с диагнозом синдром активации фосфоинозитид-3-киназы были выявлены миссенс-мутации. Синдром Диджорджи следует характеризовать отдельно так как для него типична делеция длинного плеча 22 хромосомы (del 22q11.2).

В Республике Беларусь диагноз первичный иммунодефицит установлен 470 пациентам, из них генетически подтвержденный диагноз выставлен 191 пациенту, что составляет 40,6%. Этот показатель отражает высокий уровень генетической диагностики первичных иммунодефицитов в Республике Беларусь.

Наиболее обширной группой пациентов с генетически подтвержденным диагнозом ПИД является группа комбинированных иммунодефицитов, ассоциированных с синдромами 59,2 % (n=113).

Группа комбинированных иммунодефицитов, ассоциированных с синдромами весьма неоднородна и представлена большим количеством назологий. В ходе проведения исследования была отмечена разнородность в распределении пациентов с ПИД по территории Беларуси.

ЛИТЕРАТУРА

1. *McCusker, C.* Primary immunodeficiency / C. McCusker, J. Upton, R. Warrington // *Allergy Asthma Clin. Immunol.* – 2018. – Vol. 14, № 2. – P. 141–152.
2. *Delmonte, O. M.* RAG deficiency: two genes, many diseases / O. M. Delmonte, C. Schuetz, L. D. Notarangelo // *J. Clin. Immunol.* – 2018. – Vol. 38, № 6. – P. 646–655.
3. Руководство по интерпретации данных последовательности ДНК человека, полученных методами массового параллельного секвенирования (MPS) (редакция 2018, версия 2) / О. П. Рыжкова [и др.] // Медицинская генетика. – 2019. – Т. 18, № 2. – С. 3–23.
4. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity / C. Picard [et al.] // *J. Clin. Immunol.* – 2018. – Vol. 38, №1. – P. 96–128.

БИОИНФОРМАТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ МЕТАГЕНОМНОГО СЕКВЕНИРОВАНИЯ BIOINFORMATICS TOOLS FOR METAGENOME SEQUENCING

Е. Я. Скоповец¹, Р. М. Смолякова²
K. Skapovets¹, R. Smolyakova²

¹ГУ «Республиканский научно-практический центр детской онкологии, гематологии и иммунологии»,
г. Минск, Республика Беларусь

²Белорусский государственный университет, МГЭИ им. А. Д. Сахарова БГУ,
г. Минск, Республика Беларусь
Skopovets@yandex.ru

¹Belarussian Research Center for Pediatric Oncology, Hematology and Immunology,
Minsk, Republic of Belarus

²Belarusian State University, ISEI BSU,
Minsk, Republic of Belarus

Благодаря технологическим усовершенствованиям в методах секвенирования, практически все представители микробиоты могут быть быстро проанализированы, избегая этапов культивирования. В частности, процедуры, основанные на секвенировании 16S рРНК следующего поколения, которые позволяют проводить микробную идентификацию с высокой пропускной способностью в конкретном метагеноме, представляют собой мощное средство для изучения состава и биоразнообразия микробных сообществ. Огромное количество метагеномных данных следующего поколения, генерируемых такими процедурами, требует биоинформационных инструментов, способных их анализировать.

Due to technological improvements in sequencing methods, virtually all the microbiotas from a given environment can be analyzed in a single run, avoiding cultivation steps. In particular, procedures based on 16S rRNA next-generation sequencing, which allow to hold the high throughput microbial identification within a specific metagenome, represent a powerful means to investigate the composition and the biodiversity of microbial communities. The enormous amount of next-generation metagenomic data generated by such procedures requires bioinformatics tools capable of analyzing them.

Ключевые слова: секвенирование, микробиота, метагеномика, биоинформатика.

Keywords: sequencing, microbiota, metagenomics, bioinformatics.

<https://doi.org/10.46646/SAKH-2020-2-259-261>

Быстрые достижения в технологиях секвенирования ДНК и биоинформатической обработки за последние два десятилетия существенно улучшили понимание мира микробов. В настоящее время стало возможным узнать, как состав микробиоты влияет на здоровье человека, развитие заболеваний, их лечение, появляются новые исследования в области медицины, судебной экспертизы, окружающей среды и сельского хозяйства, развиваются новые направления биотехнологии. Большая часть исследований проводится с помощью исследований маркерных генов (например, бактериальных / архейных генов 16S рДНК, эукариотических генов 18S рДНК), которые описывают микробиоту с различной степенью таксономической специфичности и филогенетической информацией.

