

Local Run Manager

Модуль анализа Соматический вариант

Руководство по рабочему процессу для NextSeq 550Dx

ДЛЯ ДИАГНОСТИКИ IN VITRO

Обзор	3
Введение информации о запуске	4
Методы анализа	6
Просмотр данных о запуске и образце	8
Отчеты об анализе	8
Файлы выходных данных анализа	10
Определение последовательности нуклеотидов и различие индексов	17
История изменений	18
Техническая помощь	19



Настоящий документ и его содержание являются собственностью компании Illumina, Inc. и ее аффилированных лиц (Illumina) и предназначены для использования исключительно в рамках договора заказчиком при эксплуатации изделия (-й), описанного (-ых) в настоящем документе, и ни для какой иной цели. Настоящий документ и его содержание не подлежат использованию или распространению не по назначению и/или передаче, раскрытию или воспроизведению каким-либо способом без предварительного письменного согласия компании Illumina. Настоящим документом компания Illumina не передает никаких лицензий на свои патенты, товарные знаки, авторские права или права, признаваемые общим правом, или аналогичные права третьих лиц.

Инструкции, изложенные в настоящем документе, должны строго и точно соблюдаться квалифицированным и прошедшим соответствующее обучение персоналом для обеспечения правильной и безопасной эксплуатации изделий, описанных в настоящем документе. Перед началом работы с оборудованием пользователь обязан тщательно изучить данный документ.

НЕВЫПОЛНЕНИЕ ТРЕБОВАНИЙ ПО ПОЛНОМУ ПРОЧТЕНИЮ И ТОЧНОМУ ВЫПОЛНЕНИЮ ВСЕХ ИНСТРУКЦИЙ, СОДЕРЖАЩИХСЯ В НАСТОЯЩЕМ ДОКУМЕНТЕ, МОЖЕТ ПРИВЕСТИ К ПОВРЕЖДЕНИЮ ИЗДЕЛИЙ, ТРАВМАМ (ПОЛЬЗОВАТЕЛЯ ИЛИ ИНЫХ ЛИЦ) И ПОВРЕЖДЕНИЮ ИМУЩЕСТВА И ПРИВЕДЕТ К ОТМЕНЕ ЛЮБЫХ ГАРАНТИЙНЫХ ОБЯЗАТЕЛЬСТВ, ПРИМЕНИМЫХ К ИЗДЕЛИЯМ.

КОМПАНИЯ ILLUMINA НЕ НЕСЕТ НИКАКОЙ ОТВЕТСТВЕННОСТИ, ВОЗНИКАЮЩЕЙ ВСЛЕДСТВИЕ НЕНАДЛЕЖАЩЕГО ИСПОЛЬЗОВАНИЯ ИЗДЕЛИЙ, ОПИСАННЫХ В НАСТОЯЩЕМ ДОКУМЕНТЕ (ВКЛЮЧАЯ ИХ ЧАСТИ ИЛИ ЧАСТИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ).

© 2021 г. Illumina, Inc. Все права защищены.

Все товарные знаки являются собственностью компании Illumina, Inc. или их соответствующих владельцев. Информацию о конкретных товарных знаках см. на веб-сайте по адресу www.illumina.com/company/legal.html.

Обзор

Модуль анализа Local Run Manager Somatic Variant предназначен для использования с пользовательским набором Illumina TruSeq Custom Amplicon Kit Dx и NextSeq 550Dx. При использовании с модулем Somatic Variant анализ предназначен для приготовления библиотек, используемых для секвенирования ДНК из зафиксированной в формалине и залитой парафином (FFPE) ткани. Анализ выявляет соматические мутации с низкой частотой вариантов.

Модуль анализа оценивает короткие участки амплифицированной ДНК, или ампликоны, на наличие вариантов. Сфокусированное секвенирование ампликонов обеспечивает высокий охват определенных областей в большом количестве образцов. Модуль анализа выполняет вторичный анализ и создание отчетов из запусков секвенирования с использованием двухцепочечного подхода, который включает пулы прямых и обратных олигонуклеотидов (oligo pools). См. листок-вкладыш *Комплект TruSeq Custom Amplicon Kit Dx (документ No 1000000029772)*.

