ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р ИСО 20397-2— 2023

Биотехнология

МАССОВОЕ ПАРАЛЛЕЛЬНОЕ СЕКВЕНИРОВАНИЕ

Часть 2

Оценка качества данных секвенирования

(ISO 20397-2:2021, IDT)

Издание официальное

Москва
Российский институт стандартизации
2023

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Федеральным государственным бюджетным учреждением «Российский институт стандартизации» (ФГБУ «Институт стандартизации») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 326 «Биотехнологии»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 10 августа 2023 г. № 633-ст

4 Настоящий стандарт идентичен международному стандарту ИСО 20397-2:2021 «Биотехнология. Массовое параллельное секвенирование. Часть 2. Оценка качества данных секвенирования» (ISO 20397-2:2021 «Biotechnology — Massively parallel sequencing — Part 2: Quality evaluation of sequencing data», IDT).

Международный стандарт разработан Техническим комитетом ТК 276 «Биотехнология» Международной организации по стандартизации (ИСО)

5 ВВЕДЕН ВПЕРВЫЕ

6 Некоторые элементы настоящего стандарта могут являться объектами патентных прав

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© ISO, 2021
© Оформление. ФГБУ «Институт стандартизации», 2023

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 1
4 Исходные данные 5
5 Выравнивание последовательностей и картирование 6
6 Идентификация вариантов 9
7 Валидация 10
8 Документирование 11
Приложение А (справочное) Показатели качества для конкретных примеров платформ MPS 12
Приложение В (справочное) Охват и рекомендации по прочтению по приложениям 13
Приложение С (справочное) Программное обеспечение для выравнивания и сопоставления последовательностей 15
Библиография 16

Введение

Массовое параллельное секвенирование (MPS) — это высокопроизводительный аналитический подход к секвенированию нуклеиновых кислот с использованием массово-параллельной архитектуры, которая позволяет исследовать целые геномы, транскриптомы и специфические мишени из нуклеиновых кислот различных организмов за относительно короткое время.

MPS используется во многих областях биологии, позволяя определять и осуществлять высокопроизводительный анализ миллионов и тысяч миллионов нуклеотидных оснований. Биологическая изменчивость полимеров дезоксирибонуклеиновых и рибонуклеиновых кислот живых организмов приводит к трудностям в точном определении их последовательностей. Качество определения последовательности с помощью MPS зависит от многих факторов, в том числе от качества образца, подготовки библиотеки, выбора платформы и качества данных секвенирования.

Анализ данных секвенирования создает значительные сложности в биоинформатике, связанные с хранением данных, временем вычислений и точностью определения вариантов. Одной из основных проблем, возникающих при работе с данными секвенирования, которую в ряде случаев упускают из виду, является мониторинг показателей контроля качества на всех этапах обработки конвейера. Информация о качестве данных необходима для последующего анализа последовательностей. Контроль качества при обработке и анализе данных секвенирования нуклеиновых кислот можно разделить на три этапа: исходные данные, выравнивание и поиск вариантов. В настоящем стандарте приведен список критериев для оценки качества данных секвенирования MPS, а также конкретные рекомендации для различных платформ MPS.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Биотехнология

МАССОВОЕ ПАРАЛЛЕЛЬНОЕ СЕКВЕНИРОВАНИЕ

Часть 2

Оценка качества данных секвенирования

Biotechnology. Massively parallel sequencing. Part 2. Quality evaluation of sequencing data

Дата введения — 2024—03—01

1 Область применения

Настоящий стандарт устанавливает общие требования и рекомендации по оценке и контролю качества данных массового параллельного секвенирования (MPS). Он охватывает процедуры генерации исходных данных, выравнивания последовательностей и поиска вариантов.

Настоящий стандарт также содержит общие рекомендации по валидации и документированию данных MPS.

Настоящий стандарт не относится к процессам, связанным со сборкой *de novo*.

2 Нормативные ссылки

В настоящем стандарте нормативные ссылки отсутствуют.

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями.

ИСО и МЭК ведут терминологические базы данных для использования в области стандартизации по следующим адресам:

- Электропедия МЭК: доступна по адресу: gost.e-ecolog.ru ;
- платформа онлайн-просмотра ИСО: доступна по адресу: gost.e-ecolog.ru .

3.1 **последовательность адаптера; адаптер** (adapter sequence, adapter): Синтетический олигонуклеотид известной последовательности, который может быть добавлен к 3' или 5' концам фрагмента нуклеиновой кислоты.

Примечание — Он обеспечивает участок праймера, а также другие необходимые последовательности для секвенирования вставки.

3.2 **алгоритм** (algorithm): Полностью определенная конечная последовательность инструкций, с помощью которой значения выходных переменных могут быть вычислены из значений входных переменных.

[МЭК 60050-351:2013, 351-42-27, изменено — примечания удалены]

3.3 **распознание нуклеотидов** (base calling): Процесс вычисления в массивно-параллельном секвенировании перевода необработанных электрических сигналов в последовательность нуклеотидов.

Примечание — Применение распознавания нуклеотидов и алгоритмов характеризуется точностью считывания и консенсуса.

Издание официальное

3.4 **биоинформатический конвейер** (bioinformatics pipeline): Отдельные программы, сценарии или части программного обеспечения, связанные между собой, в которых необработанные данные или выход одной программы используют в качестве входных данных для следующего этапа обработки данных.

Пример — Результаты работы программы тримминга базового качества могут быть использованы в качестве входных данных для ассемблера de novo.

3.5 **эффективность захвата** (capture efficiency): Процентное содержание всех секвенированных или картированных ридов, которые перекрывают целевые области.

3.6 **покрытие; глубина покрытия** (coverage, coverage depth): Количество раз, когда определенное положение основания считывается в ходе секвенирования.

Примечание — Количество прочтений, покрывающих определенную позицию.

3.7 **ширина покрытия** (coverage breadth): Доля генома в собранном/целевом размере генома в циклах секвенирования.

3.8 **плотность кластеров** (cluster density): Количество кластеров для каждой области сканирования.

Примечание 1 — Плотность кластеров, применяемая к платформам *MPS* (3.30), требует амплификации.

