

Пятницкий Михаил Алексеевич

**ВЫСОКОПРОИЗВОДИТЕЛЬНОЕ СЕКВЕНИРОВАНИЕ В
МОЛЕКУЛЯРНОЙ ОНКОЛОГИИ: ПОИСК МИШЕНЕЙ И
СТРАТИФИКАЦИЯ ПАЦИЕНТОВ ДЛЯ ПЕРСОНАЛИЗАЦИИ
ПРОТИВООПУХОЛЕВОЙ ТЕРАПИИ**

1.5.8. – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

доктора биологических наук

Москва – 2022 г.

Работа выполнена в Федеральном государственном бюджетном научном учреждении
«Научно-исследовательский институт биомедицинской химии
имени В. Н. Ореховича» (ИБМХ)

Научные консультанты: доктор биологических наук, академик РАН
Лисица Андрей Валерьевич

доктор биологических наук, профессор РАН
Мошковский Сергей Александрович

Официальные оппоненты: **Аульченко Юрий Сергеевич**,
доктор биологических наук, ФГБНУ «Федеральный
исследовательский центр «Институт цитологии и
генетики СО РАН», главный научный сотрудник.

Орлов Юрий Львович,
доктор биологических наук, профессор РАН,
«ФГАОУ ВО Первый МГМУ им. И.М. Сеченова
Минздрава России (Сеченовский Университет)»,
профессор.

Стрельников Владимир Викторович,
доктор биологических наук, ФГБНУ «Медико-
генетический научный центр имени академика Н.П.
Бочкова», заведующий лабораторией.

Ведущая организация: Федеральное государственное бюджетное
учреждение науки Институт общей генетики им.
Н.И. Вавилова Российской академии наук

Защита состоится «20» октября 2022 г. в 11.00 часов на заседании диссертационного
совета 24.1.172.01 (Д 001.010.01) при Федеральном государственном бюджетном
научном учреждении «Научно-исследовательский институт биомедицинской химии
имени В.Н. Ореховича» по адресу: 119121, Москва, ул. Погодинская, д.10, стр.8.

С диссертацией можно ознакомиться в ФГБНУ «Научно-исследовательский институт
биомедицинской химии имени В.Н. Ореховича» и на сайте www.ibmc.msk.ru

Автореферат разослан « ____ » _____ 2022 г.

Ученый секретарь Диссертационного совета,
кандидат химических наук

Е.А. Карпова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Если суммировать в одном предложении ключевую информацию, которую исследователи узнали о злокачественных опухолях за последние 100 лет, то вероятно это будет фраза “Рак – это болезнь генов”. Систематическое развитие онкогеномики как области знаний по исследованию геномных aberrаций в раковых клетках стало возможным с 2003 года, ознаменовавшего успешное завершение международного проекта “Геном человека”. Следующей вехой стало создание революционной технологии высокопроизводительного параллельного секвенирования ДНК (Next Generation Sequencing, NGS). Низкая себестоимость и относительная простота лабораторных методик NGS позволила быстро накапливать полногеномные выборки данных. Благодаря организации масштабных проектов, таких как The Cancer Genome Atlas и International Cancer Genome Consortium, сейчас в открытом доступе находятся сведения о геномах, транскриптомах и протеомах десятков тысяч образцов опухолей.

Совокупность молекулярных профилей опухолевых клеток и клинической информации может рассматриваться как существенно новый тип данных. Аналогично тому, как одновременное сопоставление геномов различных организмов привело к появлению сравнительной геномики и позволило пролить свет на многие вопросы эволюционной биологии, так и накопление больших объемов данных о мутациях в опухоли стимулировало изучение множества различных аспектов молекулярной онкологии.

Во-первых, анализируя закономерности в распределении мутаций, возможно выявить ключевые гены, ответственные за процессы малигнизации и прогрессии опухоли. Примерами могут служить обнаружение мутаций в гене *BRAF* в ~50% случаев меланомы, мутаций в гене *PIK3CA* в 25-30% случаев рака молочной железы и колоректального рака, мутаций в гене *EGFR* в 10-15% случаев немелкоклеточного рака легкого и т.д. Анализ наблюдаемого спектра мутаций в раковых клетках может быть использован для идентификации новых генов-мишеней и разработки таргетных препаратов, селективно воздействующих только на клетки опухоли (Hoelder et al., 2012).

Другой важной задачей является выделение отдельных подтипов онкологических заболеваний. Для этого вся когорта пациентов разбивается на группы (стратифицируется) на основе различных молекулярных характеристик опухоли. Подразумевается, что внутри каждой группы, объединенной общими механизмами онкогенеза, возможна выработка более эффективной стратегии противоопухолевой терапии по сравнению с традиционным неадаптивным подходом. Примером может

служить рак молочной железы, который два десятилетия назад считался относительно «простым» заболеванием, а в настоящее время разделен на несколько молекулярных подтипов с различными патологическими особенностями и клиническими проявлениями (Kittaneh et al., 2013).

Особый интерес представляет прецизионная онкология – изучение молекулярных особенностей опухоли одного пациента. Поскольку спектр выявляемых мутаций в настоящее время рассматривается как основная причина эффективности ответа на схему лечения, то для персонализации подбора противоопухолевой терапии необходимо проанализировать соответствующим образом отобранные и проаннотированные генетические варианты. Задача прикладной клинической биоинформатики состоит в разработке программного обеспечения, облегчающего интерпретацию результатов высокопроизводительного секвенирования генома и транскриптома опухоли для определения специфических механизмов онкогенеза и выявления биомаркеров, обладающих тераностической ценностью. Создание таких инструментов с дружественным пользовательским интерфейсом, адаптированным под использование практикующими химиотерапевтами, способствует внедрению передовых технологий прецизионной онкологии в клиническую практику (Pishvaian, et al., 2019).

Таким образом, применение методов высокопроизводительного секвенирования ДНК в онкологии и всесторонний анализ полученных результатов позволяет решать различные научно-практические задачи, нацеленные как на долгосрочную перспективу (поиск генов-мишеней для противоопухолевой терапии), так и обладающие потенциальной возможностью быстрого транслирования в клинику (стратификация пациентов, анализ генетических вариантов для оптимизации схемы лечения). Представленная диссертационная работа призвана проиллюстрировать это утверждение.

Цель работы состояла в том, чтобы разработать и применить вычислительные методы для предсказания критически важных белков в протеоме опухоли, стратификации пациентов и подбора персонализированной противораковой терапии, опираясь на данные по высокопроизводительному секвенированию геномов и транскриптомов малигнизированных клеток.

Для достижения поставленной цели решались следующие **задачи**:

1. Разработать метод по идентификации ключевых белков в протеоме опухоли, чья активность необходима для функционирования раковых клеток.
2. Применить разработанный метод для данных о соматических мутациях в клетках меланомы, получив список потенциальных генов-мишеней.

3. Разработать метод стратификации пациентов в соответствии с выявляемыми особенностями регуляции генной экспрессии в образцах тканей.
4. Сравнить разработанный метод стратификации пациентов с аналогичными алгоритмами, применить его для изучения молекулярных механизмов, ассоциированных с эффективностью противораковой терапии на примере цетуксимаба.
5. Разработать программный конвейер по аннотации и приоритизации генетических вариантов, полученных в результате высокопроизводительного секвенирования геномов опухолевых клеток.
6. Проанализировать экзомы образцов колоректального рака и гепатоцеллюлярной карциномы из российской популяции, оценить пригодность разработанного конвейера для выявления вероятных генов-драйверов онкологической трансформации и определить потенциально эффективные препараты.
7. Спроектировать и разработать интернет-платформу для интерпретации молекулярного профиля онкологического пациента с целью оптимизации персонализированной противораковой терапии.

Научная новизна и практическая значимость. В диссертационной работе последовательно решаются три задачи, каждая из которых имеет научно-практическое применение.

Представлен новый подход к поиску белков, критически важных для функционирования опухолевых клеток, и основанный на анализе данных большого массива данных о соматических мутациях. Выявленные белки могут рассматриваться как перспективные мишени при создании новых препаратов для противоопухолевой терапии. Нами была проведена экспериментальная работа по валидации вычислительных предсказаний критически важных белков для клеток меланомы.

Разработан новый метод для стратификации пациентов, основанный на анализе транскриптомных данных и выявлении активированных генов-регуляторов экспрессии. Предложенный метод позволяет не только классифицировать образцы без учителя (осуществлять стратификацию), но и может рассматриваться как биологически мотивированное уменьшение размерности пространства признаков, позволяя переходить от изучения экспрессии десятков тысяч генов к анализу активности десятков кластеров регуляторов. Сравнение с некоторыми альтернативными подходами к решению задачи стратификации показало преимущество разработанного нами метода. Идентифицированные кластеры генов-регуляторов позволяют интерпретировать сигнальные каскады, задействованные у каждого отдельного пациента. Применение разработанного метода для выдвижения

гипотез о задействованных молекулярных механизмах проиллюстрировано на примере изучения случаев резистентности колоректального рака к терапии цетуксимабом.

В работе описан программный конвейер, предназначенный для аннотации генетических вариантов большим количеством общебиологической и онкоспецифической информации. Практическая применимость разработанной системы для оптимизации противоопухолевой терапии проиллюстрирована на примере анализа результатов полноэкзомного секвенирования образцов колоректального рака и гепатоцеллюлярной карциномы. Программный конвейер является ключевым компонентом разработанной веб-платформы по автоматической интерпретации молекулярного профиля онкологического пациента. Такая платформа может использоваться для облегчения работы врача-химиотерапевта, предоставляя доступ к базе данных о клинической значимости биомаркеров и генерируя отчет о потенциально эффективной или неэффективной терапии на основе загруженных данных о молекулярном портрете опухоли. С помощью разработанной веб-платформы в ООО «Онко Генотест» были проанализированы более 50 образцов различных солидных опухолей.

Личный вклад автора. Автор участвовал в разработке метода поиска уязвимых мест в раковом протеоме, основываясь на результатах геномного секвенирования. Автор реализовал указанный метод, а также участвовал в проведении биоинформатического анализа данных по мутациям в меланоме и здоровых тканях. Автор участвовал в разработке подхода к стратификации пациентов на основе данных о регуляции генов. Автор осуществил программную реализацию указанного метода, провел сопоставление с аналогичными алгоритмами, проанализировал данные по чувствительности пациентов к терапии цетуксимабом, а также участвовал в биологической интерпретации и выдвижении гипотезы об активации эпителиально-мезенхимального перехода. Автором был разработан и реализован программный конвейер по аннотации генетических вариантов, проведена обработка NGS-данных для образцов колоректального рака и гепатоцеллюлярной карциномы, а также изучена применимость конвейера для персонализации противоопухолевой терапии. Автор участвовал в разработке программной архитектуры и интерфейса пользователя интерактивной веб-платформы для прецизионной онкологии. Под руководством автора была спроектирована структура базы данных биомаркеров, формат взаимодействия между фронтенд- и бэкенд-компонентами платформы, программные модули для оценки качества результатов секвенирования и визуализации найденных взаимосвязей. Основная

биоинформатическая составляющая бэкенд-компонента системы также была разработана автором.

Апробация работы. Основные результаты работы доложены и обсуждены на конференциях и конгрессах: ASCO Annual Meeting 2017, «Московская конференция по вычислительной молекулярной биологии» (Москва, Россия, 2013 и 2015), «Международная конференция по биоинформатике регуляции и структуры генома» (Новосибирск, Россия, 2012 и 2018), HUPO 13th Annual World Congress (Мадрид, Испания, 2014), HUPO 12th Annual World Congress (Йокогама, Япония, 2013), HUPO 11th Annual World Congress (Бостон, США, 2012), 6th Congress АОНУРО (Пекин, Китай, 2012), HUPO 9th Annual World Congress (Сидней, Австралия, 2010), HUPO 7th Annual World Congress (Амстердам, Нидерланды, 2008), 4th RECOMB Conference on Regulatory Genomics, Systems Biology and DREAM Challenges (Барселона, Испания, 2011), 38th FEBS Congress (Санкт-Петербург, Россия, 2013), 62nd ASMS Conference on Mass Spectrometry and Allied Topics (Балтимор, США, 2014), Genetics of Aging and Longevity (Сочи, Россия, 2014), «Молекулярная диагностика» (Москва, Россия, 2014), «Российский конгресс лабораторной медицины» (Москва, Россия, 2015 и 2021), 65th ASMS Conference on Mass Spectrometry and Allied Topics (Индианаполис, США, 2017), Big Data in Pharma (Москва, Россия, 2018), XVIII Ассамблея «Здоровая Москва» (Москва, Россия, 2020), Startup Village (Сколково, Россия, 2016 и 2021).

Основные положения, выносимые на защиту.

1. Результаты крупномасштабных проектов по секвенированию геномов опухолевых клеток могут использоваться для поиска белков, чье функционирование необходимо для выживания раковых клеток. Такие белки являются перспективными мишенями для новых противоопухолевых препаратов.
2. Предложенный в работе метод позволяет решать задачу стратификации пациентов с помощью выявления кластеров активированных генов-регуляторов экспрессии. Идентифицированные кластеры регуляторов имеют биологическую интерпретацию и позволяют выдвигать гипотезы о молекулярных механизмах, характерных для выделенных подтипов заболеваний.
3. Разработанная веб-платформа автоматизирует клиническую интерпретацию молекулярно-генетического профиля пациента, включая результаты высокопроизводительного секвенирования ДНК образцов опухоли. Использование подобных программных продуктов позволяет реализовывать принципы персонализированной медицины с целью оптимизации противораковой терапии.

Публикации. Материалы диссертационной работы отражены в 46 публикациях, включая 24 статьи (5 и 19 статей в российских и международных научных изданиях соответственно), 17 публикаций в докладах российских и международных научных конференций, а также в трёх главах коллективных монографий. По итогам работы получен патент №2577107 «Способ поиска белков-мишеней, запускающих процесс канцерогенеза, в образцах тканей индивидуального пациента, для последующей противоопухолевой фармакотерапии», а также выдано свидетельство о государственной регистрации программы для ЭВМ №2020618223 «Платформа для обработки и интерпретации клинических данных и молекулярно-генетических исследований опухолей человека для подбора персонализированной терапии». Индекс Хирша соискателя ученой степени согласно данным системы Scopus составляет 12 баллов.

Объем и структура диссертации. Диссертационная работа изложена на 239 страницах машинописного текста, включая 57 рисунков и 23 таблицы, и состоит из введения, шести глав основного текста, заключения, выводов и списка литературы, содержащего 446 источников. Первая глава диссертации посвящена обзору литературы, описание исходных данных и методов обработки дано во второй главе. В остальных главах излагаются и обсуждаются полученные результаты.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Предсказание критически важных генов

Исходными данными для анализа являлись сведения о встречаемости соматических мутаций в клетках меланомы, которые были загружены из веб-ресурса проекта The Cancer Genome Atlas [<https://portal.gdc.cancer.gov/>] и профильтрованы с целью удаления ложно-положительных и герминативных вариантов согласно работе (Kandoth et al., 2013). Суммарно было определено 181 175 соматических вариантов. Данные о полиморфизмах были загружены с сайта проекта «1000 Genomes» [www.internationalgenome.org] и аннотированы с использованием программного обеспечения ANNOVAR (Wang, et al., 2010). Итоговая таблица содержала 425 069 генетических вариантов, расположенных в кодирующих областях генома. Данные по экспрессии генов были загружены из ресурса The Cancer Genome Atlas и содержали информацию об уровнях экспрессии 20 531 генов для 374 образцов меланомы.

Величину dN/dS (отношение доли несинонимичных замен к доле синонимичных замен) для всех белок-кодирующих генов человека рассчитывали согласно методике описанной Ostrow и коллегами (Ostrow et al., 2014). Для каждого

гена рассматривали самый длинный транскрипт. Патогенность миссенс-мутаций оценивали согласно ресурсу dbNSFP (Liu, et al., 2016), интегрирующем результаты нескольких алгоритмов в два ансамблевых значения, MetaSVM и MetaLR. Считали, что мутация является патогенной, если значение MetaSVM или MetaLR превышало 70% перцентиль. Также патогенными считали следующие классы генетических aberrаций: инсерция/делеция, потеря стоп-кодона, возникновение стоп-кодона, мутация сайта сплайсинга.