Модуль анализа Somatic Variant требует расходных материалов для секвенирования из 300 циклов. Для получения дополнительной информации см. листок-вкладыш к *NextSeq 550Dx High Output Reagent Kit v2* или *NextSeq 550Dx High Output Reagent Kit v2.5*.

О данном руководстве

В этой документации содержатся инструкции по настройке параметров запуска для параметров секвенирования и анализа для модуля анализа Соматический вариант. Для получения информации об информационной панели Local Run Manager и настройках системы, см. *NextSeq 550Dx Instrument Reference Guide (Справочное руководство к прибору NextSeq 550Dx) (документ № 1000000009513)*.

Просмотр Local Run Manager

Интерфейс Local Run Manager просматривается в Программное обеспечение прибора NextSeq 550Dx Operating Software (NOS) или через веб-браузер. Поддерживаемый веб-браузер – Chromium.



ПРИМЕЧАНИЕ.

Если вы используете неподдерживаемый браузер, загрузите поддерживаемый браузер, когда появится сообщение Confirm Unsupported Browser (Подтвердить неподдерживаемый браузер). Выберите «[here](#)» (здесь), чтобы загрузить поддерживаемую версию Chromium.

Просмотр на мониторе прибора

- 1 Для просмотра интерфейса Local Run Manager на мониторе прибора выберите один из следующих вариантов:
 - ▶ На экране Home (Главная страница) NOS выберите **Local Run Manager**. Нажмите X в правом верхнем углу, чтобы вернуться к NOS после завершения.
 - ▶ Выберите значок Minimize (Свернуть) NOS, откройте веб-браузер Chromium, который находится на приборе, и введите **http://localhost** в адресной строке. Только администраторы могут свернуть NOS.

Просмотр с сетевого компьютера

- 1 Откройте веб-браузер Chromium на компьютере с доступом к той же сети, что и прибор, и подключитесь, используя IP-адрес или имя прибора. Например, <http://myinstrument>.

Введение информации о запуске

Настройка параметров

- 1 Войдите в Local Run Manager.
- 2 Выберите **Create Run** (Создать запуск) и выберите **Somatic Variant** (Соматический вариант).
- 3 Укажите название запуска, позволяющее идентифицировать запуск в секвенировании в ходе анализа.
Используйте буквенно-цифровые символы, пробелы, символы подчеркивания или дефисы.
- 4 **[Необязательно]** Введите описание запуска, чтобы идентифицировать его.
Используйте буквенно-цифровые символы, пробелы, символы подчеркивания или дефисы.
- 5 Выберите количество образцов и набор индексов из раскрывающегося списка.
При выборе нужно учитывать следующую информацию.
 - ▶ В раскрывающемся списке содержится количество образцов с набором индексов. Например, 24-Set 1 указывает на 24 образца, которые необходимо протестировать, с индексами из набора индексов 1.
 - ▶ Номера наборов индексов относятся к различным наборам индексов i5. Set 1 и Set 2 обеспечивают разнообразие индексов. Для предотвращения истощения одного набора предлагаются два набора индексов.
 - ▶ Выберите количество образцов, которое ближе всего к количеству образцов, которые вы тестируете. Если точного количества образцов нет в списке, выберите ближайшее число, но меньше, чем число, которое вы тестируете. Например, если вы хотите протестировать 18 образцов, выберите 16 образцов.
 - ▶ Лунки образцов и индексных комбинаций, которые соответствуют требованиям к разнообразию индексов, выделены зеленым цветом. Если вы выберете другие комбинации лунок и индексов, при сохранении запуска вы получите уведомление, если требования к разнообразию индексов не выполнены.