Примечание 2 — Плотность отдельных кластеров последовательностей, каждый из которых возникает из одной молекулы на некоторых платформах секвенирования.

Примечание 3 — Плотность кластеров, как правило, выражается в тысячах на мм^2.

3.9 **секвенирование кольцевых консенсусных последовательностей;** CCS (circular consensus sequencing, CCS): Режим секвенирования, при котором размер вставки секвенируется несколько раз в амплификации по типу «катящегося кольца», что обеспечивает высокую точность.

Примечание — В этом режиме допускается использовать несколько проходов от одной и той же молекулы для достижения более высокой точности определения одной молекулы.

3.10 **диапазон покрытия** (coverage range): Диапазон глубины покрытия по всему геному при проведении секвенирования.

3.11 **вариация числа копий; вариант числа копий;** CNV (copy number variation, copy number variant, CNV): Вариант числа копий одного или нескольких участков ДНК или присутствующих в геноме организма.

Примечание — CNVs — это вставки, делеции, инверсии и дупликации, содержащие не менее 1000 оснований в длину.

3.12 **дезоксирибонуклеиновая кислота;** ДНК (deoxyribonucleic acid, DNA): Полимер дезоксирибонуклеотидов, существующий в двухцепочечной (дцДНК) или одноцепочечной (оцДНК) форме.

[ИСО 22174:2005, 3.1.2]

3.13 **делеция** (deletion): Потеря одной (или более) пары нуклеотидных оснований из последовательности нуклеиновой кислоты по сравнению с ее эталонной последовательностью.

3.14 **уровень дупликации** (duplication level): Количество идентичных повторов для каждой последовательности в библиотеке.

Примечание — Уровень дупликации, как правило, отображается в виде графика, показывающего относительное количество последовательностей с разной степенью дупликации.

3.15 **гуанин-цитозиновый состав;** GC-состав (GC content): Процентное содержание гуанина и цитозина в одной или более последовательности(ях) нуклеиновой кислоты.

Примечание — Количество гуанина и цитозина в полинуклеиновой кислоте, как правило, выражается в мольной доле (или процентах) от общего количества азотистых оснований. Общее количество азотистых оснований включает в себя общее количество нуклеотидных оснований ридов в результате одного или нескольких циклов MPS.

3.16 **ген** (gene): Последовательность нуклеотидов в ДНК или РНК, кодирующая РНК либо белковый продукт.

Примечание 1 — Гены признаны основной единицей наследственности.

Примечание 2 — Ген может состоять из несмежных сегментов нуклеиновых кислот, которые перестраиваются в результате этапа ядерного процессинга.

Примечание 3 — Ген может включать или быть частью оперона, который содержит элементы для экспрессии гена.

3.17 **вставка-делеция (индел)** (indel): Вставка(3.18) и/или *делеция* (3.13) нуклеотидов в геноме ДНК.

Примечание — Инделы имеют длину менее 1000 оснований.

3.18 **вставка** (insertion): Добавление одной или нескольких пар нуклеотидных оснований в последовательность ДНК.

[ISO/TS 20428:2017, 3.19, изменено — «ДНК» заменено «нуклеиновой кислотой»]

3.19 **секвенирование** (sequencing): Определение порядка и содержания нуклеотидных оснований (аденина, гуанина, цитозина, тимина и урацила) в молекуле нуклеиновой кислоты.

Примечание — Последовательность, которая, как правило, описывается с 5-го конца по 3-й.

[ISO/TS 17822-1:2020, 3.19, изменено — «ДНК» исключена из термина; в определении «ДНК» заменено «нуклеиновой кислотой» и добавлен «урацил»]

3.20 **выравнивание последовательностей** (sequence alignment): Расположение последовательностей нуклеиновых кислот в соответствии с областями сходства.

Примечание — Выравнивание последовательностей может не требовать эталонного генома/эталонной целевой области нуклеиновой кислоты, и создание сборки может не быть его целью.

3.21 **исходные данные** (raw data): Первичные данные секвенирования, полученные секвенатором без предварительной фильтрации на основе программного обеспечения для целей анализа.

3.22 **рибонуклеиновая кислота;** РНК (ribonucleic acid; RNA): Полимер рибонуклеотидов, возникающий в двухцепочной или одноцепочной форме.

Примечание — Синтез белков в клетках регулируется генетической информацией, содержащейся в последовательности нуклеотидов в классе РНК, известном как РНК-посредник (мРНК).

3.23 **рибонуклеотид** (ribonucleotide): Нуклеотид, содержащий рибозу в качестве пентозного компонента, образующий основные строительные блоки для РНК.

Примечание — Рибонуклеотиды состоят из аденилата (AMP), гуанилата (GMP), цитидилата (CMP) или уридилата (UMP).

3.24 **прочтение (рид); прочтенная последовательность** (read, sequence read): Нуклеотидная последовательность, генерируемая устройством секвенирования.

Примечание — Прочтение (рид) — это выведенная последовательность пар оснований нуклеиновых кислот (или вероятностей пар оснований), соответствующая всему (или части) одному фрагменту нуклеиновой кислоты. Рид допускается использовать для обозначения последовательностей, полученных в результате экспериментов MPS.

3.25 **тип прочтения** (read type): Категория последовательности, которая зависит от того, как разработан и проведен эксперимент по прочтению последовательности.

Пример — Тип считывания (рида) может быть одноконцевым, парно-концевым, спаренным концевым, непрерывным длинным считыванием, циркулярным консенсусом.

3.26 **референсная последовательность** (reference sequence): Последовательность нуклеиновой кислоты, используемая либо для выравнивания путем картирования ридов последовательности, либо в качестве основы для аннотаций, таких как гены и вариации последовательности.

3.27 **демультиплексирование** (demultiplexing): Вычислительный процесс, обратный процессу мультиплексирования, смешивание двух или более образцов вместе таким образом, чтобы их можно было секвенировать за один цикл на приборе MPS.

Примечание 1 — Образцы, подлежащие объединению, должны иметь штрихкоды/индексы до смешивания.

Примечание 2 — Демультиплексирование — это вычислительный алгоритм, который разделяет пул ридов в соответствии с их исходным образцом на основе штрихкода.

