а также с активирующими мутациями *PIK3CA* при раке молочной железы ($p = 2 \times 10^{-5}$). В то же время для активирующих мутаций *EGFR* при раке легкого и для мутаций *KRAS* при колоректальном раке связи обнаружено не было (табл. 2).

На уровне значений алгоритмически сконструированных генных подписей удалось обнаружить значимую связь с мутацией *BRAF* при раке щитовидной железы ($p < 1 \times 10^{-8}$) и при меланоме ($p < 1 \times 10^{-8}$); с мутацией *EGFR* – при плоскоклеточной карциноме легкого ($p = 0,0002$); с мутацией *KRAS* – при колоректальном раке ($p = 1 \times 10^{-8}$); с мутацией *PIK3CA* – при раке молочной железы ($p = 1 \times 10^{-9}$). И только для мута-

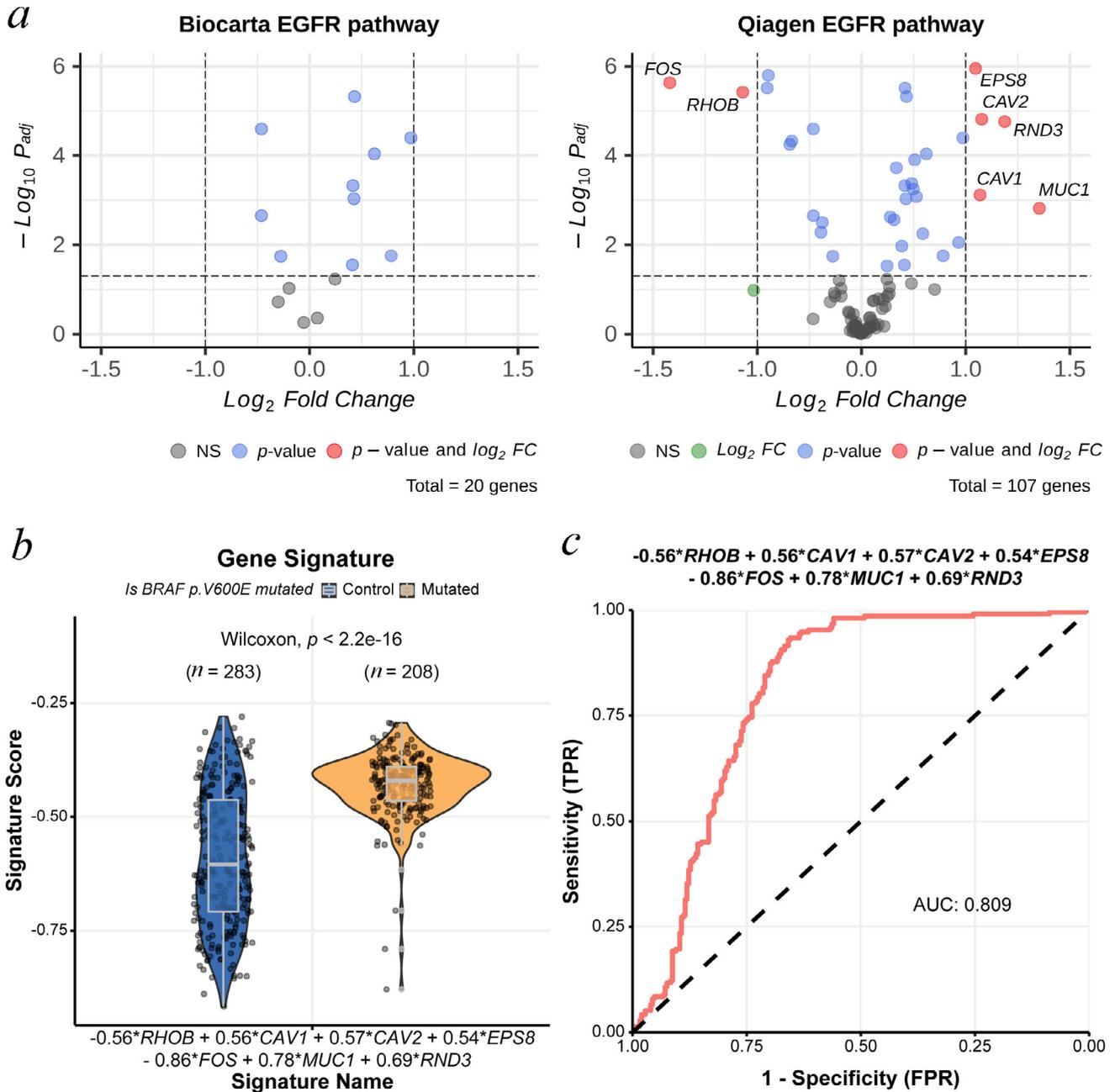


Рис. 1. *a* – Результаты анализа дифференциальной экспрессии генов пути EGFR (Biocarta 1.2 и Qiagen Pathway Central) между *BRAF*^{V600E} мутированными и образцами дикого типа РЩЖ из набора данных TCGA; *b* – значения генной подписи; *c* – биомаркерная значимость (ROC AUC) для генной подписи, сформированной из дифференциально экспрессируемых генов пути Qiagen EGFR

ций *EGFR* при аденокарциноме легкого связи обнаружено не было. Таким образом, мы предполагаем, что в целом текущая версия подхода, основанного на генных подписях, оказалась более эффективной в поиске ассоциаций с мутантным статусом, чем другие подходы, протестированные в настоящей работе. Среди всех проверенных ассоциаций только 7-генная подпись, разработанная для мутации *BRAF*^{V600E} при раке

щитовидной железы, оказалась надежным биомаркером (AUC 0,809–0,912), что подтверждается ее валидацией на трех независимых наборах данных (рис. 2). Интересно, что при этом в тех же самых наборах сходными были и особенности активации пути EGFR (рис. 3).

В двух видах рака (РЩЖ и меланома) мы реконструировали разные генные подписи, связанные с мутацией *BRAF*^{V600E}. Эти подписи