Импорт файлов манифеста для запуска

- 1 Убедитесь, что все манифесты, которые необходимо импортировать, доступны в сетевом расположении, подключенном к прибору, или на USB-носителе.
- 2 Нажмите **Import Manifests** (Импорт манифестов).
- 3 Перейдите к файлу манифеста и выберите манифесты, которые вы намерены добавить.



ПРИМЕЧАНИЕ.

Чтобы открыть доступ ко всем файлам манифеста для всех запусков с помощью модуля анализа Соматический вариант, добавьте манифесты с помощью функции Module Settings (Параметры модуля). Для данной функции требуются разрешения пользователя на уровне администратора. Для получения дополнительной информации см. *NextSeq 550Dx Instrument Reference Guide (Справочное руководство к прибору NextSeq 550Dx) (документ № 1000000009513)*.

Порядок указания образцов для запуска


Укажите образцы для запуска с помощью одного из вариантов и в соответствии с указаниями, приведенными далее.

- ▶ **Ввести образцы вручную** – воспользуйтесь пустой страницей на экране Create Run (Создать запуск).
- ▶ **Импортировать образцы** – перейдите ко внешнему файлу в формате (*.csv), в котором значения разделены запятыми. Шаблон для загрузки находится на экране Create Run (Создать запуск).

После заполнения таблицы с образцами вы можете экспортировать информацию об образцах во внешний файл. Используйте файл в качестве справочной информации при подготовке библиотек или импорте файла для другого запуска.

Введение образцов вручную

- 1 Введите уникальное название образца в поле Sample Name (Название образца). Используйте буквенно-цифровые символы, тире или символы подчеркивания. Название образца автоматически указывается в соответствующей лунке в другом пуле.
- 2 **[Необязательно]** Для образцов положительных или отрицательных контролей нажмите правую кнопку мыши и выберите тип контроля. Название контроля в одной лунке для образца автоматически указывается в названии соответствующей лунки в другом пуле с тем же контролем.
- 3 **[Необязательно]** Введите описание образца в поле Sample Description (Описание образца). Используйте буквенно-цифровые символы, тире или символы подчеркивания. Описание образца автоматически указывается в соответствующей лунке в другом пуле. Описания образцов связаны с идентификатором образца. Описания образцов перезаписываются, если такой же идентификатор образца используется снова в одном из последующих запусков.
- 4 Выберите индексный адаптер 1 из раскрывающегося списка Индекс 1 (i7). При использовании предложенных лунок для образцов программное обеспечение автоматически заполняет индексные адаптеры i7 и i5, которые соответствуют требованиям к различию индексов. Если в списке не указано точное число образцов, которые вы анализируете, обязательно выберите индексные адаптеры для дополнительных лунок. Если вам необходимо выбрать индексы для дополнительных лунок или вы не используете сочетания рекомендованных индексных адаптеров, то, прежде чем выбрать индексы, обязательно прочтите раздел *Определение последовательности нуклеотидов и различие индексов на странице 17*.
- 5 Выберите индексный адаптер 2 из раскрывающегося списка Индекс 2 (i5).

- 6 Выберите файл манифеста из раскрывающегося списка «Манифест». Для образцов в пуле A требуется манифест, отличный от манифеста для образцов в пуле B.
- 7 Выберите соответствующий параметр, чтобы просматривать, распечатывать или сохранять макет планшета в качестве справочника для подготовки библиотек.
 - ▶ Нажмите значок  **Print** (Печать), чтобы отобразить макет планшета. Нажмите **Print** (Печать), чтобы распечатать макет планшета.
 - ▶ Нажмите **Export** (Экспорт), чтобы экспортировать информацию об образце во внешний файл.

Проверьте правильность информации в манифестах и информации об образце. Неправильная информация может повлиять на результаты.
- 8 Нажмите **Save Run** (Сохранить запуск).