Функциональный анализ обогащения генов с помощью гипергеометрического теста проводили с использованием ресурса WebGestalt [www.webgestalt.org]. В качестве референсного набора использовали список из 16 172 генов со средней экспрессией в клетках меланомы, превышающей 20%-перцентиль. Коррекцию на множественное тестирование проводили с использованием метода Бенджамини-Хохберга.

Подавление экспрессии генов и определение функциональной активности клеток меланомы

Клеточную линию злокачественной меланомы человека A375 культивировали в стандартных условиях. Использовали культуральную среду DMEM (Dulbecco's Modified Eagle Medium) с добавлением 10% фетальной бычьей сыворотки и L-глутамин при 5% CO₂ и 37° C. Перед трансфекцией клетки трипсинизировали, ресуспендировали, разбавляли до 1 мл и анализировали с помощью проточного цитометра Partec (Кент, Великобритания). Для каждого из четырех отобранных генов (*ZNFx1*, *RHPN2*, *STK11IP* и *UNC45A*) были подобраны от одной до трех миРНК согласно рекомендациям работы Lagana и соавторов (Lagana, et al., 2015). После обработки миРНК клетки инкубировали 24 часа, затем лизировали и выделяли РНК. Далее проводили реакцию обратной транскрипции с последующей ПЦР в реальном времени. Для нормализации использовали уровни экспрессии четырех референсных генов (*B2M*, *HPRT*, *TFRC*, *TBP*). Эксперимент проводили в пяти биологических повторностях. Анализ жизнеспособности клеток проводили с использованием CellTiter-Blue Cell Viability Assay (Promega, США) через 12, 24, 36, 48 и 60 часов после обработки клеток меланомы соответствующей миРНК. Экспериментальные методики приведены согласно оригинальной работе Кривошеевой и соавторов (Krivosheeva, et al., 2020)

Стратификация пациентов на основе регуляторов генной экспрессии

Данные по экспрессии генов загрузили из ресурса NCBI GEO [www.ncbi.nlm.nih.gov/geo]. Разработанный метод применили к экспериментам GSE4183 и GSE5851. Первый набор данных включал транскриптомные профили для 15 образцов колоректального рака (KPP), 15 образцов с предраковой аденомой, 15 образцов с воспалительными заболеваниями кишечника и восемь образцов интактного кишечника. В эксперименте GSE5851 изучалась генная экспрессия 80 образцов распространенного метастатического KPP.

Для определения молекулярных механизмов ответа на терапию цетуксимабом анализировали девять экспериментов по транскриптомному профилированию KPP: GSE13294, GSE14333, GSE17538, GSE18105, GSE4107, GSE4183, GSE9348, GSE8671, GSE5851. Общее количество образцов составило 863, из которых 794 образца были получены от больных KPP, а 79 образцов интактного кишечника — от здоровых доноров. Для нормализации данных использовали алгоритм FARMS (Hochreiter, et al., 2006). При сопоставлении зондов с генами, в тех случаях, когда одному гену соответствовало несколько зондов, отбирали только один зонд с максимальной интенсивностью. Итоговый набор данных состоял из 863 образцов и 6377 генов.

Изучали активированные регуляторные сети, детектированные при помощи алгоритма SNEA (Sivachenko et al., 2007), состоящие из центрального гена-регулятора и набора регулируемых генов. Использовали программное обеспечение Pathway Studio 9.0 (Elsevier, Нидерланды) со следующими параметрами: тип регулятора «белок», «функциональный класс» или «комплекс», тип соседей по генно-регуляторной сети «белок», типы учитываемых взаимосвязей «Экспрессия» или «Связывание с промотором».

Найденные регуляторы объединяли в кластеры регуляторов на основании схожести набора генов-мишеней. Для этого использовали расстояние Жаккара, вычисляя для каждой пары регуляторов r_1 и r_2 , регулирующих наборы генов A_1 и A_2 соответственно, меру $d(r_1, r_2) = 1 - \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|}$. Вычисляли матрицу попарных расстояний для всех пар регуляторов, после чего выделяли кластеры регуляторов согласно методу Уорда. Оптимальное количество кластеров определяли согласно методу “максимизации силуэта” (Kaufman&Rousseeuw, 2005). Для каждого j -го кластера регуляторов вычисляли его меру активности $C_j = \sum_{i=1}^{N_j} K_i$, где N_j — число регуляторов принадлежащих кластеру, $K_i = \text{median}(r_i) \times |r_i|$, r_i — вектор величин logratio для всех генов, контролируемых i -м регулятором. Величину logratio для каждого гена вычисляли как $\log_2$ (экспрессия в опухоли)/(экспрессия в нормальной ткани).

Полученные значения C_j использовали для кластеризации пациентов согласно методу Уорда, а в качестве меры расстояния между образцами использовали коэффициент корреляции Пирсона.

Для сопоставления полученных разбиений с клинико-патологическими характеристиками образцов и сравнения с другими методами использовали индекс Рэнда. Оценивали воспроизводимость получаемых разбиений образцов на группы. Выполняли 1000 бутстрэп-циклов, каждый раз используя случайным образом отобранные 90% образцов из исходного набора данных, и рассчитывали индекс Рэнда, оценивая степень соответствия между разбиением согласно активностям кластеров регуляторов и референсным разбиением. В результате получали распределения индексов Рэнда, что позволяло статистически оценить достоверность наблюдаемых различий в соответствии разбиений.

Вычисляли дискриминант Фишера $J_i = \frac{(\mu_{i,a} - \mu_{i,b})^2}{\sigma_{i,a}^2 + \sigma_{i,b}^2}$, чтобы оценить дискриминаторную силу i -го кластера регуляторов между классами a и b , имеющими среднюю активность $\mu_{i,a}$ и $\mu_{i,b}$, а также дисперсии активности $\sigma_{i,a}^2, \sigma_{i,b}^2$. Кластер регуляторов считали достоверно коррелированным с меткой класса, если уровень значимости для перестановочного теста (1000 случайных перестановок меток классов) не превышал значения 0,05 в более чем 90 из 100 циклов 10-кратного сэмпирования без возвращения. Вычисляли площадь под кривой чувствительность-специфичность для оценки дискриминирующих признаков.

В качестве референсных методов стратификации пациентов использовали алгоритм GSVA с настройками по умолчанию (Hanzelmann et al., 2013), а также кластеризацию профилей экспрессии методом РАМ (Maechler et al., 2014). Программное обеспечение, реализующее разработанный подход к стратификации пациентов, доступно по адресу [<https://sourceforge.net/projects/bsnea>].

Высокопроизводительное секвенирование образцов

Провели полноэкзомное секвенирование пяти образцов КРР и пяти соответствующих образцов интактного толстого кишечника, а также одного образца гепатоцеллюлярной карциномы и прилегающей здоровой ткани печени. Среднее содержание опухолевых клеток в образцах составило 70%. Образцы опухоли и интактной ткани у пациентов использовали с их информированного согласия.

Образцы геномной ДНК фрагментировали на ультразвуковом приборе Covaris (США), с получением библиотеки фрагментов, размеры которых распределялись между 150 и 200 п.о. После лигирования адаптеров, фрагменты очищали на гранулах

Agencourt AMPure SPRI beads (Agilent, США), амплифицировали путем опосредованной лигированием ПЦР и обогащали посредством чипа Agilent SureSelect Human All Exon Enrichment Kit V4 (Agilent, США). Захваченную на чип библиотеку фрагментов секвенировали на приборе Illumina HiSeq2000 (США). Использовали режим парного секвенирования с длиной прочтения, равной 100 оснований.

Полученные прочтения картировали на референсный геном человека (версия hg19) с помощью программ BWA (образцы KPP) или с помощью bowtie2 (гепатоцеллюлярная карцинома). Полученные выравнивания были дедулицированы, перевыравнены в областях инделов, а также рекалиброваны по качеству согласно протоколу GATK Best Practices. Поиск генетических вариантов проводили с помощью программ SOAPsnp и MuTect2 с настройками по умолчанию (образцы KPP) или с помощью VarScan2 (гепатоцеллюлярная карцинома).

Аннотация генетических вариантов

Аннотацию проводили, объединяя результаты различных программных инструментов и баз данных, включая dbNSFP (Liu, et al., 2016), Oncotator (Ramos, et al., 2015), MutationAssessor (Reva et al., 2011), FunSeq (Fu, et al., 2014), CHASM (Douville et al., 2013), CRAVAT (Masica, et al., 2017), PharmGKB (Whirl-Carrillo, et al., 2012), ConsensusPathDB (Kamburov, et al., 2013), gnomAD (Karczewski, et al., 2020), CIViC (Griffith, et al., 2017). Информацию из некоторых ресурсов (например, dbNSFP) использовали как для аннотации на уровне генов, так и для аннотации на уровне генетических вариантов. Некоторые из использованных ресурсов (например, Oncotator и ConsensusPathDB) представляют собой компиляцию сведений из ряда других баз данных.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Уязвимые точки протеома клеток меланомы.

Был разработан новый подход к идентификации уязвимых точек в протеоме опухоли («essential proteins», критически важных белков), чья активность необходима для функционирования раковых клеток. В основе предложенного метода лежит поиск белков, которые находятся под воздействием отрицательного отбора, то есть обедненных функционально важными заменами аминокислотных остатков. Предполагается, что именно консервативность таких белков является следствием их необходимости для выживания раковых клеток. В данном разделе не делали различия между белками и кодирующими их генами.

При разработке метода опирались на известный факт из истории прикладной статистики, а именно — на работу Абрахама Вальда по проблемам оценки уязвимости бомбардировщиков. Согласно условиям задачи самолет возвращается с боевого задания, поврежденный зенитным огнем противника; необходимо определить для каких частей воздушного судна следует усилить бронирование для улучшения выживаемости. Важно отметить, что данные являются цензурированными — т.е. нельзя изучить распределение попаданий для самолетов, которые не вернулись на базу (т.н. ошибка выжившего). Максимально упрощая, полученные Вальдом результаты можно сформулировать как предложение усиливать бронирование свободных от попаданий элементов: отсутствие пробоин говорит о том, что именно такие места являются критически важными узлами самолета. В противном случае попадание в эти места приводит к уничтожению воздушного судна и делает данные об уязвимости ненаблюдаемыми.

Используя схожую логику, мы предположили, что если вместо успешного возвращения самолета с задания рассматривать выживаемость раковой клетки, то аналогом зенитного огня являются возникающие *de novo* соматические патогенные мутации, нарушающие функционирование того или иного белка. Таким образом, гены, обедненные патогенными соматическими мутациями, вероятно играют важную роль в раковой клетке: клетки с мутациями в таких генах погибают, а их геномы становятся ненаблюдаемыми и не могут быть просеквенированы. Аналогия между работой Вальда и гипомутированными белками, критически важными для выживания раковых клеток, проиллюстрирована на Рисунке 1.

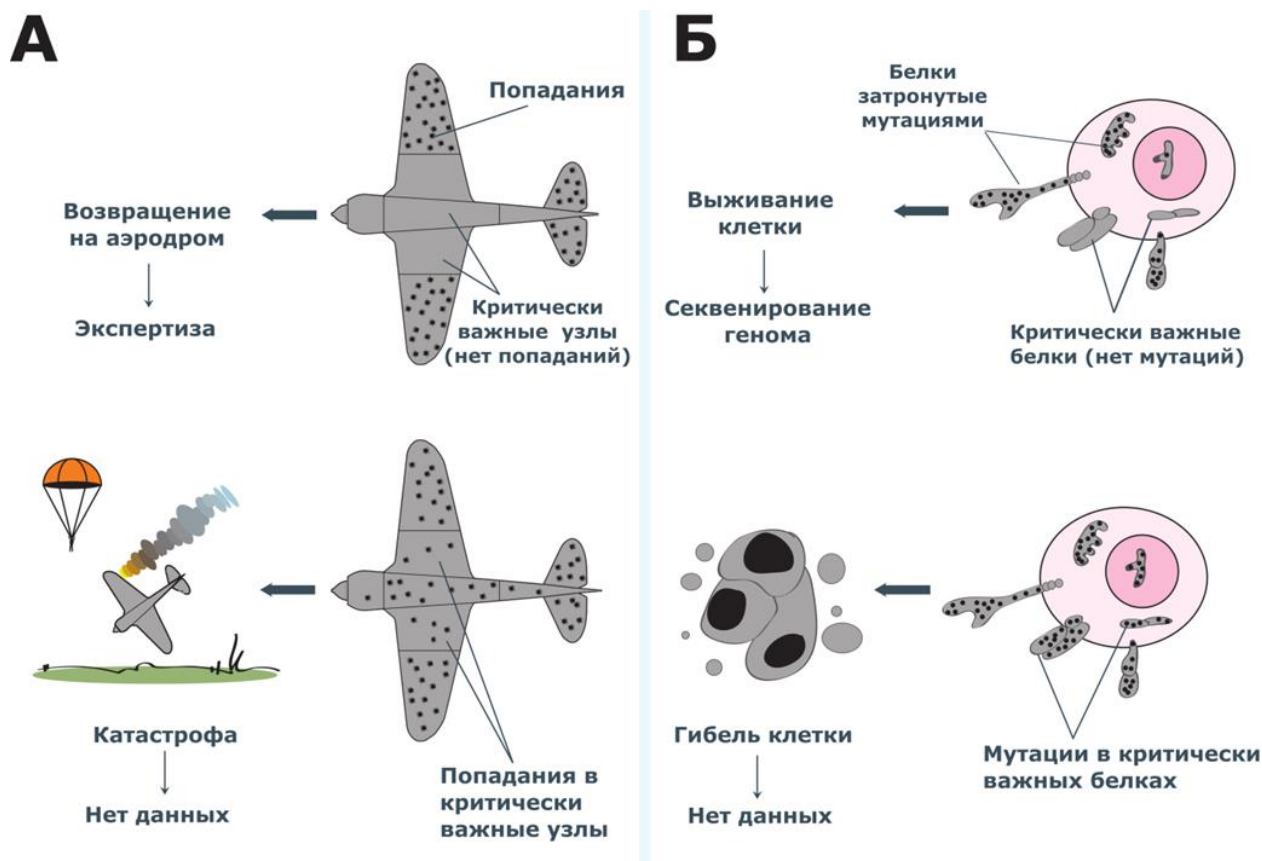


Рисунок 1. Аналогия между критическими важными узлами самолетов (работа А. Вальда) и гипомутированными белками, критически важными для выживания раковых клеток. А) Узлы, которые остались неповрежденными на вернувшихся с задания самолетах, важны для живучести воздушного судна. Б) Аналогичным образом предположили, что белки с малым числом патогенных соматических мутаций необходимы для выживания раковых клеток

В качестве объекта исследования была выбрана меланома, как опухоль, характеризующаяся наибольшей частотой соматических мутаций по сравнению с другими типами рака. Чтобы найти белок-кодирующие гены, находящиеся под отрицательным отбором при развитии опухоли, рассчитывали соотношение dN/dS в качестве индикатора давления отбора для всех белок-кодирующих генов человека. С целью повысить достоверность оценки этой величины из анализа исключали гены, для которых было известно менее чем 10 генетических вариантов. Считали, что ген находится под отрицательным отбором в том случае, если соответствующая величина dN/dS была меньше порогового значения 0,25. Величину порога получили, проанализировав распределение значений dN/dS для трех типов рака с наибольшим количеством точечных мутаций: меланомы кожи, карциномы эндометрия матки и аденокарциномы легких.

Из последующего анализа были удалены те гены, которые незначительно экспрессируются в клетках меланомы и, следовательно, скорее не могут являться критически важными для этого вида рака. Для этого по данным транскриптомного профилирования (374 образца) рассчитали среднюю экспрессию каждого гена в опухоли и исключили гены с уровнями экспрессии, не превышавшими 20%-перцентиля.