3.28 **картирование** (mapping): Сборка последовательностей нуклеиновых кислот против существующей базовой (эталонной) последовательности, чтобы построить консенсусную последовательность.

3.29 **спаренные концевые риды** (mate pairs, mate pair reads): Парные риды, которые соответствуют концам длинного фрагмента последовательности нуклеиновой кислоты, полученного путем сжатия образца на большие фрагменты (более 2 кб или не менее 2 кб).

3.30 **массовое параллельное секвенирование;** MPS (massively parallel sequencing; MPS): Метод секвенирования, основанный на определении полимеризации инкремента на основе шаблона множества независимых молекул ДНК одновременно.

Примечание — Технология массивно-параллельного секвенирования может обеспечить миллионы или миллиарды коротких прочтений за один цикл.

3.31 **парно-концевые риды** (paired-end reads): Секвенирование ридов обоих концов фрагмента ДНК.

Примечание — При парно-концевом секвенировании прибор секвенирует оба конца коротких вставок, как правило, от 200 до 800 бит/с.

3.32 **оценка качества секвенирования; показатель качества *Q*** (quality score, *Q* score, Phred quality score): Мера качества секвенирования данного нуклеотидного основания.

Примечание 1 — *Q* вычисляют по формуле

$$Q = -10 \log10(p),$$

где p — оценочная вероятность того, что основание названо неверно.

Примечание 2 — Оценка качества, равная 20, представляет собой вероятность ошибки 1 к 100, что соответствует точности вызова 99 %.

Примечание 3 — Более высокие баллы качества указывают на меньшую вероятность ошибки. Более низкие показатели качества могут привести к тому, что значительная часть ридов окажется непригодной для использования. Низкие показатели качества могут также указывать на ложноположительные вызовы вариантов, что приводит к неточным выводам.

3.33 **цикл (запуск)** (run): Единый технологический цикл секвенсора от момента инициирования до получения исходных данных.

3.34 **аннотация секвенированной последовательности** (sequence annotation): Процесс добавления пояснения, комментария или ссылки на конкретные особенности в последовательности ДНК, РНК или белка с описательной информацией о структуре или функции.

Примечание — Процесс аннотации секвенированной последовательности можно рассматривать как присвоение метаданных последовательности.

3.35 **одноконцевой рид (одноконцевое прочтение)** (single-end read): Эквивалентное прочтение, получаемое при прочтении фрагмента ДНК с одного конца до другого.

3.36 **однонуклеотидный вариант;** SNV (single nucleotide variant, SNV): Изменение в одном нуклеотиде молекулы нуклеиновой кислоты.

3.37 **структурная вариация;** SV (structural variation, SV): Фрагмент ДНК размером примерно 1000 оснований или больше, который может включать инверсии и сбалансированные транслокации или геномный дисбаланс.

Примечание — Общие типы структурных вариантов включают в себя варианты числа копий (делеции, вставки, амплификации, дупликации), делеции с нейтральным числом копий (потеря гетерозиготности), инверсии, сегментные дупликации и транслокации (сбалансированные или несбалансированные).

3.38 **субчтение** (subread): Доля прочтения, находящаяся между адаптерами шпильки.

3.39 **тримминг исходных прочтений** (trimming of raw reads): Процедура, направленная на удаление низкокачественных частей или загрязнений последовательности с сохранением самой длинной высококачественной части чтения MPS.

3.40 **вариации** (variation): Отличия одного или нескольких оснований нуклеиновой кислоты в последовательности по отношению к ожидаемому(ым) основанию(ям).

3.41 **идентификация вариантов (коллинг)** (variant calling): Логическое заключение о том, что в определенной позиции нуклеотид исследуемой ДНК отличается от нуклеотида референсной последовательности.

3.42 **волновод с нулевым режимом;** ZMW (zero mode waveguide, ZMW): Оптический волновод, который направляет световую энергию в объем, который по всем измерениям мал по сравнению с длиной волны света.

Примечание — Полимераза закрепляется на дне этой ZMW, и включение нуклеотидов измеряется по увеличению флуоресценции во время связывания с последующим уменьшением после включения.

4 Исходные данные

4.1 Общие положения

Каждому нуклеотиду в последовательности необходимо присвоить числовое значение (основная оценка качества), которое соотносится с предполагаемой точностью процесса распознавания нуклеотидных оснований (base calling), если это применимо.

4.2 Файл с исходными данными

Для генерации файлов чтения последовательностей следует использовать программное обеспечение и/или конвейеры, специфичные для конкретного прибора. Контролируемые физические параметры, такие как соотношение сигнал/шум, должны быть задокументированы. Также эти физические параметры необходимо контролировать во время каждого эксперимента по секвенированию.

Файлы чтения последовательностей должны быть сконфигурированы в соответствующий формат файла, содержащий компиляцию отдельных чтений последовательностей, каждое из которых имеет свой собственный идентификатор, и ассоциированную оценку качества основания для каждого нуклеотида.

Примечание — Формат FASTQ (или конвертируемый в формат FASTQ) может использоваться в качестве стандартного формата де-факто для последующего анализа качества наборов данных MPS. FASTQ широко признан в качестве кроссплатформенного формата обмена файлами.

Файл выходных данных, полученных после проведения секвенирования, и соответствующие метрики качества должны быть проанализированы в последующем биоинформатическом конвейере с использованием соответствующего программного обеспечения.

4.3 Оценка качества исходных данных

4.3.1 Общие требования

Показатели контроля качества могут отличаться в зависимости от платформы MPS, метода подготовки библиотеки и предполагаемого использования анализа.

Результаты секвенирования должны интерпретироваться компетентным персоналом. Интерпретацию выполняют таким образом, чтобы соответствовать уровню качества, соответствующему предполагаемой цели анализа, с учетом статистически достоверного числа повторных прочтений.

Инструменты обработки ридов применяют с учетом оценки качества и обрезки необработанных прочтений.

4.3.2 Базовая статистика

Необходимо зарегистрировать основные статистические данные, включая, но не ограничиваясь ими:

a) тип платформы;
b) тип рида;
c) набор для подготовки библиотеки;
d) длину рида;
e) количество ридов;
f) общее содержание GC;
g) общую длину последовательностей.