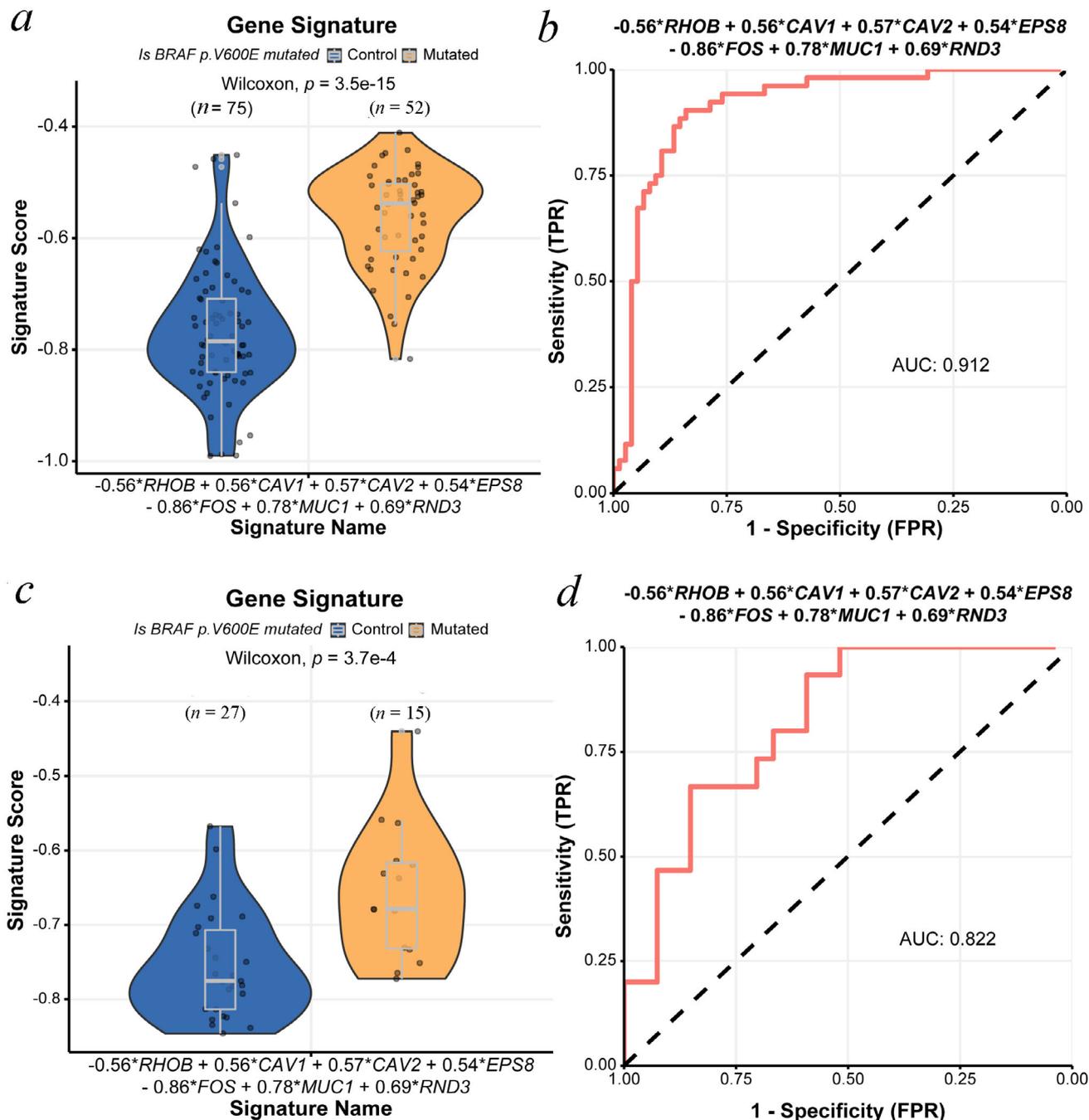


Рис. 2. Валидация 7-генной подписи на литературном (вверху) и экспериментальном (внизу) наборах данных. *a, c* – Значения генной подписи; *b, d* – её биомаркерная значимость (ROC AUC)

включали в себя 8 генов с пониженной и 8 генов с усиленной экспрессией при меланоме, 2 гена с пониженной и 5 генов с повышенной экспрессией при РЦЖ. Между этими двумя подписями было только одно совпадение по генному составу: ген *FOS* с пониженной экспрессией. Это доказывает, что одна и та же мутация *BRAF*^{V600E} при этих двух видах рака имеет различную молеку-

лярную реализацию даже среди генных продуктов пути EGFR. Возможно, эти различия могут лежать в основе успешной терапии блокаторами мутантного *BRAF* при меланоме [47] и отсутствием клинического интереса к ним при РЦЖ [17].

Клинические подходы на основе метода РНК-секвенирования в будущем могут стать



Рис. 3. Усредненный профиль активации пути Qiagen EGFR для образцов *BRAF*^{V600E} мутированного папиллярного рака щитовидной железы по наборам данных (a) The Cancer Genome Atlas Thyroid Cancer (TCGA-THCA), (b) исследования Yoo et al. [28] (PRJEB11591) и (c) экспериментального набора GSE138042 [38]. Гены, включенные в 7-генную подпись для прогнозирования статуса *BRAF*^{V600E} при раке щитовидной железы, обведены жирным кружком

прорывными технологиями, адаптированными к модели «все в одном тесте», поскольку в настоящее время в онкологии активно реконструируются многочисленные клинически значимые генные подписи [48]. Таким образом, обнаружение специфических для опухолей генных подписей для важных онкогенных мутаций имеет решающее значение для расширения функциональности тестов на основе РНК. Наши результаты доказывают, что это реалистичная задача: реконструирована и валидирована первая экспрессионная подпись, которая может надежно предсказать статус мутации *BRAF*^{V600E} при раке щитовидной железы. В то же время для того, чтобы обогатить репертуар клинических РНК-тестов и приложений РНК-секвенирования, необходимы дальнейшие исследования.

Финансирование. Выполнение данной работы было поддержано грантом Российского научного фонда № 18-15-00061.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов в какой-либо сфере деятельности.

Соблюдение этических норм. При выполнении настоящей работы были соблюдены все требования международных, национальных и/или ведомственных руководств по работе с лабораторными животными.