На следующем шаге изучали степень патогенности генетических вариантов встречающихся как среди соматических мутаций (данные TCGA), так и среди герминативных полиморфизмов (данные проекта 1000 Genomes). Все генетические варианты были разделены на два класса по степени воздействия на функционирование белка. К первому классу относили мутации приводящие к сдвигу рамки считывания, нонсенс-мутации и нарушения сайтов сплайсинга. Синонимичные варианты относили ко второму классу. Несинонимичные замены были разделены на два класса по функциональному эффекту согласно вычислительным предсказаниям ресурса dbNSFP.

Отбирали гены, которые были обеднены функционально важными соматическими мутациями в опухоли по сравнению с вариантами, встречающимися в здоровых тканях. Для каждого гена вычисляли долю мутаций с высоким функциональным воздействием для клеток меланомы (f_{SKCM}) и для интактной кожи (f_{1KG}) и отбирали гены где выполнялось условие $f_{1KG} > f_{SKCM}$.

После применения вышеописанного набора фильтров получили список из 91 белок-кодирующих гена, предположительно важных для выживаемости меланомы (Таблица 1). Многие белки оказались ассоциированными с клеточной мембраной, включая ионные каналы и транспортёры различных малых молекул, нейрональные белки, молекулы клеточной адгезии и т.д. Для того чтобы подтвердить эту гипотезу, провели анализ обогащения списка белков категориями Gene Ontology, сигнальными путями из баз данных KEGG и PharmGKB.

Таблица 1. Список из 91 белок-кодирующего гена, которые были предсказаны как критически важные для выживания клеток меланомы. Функциональные категории определены с помощью ручной биокурации

Категория	Названия генов
Мембранный транспорт	ABCA3, CATSPER1, SLC12A8, SLC24A1, SLC26A7, SLC27A5, SLC45A1, SLC5A6, SLC9A3, TMC7
Межнейронные контакты, синапсы	CADPS2, EPB41L1, MRGPRX3, MRGPRX4, NRP2, PLXNA2, SEMA4F, SEMA6C, THBS4, UNC5B
Межклеточные контакты, адгезия	CDH11, GJB4, GJB5, ITGB5, LRFN1, LRG1, PCDH13, PCDHGA12, PCDHGC5
Метаболические ферменты	AKR1B1, ENGASE, GCNT3, INPP5B, NMNAT3, NNT, TGM5, UBIAD1
Протеазы и пептидазы	ADAMTS15, BMP1, CAPN11, CAPN12, CASP10, PM20D1, TMPRSS9
Связывание с нуклеиновыми кислотами	ADAR, ERCC6, NFATC2, ZNF195, ZNF493, ZNFX1
Рецепторы цитокинов и гормонов	CCR5, CXCR1, EDAR, MC2R, TNFRSF10A
Киназы	DAPK2, MLKL, PTK2B
Сигнальные каскады в опухолях	BCL2L12, MYCT1, RSPO1
Прочее	CRISPLD2, CYP2J2, EPB41L4B, FGF5, GPR115, HBB, IPO13, LRRC15, MIB2, MYH9, MYO18A, NVL, OR2C3, PDIA4, PDPR, PLD2, PVRL4, RHPN2, SH3BP4, SMYD1, STEAP3, STK11IP, STOX1, TBC1D9B, TMEM104, TMEM63B, TTC7A, UNC45A, VPS18, XPO6

По результатам анализа было найдено статистически достоверное обогащение списка белками плазматической мембраны ($p < 0,001$). Мы предположили, что одним из механизмов, ведущих к наблюдаемой консервативности аминокислотной последовательности белков плазматической мембраны у клеток меланомы, может быть уход опухоли от иммунного ответа посредством клональной эволюции. Мутации, влияющие на функционирование белков плазматической мембраны, могут приводить к формированию неоантигенов, зависящих от основного комплекса гистосовместимости второго типа, МНС II. Белковые эпитопы, связанные с МНС II и распознающиеся субпопуляцией Т-хелперных лимфоцитов, образуются путем протеолиза фагоцитированных внеклеточных белков и периферических белков, сопровождающих интернализированные части мембран. Тот факт, что такие ассоциированные с мембранами белки обеднены патогенными мутациями, может указывать на уход клеток меланомы от иммунитета, опосредованного CD4⁺ Т-клетками. Частично это подтверждается фактом наблюдения значительного ответа CD4⁺ Т-клеток на персонализированные неоэпитопы опухоли, который был зарегистрирован у пациентов с метастатической меланомой (Linnemann, et al., 2015).

Таким образом, присутствие среди критически важных белков значительного количества белков клеточной периферии и мембранных белков может быть результатом отрицательного отбора против образования неоантигенных Т-хелперных эпитопов.

Интересно отметить, что хотя молекулы МНС-II экспрессируются только в ограниченном наборе тканей, таких как антигенпредставляющие клетки, именно клетки меланомы способны к экспрессии различных типов рецептора (Mendez et al., 2009). Это можно рассматривать как косвенное подтверждение предложенной нами биологической интерпретации.

Мы обнаружили, что для некоторых из выявленных белков ранее уже было показано их участие в онкогенезе меланомы. Одним из таких белков являлся хемокиновый рецептор типа 5 (CCR5), широко изученный благодаря его способности обеспечивать слияние вируса иммунодефицита человека с клеткой-мишенью. Было обнаружено, что функциональный CCR5 важен для прогрессирования меланомы: у мышей CCR5 -/- наблюдалось снижение объема опухоли и увеличение выживаемости по сравнению с мышами дикого типа (Song et al., 2012). В работе (Vivanco et al., 2014) сообщалось, что рост и метастазирование меланомы ингибировалось у CCR5 -/- мутантных мышей. Авторы предположили, что сигнализация через CCR5/CCL5 отрицательно влияет на эффективность CD8 Т-лимфоцитов. Таким образом, клетки меланомы, экспрессирующие функциональный ген CCR5, могут способствовать уходу опухоли от иммунного ответа. Тот факт, что предсказание CCR5 как белка, важного для выживания меланомы, было подтверждено экспериментально, иллюстрирует работоспособность представленного подхода. Другим рецептором цитокинов в списке является ген *CXCR1*. Известно, что этот рецептор участвует в процессах роста и метастазирования, а его повышенная экспрессия коррелирует с малигнизацией клеток меланомы (Uen et al., 2015). Можно предположить, что экспрессия интактных белков CCR5 и CXCR1 используется клетками опухоли для выживания в провоспалительном микроокружении.

Еще одно обогащение критически важных белков меланомы связано с комплексом рецепторов семафорина. Примечательно, что оба типа семафорин-связывающих рецепторов (плексины и нейропилины) представлены в списке и могут участвовать в выживании опухолевых клеток (Nasarre et al., 2014). Плексин A2 (PLXNA2) защищает раковые клетки от гибели после инфицирования папилломавирусом человека (Shostak et al., 2014). Экспрессия нейропилина-2 (NRP2) коррелирует с прогрессированием меланомы (Rossi et al., 2014). Эти результаты

служат основой для дальнейших экспериментов по обнаружению мишеней противораковых лекарственных средств среди класса семафоринов.

Провели экспериментальную проверку значимости функционирования некоторых найденных генов для выживания клеток меланомы (Krivosheeva et al., 2020). В качестве метода подавления экспрессии генов была использована технология РНК-интерференции.

В качестве объекта исследования использовали клеточную линию меланомы человека A375. Это наложило существенные ограничения на выбор валидируемых генов-мишеней, поскольку модели клеточных линий не предусматривают межклеточных контактов и не моделируют прогрессию опухолей за счет взаимодействия с микроокружением. Вследствие этого было признано нецелесообразным блокировать мембранные белки, которые на предыдущем этапе были идентифицированы как наиболее перспективные потенциальные мишени. В результате остановились на подавлении экспрессии менее приоритетных внутриклеточных мишеней.

После проведения литературного и экспрессионного анализа были отобраны четыре гена-кандидата, наиболее подходящих для дальнейшего функционального исследования: *UNC45A*, *RHPN2*, *ZNFX1* и *STK11IP*. Для каждого из исследуемых генов было подобрано несколько миРНК (по три миРНК для генов *ZNFX1* и *RHPN2*, и по одной миРНК для генов *UNC45A* и *STK11IP*).

Удалось добиться приблизительно двухкратного уменьшения экспрессии генов-мишеней. Для гена *ZNFX1* только одна миРНК из трех продемонстрировала эффективное подавление экспрессии транскрипта до 42% по сравнению с контрольным образцом, обработанным неспецифической миРНК. Для гена *RHPN2* также использование только одной миРНК привело к снижению экспрессии гена до уровня 46% относительно контроля. Для гена *UNC45A* экспрессию удалось снизить только до 53% по сравнению с контролем. При этом удалось добиться относительно эффективного подавления экспрессии только для гена *STK11IP* — до 13%. В дальнейшем для подавления экспрессии каждого гена использовали одну наиболее эффективную миРНК.

Было найдено, что снижение уровня экспрессии отобранных генов-мишеней не приводит к значимым изменениям жизнеспособности клеточной линии меланомы человека A375. Только для гена *STK11IP* наблюдался слабый эффект снижения жизнеспособности клеток с подавленной экспрессией, однако статистической достоверности достичь не удалось.

Проанализировали изменение уровня апоптоза в клетках со сниженным уровнем экспрессии исследуемых генов по изменению активности каспаз 3/7. По результатам экспериментов обнаружили значимое увеличение активности каспаз 3/7, в два раза — для гена *ZNFx1*, в 1,5 раза — для гена *UNC45A* и в 1,2 раза — для гена *RHPN2*.

Итого, проведенные экспериментальные исследования показывают, что, хотя подавление экспрессии генов *UNC45A*, *RHPN2* и *ZNFx1* приводит к усилению процессов, связанных с апоптозом, но не влияет на уровень жизнеспособности и скорость деления клеток меланомы. В качестве возможного объяснения можно предложить недостаточный уровень подавления экспрессии — было достигнуто только двухкратное обратимое уменьшение уровня экспрессии генов-мишеней. Другим объяснением может быть невозможность воссоздания на клеточной культуре условий, действующих на клетки меланомы *in vivo*.

Стратификация пациентов на основе регуляторов генной экспрессии.

Данный раздел посвящен разработке метода, позволяющего разделять на группы (стратифицировать) образцы в соответствие с выявляемыми особенностями регуляции генной экспрессии.

Общая схема предложенного метода представлена на Рисунке 2. Входными данными служат результаты транскриптомного профилирования: образцы от пациентов, страдающих от определенного заболевания, и контрольная группа образцов. Предлагаемый подход заключается в последовательной «свертке» экспрессионных данных, переходя от полнотранскриптомных профилей к активированным регуляторам экспрессии, а затем и к кластерам таких регуляторов. Применение алгоритма анализа обогащения подсетями (Subnetwork Enrichment Analysis, SNEA) позволяет преобразовать чрезмерно детализированные и, как следствие, мало информативные дифференциальные профили экспрессии генов в информацию о существенно меньшем количестве ключевых регуляторов экспрессии. Разработанный подход позволяет не только классифицировать выборку без учителя (осуществлять стратификацию), но и обеспечивает базис для релевантной биологической интерпретации благодаря идентификации активированных сигнальных каскадов для каждого отдельного пациента. Результатом работы алгоритма является определение подклассов заболевания, причем каждый подкласс характеризуется собственным паттерном изменений в регуляции генной экспрессии.

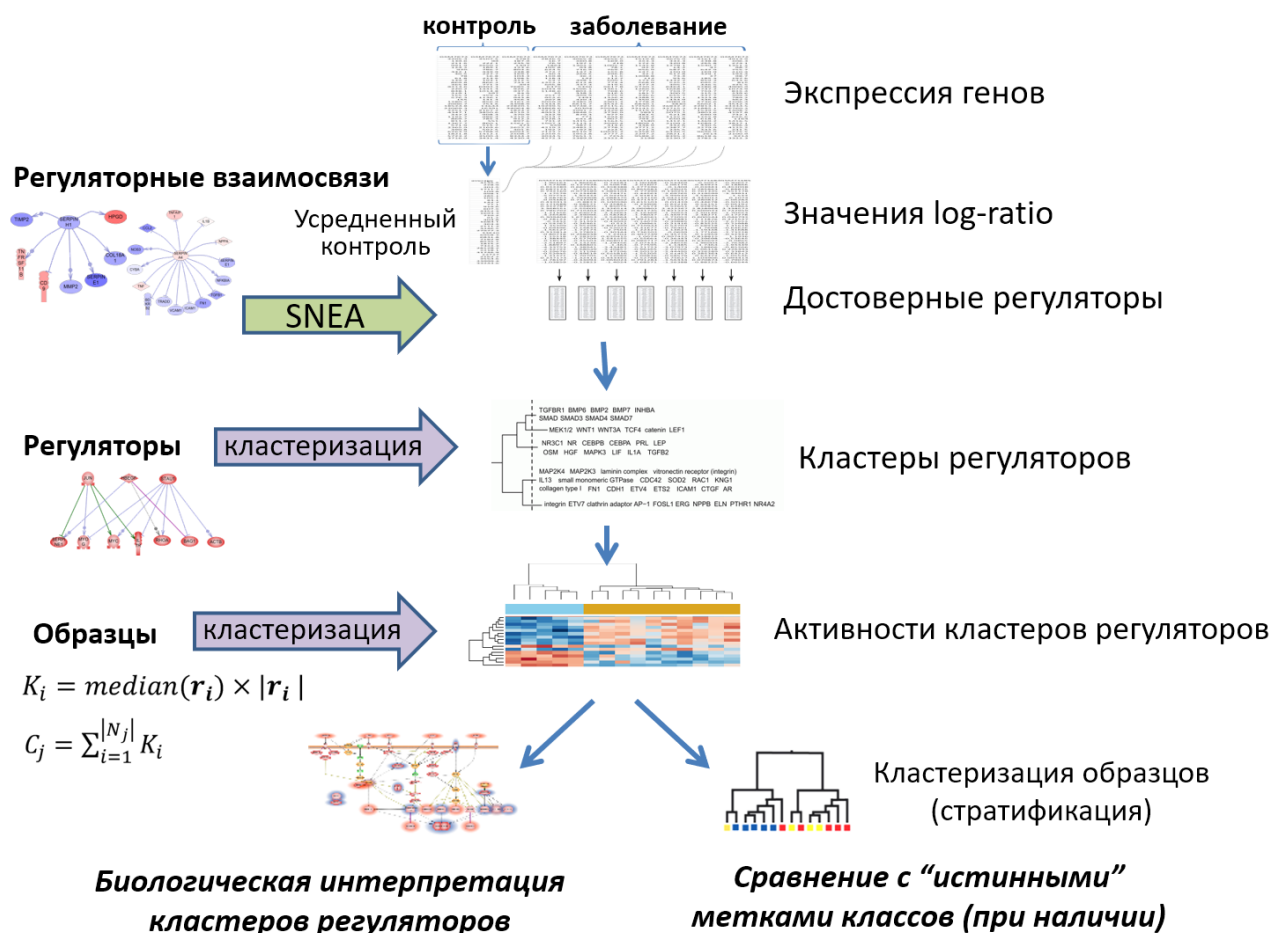


Рисунок 2. Схема предложенного метода стратификации пациентов на основе транскриптомных данных. SNEA — Subnetwork Enrichment Analysis, алгоритм выявления активированных генов-регуляторов

Реализацию разработанного метода проводили в несколько этапов. Образцы из группы изучаемого заболевания сравнивали с усредненным контрольным образцом и рассчитывали т.н. величину logratio для каждого гена. Величину logratio определяли как логарифм отношения средней экспрессии каждого гена при заболевании деленное на среднюю экспрессию гена в контрольной выборке. Полученные величины logratio поступали на вход алгоритму SNEA, получая таким образом список активированных генов-регуляторов. Для каждого образца сохраняли первые 100 наиболее достоверных регуляторов, для которых рассчитанный согласно SNEA уровень значимости не превышал значения 0,05.