4.3.3 Показатели качества

К показателям контроля качества для оценки исходных данных относятся, но не ограничиваются:

a) распределение длины последовательностей;
b) на содержание GC последовательности;
c) показатель качества;

1) качество последовательности оснований;
2) показатель качества последовательности.

Примечание 1 — Низкое качество оценки может указывать на увеличение числа ложноположительных вызовов вариантов;

3) все последовательности должны быть помечены как «предупреждение» или «пройдено» для качества последовательности по каждому основанию;

d) состав последовательности по нуклеотидам;
e) допустимость соотношения сигнал/шум;
f) уровни дупликаций последовательности;
g) уровень перепредставленности последовательностей;
h) плотность кластера;
i) отношение транзиция/трансверсия для секвенирования целого экзома или целого генома или секвенирования больших ампликонов;
g) доля адаптеров/загрязнение (контаминация) последовательности адаптерами;
k) загрязняющие вещества (идентификация, количественное определение);
l) частота ошибок.

Примечание 2 — Включает в себя ошибки гомополимера: ошибки в количестве оснований, вызванные тем, когда один нуклеотид встречается в последовательности более одного раза в последовательном порядке;

m) анализ *k*-меров.

Примечание 3 — В вычислительной геномике *k*-меры относят ко всем возможным подпоследовательностям (длиной *k*) из последовательности нуклеиновой кислоты. Перепредставленность *k*-меров может быть проанализирована для обнаружения потенциальной неправильной сборки генома, если повторяющиеся последовательности ДНК, возможно, были объединены;

n) N фрагмент.

Примечание 4 — Количество и/или процент неоднозначных вызовов;

o) повторение растяжки и повторение последовательности;
p) распределение нуклеотидов между циклами.

4.4 Предварительная обработка исходных данных

Предварительная обработка исходных данных может включать следующие вычислительные этапы, если они применимы, но не ограничивается ими:

a) удаление/обрезку некачественных последовательностей/баз;
b) демультиплексирование;
c) удаление адаптеров/праймеров и загрязнений;
d) исправление ошибок;
e) фильтрацию дуплицированных ридов;
f) обрезку ридов до фиксированной длины;
g) вызов CCS-ридов.

При использовании данных CCS необходимо получить и отфильтровать риды CCS перед последующим анализом.

5 Выравнивание последовательностей и картирование

5.1 Общие требования

Стратегия выравнивания последовательностей и картирования должна быть выбрана на основании приложения.

Пример — Существует сплайсированное картирование для РНК и несплайсированное картирование для стратегии картирования секвенирования РНК.

Для выравнивания допускается использовать программное обеспечение и инструменты для выравнивания и картирования.

Качество выравнивания допускается оценивать визуально, используя соответствующие виды выравнивания, а также используя информацию, представленную в файле выравнивания.

Примеры программного обеспечения для выравнивания и картирования последовательностей различного назначения приведены в приложении С.

Для картирования используются референсные геномы/референсные целевые области нуклеиновых кислот, которые должны быть тщательно подобраны в зависимости от плана эксперимента.

Примечание 1 — Изучение содержит версию референсного генома/референсной целевой области нуклеиновой кислоты, выбор различных штаммов в одном организме и выбор геномов с маской, мягкой маской или без маски.

Примечание 2 — Программное обеспечение для выравнивания и картирования секвенирования с открытым исходным кодом доступно онлайн.

5.2 Последовательность и формат файла картирования

Данные выравнивания последовательностей всегда хранятся в следующих форматах файлов.

a) Карта выравнивания последовательностей (SAM) [17], [24].

Примечание 1 — SAM представляет собой текстовый формат с разделителем TAB, состоящий из заголовка, который является необязательным, и секции выравнивания. Каждая строка выравнивания имеет 11 обязательных полей для основной информации о выравнивании, такой как положение сопоставления, и переменное количество необязательных полей для гибкой или специфической для выравнивателя информации.

b) Карта двоичного выравнивания (BAM) [15], [17].

Примечание 2 — Это сжатый формат, аналогичный формату SAM в двоичном виде.

c) Сжатая карта выравнивания, ориентированная на ссылку (CRAM) [16].

Примечание 3 — CRAM — это формат файла чтения секвенирования, который экономит место за счет сжатия данных последовательностей на основе ссылок и предлагает режимы сжатия без потери информации и с потерями.

d) Группа экспертов по движущимся изображениям в геномике (MPEG-G) [3]—[8].

Примечание 4 — MPEG-G — это формат представления геномики, основанный на концепции геномной записи, структуре данных, состоящей либо из одного прочтения последовательности, либо из парных прочтений последовательности и связанной с ними информации о секвенировании и выравнивании; она может содержать подробные данные сопоставления и выравнивания, идентификатор одиночного или парного чтения (имя чтения) и значения качества. Геномные записи агрегируются и кодируются в структурах, называемых единицами доступа. Эти структуры представляют собой единицы закодированной геномной информации, к которым можно получить отдельный доступ и проверить их.

Примечание 5 — MPEG-G определен в серии стандартов ИСО/МЭК 23092.

Файл выравнивания должен содержать информацию о расположении, ориентации и качестве каждого чтения (рида) в выравнивании.

Для работы с файлами выравнивания могут применяться алгоритмы и инструменты в зависимости от их применения.

5.3 Контроль качества выравнивания последовательностей и картирование

5.3.1 Базовая статистика выравнивания

5.3.1.1 Общие требования

Необходимо получить и записать основные статистики выравнивания или картирования.

Базовая статистика выравнивания или картирования может отличаться в зависимости от плана эксперимента и типа считывания.

5.3.1.2 Статистика картирования для одноконцевых ридов

a) Общее количество ридов относится к количеству прочтений, которые сопоставлены с эталонной последовательностью или геномом.

b) Несопоставленные риды означают количество прочтений, которые не удалось сопоставить с эталонной последовательностью или геномом.

c) Сопоставленные риды означают количество прочтений, выровненных с эталонной последовательностью или геномом.

d) Уникально сопоставленные риды означают количество прочтений, точно выровненных с эталонной последовательностью или геномом.