Дополнительные материалы. Приложение к статье на английском языке опубликовано на сайте журнала «Biochemistry» (Moscow) (<http://protein.bio.msu.ru/biokhimiya/>) и на сайте издательства Springer (www.springer.com/journal/10541), том 86, вып. 11, 2021.

СПИСОК ЛИТЕРАТУРЫ

- Masuda, H., Zhang, D., Bartholomeusz, C., Doihara, H., Hortobagyi, G. N., and Ueno, N. T. (2021) Role of epidermal growth factor receptor in breast cancer, *Breast Cancer Res. Treat.*, **136**, 331-345.
- Bardeesy, N., Kim, M., Xu, J., Kim, R.-S., Shen, Q., et al. (2005) Role of epidermal growth factor receptor signaling in russian academy of sciences-driven melanoma, *Mol. Cell. Biol.*, **25**, 4176-4188, doi: 10.1128/mcb.25.10.4176-4188.2005.
- Zhou, C., and Yao, L. (2016) Di Strategies to improve outcomes of patients with EGRF-Mutant Non-Small cell lung cancer: review of the literature, *J. Thorac. Oncol.*, **11**, 174-186.
- Greaves, W. O., Verma, S., Patel, K. P., Davies, M. A., Barkoh, B. A., et al. (2013) Frequency and spectrum of BRAF mutations in a retrospective, single-institution study of 1112 cases of melanoma, *J. Mol. Diagnostics*, **15**, 220-226, doi: 10.1016/j.jmoldx.2012.10.002.
- Akbani, R., Akdemir, K. C., Aksoy, B. A., Albert, M., Ally, A., et al. (2015) Genomic classification of cutaneous melanoma, *Cell*, **161**, 1681-1696, doi: 10.1016/j.cell.2015.05.044.
- Kebebew, E., Weng, J., Bauer, J., Ranvier, G., Clark, O. H., et al. (2007) The prevalence and prognostic value of BRAF mutation in thyroid cancer, *Ann. Surg.*, **246**, 466-470, doi: 10.1097/SLA.0b013e318148563d.
- Holderfield, M., Deuker, M. M., McCormick, F., McMahon, M. (2014) Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond, *Nat. Rev. Cancer*, **14**, 455-467.
- Silva, J. M., Bulman, C., and McMahon, M. (2014) BRAFV600E cooperates with PI3K signaling, independent of AKT, to regulate melanoma cell proliferation, *Mol. Cancer Res.*, **12**, 447-463, doi: 10.1158/1541-7786.MCR-13-0224-T.
- Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., et al. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer, *Gene*, **366**, 2-16.
- Frampton, J. E. (2010) Cetuximab: a review of its use in squamous cell carcinoma of the head and neck, *Drugs*, **70**, 1987-2010.
- Messersmith, W. A., and Ahnen, D. J. (2008) Targeting EGFR in colorectal cancer, *N. Engl. J. Med.*, **359**, 1834-1836.
- Alexander, M., Lin, E., and Cheng, H. (2020) Leptomeningeal metastases in non-small cell lung cancer: optimal systemic management in NSCLC with and without driver mutations, *Curr. Treat. Options Oncol.*, **21**, doi: 10.1007/s11864-020-00759-3.
- Garg, A., Batra, U., Choudhary, P., Jain, D., Khurana, S., et al. (2020) Clinical predictors of response to EGFR-tyrosine kinase inhibitors in EGFR-mutated non-small cell lung cancer: a real-world multicentric cohort analysis from India, *Curr. Probl. Cancer*, **44**, 100570, doi: 10.1016/j.cuprprobcancer.2020.100570.
- Lazzari, C., Gregorc, V., Karachaliou, N., Rosell, R., Santarpia, M. (2020) Mechanisms of resistance to osimertinib, *J. Thorac. Dis.*, **12**, 2851-2858, doi: 10.21037/jtd.2019.08.30.
- Odogwu, L., Mathieu, L., Blumenthal, G., Larkins, E., Goldberg, K. B., et al. (2018) FDA approval summary: dabrafenib and trametinib for the treatment of metastatic non-small cell lung cancers harboring BRAF V600E mutations, *Oncologist*, **23**, 740-745, doi: 10.1634/theoncologist.2017-0642.
- Kim, G., McKee, A. E., Ning, Y. M., Hazarika, M., Theoret, M., et al. (2014) FDA approval summary: Vemurafenib for treatment of unresectable or metastatic melanoma with the BRAFV600E mutation mutation, *Clin. Cancer Res.*, **20**, 4994-5000, doi: 10.1158/1078-0432.CCR-14-0776.
- Tang, K. T., and Lee, C. H. (2010) BRAF mutation in papillary thyroid carcinoma: pathogenic role and clinical implications, *J. Chinese Med. Assoc.*, **73**, 113-128.
- Narayan, P., Prowell, T. M., Gao, J. J., Fernandes, L. L., Li, E., et al. (2020) FDA Approval summary: alpelisib plus fulvestrant for patients with HR-positive, HER2-negative, PIK3CA-mutated, advanced or metastatic breast cancer, *Clin. Cancer Res.*, **27**, 1842-1849, doi: 10.1158/1078-0432.ccr-20-3652.
- Rodon, J., Soria, J. C., Berger, R., Miller, W. H., Rubin, E., et al. (2019) Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial, *Nat. Med.*, **25**, 751-758, doi: 10.1038/s41591-019-0424-4.
- Poddubskaya, E., Sorokin, M., Garazha, A., Glusker, A., Moisseev, A., et al. (2020) Clinical use of RNA sequencing