На следующем шаге производили группировку регуляторов в кластеры. Этот шаг необходим по следующей причине: было найдено, что из-за биологической изменчивости в различных образцах алгоритм SNEA часто идентифицирует разные регуляторы, которые принадлежат к одному и тому же сигнальному пути. Например, пусть белок SMAD3 будет найден как достоверно активируемый регулятор в одном

образце, тогда как в другом образце был идентифицирован белок SMAD4. Оба регулятора связаны с одним и тем же сигнальным путем TGF- β /SMAD и можно предположить, что данный сигнальный каскад активирован в обоих образцах. Таким образом, представляется целесообразным объединять в группы регуляторы, обладающие сходными наборами генов-мишеней. Это предложение имеет биологический смысл, поскольку факторы транскрипции (составляющие большинство детектируемых регуляторов), как известно, являются избыточными — один член семейства может компенсировать инактивацию другого. Кластеризация регуляторов может рассматриваться как еще один шаг в биологически мотивированном отборе признаков, поскольку из тысяч генов отбираются сотни регуляторов, которые затем группируются в десятки кластеров регуляторов.

После группирования схожих регуляторов, каждому кластеру присваивали значение, описывающее его активность в данном образце на основе экспрессии составляющих кластер генов (см. Материалы и методы). Тем самым исходная матрица транскриптомных профилей размера $N \times m$, где N — число генов, а m — число пациентов, преобразовывали в матрицу размерности $r \times m$, где r — количество кластеров регуляторов, причем $r \ll N$. Полученная матрица активности регуляторов являлась входными данными для кластерного анализа. Использовали иерархическую кластеризацию по Уорду и корреляцию Пирсона как меру сходства двух образцов, однако данный подход может быть обобщен для использования с другим методом кластеризации и метрикой расстояния между образцами. В результате кластеризации матрицы активностей кластеров регуляторов происходит разделение (стратификация) пациентов на группы, которые потенциально соответствуют молекулярным подтипам анализируемого заболевания.

Существуют два типа оценки и интерпретации полученных результатов. Во-первых, если известны истинные подтипы анализируемого заболевания, то возможно оценить точность соответствия предсказанных кластеров образцов этим подтипам с помощью т.н. индекса Рэнда, являющегося количественной мерой совпадения между двумя разбиениями. Помимо оценки точности стратификации важно было оценить и устойчивость работы метода к варьированию количества образцов. Для этого выполняли 100 прогонов 10-кратной кросс-валидации, вычисляя индекс Рэнда для каждого прогона. Сравнивая распределения индексов Рэнда, полученных в результате множества прогонов, возможно статистически оценить достоверность различий между несколькими подходами к стратификации пациентов.

Во-вторых, возможна биологическая интерпретация состава кластеров регуляторов. Этот тип анализа включает в себя различные виды поиска обогащения

кластеров сигнальными и метаболическими каскадами. Идентифицированные кластеры регуляторов, чья активность достоверно различна между подгруппами пациентов, могут быть использованы как основа для анализа молекулярных механизмов, ответственных за специфическое клиническое состояние.

Используя разработанный метод, нами был проанализирован набор данных GSE4183, содержащий информацию о транскриптоме 53 образцов, разделенных на четыре группы: интактный кишечник, колоректальный рак, предраковая аденома, воспалительные заболевания кишечника. В результате применения алгоритма SNEA суммарно было выявлено 1214 регуляторов генной экспрессии, которые были сгруппированы в 28 кластеров. Оценки активности кластеров регуляторов использовали для проведения иерархического кластерного анализа (Рисунок 3). Как следует из рисунка, при разделении полученной дендрограммы на три кластера (пунктирная линия) получаемые кластеры пациентов хорошо соотносятся с клинико-патологическими характеристиками. Так, левый кластер состоял из 15 пациентов, из которых 11 пациентов (73%) относились к классу аденомы, а четыре пациента были приписаны к классу колоректального рака. Изучение тонкой структуры дендрограммы левого кластера показало, что три из четырех образцов колоректального рака также кластеризовались совместно. Средний кластер состоял из 10 пациентов, из которых семь пациентов (70%) относились к колоректальному раку, а три пациента — к классу аденомы. Наконец, правый кластер состоял из 20 пациентов, из которых 15 пациентов (75%) соответствовали воспалительным заболеваниям кишечника, четыре пациента — колоректальному раку и один пациент — аденоме. Таким образом, каждый кластер на 70% и более состоял только из пациентов одной нозологии, что можно интерпретировать как хорошее согласие с истинной разметкой принадлежности пациентов к классам, а значит и высокую точность работы разработанного метода стратификации.

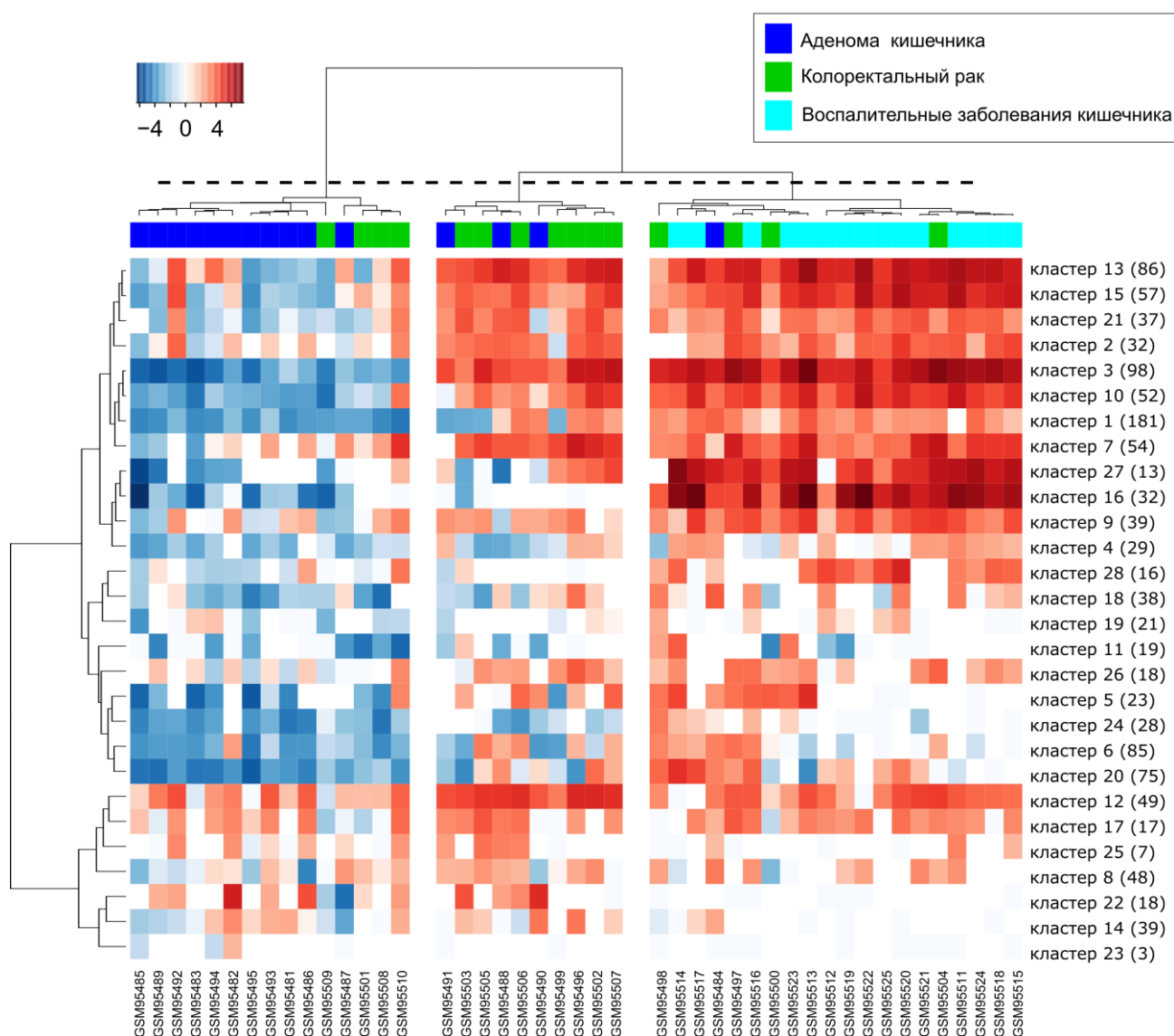


Рисунок 3. Тепловая карта активности кластеров регуляторов для транскриптомного профилирования, эксперимент GSE4183 (Galamb, et al., 2008). Каждый кластер регуляторов представлен строкой, каждый образец — столбцом. В скобках для каждого кластера регуляторов приведено количество входящих в его состав генов-регуляторов. Цвет ячейки кодирует активность кластера регуляторов. Пунктирная линия отображает деление всех образцов на три кластера. Использован иерархический кластерный анализ методом Уорда

Провели количественное сравнение с целью сопоставить точность работы предложенного метода и существующих подходов. В качестве референсных методов использовали алгоритм Gene Set Variation Analysis (GSVA) который трансформирует данные геной экспрессии в активности сигнальных каскадов, а также непосредственную кластеризацию транскриптомных данных методом Partitioning Around Medoids (PAM). Было найдено, что предложенная нами стратификация образцов на основе активностей регуляторов статистически достоверно превосходит по точности как метод PAM, так и алгоритм GSVA. Средний индекс Рэнда для 100

прогонов 10-кратной кросс-валидации для предложенного метода составлял $0,370 \pm 0,014$, тогда как аналогичные значения составили $0,320 \pm 0,009$ и $0,349 \pm 0,015$ для кластеризации методом РАМ и алгоритма GSVA соответственно. В обоих случаях уровень значимости p для непарного критерия Уилкоксона не превышал пороговое значение 0,05. Таким образом, применение разработанного метода позволило улучшить точность стратификации пациентов.

Молекулярные механизмы ответа на терапию цетуксимабом

Задача этого этапа состояла в том, чтобы установить, применим ли развиваемый подход для выявления тонкой структуры механизмов патогенеза. В качестве объекта исследования изучали ответ на терапию цетуксимабом у пациентов с колоректальным раком (КРР). Были проанализированы девять экспериментов по транскриптомному профилированию КРР, включавшие суммарно 784 образца опухоли и 79 образцов здоровой ткани. Имеющийся набор данных о КРР был разделен на обучающую и тестовую выборку. В качестве тестовой выборки использовали эксперимент GSE5851, содержащий информацию о степени чувствительности пациента к терапии цетуксимабом (полный ответ, частичный ответ, стабилизация заболевания и прогрессирование заболевания). Все прочие эксперименты составили обучающую выборку.

Для каждого образца из обучающей выборки, нормализованного на данные экспрессии интактной здоровой ткани, запускали алгоритм SNEA. Суммарно было идентифицировано 695 регуляторов экспрессии, которые являлись статистически достоверными по крайней мере, в одном образце. Регуляторы экспрессии, ассоциированные с вариабельностью транскриптома между отдельными пациентами, классифицировали как «редкие», если они были идентифицированы менее чем в 20 образцах. Обнаружили 522 таких «редких» регулятора и исключили их из дальнейшего анализа. Также были удалены четыре образца, содержащих исключительно «редкие» регуляторы. В итоге обучающая выборка включала 699 образцов и 173 регулятора, сгруппированных в 14 кластеров. Оптимальное число кластеров определили с помощью L-метода (Salvador & Chan, 2004).

Для тестовой выборки было найдено 383 регулятора экспрессии. Из дальнейшего анализа исключили регуляторы, которые не были идентифицированы в обучающем наборе, что привело к итоговому списку из 118 статистически достоверных регуляторов экспрессии. Они были отнесены к 14 кластерам, идентифицированным ранее в обучающей выборке (Таблица 2).

Таблица 2. Кластеры статистически достоверных регуляторов, найденных одновременно в обучающей и тестовой выборках

Кластер	Регуляторы в составе кластера	Функциональная роль
C1	CDX2, Homeo, GATA6, HNF1A, HNF1B, CDX1, GATA4	Регуляция экспрессии генов в печени и кишечнике
C2	IL17A, TAC1, JAK, TLR2, IL22, CALCA, IKBKB, TLR4, MYD88	Воспалительный иммунный ответ
C3	laminin-111, GAST, P-TEFb, nuclear hormone receptor, histone, PAX6, TBX1, PDX1, PKNOX1, NFYB, insulin, NR5A2, IL9, POU2AF1, NR1I2, NGF family, MAP2K7, CALR, NCOA2, POU2F1, MYOCD, THRA	Адипоцитокиновая и гонадотропная сигнализация
C4	CCL5, IL1R1, chemokine, IL11, CCL2, MIF, HMGB1, IL1F8 prostaglandin-endoperoxide synthase, TNFR	Иммунорегуляторные и воспалительные процессы
C5	CAMP, adenosine nucleotide receptor, CFTR, FGF18, TLR5, IL17F	Распознавание патогенов и активация иммунных реакций
C6	OSM, TGFB2, IL1A, PRL, HGF, CEBPB, LIF, NR3C1, LEP, MAPK3, NR, CEBPA	Регуляция массы тела
C7	WNT3A, TCF4, LEF1, MEK1/2	Сигнальный каскад Wnt
C8	F7, fibrinogen, F3, DCN, MET, C5AR1	Свертываемость крови
C9	SMAD4, SMAD, INHBA, SMAD3, SMAD7, BMP7, BMP2, TGFBR1	Сигнальный каскад TGF- β /SMAD
C10	IL1i, CD14, lipoxidase, IL20, MAP3K7, SYK, p38, TXN	Регуляция клеточного цикла
C11	laminin complex, collagen type I, CTGF, MAP2K3, MAP2K4, IL13, ETS2, small monomeric GTPase, ETV4, CDH1, FN1, KNG1, ICAM1, vitronectin receptor (integrin)	Эпителиально-мезенхимальный переход, соединительная ткань
C12	MIRN29C, ZEB1, MYBL2, Maf, MAF	Регуляция транскрипции
C13	NR4A2, PTHR1, FOSL1, NPPB	Ответ на гипоксию
C14	IRF3, DDX58	Регуляция продукции интерферона, ответ на вирусную инфекцию

Вычисляли оценки активности кластеров регуляторов согласно разработанному методу. На Рисунке 4 приведена тепловая карта активности регуляторов для образцов из тестовой выборки, а также отражена информация о чувствительности пациентов к терапии цетуксимабом.

Как следует из результатов кластеризации, 80 пациентов, получавших терапию цетуксимабом, могут быть разделены на четыре кластера. При этом для кластеров K1, K3, и K4 наблюдается приблизительно равная пропорция пациентов, ответивших (стабилизация заболевания, частичный и полный ответ) и не ответивших на анти-EGFR терапию (прогрессирование заболевания). Исключение составляет кластер пациентов K2, преимущественно состоящий из пациентов, не ответивших на терапию цетуксимабом— для 15 из 17 пациентов отмечено прогрессирование заболевания и только для двоих пациентов наблюдалась стабилизация заболевания. Эта группа пациентов была изучена более детально.