Целевая метагеномика, также называемая метагенетикой или основанная на ампликонах метагеномика, фокусируется только на таксономически информативном геномном маркере. Этот локус усиливается путем проведения ПЦР перед секвенированием, что значительно сокращает объем данных, которые нужно секвенировать и анализировать. Для исследований прокариотических микроорганизмов, мишенью выбора является часть гена 16S рДНК, состоящая из консервативных и гипервариабельных областей, специфичных для разных прокариотических таксонов. Действительно, появление секвенирования «нового поколения» позволяет проводить исследования в меньшем масштабе и по более низкой цене. Одним из основных различий технологиями секвенирования является тип и количество ошибок секвенирования, влияющих на качество последующих данных.

После того, как исходные считывания секвенирования были получены из амплифицированной целевой области гена 16S рДНК, следующим шагом является их анализ для оценки микробного разнообразия и состава таксонов. Для этих целей доступно все большее число алгоритмов биоинформатического анализа. Они специально предназначены для объединения нескольких этапов, включают предварительную обработку чтений, обнаружение химерных последовательностей, кластеризацию «прочтений» в операционные таксономические единицы (OTU) и таксономическую идентификацию. Эти конвейеры объединяют множество алгоритмов, чтобы предложить самый широкий диапазон возможностей. Как следствие, они требуют передовых навыков в области биоинформатики и вычислительных ресурсов и могут отпугивать пользователей, не знакомых с разнообразием существующих аналитических процессов. Каждый конвейер предлагает свои собственные рекомендации, с профилями конфигурации, справочными базами данных и рекомендованными аналитическими этапами. Выбор конвейера с набором параметров и алгоритмов для данного приложения может быстро стать сложной задачей без протокола оценки.

В последнее время было предпринято множество методик для оптимизации процессов обработки метагеномных данных. Однако эти сравнительные исследования в основном сосредоточены на одном аналитическом этапе, в том числе влияние предварительной обработки последовательности, кластеризации в OTU таблицы по 97% сходству и определение таксономического назначения. Чаще всего используются наборы данных с платформ секвенирования Illumina и 454 [1].

Обширный литературный обзор позволил определить биоинформатические алгоритмы, разработанные для целевого анализа метагеномики и WGS, с целью оценки таксономического состава и разнообразия образцов. Существует две различные методологии, первая – начинается с кластеризации, а вторая – с присваивания таксономической единицы последовательности. Подходы, основанные на кластеризации, также называемые подходами, основанными на выравнивании, начинаются с этапа кластеризации OTU, где считывания собираются в OTU-таблицы на основе сходства последовательностей. Из каждого кластера репрезентативная последовательность извлекается в соответствии с различными методами (консенсус, самая длинная или более распространенная последовательность и т. д.). Затем эту последовательность выравнивают с каждой из последовательностей 16S рДНК эталонной базы данных, используя инструмент поиска гомологии. Репрезентативная последовательность и OTU, к которому она принадлежит, в конечном итоге присваиваются таксономической группе путем проверки наилучшего соответствия.

Подходы, основанные на назначении, появились совсем недавно для метагеномного анализа. Эти методы не группируют чтения по их внутреннему сходству, но сначала сравнивают их все с базой данных (например, с помощью распознавания k-мер, например Kraken). На основании этих сравнений они назначают самую низкую возможную таксономию для каждого чтения или более низкого общего предка (LCA) для группы последовательностей одной и той же таксономии в справочной базе данных. После этого чтения сгруппированы в различные таксономические единицы на основе их аннотаций.

Эти два аналитических подхода различны и могут привести к разным результатам. Существует много других мощных программных инструментов с открытым исходным кодом для метагеномных исследований, в том числе Mothur, Phyloseq и связанные с ними инструменты, доступные через Bioconductor. Основное различие между Mothur и QIIME заключается в интерактивных визуализациях. Пакет для анализа микробиомов QIIME предоставляет подавляющее большинство полезных важных биоинформатических инструментов, необходимых для полного и тщательного анализа микробиомов. Недавно разработан пакет под названием QWRAP, который действует как как программная оболочка, помогая пользователю использовать общедоступные инструменты, такие как FASTQC и FASTX, для проверки качества и фильтрации качества, а затем инструменты QIIME для анализа микробиомов [1, 2, 3]. Скрипты QWRAP написаны с использованием языков bash, perl, python и R. QWRAP работает с демуплектированными файлами Illumina fastq и может использоваться как с односторонним чтением, так и с перекрывающимся парным чтением. После анализа QWRAP генерирует статический HTML-отчет, который можно просмотреть в любом веб-браузере (например, Internet Explorer, Firefox, Google Chrome).