Импорт образцов

- 1 Выберите **Import Samples** (Импортировать образцы) и перейдите к расположению файла информации об образце. Существует два типа файлов, которые можно импортировать.
 - ▶ Выберите **Template** (Шаблон) на экране Create Run (Создать запуск), чтобы создать новый макет планшета. Файл шаблона содержит правильные заголовки столбцов для импорта. Введите информацию об образцах в каждый столбец для образцов в данном запуске. Удалите пример информации в неиспользуемых ячейках, а затем сохраните файл.
 - ▶ Файл информации об образце, ранее экспортированном из модуля Соматический вариант с помощью функции экспорта.
- 2 Нажмите значок  **Print** (Печать), чтобы отобразить макет планшета.
- 3 Нажмите **Print** (Печать), чтобы распечатать схему плашки в качестве справочника для подготовки библиотек.
- 4 **[Необязательно]** Нажмите **Export** (Экспорт), чтобы экспортировать информацию об образце во внешний файл.

Проверьте правильность информации в манифестах и информации об образце. Неправильная информация может повлиять на результаты.
- 5 Нажмите **Save Run** (Сохранить запуск).

Редактирование запуска

Инструкции по редактированию информации в запуске перед секвенированием см. в разделе *NextSeq 550Dx Instrument Reference Guide (Справочное руководство к прибору NextSeq 550Dx)* (документ № 100000009513).

Методы анализа

Модуль анализа Соматический вариант выполняет следующие шаги анализа, а затем записывает файлы выходных данных анализа в папку Alignment (Выравнивание)Analysis (Анализ).

- ▶ Прочтения индекса демультиплексов
- ▶ Создает файлы FASTQ
- ▶ Выравнивается относительно референса

- ▶ Идентифицирует варианты

Демультимплексирование

Демультимплексирование позволяет сравнивать последовательность прочтения индекса с индексными последовательностями, указанными для запуска. Параметры качества на данном этапе не учитываются.

Порядок идентификации прочтения индекса указан ниже.

- ▶ Образцам присваивают номера, начиная с 1, в соответствии с порядком, в котором они указаны для запуска.
- ▶ Образец номер 0 зарезервирован для кластеров, не назначенных для образца.
- ▶ Кластеры назначают образцу, если данная индексная последовательность совпадает точно или при наличии не более одного несовпадения на прочтение индекса.

Генерирование файла FASTQ

После демультимплексирования программное обеспечение генерирует промежуточные файлы анализа в текстовом формате FASTQ, используемом для представления последовательностей. Файлы FASTQ содержат прочтения для каждого образца и связанные оценки качества. Кластеры, не прошедшие через фильтр, исключаются.

В каждом файле FASTQ содержатся прочтения только для одного образца, название такого образца включено в имя файла FASTQ. Файлы FASTQ содержат первичные входные данные для выравнивания. На образец на пул олигонуклеотидов (oligo pool) создаются восемь файлов FASTQ, четыре из прочтения 1 и четыре из прочтения 2, в результате чего на образец приходится в общей сложности 16 файлов FASTQ.

Выравнивание

На этапе выравнивания попарный алгоритм Smith-Waterman позволяет выравнивать кластеры из каждого образца по отношению к последовательности ампликона, указанной в файле манифеста.

Попарный алгоритм Smith-Waterman выполняет полуглобальное выравнивание последовательностей для определения подобных участков в двух последовательностях. Вместо сравнения всей последовательности алгоритм Смита – Уотермана сравнивает сегменты всех возможных длин.

Каждое парноконцевое прочтение оценивается с точки зрения его выравнивания относительно соответствующих последовательностей зондов для этого прочтения.

- ▶ Прочтение 1 оценивается относительно обратнoкомплементарного участка нисходящего локус-специфического олигонуклеотида (DLSO, downstream locus-specific oligo).
- ▶ Прочтение 2 оценивается относительно восходящего локус-специфического олигонуклеотида (ULSO, Upstream Locus-Specific Oligo).
- ▶ Если начало прочтения совпадает с последовательностью зонда, а количество отличий при этом не превышает три (несовпадения или смещения в связи с ведущими вставками-делециями), полная длина прочтения выравнивается по отношению к мишени ампликона для данной последовательности.
- ▶ Вставки-делеции в DLSO и ULSO не наблюдаются, учитывая химическую реакцию анализа.