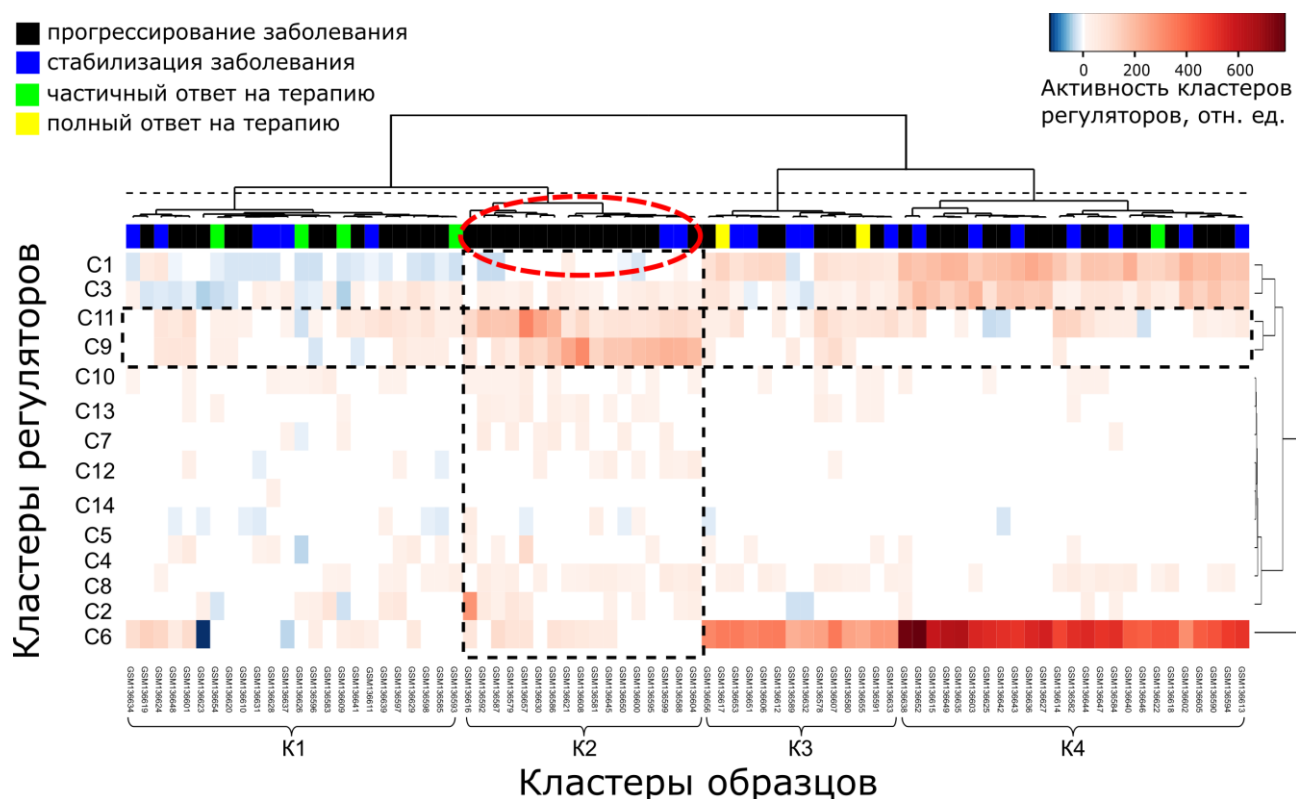


Рисунок 4. Тепловая карта активности 14 кластеров регуляторов экспрессии для 80 пациентов KPP, получавших цетуксимаб. Верхняя шкала кодирует тип ответа опухоли на терапию цетуксимабом (полный ответ, частичный ответ, стабилизация заболевания, прогрессирования заболевания). Пунктиром выделен кластер образцов K2, состоящий в основном из пациентов, характеризовавшихся отсутствием ответа на терапию цетуксимабом и одновременно повышенной активностью сигнального каскада TGF- β / SMAD (кластер регуляторов C9), а также маркерами эпителиально-мезенхимального перехода (кластер регуляторов C11)

Определяли вклад различных признаков, обуславливающих различие между кластерами пациентов, с целью идентифицировать молекулярные механизмы, ассоциированные с отсутствием ответа на терапию цетуксимабом. Обнаружили, что за выделение кластера пациентов K2 в наибольшей степени «отвечают» кластеры регуляторов C9 и C11. Было найдено, что активность указанных кластеров регуляторов достоверно различалась для пациентов из кластера K2 по сравнению с остальной выборкой (непарный тест Уилкоксона, $p < 0,001$).

Обратились к биологической интерпретации кластеров регуляторов C9 и C11. Кластер C9 содержал гены, относящиеся к сигнальному каскаду TGF- β /SMAD. Известно, что нарушения сигнализации TGF- β играют роль в патогенезе солидных опухолей, включая KPP. Трансформирующий ростовой фактор TGF- β уменьшает пролиферацию эпителиальных клеток путем индукции ингибиторов CDK и способствует клеточно-матричным взаимодействиям. На поздних стадиях онкогенеза сигнальный каскад TGF- β становится провоспалительным и

прометастатическим. Предполагается, что значительную роль в этом процессе играет эпителиально-мезенхимальный переход (ЭМП). Кластер C11 содержал гены, непосредственно участвующие в ЭМП: эпителиальные маркеры (комплекс ламинина, коллаген I, Е-кадгерин) и мезенхимальные маркеры (интегрин, фибронектин, CTGF). Другие регуляторы из кластера C11 (транскрипционные факторы ETV4 и ETS2) активируют гены кодирующие матриксные металлопротеиназы, связанные с инвазией и метастазированием опухолевых клеток.

В совокупности, результаты анализа активированных регуляторов позволяют сформулировать гипотезу о том, что причиной отсутствия чувствительности к лечению цетуксимабом является TGF- β -индуцированный ЭМП.

Провели более детальный анализ с целью глубже изучить потенциальные механизмы чувствительности к анти-EGFR терапии. Вычислили профиль дифференциальной экспрессии между пациентами, которые показали частичный или полный ответ против пациентов с прогрессирующим заболеванием. Список достоверно активированных регуляторов, полученный с помощью алгоритма SNEA, включал транскрипционные факторы FOXM1 ($p=0,001$) и NKX2-1 ($p=0,01$), а также цитокин TGFB1 ($p=0,02$). Известно, что NKX2-1 ингибирует TGF- β -индуцированный ЭМП и активируется киназами ERK1/2. Было показано, что FOXM1 также подавляет ЭМП. Можно предположить, что наблюдаемая активация FOXM1 / NKX2-1 у чувствительных к цетуксимабу пациентов, указывает либо на снижение активности сигнального пути TGF- β , либо на неканоническую активность сигнального пути TGF- β , не включающую NKX2-1 и FOXM1.

Проанализировали два эксперимента GSE4342 и GSE20854, в которых изучали изменения транскриптного профиля в клетках немелкоклеточного рака легкого и опухоли эндометрия при ингибировании рецептора эпидермального фактора роста EGFR с помощью gefitinib. В результате были выявлены регуляторы генной экспрессии, ассоциированные с чувствительностью к анти-EGFR терапии. Обнаружили, что в обоих экспериментах выявляемые регуляторы либо принадлежат к сигнальному каскаду TGF- β (INHBA, BMP7), либо участвуют в регуляции ЭМП (ZEB1, SNAI1). Полученные результаты подтверждают вышеизложенную гипотезу о взаимосвязи между ЭМП, сигнальным каскадом TGF- β и чувствительности к анти-EGFR терапии.

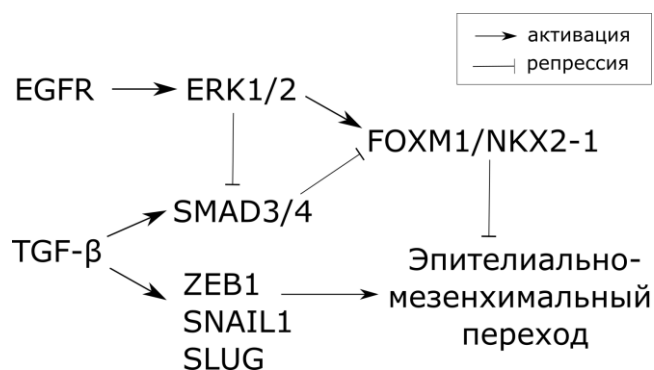


Рисунок 5. Предполагаемая схема устойчивости к терапии цетуксимабом на основе перекреста между сигнальными путями TGF- β и EGFR (упрощенный вариант). Ингибирование сигнального пути EGFR приводит к дерепрессии ЭМП, что в конечном итоге может привести к повторной дифференцировке опухоли и ее независимости от активности EGFR для роста и пролиферации клеток.

На Рисунке 5 приведена схема потенциальной резистентности опухоли к терапии цетуксимабом, вкратце резюмирующая вышеизложенные рассуждения. Показаны перекресты между сигнальными путями TGF- β и EGFR, которые происходят на уровне SMAD3/4 и ERK1/2, а также имеют транскрипционные факторы NKX2-1 и FOXM1 в качестве общей точки пересечения. Как следует из предложенной гипотезы, лекарственное ингибирование сигнального каскада EGFR может привести к дерепрессии процесса ЭМП, что в свою очередь ассоциировано с повторной дифференцировкой опухоли и ее дальнейшей независимости от активности рецептора EGFR для роста и пролиферации клеток.

Разработка системы аннотации генетических вариантов

В данном разделе рассматривается задача анализа генетических вариантов с целью выявления вероятных генов-драйверов онкологической трансформации и определения потенциально эффективной противоопухолевой терапии. Для этого разработали программный конвейер, интегрирующий различные сторонние инструменты и базы данных (Рисунок 6). Итоговая цель работы системы состоит в компилировании максимального количества информации о генах и генетических вариантах, которая призвана облегчить исследователю интерпретацию результатов NGS для определения их тераностической ценности или установления задействованных механизмов прогрессии опухоли.

Входными данными для программного конвейера служит файл формата VCF. Для первичной аннотации, т.е. определения эффектов мутации на уровне генов (синонимичная, миссенс и т.д.), используется программа SnpEff. Дальнейшая аннотация проводится как на уровне генов, так и на уровне генетических вариантов, и

состоит из добавления общебиологической информации, а также сведений о потенциальной роли гена/мутации в онкогенезе.

Для неспецифической аннотации генов из баз данных запрашивается информация о функции, компилируется сводка об уровнях экспрессии в различных тканях, определяются партнеры по белок-белковым взаимодействиям, составляется список сигнальных каскадов, в которых участвует продукт гена. Затем каждый ген классифицируется согласно системе Cancer Hallmarks, проверяется его присутствие в списках известных онкогенов и опухолевых супрессоров. Для генетических вариантов определяется влияние на функционально-важные участки белка (сайты фосфорилирования, активные центры и т.д.), предсказывается патогенность мутации, собираются данные о популяционных частотах. Далее добавляется информация о клинической и фармакогеномной значимости мутации, предсказывается степень “драйверности”, компилируются данные о встречаемости мутации в опухолях различной локализации.

В качестве тестовых данных использовали результаты полноэкзомного секвенирования образцов от пяти пациентов с колоректальным раком из российской популяции. В среднем для каждого из десяти образцов получили 105 миллионов парноконцевых прочтений длиной 100 оснований. После картирования на референсный геном средняя глубина покрытия прочтениями оказалась более чем 100-кратной для всех образцов кроме образца K31, где она оказалась 84-кратной. По итогам идентификации соматических мутаций для каждого образца в среднем было найдено 69,6 миссенс-мутаций, 2,8 инсерций и делеций (инделов) со сдвигом рамки считывания, 0,6 инделов не приводящих к сдвигу рамки считывания, 5,4 нонсенс-мутаций и 1,0 мутаций участков сплайсинга.

По результатам анализа для трех из пяти пациентов была выявлена мутация кодона 12 гена *KRAS*. Наличие этой мутации ассоциировано с потенциальной неэффективностью терапии ингибиторами EGFR (цетуксимаб, панитумумаб). Мутации классических генов-супрессоров опухоли *APC* и *TP53* были обнаружены в четырех случаях из пяти. Однако присутствие этих, безусловно, “драйверных” мутаций носит скорее информативный характер – на сегодняшний день практически нет эффективных способов транслировать это знание в клиническую практику.

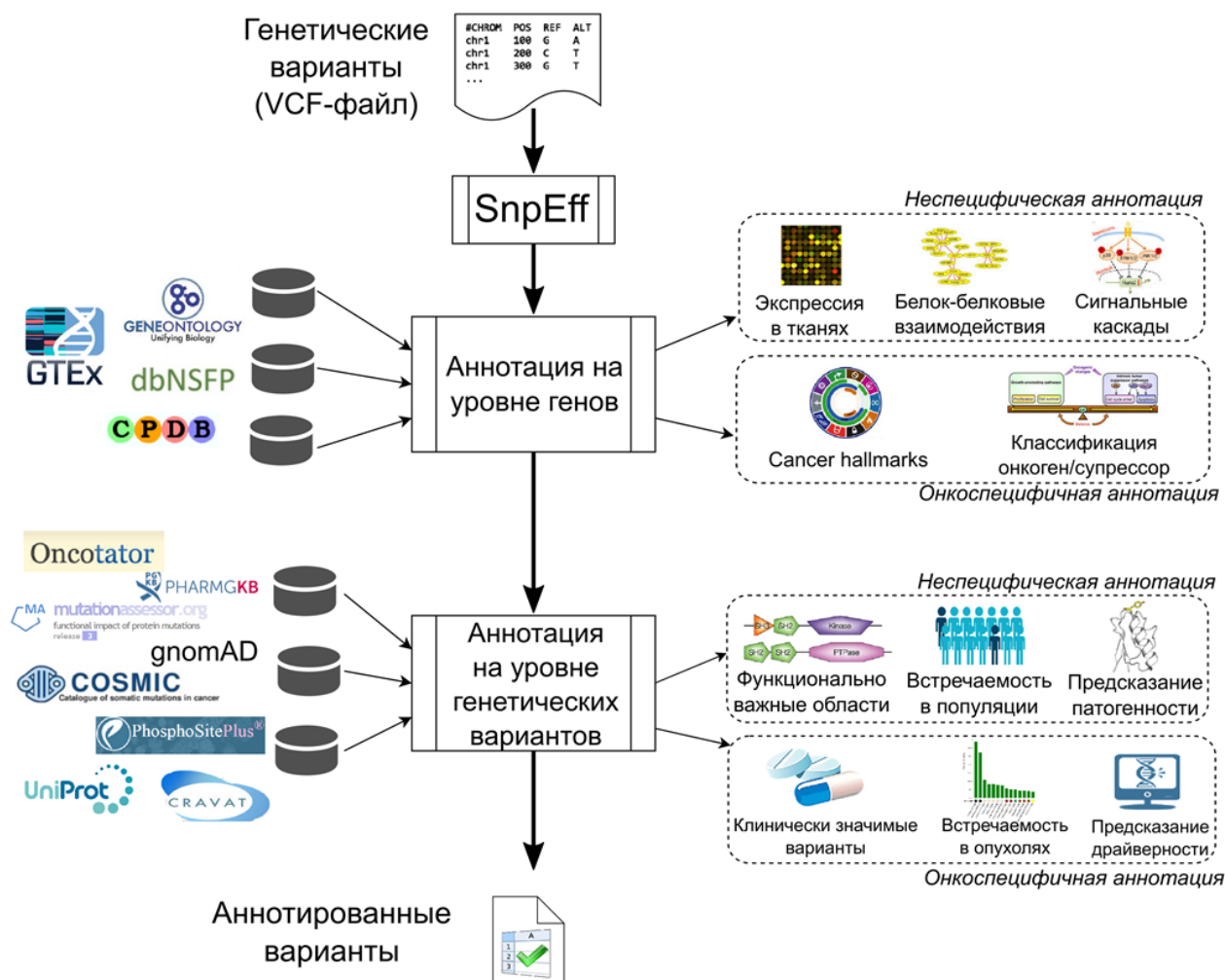


Рисунок 6. Общая схема разработанного программного конвейера по аннотации генетических вариантов

При изучении генома опухоли пациента K25 были обнаружены мутации в генах-супрессорах опухоли *SMAD4* и *BRCA1*. Мутации в гене *SMAD4* при KPP ассоциированы с ухудшением выживаемости. Выявленная мутация в гене *BRCA1*, согласно вычислительному предсказанию, нарушает функцию белка, что может нарушать работу системы репарации двухцепочечных разрывов ДНК.

У пациента K26 была найдена мутация, приводящая к сдвигу рамки считывания для гена деубиквитирующего фермента *USP9X*, что может свидетельствовать о чувствительности к воздействию 5-фторурацила. Также была обнаружена мутация в гене *XPO1*. В работе (Ferreiro-Neira et al., 2016) сообщается о повышенной чувствительности к облучению для клеточных линий KPP и ксенотрансплантатов при воздействии селективным ингибитором *XPO1* селинексором. Эти данные свидетельствуют в пользу химиолучевой терапии на основе 5-фторурацила, как возможной стратегии лечения для пациента K26.

У пациента K27 была найдена нонсенс-мутация гена *CDKN1B*. Отсутствие или уменьшение экспрессии *CDKN1B* может быть связано с неблагоприятным прогнозом во многих типах рака, включая КРР. Также была обнаружена мутация старт-кодона в транспортёре ионов меди *ATP7A*, что может способствовать восстановлению чувствительности к терапии препаратами платины.