Информация о качестве, представленная в файлах FASTQ, используется для фильтрации качества. Качественная фильтрация значительно улучшает оценки разнообразия для образцов, обработанных с помощью

секвенирования Illumina. Исходя из качества операций чтения, можно решить отфильтровать необработанные данные для удаления операций чтения низкого качества и / или баз низкого качества. Можно обрезать отдельные (или все) показания с 3'-конца, где качество падает ниже определенного показателя Q (например, показатель Q 30), или полностью удалить считывание, если его общее качество падает ниже определенного порога. Другим этапом контроля качества является удаление химерных последовательностей. Химеры - это последовательности, которые создаются в процессе амплификации ПЦР из двух или более шаблонов вместо одного родительского шаблона. Химерные последовательности можно удалять с помощью таких программ, как BLAST, ChimeraSlayer или UCHIME.

Последовательности, прошедшие этап фильтрации контроля качества, сначала группируются с заданным уровнем сходства последовательностей, обычно 97% для поддержки их таксономической классификации. QIIME предоставляет несколько алгоритмов и методов для выполнения кластеризации OTU. Получающиеся кластеры почти идентичных последовательностей называются оперативными таксономическими единицами (OTU). Создается таблица OTU, которая включает в себя счетчик числа считываний последовательности, представляющих каждый OTU, а также таксономическую классификацию этого OTU. Эта таблица OTU используется для последующих анализов. Относительная численность каждого таксона, присутствующего в наборе образцов, может быть определена из таблицы OTU и использована для генерации цифр, представляющих численность таксонов. Эти сводные диаграммы могут быть созданы для каждого конкретного таксономического уровня в иерархии (например, типа, семейства, рода и / или вида) в зависимости от глубины таксономического назначения OTU. Эта информация об изобилии может быть представлена в различных форматах, включая гистограммы и тепловые карты.

Филогенетическое древо всех идентифицированных последовательностей OTU может использоваться для помощи в визуализации и кластеризации микроорганизмов из образцов. Доступно несколько программ множественного выравнивания последовательностей, таких как PyNAST, Infernal, MUSCLE. Филогенетическое древо может быть сгенерировано из выровненных последовательностей с использованием таких программ, как FastTree, clustalw, clearcut, raxml или MUSCLE.

Различные статистические анализы могут быть выполнены для оценки различий в распределении OTU и численности между выборками и группами. Альфа-разнообразие - это мера разнообразия в образце или окружающей среде. Различные методы, доступные для расчетов альфа-разнообразия, могут учитывать такие факторы, как богатство (количество OTU / видов, присутствующих в выборке) и равномерность (относительное изобилие различных OTU / видов и их равномерное распределение). Наблюдаемые виды: измеряет уникальные OTU в образце. Основные показатели альфа-разнообразия:

- Chao1: оценка видового богатства.
- Индекс Шеннона: измеряет богатство и равномерность.
- Индекс Симпсона: измеряет как богатство, так и равномерность, но меньше влияет присутствие редких видов по сравнению с индексом Шеннона.

- PD (Phylogenetic Distance): включает филогенетическое расстояние в расчет разнообразия.

Бета-разнообразие обеспечивает измерение расстояния или различий между каждой парой выборок. Когда используется более двух выборок, бета-разнесение рассчитывается для каждой пары выборок для генерации матрицы расстояния / различия. Вот некоторые из наиболее часто используемых метрик бета-разнообразия:

- Брей-Кертис: метод, не основанный на филогении, который учитывает изобилие;
- Невзвешенный UniFrac: использует наличие и отсутствие OTU между образцами вместе с их филогенетическими расстояниями для определения бета-разнообразия;
- Взвешенный UniFrac: использует информацию об изобилии для каждого OTU вместе с их филогенетическими расстояниями для определения бета-разнообразия.

Поскольку веб-страницы представлены в виде статических HTML-файлов, вся папка результатов может быть доступна другим пользователям, чтобы они могли просматривать данные на своих локальных рабочих станциях.

Все инструменты, описанные в предыдущем абзаце, могут взаимодействовать посредством плагинов, обмена файлами в стандартных форматах или использования многоязычных сред, таких как Jupyter Notebooks. Например, формат BIOM поддерживается всеми из них. Таким образом, разработка различных новых программных обеспечений для обработки данных, полученных в результате метагеномного секвенирования, значительно облегчила и ускорила получение результатов о таксономии большого количества представителей микрофлоры [2].

ЛИТЕРАТУРА

1. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy / Q Wang [et. al] // *Applied and Environmental Microbiology*. – 2007. – № 73(16). – P.5261–5267.
2. QIIME allows analysis of high-throughput community sequencing data / J. Caporaso [et. al.] // *Nature Methods*. – 2010. – № 7(5). – P.335–336.
3. Ribeca, P. Computational challenges of sequence classification in microbiomic data / P. Ribeca, G. Valiente // *Briefings in Bioinformatics*. – 2011. – № 12(6). – P.614–625.