Фильтрацию выравниваний выполняют по результатам выравнивания на основании величин несовпадения на участке, представляющем интерес, или в полном ампликоне – в зависимости от длины ампликона. Отфильтрованные выравнивания записываются в файлах выравнивания как невыровненные и не используются при определении вариантов.

Распознавание вариантов

Разработанное Illumina, ПО для распознавания PISCES идентифицирует варианты, присутствующие с низкой частотой в образце ДНК.

ПО для распознавания вариантов PISCES идентифицирует SNV, MNV и небольшие делеции в три этапа:

- ▶ Рассматривает каждую позицию в референсном геноме отдельно
- ▶ Подсчитывает основания в заданной позиции для совмещенных прочтений, которые перекрывают позицию
- ▶ Вычисляет балл варианта, который измеряет качество распознавания с помощью модели Пуассона. Варианты с оценкой качества ниже Q30 исключаются.

Варианты сначала распознаются для каждой пулы отдельно. Затем варианты из каждой пулы сравниваются и объединяются в один файл выходных данных. Если вариант присутствует в обеих пулах и проходит все фильтры, перечисленные в [Аннотации к файлам VCF на странице 13](#), вариант отмечается как PASS (ПРОЙДЕН) в файле распознавания варианта (VCF).

Просмотр данных о запуске и образце

- 1 На информационной панели Local Run Manager щелкните имя запуска.
- 2 На вкладке Run Overview (Обзор запуска) просмотрите показатели запуска секвенирования.
- 3 **[Необязательно]** Щелкните значок **Copy to Clipboard** (Копировать в буфер обмена) , чтобы скопировать путь папки выходных данных запуска.
- 4 Перейдите на вкладку Sequencing Information (Информация о секвенировании), чтобы просмотреть параметры запуска и информацию о расходных материалах.
- 5 Щелкните вкладку Samples and Results (Образцы и результаты), чтобы просмотреть расположение отчета об анализе.
 - ▶ Если анализ был повторен, разверните раскрывающийся список Select Analysis (Выбрать анализ) и выберите соответствующий анализ.
- 6 Щелкните значок **Copy to Clipboard** (Копировать в буфер обмена) , чтобы скопировать путь к папке Analysis (Анализ).

Для получения дополнительной информации о вкладках Run Overview (Обзор запуска) и Sequencing Information (Информация о секвенировании), а также о том, как перезапустить анализ, см. *NextSeq 550Dx Instrument Reference Guide (Справочное руководство к прибору NextSeq 550Dx) (документ № 1000000009513)*.

Отчеты об анализе

Результаты анализа обобщены на вкладке Samples and Results (Образцы и результаты), а также в виде совокупного отчета в папке Alignment (Выравнивание). Отчет по каждому образцу также доступен в формате PDF для каждого образца.

Информация на вкладке Samples and Results (Образцы и результаты)

1 Щелкните образец в списке, чтобы просмотреть отчет о образце.

Таблица: 1 Информация о запуске и пробе

Заголовки столбцов	Описание
Статус запуска	Указывает, удался ли запуск секвенирования или нет.
Общий выход (GB)	Количество оснований, вызываемых в запуске секвенирования. Показывает порог прохождения, состояние прохождения или сбоя.
% ≥ Q30	Процент прочтений в запуске секвенирования с показателем качества 30 (Q30) или выше. Показывает порог прохождения, состояние прохождения или сбоя.
Название образца	Имя образца, предоставленное при создании запуска.
Общее количество PF-прочтений	Общее количество прочтений, проходящих фильтр.
Прочтение 1 % ≥ Q30	Процент прочтений в Read 1 (Прочтение 1) с показателем качества 30 (Q30) или выше для образца.
Прочтение 2 % ≥ Q30	Процент прочтений в Read 2 (Прочтение 2) с показателем качества 30 (Q30) или выше для образца.
Частота распознаваний Autosome	Количество геномных положений в аутосомах (хромосомах с 1 по 22), которые соответствуют предварительно заданному пороговому значению доверительного интервала, деленное на общее количество исследованных аутосомных геномных положений. Частота распознаваний описывается на основе выборки и регистрируется в процентах, который рассчитывается как 1 минус (количество аутосомных позиций с неполными распознаваниями, деленное на общее количество секвенированных аутосомных позиций).