В процессе анализа данных пациента K28 не удалось идентифицировать дополнительных мутаций-“драйверов”, помимо уже описанных нарушений в генах *APC* и *KRAS*. Интересной находкой явилась мутация p.R962H в гене *KDR*, кодирующем рецептор сосудистого эндотелиального фактора роста. Было показано, что у пациентки, страдающей КРР и характеризующегося мутацией соседней аминокислоты p.R961W, наблюдалась хорошая переносимость и исключительный ответ на терапию регорафенибом, включавший значимое клиническое и радиологическое улучшение (Loaiza-Bonilla, et al., 2016).

Единственная связанная с опухолями мутация, обнаруженная у пациента K31, была локализована в гене *TET2*. Ген *TET2* обладает свойством онкосупрессора в миелоидных опухолях и регулирует миелопоэз, а также играет роль в активном деметилировании ДНК. Специфическая роль гена *TET2* при КРР требует дальнейшего исследования.

Программный конвейер также был использован для аннотации результатов полноэкзомного секвенирования образца гепатоцеллюлярной карциномы из российской популяции. Всего было выявлено 9250 однонуклеотидных замен в кодирующих участках, включая 77 соматических и 9173 герминативных генетических варианта. В частности, была обнаружена соматическая мутация *ERBB3* p.D1014N, которая является драйверной согласно методу CHASM. Вкупе с обнаруженной гиперэкспрессией гена *EGF*, это может свидетельствовать об aberrантной активации сигнального каскада *ERBB3*, что в свою очередь позволяет предположить о потенциальной эффективности таких препаратов как эрлотиниб и лапатиниб. Также в образце была детектирована соматическая нонсенс-мутация *BTG1* p.K150*, с большой вероятностью приводящая к утрате функции данного белка. Среди белков-партнеров, чья активность подавляется *BTG1*, важную роль играют антиапоптотические гены *MMP9*, *BCL2* и *CCND1*. Это позволяет выдвинуть гипотезу о том, что в исследуемом случае гепатоцеллюлярной карциномы ген *BTG1* является «драйвером» онкогенеза. Был выявлен ряд герминативных генетических вариантов, связанных с побочными реакциями на лекарственную терапию. Так была обнаружена замена *DPYD* p.R29C, ассоциированная с повышенной токсичностью при терапии 5-фторурацилом. Наличие у пациента генетических вариантов *XPC* p.Q902K, *SLC22A2*

p.S270A, XRCC1 p. R194W, и LRP2 p. K4094E связано с повышенным риском лекарственной токсичности при лечении цисплатином, а наличие полиморфизма MTHFR p.E429A коррелирует с повышенным риском миелосупрессии у пациентов, получавших метотрексат.

В ходе создания и применения программного конвейера по аннотации генетических вариантов, был получен ценный опыт по использованию и т.н. гармонизации (корректному объединению) различных баз знаний, потенциально пригодных для клинического использования. Этот опыт был использован при разработке описанной ниже программной системы.

Веб-платформа для прецизионной онкологии

Нами была спроектирована и реализована программная платформа, позволяющая проводить автоматическую интерпретацию молекулярного профиля онкологического пациента. Система предназначена для автоматической генерации списка потенциально эффективных (а также неэффективных и токсичных) лекарственных препаратов, исходя из молекулярно-генетических анализов опухоли полученных с помощью высокопроизводительного секвенирования (NGS), иммуногистохимического анализа (ИГХ), флуоресцентной гибридизации *in situ* (FISH), определения микросателлитной нестабильности. Результаты интерпретации предоставляются на веб-ресурсе, где в интерактивной форме приводятся описания всех найденных препаратов, биомаркеров и клинических испытаний. Пользователю доступны ссылки на оригинальные публикации, в которых описаны взаимосвязи между биомаркером (генетическим вариантом, экспрессией белка, микросателлитной нестабильностью и т.д.) и эффективностью препарата. Помимо режима автоматического анализа, поддерживается и ручное редактирование списка найденных препаратов.

Для решения задачи выбора противораковой терапии не применялись методы машинного обучения и искусственного интеллекта. Отнесение каждого препарата в категорию потенциально эффективных или неэффективных было обусловлено исключительно наличием ссылок на публикации в рецензируемых изданиях, а не является результатом вычислительного предсказания. Разработанная веб-платформа не позиционируется как медицинское изделие. Предоставляемая системой информация о молекулярном профиле опухоли и его потенциальной связи с ответом на терапию носит исключительно ознакомительный характер, а вся ответственность за принятие клинических решений лежит на враче-онкологе.

Архитектура веб-платформы для прецизионной онкологии представлена на Рисунке 7. Входными данными являются результаты молекулярно-генетических анализов образца опухоли. Генетические варианты, идентифицированные с помощью технологий NGS, загружаются на платформу в виде файла формата VCF. Результаты прочих исследований (в первую очередь, данные ИГХ и FISH) предоставляются как текстовый файл определенного формата. Данные о пациенте и конфигурационные параметры (расположение справочных таблиц, настройки фильтрации генетических вариантов и т.д.) хранятся в виде файла формата TOML.

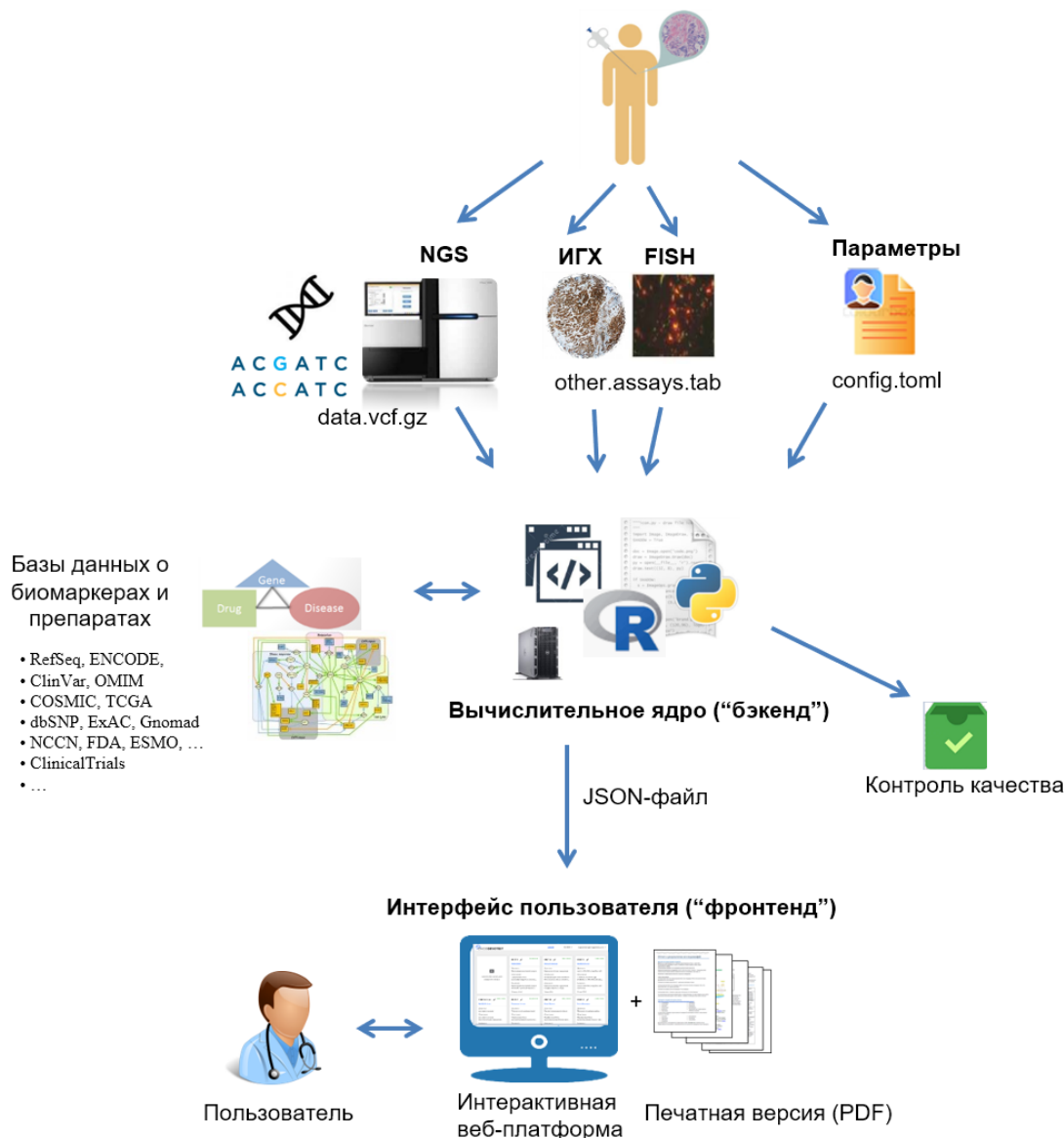


Рисунок 7. Архитектура веб-платформы для прецизионной онкологии. Результаты молекулярного профилирования опухоли передаются в специализированных форматах в вычислительное ядро (т.н. "бэкенд"-компонент), реализованный в виде набора R- и Python-скриптов. Бэкенд-компонент генерирует результаты (аннотированные генетические варианты, списки препаратов и т.д.) которые передаются т.н. "фронтенд"-компоненту. Задачей последнего является отображение результатов подбора противоопухолевой терапии в виде интерактивной веб-страницы, а также генерирование печатной версии отчета

Программная архитектура платформы состоит из двух компонентов: 1) “бэкенд”, вычислительное ядро, предназначенное для обработки входных данных и 2) “фронтенд”, обеспечивающий веб-интерфейс и взаимодействие с пользователем. В результате работы бэкенда формируется файл формата JSON, который загружается в базу данных сервера. Фронтенд-компонент реализован на языках Python и JavaScript, и отображает результаты подбора противоопухолевой терапии в виде интерактивной веб-страницы, обращаясь к базе данных сервера.

Бэкенд-компонент платформы представляет собой набор скриптов на языках Python и R. Основной задачей бэкенд-компонента является сопоставление выявленных у данного пациента биомаркеров с базой знаний по эффективности противоопухолевых препаратов. Например, пусть в базе знаний содержится информация об эффективности BRAF-ингибиторов (в частности, вемурафениба) для пациентов с меланомой, у которых выявлена замена V600E в гене BRAF (Chapman, et al., 2011). В случае, если у анализируемого пациента с диагнозом “меланома” в результате секвенирования генома опухоли будет обнаружен генетический вариант BRAF p.V600E, то по итогам работы бэкенд-компонента вемурафениб будет занесен в список потенциально эффективных препаратов при данном заболевании. Если у пациента злокачественная опухоль иной природы (не меланома), то вемурафениб будет отнесен к препаратам, потенциально эффективным при других заболеваниях. Отдельно формируются такие списки препаратов как “Потенциально неэффективные препараты”, “Токсичные препараты”, “Другие препараты”. Например, известно, что наличие мутаций в гене KRAS при колоректальном раке ассоциировано с неэффективностью применения тамоксифена (Benson, et al., 2017). Если у пациента с колоректальным раком имеется генетический вариант UGT1A1 p.G71R, то алгоритмы веб-платформы добавят к списку “Токсичных препаратов” иринотекан согласно работе (Rouits, et al., 2004).

Наиболее частым сценарием использования платформы является секвенирование образца опухоли при отсутствии экспериментальных данных об образце заведомо здоровой ткани. В таком случае затруднено достоверное разделение генетических вариантов на герминативные и соматические, поэтому использовали следующую процедуру для приблизительной классификации вариантов. Считали, что вариант является герминативным, если его популяционная частота в европейской популяции согласно ресурсу gnomAD превышала 0,1%, а также если вариант не встречался в базе данных COSMIC (Tate, et al., 2019). Отдельно сообщали о наличии мутаций в 59 генах, ассоциированных с высокопенетрантными генетическими заболеваниями согласно рекомендациям Американской коллегии медицинской

генетики (Kalia, et al., 2017). Обнаружение таких вариантов (т.н. “случайных находок”), которые не отмечены в базе ClinVar (Landrum, et al., 2014) как доброкачественные может служить основанием для консультации пациента с медицинским генетиком.

Каждый генетический вариант относили к одному из трех классов. В том случае, если данный генетический вариант обладал тераностической ценностью — встречался в базе данных биомаркеров — его относили к категории “клинически значимых вариантов”. В категорию “Полиморфизмы” относили все генетические варианты, чья популяционная частота превышала 0.1% и для которых не была показана их клиническая значимость. Все прочие генетические варианты относили к категории “Варианты с неизвестной значимостью”.

Ключевым элементом веб-платформы является база данных, содержащая сведения о клинической значимости биомаркеров. Клиническая ценность биомаркера может быть описана следующим образом: “в исследовании с Pubmed-идентификатором [PMID:16236737] [источник], показано, что наличие экспрессии *HER2* [биомаркер] при раке молочной железы [заболевание] ассоциировано с чувствительностью [тип эффекта] к трастузумабу [препарат]”. База данных состоит из множества подобных записей, связывающих биомаркер, препарат, тип эффекта и заболевание. В качестве биомаркеров поддерживаются любые молекулярно-генетические события – соматические и герминативные генетические варианты, изменения генной экспрессии или копийности, структурные перестройки, уровень опухолевой нагрузки, наличие микросателлитной нестабильности, и т.д.

Клинически значимые биомаркеры в базе данных были разделены на четыре класса: предиктивные, диагностические, предрасполагающие и прогностические. Принадлежность к классу “предиктивные биомаркеры” означает, что существует взаимосвязь с каким-либо лекарственным препаратом. При этом эффект препарата может быть как положительным (опухоль чувствительна к терапии), так и отрицательным (опухоль резистентна к воздействию терапии). Биомаркер также может иметь диагностическую ценность: мутации в гене *TERT* чаще встречаются при глиобластомах и олигодендроглиомах, нежели при астроцитомах (Yuan, et al., 2016). Предрасполагающие биомаркеры означают существование фактора риска, который может привести к развитию заболевания: известно, что наличие определенных мутаций в гене *MUTYH* является фактором риска для развития колоректального рака (Yanus, et al., 2018). В том случае, если заболевание уже развилось, биомаркер может иметь прогностическую ценность: показано, что наличие мутаций в гене *DNMT3A*

ассоциировано с неблагоприятным прогнозом при остром миелобластном лейкозе (Metzeler, et al., 2016).

Важным атрибутом клинической значимости биомаркера является достоверность — степень уверенности в корректности информации. Опираясь на подход, использованный в ресурсе CIViC (Griffith, et al., 2017), нами была принята следующая классификация уровней доказательности (по мере уменьшения надежности):

1. Уровень А: одобрение FDA, рекомендации NCCN, ASCO, ESMO, Минздрава России, RUSSCO.
2. Уровень В: клинические испытания и другие исследования
 - 1) В1: клинические исследования поздних фаз (III/IV фазы)
 - 2) В2: клинические исследования ранних фаз (I/II фазы)
 - 3) В3: мета-анализ нескольких публикаций
 - 4) В4: когортное исследование
3. Уровень С: отдельные клинические случаи
4. Уровень D: доклинические исследования (клеточные линии, животные модели)
5. Уровень E: исследования *in silico* (вычислительные предсказания)

Для каждой записи в базе данных клинической значимости биомаркеров указан источник информации (ссылка на публикацию или клиническое исследование) и уровень доказательности согласно приведенной шкале.

Исходной точкой входа в платформу является веб-страница <https://youroncohelpr.oncogenotest.com> (требуется предварительная авторизация). На стартовой странице отображаются карточки пациентов, для которых у текущего пользователя имеется доступ. При щелчке по соответствующей карточке система переходит в режим работы с отдельным пациентом: в левой части экрана отображается меню навигации по разделам, а правая часть экрана используется для визуализации имеющихся данных (Рисунок 8). Пользователю доступны следующие разделы: «Информация», «Гистология», «Секвенирование», «Дополнительные исследования», «Препараты», «Клинические испытания» и «Заключение».