ВЫБОР ФАНТОМА ДЛЯ ГАРАНТИИ КАЧЕСТВА СИСТЕМ ДОЗИМЕТРИЧЕСКОГО ПЛАНИРОВАНИЯ ЛУЧЕВОЙ ТЕРАПИИ

CHOICE OF PHANTOM FOR QUALITY ASSURANCE OF COMPUTERIZED PLANNING SYSTEMS FOR RADIATION TREATMENT OF CANCER

Н. В. Срибная, В. Ф. Малишевский
N. Sribnaya, V. Malishevskiy

*Белорусский государственный университет, МГЭИ им. А.Д. Сахарова БГУ,
г. Минск, Республика Беларусь
nikol.sribnaya94@gmail.com
Belarusian State University, ISEU BSU, Minsk, Republic of Belarus*

Лучевая терапия относится к области высоких технологий, точность которой так же обеспечивает гарантию качества системы планирования. При проведении испытаний огромную роль играет выбор фантома. Выделяют фантомы для измерений на КТ, антропоморфные и пластинчатые. Для калибровки КТ у фантома должны быть вставки из материалов с сертифицированной плотностью. Антропоморфную схему можно создать при наличии сменных пластин, выполненных из материала, эквивалентного органам человека. Фантомы различаются по многим характеристикам, таким как форма, материал и т.д. При выборе фантома для проведения гарантии качества в полном объёме, он должен обладать возможностями калибровки КТ, создания антропоморфной геометрии, проведения измерений в неоднородной среде, моделирование кривизны поверхности. Фантом с характеристиками CIRS модель 002LFC оптимально подходит для проведения испытаний.

Radiation therapy belongs to the field of high technology, the accuracy of which also guarantees the quality of the planning system. In carrying out tests of the quality assurance of the planning system, the choice of the phantom plays a huge role. There are phantoms for CT-calibration, with anthropomorphic geometry and laminar. For calibration, the CT phantom has to inserts consisting materials with certified density. Phantom has the ability for modeling anthropomorphic geometry if it has replaceable plates made of material equivalent to human organs. Phantoms differ by material, shape and etc. For quality assurance of planning, system phantom should have the ability for calibration CT, modeling anthropomorphic geometry, take measurements in a non-uniform environment, modeling of curvature of a surface. The characteristics of phantom CIRS 002LFC is optimally suited for testing.

Ключевые слова: дозиметрия, фантом, гарантия качества, лучевая терапия, ввод в эксплуатацию, система дозиметрического планирования.

Key words: dosimetry, phantom, quality assurance, radiation treatment of cancer, commissioning, computerized planning system.

<https://doi.org/10.46646/SAKH-2020-2-262-265>

Лучевая терапия является одной из наиболее эффективных способов лечения опухолей головы, шеи, шейки матки, мочевого пузыря, простаты, молочной железы и кожи. Она относится к области высоких технологий, потенциал которой реализуется только через детальное планирование облучения и тщательного выполнения всех регламентов в процессе всего лечения [1].

Гарантия качества компьютерной системы дозиметрического планирования является обязательной процедурой при её эксплуатации. Она обеспечивает точность подведения лечебной дозы к опухоли и минимизирует шанс возникновения случайной ошибки на всех этапах подготовки, планирования и проведения лечения [2].

Ввод в клиническую эксплуатацию является одной из наиболее важных частей всей программы гарантии качества для систем дозиметрического планирования и включает в себя: тестирование функций системы, проверку способности алгоритмов расчета доз воспроизводить расчеты для измеренных доз и документальное оформление различных возможностей системы.

При вводе в эксплуатацию существуют анатомические и дозиметрические тесты. Анатомические тесты связаны с созданием анатомической модели и планирования облучения. Дозиметрические тесты проводятся для охвата широкого диапазона методик облучения, применяемых в клинической практике. Целью всех этих испытаний является подтверждение точности и корректности лечения пациента. Во время проведения испытаний проверяются показатели системы дозиметрического планирования для типовых традиционных и конформных методик лучевой терапии, включая сравнение результатов расчета и измеренных значений для фантома и проверки расчета значений ME/время [3].

Огромную роль играет выбор фантома для проведения испытаний при вводе в клиническую эксплуатацию: проверки анатомии и входных данных, проведения дозиметрических тестов. К важнейшим проверочным процедурам относятся проверка передачи КТ-изображений, передача анатомических данных пациентов.