- and oncoPrint analytics to predict personalized targeted therapeutic efficacy, *J. Clin. Oncol.*, **38**, e13676-e13676, doi: 10.1200/jco.2020.38.15_suppl.e13676.
21. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nat. Biotechnol.*, **36**, doi: 10.1038/nbt.4096.
 22. Buzdin, A., Sorokin, M., Garazha, A., Sekacheva, M., Kim, E., et al. (2018) Molecular pathway activation – new type of biomarkers for tumor morphology and personalized selection of target drugs, *Semin. Cancer Biol.*, **53**, 110-124.
 23. Borisov, N., Sergeeva, A., Suntsova, M., Raevskiy, M., Gaifullin, N., et al. (2021) Machine learning applicability for classification of PAD/VCD chemotherapy response using 53 multiple myeloma RNA sequencing profiles, *Front. Oncol.*, **11**, 652063, doi: 10.3389/fonc.2021.652063.
 24. Borisov, N., Suntsova, M., Sorokin, M., Garazha, A., Kovalchuk, O., et al. (2017) Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data, *Cell Cycle*, **16**, 1810-1823, doi: 10.1080/15384101.2017.1361068.
 25. Borisov, N. M., Terekhanova, N. V., Aliper, A. M., Venkova, L. S., Smirnov, P. Y., et al. (2014) Signaling pathways activation profiles make better markers of cancer than expression of individual genes, *Oncotarget*, **5**, 10198-10205, doi: 10.18632/oncotarget.2548.
 26. Tsimberidou, A. M., Fountzilas, E., Bleris, L., and Kurzrock, R. (2020) Transcriptomics and solid tumors: the next frontier in precision cancer medicine, *Semin. Cancer Biol.*, doi: 10.1016/j.semcancer.2020.09.007.
 27. Petrov, I., Suntsova, M., Ilnitskaya, E., Roumiantsev, S., Sorokin, M., et al. (2017) Gene expression and molecular pathway activation signatures of MYCN-amplified neuroblastomas, *Oncotarget*, **8**, 83768-83780, doi: 10.18632/oncotarget.19662.
 28. Yoo, S. K., Lee, S., Kim, S. J., Jee, H. G., Kim, B. A., et al. (2016) Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers, *PLoS Genet.*, **12**, doi: 10.1371/journal.pgen.1006239.
 29. Titov, S. E., Ivanov, M. K., Karpinskaya, E. V., Tsivlikova, E. V., Shevchenko, S. P., et al. (2016) miRNA profiling, detection of BRAF V600E mutation and RET-PTC1 translocation in patients from Novosibirsk oblast (Russia) with different types of thyroid tumors, *BMC Cancer*, **16**, doi: 10.1186/s12885-016-2240-2.
 30. Suntsova, M., Gaifullin, N., Allina, D., Reshetun, A., Li, X., et al. (2019) Atlas of RNA sequencing profiles for normal human tissues, *Sci. Data*, **6**, 1-9, doi: 10.1038/s41597-019-0043-4.
 31. Amin, S. B., Yip, W. K., Minvielle, S., Broyl, A., Li, Y., et al. (2014) Gene expression profile alone is inadequate in predicting complete response in multiple myeloma, *Leukemia*, **28**, 2229-2234, doi: 10.1038/leu.2014.140.
 32. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., and Madden, T. L. Magic-BLAST, an accurate RNA-seq aligner for long and short reads, *BMC Bioinformatics*, doi: 10.1186/s12859-019-2996-x.
 33. Zhao, Y., Wong, L., and Goh, W. W. B. (2020) How to do quantile normalization correctly for gene expression data analyses, *Sci. Rep.*, **10**, 1-11, doi: 10.1038/s41598-020-72664-6.
 34. Huang, X., Stern, D. F., and Zhao, H. (2016) Transcriptional profiles from paired normal samples offer complementary information on cancer patient survival – evidence from TCGA pan-cancer data, *Sci. Rep.*, **6**, 1-9, doi: 10.1038/srep20567.