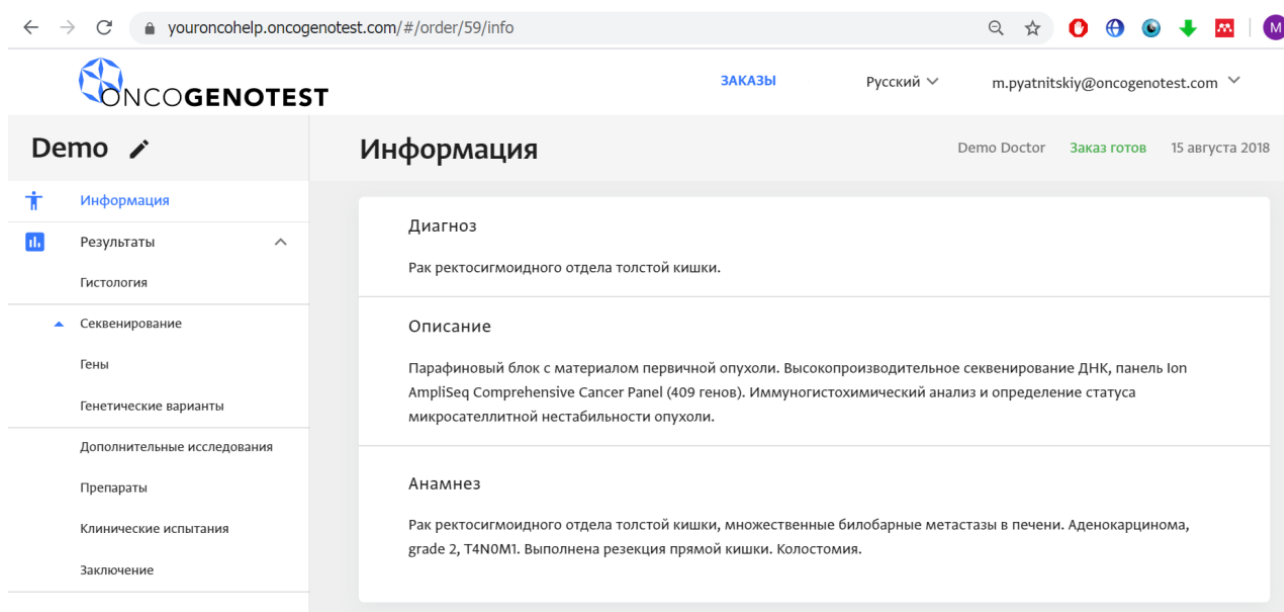


Рисунок 8. Платформа для прецизионной онкологии

В разделе «Информация» указаны общие сведения о пациенте, включая диагноз и анамнез, а также описание имеющихся биоматериалов и проведенных молекулярно-генетических анализов.

В разделе «Гистология» поддерживается добавление цифровых изображений, полученных при гистологическом изучении препарата. Приводится описание гистологического типа опухоли, указывается оценка доли опухолевых клеток. Эти значения заполняются при патоморфологическом анализе. Дальнейшие исследования (в том числе секвенирование генома) целесообразно проводить только в том случае, если доля опухолевых клеток в препарате составляет не менее 30%.

Раздел «Секвенирование» предназначен для отображения информации о результатах NGS-анализа и состоит из двух подразделов. В подразделе «Гены» перечислены все гены, которые были частично или полностью покрыты при секвенировании. В подразделе «Генетические варианты» приведен список всех однонуклеотидных замен и коротких инделов, которые были выявлены в процессе анализа. Для каждого генетического варианта в результате работы вычислительного ядра автоматически заполняются следующие поля: предположительная классификация как соматического/герминативного варианта, геномные координаты (для версии сборки генома человека hg19), количество прочтений, поддерживающих референсный и альтернативный аллель, аллельная частота, предсказание влияния на функцию белка (согласно базе dbNSFP), частота встречаемости варианта в популяции (согласно ресурсу gnomAD), а также принадлежность к одному из трех классов

(клинически значимые варианты, полиморфизмы, варианты с неизвестной значимостью). Другие поля заполняются или редактируются экспертом для генетических вариантов, представляющих наибольший интерес с клинической точки зрения. В поле «Биологический эффект» в свободной форме может указываться информация о локализации варианта относительно доменной структуры белка и известных экспериментальных работах по изучению данного генетического варианта. Поле «Комментарий» используется для текстового описания клинических проявлений наличия данного варианта (Рисунок 9).

Описание генетического варианта KRAS p.G12D 	
Ген	KRAS
Тип	Замена, missense_variant, Соматическая
Координаты	chr12:25398284:C>T
Кол-во прочтений (реф./альт.)	390 / 86
Аллельная частота	18.1%
Эффект на уровне белка / ДНК	p.Gly12Asp / с.35G>A , Возможно патогенная
Биологический эффект	Данный генетический вариант располагается в GTF-связывающем центре гена KRAS. G12D приводит к снижению GTF-азной активности KRAS и активации сигнальных путей PI3K/AKT/mTOR и RAS/RAF/MEK/ERK (PMID: 23622131).
Частота в популяции	0%
Комментарий	Согласно последним рекомендациям NCCN, основанных на данных клинических исследований CRYSTAL и OPUS, наличие активирующих мутаций в гене KRAS является противопоказанием для добавления анти-EGFR терапии (Цетуксимаб, Панитумумаб) к стандартной схеме лечения (PMID: 24024839 , https://meetinglibrary.asco.org/record/89575/abstract).

Рисунок 9. Информация о генетическом варианте

Ключевой информацией о клинически значимых генетических вариантах является список ассоциированных препаратов, а также прогностических и диагностических последствий (Рисунок 10). На этой странице перечисляются все найденные в базе данных сведения о клинической значимости данного варианта, включая текстовое описание, уровень достоверности, ссылку на источник. Все лекарственные препараты разделены на пять категорий: потенциально эффективные при данном заболевании, эффективные при другом заболевании, неэффективные, токсичные и прочие. Активны гиперссылки на название лекарственного препарата

(открывается окно с детальным описанием препарата) и первоисточник (открывается соответствующая запись базы данных Pubmed или другая веб-страница). Таким образом пользователь веб-платформы может самостоятельно перепроверить опубликованные данные о взаимосвязи генетического варианта и препарата, а значит, и корректность отнесения препарата в ту или иную категорию. В отдельную таблицу вынесены прочие клинические последствия генетического варианта, не связанные с препаратами. Указан тип клинического следствия (прогностический, диагностический, предрасполагающий), текстовое описание, уровень достоверности, и ссылка на первоисточник.

Препараты для биомаркера KRAS:p.Gly12Asp

Потенциально эффективные при данном заболевании

Адоптивная иммунотерапия

С: Отд. клин. случаи

Описание

В исследование включены наблюдение одного пациента с метастатическим колоректальным раком (в клиническом испытании 2 фазы, NCT0174121) и эксперименты in vitro для 24 культур лимфоцитов, инфильтрирующих опухоли (TIL), выделенные из опухолевого материала метастазов в легких. После того, как культуры были испытаны на кросс-реактивность, пациенту вводили 1,48 × 10¹¹ TIL (75% CD8 + Т-клеток), которые были специфически реакционноспособны к клеткам с вариантом KRAS G12D. Все 7 метастатических поражений в легких регрессировали через 40 дней после терапии, и у пациента наблюдался 9-месячный частичный ответ на терапию.

N. Engl. J. Med., 2016

Потенциально эффективные при других заболеваниях

Абемациклиб

В: Клин. данные

Описание

В многоцентровом исследовании I фазы применения Абемациклиба было зарегистрировано в общей сложности 225 пациентов: 33 в режиме эскалации дозы и 192 в опухолеспецифичных когортах. Ответы на терапию были достигнуты у ранее леченных пациентов с раком молочной железы, немелкоклеточным раком легкого и меланомой. Частота контроля заболевания (disease control rate) для подгруппы пациентов с мутацией KRAS составила 55% (16/29) по сравнению с 39% (13/33) в когорте пациентов с опухолями с диким типом KRAS. Стабилизация заболевания в течение 24 недель было достигнуто для 31% (9/29) пациентов с KRAS-мутантными опухолями против 12% (4/33) пациентов с KRAS дикого типа. Медиана выживаемости без прогрессирования составила 2,8 месяца и 1,9 месяца для подгрупп пациентов с мутацией KRAS и KRAS дикого типа, соответственно.

Cancer Discov, 2017

Неэффективные при всех заболеваниях

Цетуксимаб

А: Подтверждено

Описание

Статус мутации KRAS оценивали в 394 образцах опухолей у пациентов с колоректальным раком, которые были рандомизированы для получения Цетуксимаба в сочетании с поддерживающей терапией или только поддерживающей терапии (NCT00079066). 42% оцениваемых опухолей имели по крайней мере одну мутацию в 2 экзоне KRAS. Для пациентов с колоректальным раком с мутантным вариантом K-Ras терапия Цетуксимабом была неэффективна, в отличие от пациентов с диким типом K-Ras. Эффективность цетуксимаба была статистически достоверно ассоциирована со статусом мутации KRAS (p= 0,01 для общей выживаемости и P <0,001 для выживаемости без прогрессирования). У пациентов с опухолями с KRAS дикого типа схема лечения с Цетуксимабом по сравнению с поддерживающей терапией значительно улучшало общую выживаемость.

N. Engl. J. Med., 2008

Селуметиниб + Цетуксимаб

В2: Клин. испытания, ранняя фаза

Прогностические и диагностические следствия для биомаркера KRAS:p.Gly12Asp

Тип	Описание	Ссылка	Доказательность
Прогностический	KRAS mutations were significantly associated with reduced progression-free and overall survival in patients with advanced colorectal cancer.	J. Clin. Oncol., 2008	D: Доклин. исследования
Прогностический	"In 119 metastatic colorectal cancer patients treated with cetuximab-based first-line chemotherapy, KRAS G12D mutations (N=44) were associated with longer overall patient survival when compared to other KRAS G12 mutations (23.3 vs 14-18 months; G12V, A, C, S, R)."	Oncology, 2012	В: Клин. данные

Маркеров на странице 10 1 - 2 из 2

Рисунок 10. Препараты и клинические следствия для генетического варианта

Следующим элементом платформы является раздел «Дополнительные исследования». Этим термином обозначаются все молекулярно-генетические исследования, которые не относятся к высокопроизводительному секвенированию: результаты ИГХ, FISH и подобных методов. В этом разделе указывается информация об исследуемом материале (чаще всего это парафиновый блок), экспериментальном методе и собственно результате исследования. Для тех дополнительных исследований, для которых была найдена клиническая значимость, приводится общее количество ассоциированных препаратов (по всем предиктивным связям) и прочих маркеров (по всем диагностическим, прогностическим и предрасполагающим связям), и максимальный уровень достоверности для всех имеющихся сведений о клинической значимости о биомаркера.

В разделе «Препараты» скомпилированы все лекарственные препараты, которые были найдены как ассоциированные по крайней мере с одним биомаркером (генетический вариант или результат дополнительного исследования). В таблице перечислены названия препаратов, указана возможность применения препаратов при данном заболевании, ассоциированный биомаркер и уровень доказательности. Раздел компонуется из следующих подразделов:

- а) Потенциально эффективные препараты: опубликованы данные об эффективности препарата в случае обнаружения специфического биомаркера при этом заболевании (диагноз у текущего пациента).
- б) Препараты, потенциально эффективные при других заболеваниях: опубликованы данные об эффективности препарата в случае обнаружения специфического биомаркера при ином заболевании (опухоль другой локализации или гистологического типа).
- в) Потенциально неэффективные препараты: опубликованы данные о неэффективности препарата в случае обнаружения специфического биомаркера при этом или другом заболевании.
- г) Потенциально токсичные препараты: опубликованы данные о большей вероятности развития нежелательных явлений в случае обнаружения специфического биомаркера.
- д) Другие препараты: опубликованные данные об эффективности препарата в случае обнаружения специфического биомаркера отсутствуют или противоречивы. Препарат одобрен при данном заболевании.

В разделе «Клинические испытания» перечислены актуальные клинические исследования, участие в которых может потенциально рассматриваться пациентом и его лечащим врачом. Отображаются названия клинических исследований, их фаза и

цель, а также контакты лечебных учреждений, на базе которых проводятся данные испытания.

Раздел «Заключение» формируется экспертом вручную и обычно включает в себя следующие составляющие: описание клинической значимости каждого выявленного биомаркера, краткие выводы с перечислением потенциально эффективных и неэффективных препаратов, описание ограничений проведенного анализа, и прочая справочная информация.

Помимо функционала поиска и аннотации генетических вариантов, важным аспектом системы является доступность и дружелюбность пользовательского интерфейса. При проектировании веб-платформы стремились максимально упростить все разделы системы и минимизировать необходимые действия пользователя. В итоге, разработанный интерфейс позволяет быстро получить доступ к информации о том, какие терапии являются потенциально эффективными для данного пациента, просмотреть найденные генетические варианты в других базах данных и оперативно получить все ссылки на первоисточники. Все это помогает врачу-онкологу сформировать собственное мнение об оптимальной схеме терапии для пациента.

Разработанная программная система не свободна от недостатков и может быть усовершенствована. В частности, доступные пользователям-экспертам функции редактирования сущностей (ген, генетический вариант, дополнительное исследование, препарат) и взаимосвязей между ними осуществляется через достаточно сложный интерфейс библиотеки Django. Требуют значительного улучшения процессы загрузки в систему исходных данных, поддержка различных генных панелей и т.д. В то же время полученный опыт показал перспективность применения подобных систем в онкологии и является хорошим фундаментом для дальнейших работ в этой области.

ВЫВОДЫ

1. Разработан подход к поиску белков, критически важных для выживания опухолевых клеток, основанный на идентификации генов, находящихся под воздействием отрицательного отбора при клональной эволюции опухоли.

2. Получен список из 91 белок-кодирующего гена, потенциально важных для функционирования клеток меланомы. Показано, что список обогащен белками, ассоциированными с клеточной периферией и плазматической мембраной. Предложено биологическое обоснование наблюдаемого феномена как результат отрицательного отбора против образования неоантигенных эпитопов.

3. Разработан метод для стратификации пациентов, основанный на выявлении активированных генов-регуляторов экспрессии по данным транскриптомного профилирования. Метод реализован в виде общедоступного программного обеспечения.

4. На примере транскриптомных данных по заболеваниям толстого кишечника показано, что представленный подход превосходит по точности аналогичные методы. Выявлена группа пациентов, характеризовавшаяся отсутствием ответа на терапию цетуксимабом с одновременной активацией сигнального каскада TGF- β /SMAD и маркеров эпителиально-мезенхимального перехода, что позволило обозначить отдельные молекулярные механизмы наблюдающейся резистентности для колоректального рака.

5. Разработан программный конвейер для аннотации генетических вариантов, предназначенный для помощи эксперту в идентификации клинически значимых мутаций по результатам высокопроизводительного секвенирования генома опухоли.

6. Проанализированы данные полноэкзомного секвенирования образцов колоректального рака и гепатоцеллюлярной карциномы из российской популяции. Идентифицирован ряд онко-ассоциированных и клинически значимых мутаций, что могло быть использовано для оптимизации стратегии лечения.