Примечание 1 — Уникальность картирования зависит от обстоятельств. Риды, уникально сопоставленные на основе одного набора параметров сопоставления, могут быть многократно сопоставленными прочтениями при другом наборе параметров сопоставления.

e) Многохитовые сопоставленные риды означают количество прочтений, выровненных более одного раза с эталонной последовательностью или геномом.

Примечание 2 — Мультихит зависит от специфики картирования.

5.3.1.3 Статистика картирования для парно-концевых ридов

a) Общее количество спаренных концевых ридов относится к количеству парно-концевых прочтений, сопоставленных с эталонной последовательностью или геномом.

b) Сопоставленные спаренные концевые риды — число парных ридов, в которых оба были сопоставлены.

c) Частично сопоставленные спаренные концевые риды — число парных прочтений, в которых сопоставлен только один из них.

d) Несопоставленные спаренные концевые риды — число парных ридов, которые не удалось сопоставить с эталонной последовательностью или геномом.

e) Неправильно сопоставленные спаренные концевые риды — число парных ридов, из которых один был сопоставлен с дискордантной ориентацией.

Примечание 1 — Также известны под наименованием: несогласованно сопоставленные пары.

f) Под правильно сопоставленными спаренными концевыми ридами понимается общее число парных прочтений, из которых оба сопряженных рида были сопоставлены с согласованной ориентацией.

Примечание 2 — Также известны под наименованием: согласованно сопоставленные пары.

5.3.1.4 Длина сопоставленного субрида

Длина выравнивания субрида с целевой эталонной последовательностью не включает последовательность адаптера.

5.3.2 Индикаторы качества

В зависимости от условий применения рекомендуется использовать следующие параметры контроля качества:

a) скорость выравнивания.

Примечание 1 — Низкое качество картирования может быть результатом неспецифической амплификации, захвата загрязнения внецелевой ДНК или других причин;

b) длина фрагмента или длина ДНК/РНК, которая должна быть секвенирована;

c) статистика размера вставки для парного прочтения — длина ДНК/РНК, предназначенная для секвенирования между адаптерами.

Примечание 2 — Пик распределения размера вставки используют для оценки качества;

d) уровень дупликации только для секвенирования на основе ампликонов;

e) покрытие по назначению, включая глубину, ширину и диапазон покрытия.

Примечание 3 — В приложении В приведен список рекомендуемого покрытия для различных использований;

f) смещение AT/GC.

Примечание 4 — Оценку допускается проводить в соотношении % GC с глубиной секвенирования/покрытием;

g) оценка качества картирования;

h) эффективность захвата.

Примечание 5 — Эффективность захвата является наиболее важным параметром контроля качества для секвенирования экзома или других секвенирований на основе захвата мишени;

i) средняя или медианная глубина, процент генома, охваченного секвенированием на данной глубине;

j) количество дискордантно сопоставленных пар;

k) высокое качество выровненных прочтений;

l) частота несовпадений;

m) точность консенсуса.

Примечание 6 — Точность консенсуса основывается на выравнивании нескольких секвенированных прочтений и субридов вместе, не обязательно с эталонной последовательностью;

n) точность кругового консенсуса.

Примечание 7 — Точность кругового консенсуса основана на нескольких проходах секвенирования вокруг одной круговой молекулы шаблона. Это используется в CCS;

o) точность субридов.

Примечание 8 — Точность распознавания оснований после картирования.

5.3.3 Методы оценки качества выравнивания и картирования

Для оценки качества выравнивания следует применять подход, основанный на балльной системе оценивания.

Примечание — Выбор матрицы подсчета баллов зависит от области применения.

5.4 Постобработка выравнивания

Постобработка выравнивания может включать, но не ограничиваться:

a) локальное выравнивание вокруг инделов или расчет выравниваний по основанию;

b) удаление дубликатов;

c) перекалибровку оценок качества оснований;

d) среднюю длину ридов после обрезки по качеству оснований.

6 Идентификация вариантов

6.1 Общие требования

6.1.1 Существует четыре основных класса вариантов последовательности (SNV, инделы, CNVs и SVs). Для разных классов вариантов последовательности следует применять различные вычислительные подходы для чувствительной и специфической идентификации.

6.1.2 Диапазон программных инструментов и тип необходимой валидации зависят от плана анализа.

6.2 Файл данных для идентификации вариантов

6.2.1 Распознанные варианты должны быть аннотированы с использованием соответствующей спецификации. Спецификация должна содержать метаинформацию, строку заголовка и строки данных, каждая из которых содержит информацию о позиции в геноме и информацию о генотипе образцов для каждой позиции.

Пример 1 — Выявленные варианты аннотируются с использованием формата идентификации вариантов (VCF) [31].

Пример 2 — Существуют альтернативные спецификации для представления и хранения вариантов:

a) геномные конвенции VCF;

b) Онтология последовательностей, версия формата вариации генома 1.10;

c) Общество по изучению вариаций генома человека, Общество по изучению вариаций генома человека (HGVS), простая версия 15.11;

d) Глобальный альянс по геномике и здоровью (GA4GH), форматы файлов.

6.2.2 Файлы вариантов должны включать как спецификацию, так и используемую версию.

6.2.3 Устройства распознавания вариантов должны быть сконфигурированы на вывод; эталоны, варианты и отсутствие вызовов вместе с местной информацией, по крайней мере в области целевых областей.

6.3 Показатели качества при определении вариантов

Метрики контроля качества должны включать, но не ограничиваться (если применимо):

a) пороговые значения для глубины покрытия ридов в позиции варианта;

b) оценку качества вариантов;

c) смещение цепочек;

d) процент аллельных прочтений;

e) дополнительные конкретные метрики, относящиеся к точности и чувствительности вызова вариантов, которые могут включать, но не ограничиваются:

1) общее число вариантов,
2) количество ложноположительных результатов,
3) количество ложноотрицательных вариантов,
4) количество несовпадений аллелей и генотипов,
5) соотношение транзиций/трансверсий,
6) соотношение гетерозигот/гомозигот (het/hom);

f) анализ контаминации перекрестной выборки.