 35. BioCarta – Online maps of metabolic and signaling pathways[HLSL, Available online, URL: <https://www.hsls.pitt.edu/obrc/index.php?page=URL1151008585>.
 36. egf signaling – GeneGlobe, Available online, URL: <https://geneglobe.qiagen.com/us/explore/pathway-details/egf-signaling>.
 37. Sorokin, M., Kholodenko, R., Suntsova, M., Malakhova, G., Garazha, A., et al. (2018) Oncobox bioinformatical platform for selecting potentially effective combinations of target cancer drugs using high-throughput gene expression data, *Cancers (Basel)*, **10**, 365, doi: 10.3390/cancers10100365.
 38. Vladimirova, U., Rumiantsev, P., Zolotovskaya, M., Albert, E., Abrosimov, A., et al. (2021) DNA repair pathway activation features in follicular and papillary thyroid tumors, interrogated using 95 experimental RNA sequencing profiles, *Heliyon*, **7**, e06408, doi: 10.1016/j.heliyon.2021.e06408.
 39. Lokhandwala, P. M., Tseng, L. H., Rodriguez, E., Zheng, G., Pallavajjala, A., et al. (2019) Clinical mutational profiling and categorization of BRAF mutations in melanomas using next generation sequencing, *BMC Cancer*, **19**, 665, doi: 10.1186/s12885-019-5864-1.
 40. Hong, W., Wu, Q., Zhang, J., and Zhou, Y. (2019) Prognostic value of EGFR 19-del and 21-L858R mutations in patients with non-small cell lung cancer, *Oncol. Lett.*, **18**, 3887-3895, doi: 10.3892/ol.2019.10715.
 41. Gilson, P., Franczak, C., Dubouis, L., Husson, M., Rouyer, M., et al. (2019) Evaluation of KRAS, NRAS and BRAF hotspot mutations detection for patients with metastatic colorectal cancer using direct DNA pipetting in a fully-automated platform and Next-Generation Sequencing for laboratory workflow optimisation, **14**, e0219204, doi: 10.1371/journal.pone.0219204.
 42. Martínez-Saéz, O., Chic, N., Pascual, T., Adamo, B., Vidal, M. et al. (2020) Frequency and spectrum of PIK3CA somatic mutations in breast cancer, *Breast Cancer Res.*, **22**, doi: 10.1186/s13058-020-01284-9.
 43. Kamashev, D., Sorokin, M., Kochergina, I., Drobyshv, A., Vladimirova, U., et al. (2020) Human blood serum can donor-specifically antagonize effects of EGFR-targeted drugs on squamous carcinoma cell growth, *Heliyon*, **7**, e06394, doi: 10.1016/j.heliyon.2021.e06394.
 44. Crispo, F., Notarangelo, T., Pietrafesa, M., Lettini, G., Storto, G., et al. (2019) Braf inhibitors in thyroid cancer: clinical impact, mechanisms of resistance and future perspectives, *Cancers (Basel)*, **11**, 1388, doi: 10.3390/cancers11091388.
 45. Borisov, N., Sorokin, M., Tkachev, V., Garazha, A., and Buzdin, A. (2020) Cancer gene expression profiles associated with clinical outcomes to chemotherapy treatments, *BMC Med. Genomics*, **13**, 111, doi: 10.1186/s12920-020-00759-0.
 46. Boyd, J. C. (1997) Mathematical tools for demonstrating the clinical usefulness of biochemical markers, *Scand. J. Clin. Invest. Suppl.*, **227**, 46-63.
 47. Wong, D. J. L., and Ribas, A. (2016) Targeted therapy for melanoma, *Cancer Treat. Res.*, **167**, 251-262.
 48. Hou, J., Wang, Z., Li, H., Zhang, H., and Luo, L. (2020) Gene signature and identification of clinical trait-related m⁶A regulators in pancreatic cancer, *Front. Genet.*, **11**, 522, doi: 10.3389/fgene.2020.00522.