7. Разработана интернет-платформа для обработки и интерпретации молекулярного профиля онкологического пациента. Система позволяет автоматически определять список потенциально эффективных, неэффективных и токсичных лекарственных препаратов, основываясь на результатах молекулярно-генетических анализов опухоли, таких как высокопроизводительное секвенирование генома, иммуногистохимический анализ, флуоресцентная гибридизация *in situ*. Использование подобных программных продуктов в клинической практике позволяет реализовывать принципы персонализированной медицины и прецизионной онкологии.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Nikitina A.S., Lipatova A.V., Goncharov A.O., Kliuchnikova A.A., **Pyatnitskiy M.A.**, Kuznetsova K.G., Hamad A., Vorobyev P.O., Alekseeva O.N., Mahmoud M., Shakiba Y., Anufrieva K.S., Arapidi G.P., Ivanov M.V., Tarasova I.A., Gorshkov M.V., Chumakov P.M., Moshkovskii S.A. Multiomic Profiling Identified EGF Receptor Signaling as a Potential Inhibitor of Type I Interferon Response in Models of Oncolytic Therapy by Vesicular Stomatitis Virus // *Int. J. Mol. Sci.* 2022. V. 23. №9. P. 5244-5264.
2. **Pyatnitskiy, M.A.**, Arzumanian, V.A., Radko, S.P., Ptitsyn, K.G., Vakhrushev, I.V., Poverennaya, E.V., Ponomarenko, E.A. Oxford Nanopore MinION Direct RNA-Seq for Systems Biology // *Biology*. 2021. V. 10. №11. P. 1131-1143.
3. Ахаладзе Д.Г., Рабаев Г.С., Жуков Н.В., **Пятницкий М.А.**, Горохов Н.С., Жуликов Я.А., Рабаев Г.Г., Касацкий А.А., Володин И.В., Мякова Н.В. Начальный опыт оценки мутационной нагрузки опухоли у детей с гепатоцеллюлярной карциномой // *Педиатрия. Журнал им. Г.Н. Сперанского*. 2021. Т. 100. №3. С. 193-199.
4. Poverennaya E., Kiseleva O., Romanova A., **Pyatnitskiy M.** Predicting Functions of Uncharacterized Human Proteins: From Canonical to Proteoforms // *Genes*. 2020. V. 11. №6. P. 1-17.
5. Tarkhov A.E., **Pyatnitskiy M.**, Menshikov L.I., Fedichev P.O., Alla R., Ayyadevara S., Shmookler Reis R.J. A Universal Transcriptomic Signature of Age Reveals the Temporal Scaling of *Caenorhabditis elegans* Aging Trajectories // *Scientific Reports*. 2019. V. 9. №1.
6. **Пятницкий М.А.**, Карпов Д.С., Мошковский С.А. Методы поиска критически важных генов в раковом геноме // *Биомедицинская химия*. 2018. Т. 64. №4. С. 303-314.
7. Lobas A.A., **Pyatnitskiy M.A.**, Chernobrovkin A.L., Ilina I.Y., Karpov D.S., Solovyeva E.M., Kuznetsova K.G., Ivanov M.V., Lyssuk E.Y., Kliuchnikova A.A., Voronko O.E., Larin S.S., Zubarev R.A., Gorshkov M.V., Moshkovskii S.A. Proteogenomics of malignant melanoma cell lines: the effect of stringency of exome data filtering on variant peptide identification in shotgun proteomics. // *J Proteome Res*. 2018. V.17. №5. P. 1801-1811
8. Вахрушев И.В., Новикова С.Е., Цветкова А.В., **Пятницкий М.А.**, Ярыгин К.Н., Сравнительное протеомное профилирование ядерной и цитозольной фракций клеточных линий различного происхождения // *Биомедицинская химия*. 2018. Т. 64. №3. С. 233-240.
9. Vakhrushev I.V., Novikova S.E., Tsvetkova A.V., Karalkin P.A., **Pyatnitskii M.A.**, Zgoda V.G., Yarygin K.N. Proteomic Profiling of HL-60 Cells during ATRA-Induced

Differentiation // Bulletin of Experimental Biology and Medicine. 2018. V. 165. №4. P. 530-543

10. Khodova K., **Pyatnitskiy M.**, Ereemeeva T., Fedichev P. Application of bioinformatics approach to identify indications for PFKFB3-mediated glycolysis inhibition as potential anticancer treatment // J Clin Oncol. 2017.V.35 № 15. P. 23158.

11. Petushkova N.A., Zgoda V.G., **Pyatnitskiy M.A.**, Larina O.V., Teryaeva N.B., Potapov A.A., Lisitsa A.V. Post-translational modifications of FDA-approved plasma biomarkers in glioblastoma samples // PLoS One. 2017. V.12. №5.

12. Ponomarenko E.A., Poverennaya E.V., Ilgisonis E.V., **Pyatnitskiy M.A.**, Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I. The Size of the Human Proteome: The Width and Depth // J Anal Chem. 2016. V. 2016. P. 7436849

13. Kotelnikova E., **Pyatnitskiy M.**, Paleeva A., Kremenetskaya O., Vinogradov D. Practical aspects of NGS-based pathways analysis for personalized cancer science and medicine // Oncotarget. 2016 V. 7. №32. P. 52493-52516.

14. Kotelnikova E. A., Logacheva M. D., Nabieva E. R., **Pyatnitskiy M. A.**, Vinogradov D. V., Makarova A. S., Demin A. V., Paleeva A. G., Kremenetskaya O. S., Penin A. A., Klepikova A. V., Kasianov A. S., Shavochkina D. A., Kudashkin N. E., Patyutko Yu. I., Mugue N. S., Kondrashov A. S., Lazarevich N. L. NGS-based identification of druggable alterations and signaling pathways – hepatocellular carcinoma case report // Biopolym. Cell. 2015. V. 31. №6. P.436-446,

15. **Pyatnitskiy M.**, Karpov D., Poverennaya E., Lisitsa A., Moshkovskii S. Bringing Down Cancer Aircraft: Searching for Essential Hypomutated Proteins in Skin Melanoma. // PLoS One. 2015. V. 10. №11. P. e0142819.

16. Karpova M.A., Karpov D.S., Ivanov M.V., **Pyatnitskiy M.A.**, Chernobrovkin A.L., Lobas A.A., Lisitsa A.V., Archakov A.I., Gorshkov M.V., Moshkovskii S.A. Exome-Driven Characterization of the Cancer Cell Lines at the Proteome Level: The NCI-60 Case Study // J Proteome Res. 2014. V. 13. №12. P. 5551-5560.

17. Petushkova N.A., **Pyatnitskiy M.A.**, Rudenko V.A., Larina O.V., Trifonova O.P., Kisrieva J.S., Samenkova N.F., Kuznetsova G.P., Karuzina I.I., Lisitsa A.V. Applying of hierarchical clustering to analysis of protein patterns in the human cancer-associated liver // PLoS One. 2014. V. 9. №8. P.e103950.

18. **Pyatnitskiy M.**, Mazo I., Shkrob M., Schwartz E., Kotelnikova E. Clustering gene expression regulators: new approach to disease subtyping // PLoS One. 2014. V. 9. №1. P. e84955

19. Kotelnikova E., Shkrob M., **Pyatnitskiy M.**, Ferlini A., Daraselia N. Novel approach to meta-analysis of microarray datasets reveals muscle remodeling-related drug targets and biomarkers in Duchenne muscular dystrophy // PLoS Comp Biol. 2012. V. 8. №2. P. e1002365
20. **Pyatnitskii M.A.**, Lisitsa A.V., Moshkovskii S.A., Arnotskaya N.E., Akhmedov B.B., Zaridze D.G., Polotskii B.E., Shevchenko V.E. Differential diagnostics of squamous cell lung carcinoma by means of the mass spectrometry profiling of blood plasma // Journal of Analytical Chemistry. 2011. V. 66. №14. P. 1369-1375.
21. **Pyatnitskiy M.**, Karpova M., Moshkovskii S., Lisitsa A., Archakov A. Clustering Mass Spectral Peaks Increases Recognition Accuracy and Stability of SVM-based Feature Selection // J Proteomics Bioinform. 2010. V.3 №2. P.48-54.
22. Petushkova N.A., **Pyatnitskiy M.A.**, Lisitsa A.V., Larina O.V., Kuznetsova G.P., Skipenko O.G., Karuzina I.I., Archakov A.I. Computational approach to characterization of human liver drug-metabolizing enzymes // Eur J Pharm Sci. 2010, V. 41. №2. P. 305-311.
23. Сендерович А.И., Строганова А.М., **Пятницкий М.А.**, Вишневская Я.В., Карселадзе А.И. Молекулярно-генетическое исследование доброкачественных пигментных новообразований с помощью реакции флюоресцентной in situ гибридизации // Архив патологии. 2010. Т. 72. № 4. С. 27-30
24. Сендерович А.И., Строганова А.М., **Пятницкий М.А.**, Вишневская Я.В., Карселадзе А.И. Использование FISH-реакции для диагностики меланоцитарных новообразований // Вестник РОНЦ им. Н. Н. Блохина РАМН. 2010. Т. 21. №3. С. 54-60.

Материалы трудов конференций

25. Ponomarenko E., Kiseleva O., Ilgisonis E., **Pyatnitskii M.**, Poverennaya K. The Selection of Knockout Targets: HepG2 Multi-omics Profiling and MetaAnalysis // In: Proceedings HUPO. 2021. P. 238.
26. Ponomarenko E.A., Ekaterina I. Ilgisonis E.I., Poverennaya K.V., **Pyatnitskii M.A.**, Lisitsa A.V. Precision, personalized and digital medicine: research trends over the past 10 years // In: Proceedings HUPO. 2019. P. 229.
27. Krivosheeva I., Filatova A., **Pyatnitskiy M.**, Moshkovskii S., Skoblov M. Study of essential hypomutated genes in skin melanoma // Integrative Bioinformatics and Systems Biology (WIBSB-2018) First Sino-Russian Workshop. Abstracts the 11th International Conference. 2018. P. 46.

28. Лазаревич Н. Л., Набиева Е. Р., Виноградов Д. В., Мюге Н. С., Логачева М. Д., Пенин А. А., **Пятницкий М. А.**, Шавочкина Д. А., Кустова И. Ф., Макарова А. С., Кривцова О. М., Кудашкин Н. Е., Патютко Ю. И., Котельникова Е. А. Выявление потенциально драйверных соматических мутаций в опухолевых тканях на основании данных полнотранскриптомного секвенирования // Материалы II Всероссийской конференции по молекулярной онкологии. Успехи молекулярной онкологии. 2016. Т. 3. № 4. С. 37-38.
29. **Pyatnitsky M.**, Karpov D., Poverennaya E., Lisitsa A., Moshkovskii S. Searching For Essential Cancer Proteins: Analysis Of Hypomutated Genes In Skin Melanoma // In: Proceedings of Moscow Conference on Computational Molecular Biology. 2015. P. 36.
30. Лазаревич Н.Л., Логачева М.Д., Набиева Е.Р., **Пятницкий М.А.**, Виноградов Д.В., Макарова А.С., Кременецкая О.С., Клепикова А.В., Шавочкина Д.А., Чесноков М.С., Полуконова А.В., Кудашкин Н.Е., Мороз Е.А., Герасимов Е.С., Патютко Ю.И., Пенин А.А., Базыкин Г.А., Кондрашов А.С., Мюге Н.С., Котельникова Е.А. Комплексный анализ молекулярных нарушений при патогенезе гепатоцеллюлярной карциномы методами экзомного и транскриптомного секвенирования // Материалы VIII Всероссийской научно-практической конференции «Молекулярные механизмы роста и прогрессии злокачественных новообразований». Молекулярная биология. 2015. Т. 49. № 5. С. 857-858.
31. Vlasova M.A., Karpov D.A., Ivanov M.V., **Pyatnitskiy M.A.**, Lazarev V.N., Zgoda V.G., Lobas A.A., Archakov A.I., Gorshkov M.V., Moshkovskii S. A. Shotgun Proteomics Recognizes Cell Lines Using Exome Data: NCI-60 Panel Case Study // In: Proceedings HUPO. 2014. P.35
32. Котельникова Е.А, Мюге Н.С., **Пятницкий М.А.**, Набиева Е.Р., Виноградов Д.В., Кременецкая О.С., Макарова А.С., Клепикова А.В., Шавочкина Д.А., Юренко К.А., Чучуев Е.С., Мороз Е.А., Патютко Ю.И., Пенин А.А., Логачева М.Д., Кондрашов А.С., Лазаревич Н.Л. Комплексный анализ экзома и транскриптома в персонифицированном исследовании пациентов с гепатоцеллюлярной карциномой // Материалы VIII Всероссийской научно-практической конференции «Молекулярная диагностика 2014». Москва. 2014.
33. Nabieva E., Logacheva M., Kotelnikova E., **Pyatnitskiy M.**, Vinogradov D., Mugue N., Kondrashov A., Lazarevich N. Whole-exome sequencing in hepatocellular carcinoma // In: Proceedings of Moscow Conference on Computational Molecular Biology. 2013. Moscow, Russia. P. 235.

34. Moshkovskii S., Chernobrovkin A., **Pyatnitsky M.**, Zdoda V., Novikova S., Kuznetsova K., Vakhrushev I., Lisitsa A. Cancer-specific genome changes observed at proteome level in colorectal cancer // In: Proceedings 38th FEBS Congress. 2013. Saint Petersburg, Russia. P.631
35. Ponomarenko E., **Pyatnitskiy M.**, Poverennaya E., Lisitsa A. Comparative analysis of human chromosomes based on post-genomic data // In: Proceedings 6th Congress AOHUPO. 2012. Beijing, China. P. 322
36. Moshkovskii S., Sokolova E., Brattseva E., Karpova M., **Pyatnitsky M.**, Archakov A., Proteome and Cytokine Serum Profiling to Diagnose a Skin T Cell Lymphoma // In: Proceedings of HUPO 11-th Annual World Congress. Boston, USA. 2012. P. 149
37. **Pyatnitskiy M.**, Mazo I., Daraselia N., Shkrob M., Kotelnikova E. Cluster Analysis Of Significant Regulators As New Approach To Patients Subtyping // In: Proceedings of The 8-th International Conference On Bioinformatics Of Genome Regulation And Structure. Novosibirsk, Russia. 2012. P. 199
38. **Pyatnitsky M.**, Kotelnikova E., Shkrob M., Ferlini A., Daraselia N., Mazo I., Schwartz E. Novel approach to meta-analysis of microarray datasets for identification of new biomarkers and potential drug targets // In: Proceedings of The 8-th International Conference On Bioinformatics Of Genome Regulation And Structure. Novosibirsk, Russia, 2012. P. 58
39. Kotelnikova E, Shkrob M, **Pyatnitskiy M.**, Daraselia N. Identification of drug-resistant patient subtypes on the base of found regulatory clusters. // In: Proceedings of 4th RECOMB Conference on Regulatory Genomics, Systems Biology and DREAM Challenges. Barcelona, Spain. 2011. P. 302
40. Kotelnikova E., Shkrob M., **Pyatnitskiy M.**, Daraselia N. Meta-analysis for discovery of disease biomarkers with implicated mechanistic models // In: Proceedings of Moscow Conference on Computational Molecular Biology, Moscow, Russia. 2011. P. 176
41. Moshkovskii S., Karpova M., **Pyatnitsky M.**, Sycheva A., Simonova O., Sokolova E., Archakov A. Mass-Spectrometry-Based Plasma Proteome Barcode to Discriminate the Prostate Cancer Samples // In: Proceedings of HUPO 9-th Annual World Congress. Sydney, Australia. 2010. P. 361

Главы в коллективных монографиях

42. Moshkovskii S., **Pyatnitsky M.**, Lokhov P., Baranova A. OMICS for Tumor Biomarker Research / Biomarkers in Disease: Methods, Discoveries and Applications. Biomarkers in Cancer, ed. Preedy R., Patel V.B. Netherlands: Springer. 2015. P.3-30.

43. **Pyatnitskiy M.**, Shkrob M., Daraselia N., Kotelnikova E. Sub-Network Enrichment and Cluster Analysis Reveal Possible Pathways for Cetuximab Sensitivity / From Knowledge Networks to Biological Models, ed. Yuriev A., Daraselia N. Bentham. 2012. P. 151-172.

44. Kotelnikova E.A., **Pyatnitskiy M.A.**, Redler R.L., Dokholyan N.V. Role of Ca²⁺-mediated signaling in ALS pathology / From Knowledge Networks to Biological Models, ed. Yuriev A., Daraselia N., Bentham. 2012. P. 24-72.

Патент

45. Пат. 2577107 Российская Федерация, МПК C12Q 1/68 Способ поиска белков-мишеней, запускающих процесс канцерогенеза, в образцах тканей индивидуального пациента, для последующей противоопухолевой фармакотерапии / Котельникова Е., **Пятницкий М.**, № 2013144622/10, Заявл. 04.10.2013, опубл. 2016.

Программа для ЭВМ

46. Свидетельство о государственной регистрации программы для ЭВМ №2020618223 «Платформа для обработки и интерпретации клинических данных и молекулярно-генетических исследований опухолей человека для подбора персонализированной терапии», Пятницкий М.А., 2020.