6.4 Обработка ложноположительных вариантов

Ложноположительные варианты должны быть отмечены или отфильтрованы из исходных файлов вариантов на основе нескольких выравниваний последовательностей и метрик контроля качества, связанных с поиском вариантов.

6.5 Аннотация секвенированной последовательности

Варианты могут быть аннотированы для определения их биологической значимости, что позволяет определить функциональные приоритеты и интерпретировать последующие данные.

7 Валидация

7.1 Общие требования

7.1.1 Лаборатории, предлагающие тестирование на основе MPS, должны провести «внутреннюю» валидацию биоинформатического конвейера.

7.1.2 Требования к производительности анализа необходимо установить в ходе процедуры валидации и те же спецификации следует использовать для мониторинга производительности анализа при каждой обработке образца.

7.1.3 Конкретные параметры контроля качества и обеспечения качества должны оцениваться в ходе валидации и использоваться для определения удовлетворительных характеристик.

7.1.4 Каждая лаборатория должна определить критерии и средства мониторинга всех показателей качества для обеспечения оптимальных аналитических характеристик. Метрики качества, используемые для мониторинга, должны быть внесены в документы и периодически проверяться.

Рекомендуемые показатели качества и их конкретные значения для некоторых платформ приведены в приложении А.

7.1.5 Лаборатории должны предусмотреть конкретные меры по обеспечению сохранности каждого файла данных, созданных в биоинформатическом конвейере, и обеспечению предупреждения или предотвращению использования файлов данных, которые были изменены несанкционированным или непреднамеренным образом.

7.1.6 Дополнительная валидация требуется каждый раз, когда в любой компонент биоинформатического конвейера вносят значительные изменения.

7.2 Валидация показателей качества

7.2.1 Валидацию анализа следует проводить на основе уточненной и документированной цели анализа. Необходимо определить целевое назначение измерения и внести его в документы.

7.2.2 Лаборатории должны установить приемлемые пороговые значения оценки качества сырой базовой идентификации для анализа во время валидации.

7.2.3 Для снижения частоты ложного распознавания следует разработать методы предварительной обработки для удаления некачественных базовых вызовов.

7.2.4 Степень смещения GC во всех частях генома, включенных в анализ, должна быть определена в ходе валидации.

7.2.5 Параметры качества картирования должны быть установлены в плане валидации и должны показывать, что тест оценивает только те риды, которые отображаются на области, являющиеся мишенью анализа. При необходимости следует установить фазы для фильтрации прочтений, которые отображаются в нецелевых областях.

7.2.6 Охват определяется для достижения адекватной чувствительности и специфичности в областях, представляющих интерес.

7.2.7 Каждая лаборатория должна установить минимальные критерии глубины покрытия, характерной для конкретной области в стандартных условиях анализа, в зависимости от цели проведения секвенирования. Для однородного образца последовательность должна быть подтверждена; меньшая глубина допустима. В процессе вызова вариантов по области или редкой последовательности в смешанном образце, составляющем 1 %, необходимо глубокое секвенирование.

7.2.8 Требуемый уровень покрытия в целевых областях должен быть определен на этапе валидации (диапазон покрытия). Рекомендуемый диапазон для различных приложений описан в приложении В.

7.2.9 Для каждого анализа должны быть установлены приемлемые параметры для максимальной частоты дупликации.

7.2.10 Для увеличения количества пригодных для использования данных секвенирования и предотвращения перекосов в долях аллелей следует установить фильтрацию дуплицированных ридов с помощью аналитического конвейера.

7.2.11 Каждая лаборатория должна определить допустимый уровень смещения цепочек и изложить конкретные критерии, когда следует проводить альтернативное тестирование.

7.2.12 Показатели качества могут быть проверены с помощью соответствующих эталонных стандартов, которые были хорошо охарактеризованы и имеют надежные эталонные последовательности для точного выравнивания, поиска вариантов и т. д.

7.2.13 Рекомендуется проведение секвенирования по методу Сэнгера для подтверждения наиболее важной связующей области.

8 Документирование

8.1 Лаборатории должны документировать все алгоритмы, программное обеспечение и базы данных, используемые при анализе, интерпретации и представлении результатов MPS. Версия каждого из этих компонентов в общем биоинформатическом конвейере должна быть записана и отслеживаться для каждого результата.

8.2 В документации лаборатории должны содержаться сведения о любых пользовательских настройках, которые отличаются от конфигурации по умолчанию, или должно быть указано, какие параметры были изменены.

8.3 При необходимости следует указать номер версии эталонной последовательности и подробную информацию.

8.4 Лаборатории также должны документировать параметры контроля качества для оптимальной работы.

Пример — На первичном этапе лаборатория определяет приемлемые критерии, такие как количество прочтений, проходящих через фильтры качества, определенные прибором.

8.5 Лаборатории должны фиксировать в протоколах процессы биоинформатики, используемые для сокращения большого набора данных о вариантах до списка причинных генов и/или генов-кандидатов и/или вариантов.

8.6 Доказательства соответствия установленным требованиям должны быть также задокументированы.

Приложение А
(справочное)

Показатели качества для конкретных примеров платформ MPS

Для секвенирования нуклеиновых кислот, как правило, используют приведенные далее платформы MPS. Примеры показателей качества, используемые для оценки качества, представлены в таблице А.1.

Примечание — Секвенирование целого генома человека используют в качестве примера, чтобы обеспечить конкретные значения для каждого показателя качества.

Таблица А.1 — Показатели качества для конкретных платформ MPS

Наименование платформы	Формат файла исходных данных	Длина рида	Оценка качества (H/L)	GC-состав	Степень дупликации	Плотность кластеров	Степень адаптера
illumina®[a]HiSeq 4000	fastq.gz	От 50 до 200 bp	>Q30	От 39 % до 42 %	<10 %	5 млрд	<3 %
Thermo FisherProton™ [b]	DAT	От 50 до 200 bp	>Q20	От 39 % до 42 %	NA	От 60 до 80 млн	<3 %
BGI[c] /MGI MGISEQ-2000	fastq.gz	От 50 до 200 bp	>Q30	От 39 % до 42 %	<5 %	1,5 млрд	<3 %
Oxford Nanopore PromethION®[d]	FAST5	От 10 до 300 kbp	>Q20	От 39 % до 42 %	NA	2560 каналов[f]	<3 %
PacBio®Sequel II[e]	bam	От 10 до 100 kbp	>Q20	От 39 % до 42 %	NA	8 млн ZMWs[g]	<3 %

[a] illumina® является торговой маркой биотехнологической компании illumina, Inc. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.