PATHWAY-BASED GENE SIGNATURES OF EGFR, BRAF, PIK3CA, AND KRAS GENE MUTATIONS IN BREAST, LUNG, THYROID CANCERS AND IN MELANOMA

M. Raevskiy^{1,2}, M. Sorokin^{3,4}, U. Vladimirova³, M. Suntsova³, V. Efimov², A. Garazha¹, A. Drobyshev³, A. Moisseev³, P. Rumiantsev⁵, X. Li⁶, and A. Buzdin^{1,2,3,7*}

¹ Omicsway Corp., 340 S Lemon Ave, 6040, Walnut, 91789 CA, USA; E-mail: buzdin@oncobox.com

² Moscow Institute of Physics and Technology, Dolgoprudny, 141701 Moscow Region, Russia

³ I. M. Sechenov First Moscow State Medical University, 119991 Moscow, Russia

⁴ Oncobox Ltd., 121205 Moscow, Russia

⁵ Endocrinology Research Centre, 117312 Moscow, Russia

⁶ Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, 90095 CA, USA

⁷ Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, 117997 Moscow, Russia

EGFR, *BRAF*, *PIK3CA*, and *KRAS* genes play major roles in EGFR pathway, and accommodate activating mutations that predict response to many targeted therapeutics. However, connections between these mutations and EGFR pathway expression patterns remain unexplored. Here, we investigated transcriptomic associations with these activating mutations in three ways. First, we compared expressions of these genes in the mutant and wild type tumors, respectively, using RNA sequencing profiles from The Cancer Genome Atlas project database ($n = 3660$). Second, mutations were associated with the activation level of EGFR pathway. Third, they were associated with the gene signatures of differentially expressed genes from these pathways between the mutant and wild type tumors. We found that the upregulated EGFR pathway was linked with mutations in the *BRAF* (thyroid cancer, melanoma) and *PIK3CA* (breast cancer) genes. Gene signatures were associated with *BRAF* (thyroid cancer, melanoma), *EGFR* (squamous cell lung cancer), *KRAS* (colorectal cancer), and *PIK3CA* (breast cancer) mutations. However, only for the *BRAF* gene signature in the thyroid cancer we observed strong biomarker diagnostic capacity with AUC > 0.7 (0.809). Next, we validated this signature on the independent literature-based dataset ($n = 127$, fresh-frozen tissue samples, AUC 0.912), and on the experimental dataset ($n = 42$, formalin fixed, paraffin embedded tissue samples, AUC 0.822). Our results suggest that the RNA sequencing profiles can be used for robust identification of the replacement of Valine at position 600 with Glutamic acid in the *BRAF* gene in the papillary subtype of thyroid cancer, and evidence that the specific gene expression levels could provide information about the driver carcinogenic mutations.

Keywords: mutation, EGFR, *BRAF* V600E, transcriptomic biomarker, RNA sequencing, papillary thyroid cancer