[b] Thermo Fisher Proton™ является торговой маркой биотехнологической компании Thermo Fisher Scientific. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.

[c] MGI является торговой маркой компании, занимающейся секвенированием генома BGI. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.

[d] Oxford Nanopore PromethION® является торговой маркой Oxford Nanopore Technologies Limited. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.

[e] PacBio Sequel II® является торговой маркой биотехнологической компании Pacific Biosciences. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.

[f] Oxford Nanopore измеряется с помощью каналов.

[g] Pacific Biosciences измеряется в ZMWs (волновод с нулевой модой).

Приложение В
(справочное)

Охват и рекомендации по прочтению по приложениям

В таблице В.1 представлены примеры охвата и уровней считывания множеством различных приложений для секвенирования.

Таблица В.1 — Охват и рекомендации по прочтению по приложениям

MPS тип	Приложение	Рекомендованное покрытие (E)	Рекомендованные риды
Секвенирование целого генома	Гомозиготные однонуклеотидные варианты (SNVs) — однонуклеотидные изменения в генах, когда аллели идентичны	15×[a]	—
	Гетерозиготные SNV — однонуклеотидные изменения в генах, когда аллели отличаются друг от друга	33×	—
	Мутации вставки/делеции (INDELS) — мутации в геноме, при которых происходит вставка или удаление нуклеотидов	60×	—
	Вариация числа копий (CNV) — дисперсия в количестве копий гена у разных людей	От 1× до 8×	—
Секвенирование целого экзома	Гомозиготные SNV	100 × (3× покрытие локального рида)[b]	—
	Гетерозиготные SNVs	100× (13× покрытие локального рида)[c]	—
Целевое секвенирование	INDELs	Не рекомендовано	—
	SNVs/SVs в целевых областях	От 1000 до 10 000 раз	—
Секвенирование РНК. Секвенирование транскриптома	16S рРНК ген [23], [24]	—	Минимум 100 на образец
	Профилирование дифференциальной экспрессии — количественное измерение экспрессии генов по нескольким генам для изучения различных уровней экспрессии в образце	—	От 10 млн до 25 млн
	Альтернативный сплайсинг — идентификация различных вариантов сплайсинга из транскриптов мРНК	—	От 50 млн до 100 млн (для коротких платформ ридов) От 2 млн до 3 млн (для длинных платформ ридов)
	Аллель-специфическая экспрессия — экспрессия транскрипта, на которую влияет аллель конкретного гена	—	От 50 млн до 100 млн

Окончание таблицы В.1

MPS тип	Приложение	Рекомендованное покрытие (E)	Рекомендованные риды
MPS тип	Дифференциальная экспрессия — количественное измерение экспрессии малых РНК для изучения различных уровней экспрессии в образце	—	От ~1 млн до 2 млн
	Обнаружение новых малых РНК	—	От ~5 млн до 8 млн

Примечание 1 — Результаты могут быть подтверждены дополнительными экспериментами по протеомике.

Примечание 2 — Рекомендуемое покрытие относится к образцам генома человека.

[a] 15× означает локальное одинаковое покрытие, это не общее среднее покрытие. Числа приведены для примера.

[b] 100× — общее среднее покрытие для секвенирования целых экзомов. 3× локальное покрытие рида указывает на локальное покрытие для обнаружения SNPV. Числа приведены для примера.

[c] 00× — общее среднее покрытие для секвенирования экзома. Покрытие 15× локального рида указывает на локальное покрытие для обнаружения SNPV. Числа приведены для примера.

Приложение С
(справочное)

Программное обеспечение для выравнивания и сопоставления последовательностей

В таблице С.1 представлены примеры программного обеспечения для выравнивания и сопоставления последовательностей.

Таблица С.1 — Программное обеспечение для выравнивания и картирования последовательностей

Описание функций	Программное обеспечение/инструменты
Выравнивание или картирование	Blast, Blat, SOAP, BWA, Bowtie2 и т. д.
Оценка участков сплайсинга в анализе РНК-секвенирования	Bowtie2 [25], BWA [16], HISAT2 [14], STAR [15] и т. д.
Визуализация для представления выравнивания	Bam View [12], Integrative Genomic Viewer [30]
Примечание 1 — Программное обеспечение регулярно обновляется и значительно зависит/связано с платформами, приложениями и данными о последовательности. Эти примеры актуальны в июне 2020 г. Примечание 2 — Примеры программного обеспечения, перечисленные в данной таблице, являются подходящим доступным программным обеспечением. Данная информация приведена для удобства пользователей настоящего стандарта и не означает одобрения указанного продукта со стороны ИСО.	

Библиография

[1] ISO/TS 20428, Health informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records

[2] ISO 22174:2005, Microbiology of food and animal feeding stuffs — Polymerase chain reaction (PCR) for the detection of food-borne pathogens — General requirements and definitions

[3] ISO/IEC 23092-1:2020, Information technology — Genomic information representation — Part 1: Transport and storage of genomic information

[4] ISO/IEC 23092-2:2020, Information technology — Genomic information representation — Part 2: Coding of genomic information

[5] ISO/IEC 23092-3:2020, Information technology — Genomic information representation — Part 3: Metadata and application programming interfaces (APIs)

[6] ISO/IEC 23092-4:2020, Information technology — Genomic information representation — Part 4: Reference software

[7] ISO/IEC 23092-5:2020, Information technology — Genomic information representation — Part 5: Conformance

[8] ISO/IEC 23092-6[1)], Information technology — Genomic information representation — Part 6: Coding of genomic annotations

[9] Ardui S. et al. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic acids research. [Online]. March 2018; 46(5) 2159-2168 [viewed 2019-09–15]. Available at: gost.e-ecolog.ru

[10] Aziz N. et al. College of American Pathologiests' Laboratory Standards for Next-Generation Sequencing Clinical Tests. Arch Pathol Lab Med [Online]. April 2015, 139(4), 481-493 [viewed 2018-4–10]. Available at: gost.e-ecolog.ru ncbi.nlm.nih.gov/ pubmed/25152313

[11] Carver T. et al. Bamview: viewing mapped read alignment data in the context of the reference sequence. Bioinformatics. [Online]. March 2010; 26(5):676-677 [viewed 2019-01–15] Available at: gost.e-ecolog.ru . gov/pmc/articles/PMC2828118/

[12] Dunnen J.T. et al. HGVS recommendations for the description of sequence variants: 2016 update. March 2nd 2016. [online]. Human mutation. [viewed May 1st 2020] Available from: gost.e-ecolog.ru humu.22981

[13] Daehwan K. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology. August 2nd 2019. [online]. Springer. [viewed May 1st 2020] Available from: gost.e-ecolog.ru articles/s41587-019-0201-4

[14] Dobin A. et al. STAR: Ultrafast universal RNA-seq aligner. 25th Oct. 2012 [online] Bioinformatics. [viewed May 1st 2020] Available from: gost.e-ecolog.ru bioinformatics/article/29/1/15/272537

[15] Li H., Durbin R. Fast and accurate short read alignment with Burrows-wheeler transform. May 18th 2009. [online] Bioinformatics. [viewed May 1st 2020]. Available from: gost.e-ecolog.ru article/25/14/1754/225615

[16] European Nucleotide Archive (ENA) CRAM. [online]. EMBL-EBI 2019 [viewed 2019-01–15] Available at: https:// gost.e-ecolog.ru

[17] Github. SMA/BAM and related specificaitons. [online]. May 5th 2020 Github. [viewed May 27st 2020] Available from: gost.e-ecolog.ru

[18] Github. The Sequence Ontology Genome Variation Format Version 1.10. May 19th 2014 [viewed May 1st 2020]. Available from: gost.e-ecolog.ru

[19] Gvcftools. gVCF Conventions. September 2012 [online]. Gvcftools [viewed May 1st 2020] Available from: https:// sites.google.com/site/gvcftools/home/about-gvcf/gvcf-conventions Illumina. An introduction to next generation sequencing technology. [online]. Illumina. [viewed 2018-4–15]. Available at: gost.e-ecolog.ru content/dam/ illumina-marketing/documents/products/illumina_sequencing_ introduction.pdf

[20] Jennings L. J. et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels. Journal of Molecular Diagnostics [Online]. May 2017, 19 (3), 341-365 [viewed 2018-4–15]. Available at: gost.e-ecolog.ru org/article/S1525-1578(17)30025-9/ fulltext

[1)] В стадии подготовки.

[21] Kuczynski J. et al. «Direct sequencing of the human microbiome readily reveals community differences». Genome biology 11.5 (2010): 210.

[22] Kuczynski J. et al. «Microbial community resemblance methods differ in their ability to detect biologically relevant patterns». Nature methods 7.10 (2010): 813.

[23] Langmead B. Salzberg S.L. Fast gapped-read alignment with Bowtie 2. Nat Methods. Nat Methods. [Online]. March 2012; 9 (4) [viewed 2019-1–15] Available at: https:// gost.e-ecolog.ru

[24] LI H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. [Online]. 2009 Aug; 25(16): 2078-9. [viewed 2019-01–15] Available at: gost.e-ecolog.ru

[25] Pfeifer S. From next-generation resequencing reads to a high-quality variant data set. Heredity. [Online]. February 2017, 118 (2) [viewed 2018-4–15]. Available at: https:// gost.e-ecolog.ru

[26] Reinert K. et al. Alignment of Next-Generation Sequencing Reads. Annu Rev Genomics Human Genetics [Online]. May 2015, 16: 133-151 [viewed 2018-05-25] Available at: gost.e-ecolog.ru

[27] Rhoads A., AU K.F. PacBio Sequencing and its Applications. Genomics Proteomics Bioinformatics [Online]. October 2015; 13 (5): 278-289 [viewed 2019-09-15] Available at: gost.e-ecolog.ru

[28] Rouven N. et al. The tole of quality control in targeted next-generation sequencing library preparation. Genomics Preteomics Bioinformatics [online]. March 2016, 14, 200-206 [viewed 2018-08-01] Available at: gost.e-ecolog.ru sciencedirect.com/ science/article/pii/S1672022916301073

[29] Samtools organisation and repositories. The variant call format specification. April 2nd 2020. [online] samtoools organisation and repositories. [viewed May 1st 2020]. Available from: gost.e-ecolog.ru VCFv4.3.pdf

[30] Somak R.O.Y. et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipeline A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. The Journal of Molecular Diagnostics [Online]. Elsevier. November 2017, 20 (1) [viewed 2018-4-12]. Available at: http:// jmd.amjpathol.org/article/S1525-1578(17)30373-2/pdf

[31] Trivedi U. H. et al. Quality control of next generation sequencing without a reference. Front Gene [Online]. May 2014, 5, 111 [viewed 2018-4-10]. Available at: gost.e-ecolog.ru

[32] Yan G.U.O. et al. Three stage quality control strategies for DNA re-sequencing data. Briefing in Bioinformatics [Online]. September 2013, 15 (6), 879-889 [viewed 2018-04-13] Available at: gost.e-ecolog.ru article/15/6/879/180439

[33] ISO/TS 17822-1:2020, In vitro diagnostic test systems — Nucleic acid amplification-based examination procedures for detection and identification of microbial pathogens — Laboratory quality practice guide

УДК 615.07:006.354 ОКС 07.080

Ключевые слова: секвенирование, массовое параллельное секвенирование, оценка и контроль качества

Редактор *Е.В. Якубова*
Технический редактор *И.Е. Черепкова*
Корректор *Л.С. Лысенко*
Компьютерная верстка *А.Н. Золотаревой*

Сдано в набор 14.08.2023. Подписано в печать 18.08.2023. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 2,79. Уч.-изд. л. 2,40.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации» для комплектования Федерального информационного фонда стандартов, 117418 Москва, Нахимовский пр-т, д. 31, к. 2.

gost.e-ecolog.ru gost.e-ecolog.ru