Таблица: 2 Отчеты об Информация отчета

Заголовки столбцов	Описание
Образец	Имя образца, предоставленное при создании запуска.
Дата отчета	Дата создания отчета.
Информация об образце	ID образца, который был предоставлен при создании запуска, общее количество прочтений, которые прошли фильтр в образце, процент прочтений для образца с баллом качества 30(Q30) или выше и частота аутосомных распознаваний.
Сводка по ампликонам	Общее количество секвенированных областей ампликона и общая длина в парах оснований секвенированных ампликонов в целевых областях для образца в пуле A и пуле B, а также файл манифеста, используемый для каждого пула. В файле манифеста указываются референсный геном и заданные референсные регионы, которые должны использоваться на этапе выравнивания.
Статистика уровня прочтения	Количество и процент прочтений для образца, охватывающего каждую позицию в референсном образце, для прочтения 1 и прочтения 2 в пуле A и пуле B.
Сводка вариантов	Количество SNV, вставок и делеций, обнаруженных для образца, который прошел предложенные значения, чтобы определить, находятся ли результаты качества в приемлемом диапазоне,

Заголовки столбцов	Описание
Сводка по покрытию	Общее количество выровненных оснований, разделенное на целевой размер области, и процент областей ампликона со значениями покрытия, превышающими нижний порог охвата 0,2 * средний охват ампликона, для образца в пуле А и пуле В.
Графики покрытия	График покрытия по области ампликона показывает покрытие по областям ампликона для образца. Области со значениями покрытия ниже порога покрытия выделены красным цветом. Среднее значение всех значений обозначается оранжевой линией. График представлен для охвата пула А и пула В.
Версии программного обеспечения	Версии программного обеспечения при секвенировании образца. Включает Программное обеспечение прибора NextSeq 550Dx Operating Software (NOS), программное обеспечение Local Run Manager, программное обеспечение RTA и версию модуля SomaticGermline Variant.

Файлы выходных данных анализа

Для модуля анализа Соматический вариант создаются следующие файлы выходных данных анализа, которые предоставляют результаты анализа для выравнивания и распознавания вариантов. Файлы выходных данных анализа находятся в папке Alignment (Выравнивание).

Имя файла	Описание
Демультимплексирование (*.txt)	Промежуточные файлы, содержащие результаты демультимплексирования.
FASTQ (*.fastq.gz)	Промежуточные файлы, содержащие определения оснований с оценкой качества. Файлы FASTQ содержат первичные входные данные для этапа выравнивания.
Файлы выравнивания в формате BAM (*.bam)	Содержит выровненные прочтения для данного образца.
Файлы определения вариантов для каждой партии (Per-Pool) в формате VCF (*.vcf)	Содержит варианты, распознанные в каждой позиции из прямого или обратного пула.
Файлы распознаваний вариантов в геномном формате VCF (*.genome.vcf.gz)	Содержит генотип для каждой позиции, распознанный как вариант или референс.
Файлы консенсусных распознаваний вариантов в формате VCF (*.vcf.gz)	Содержит варианты, распознаваемые в каждой позиции из обоих пулов.
AmpliconCoverage_M1.tsv	Содержит информацию о покрытии на ампликон на образец для каждого представленного манифеста. M# представляет собой номер манифеста.

Формат файла демультимплексирования

Процесс демультимплексирования считывает индексную последовательность, прикрепленную к каждому кластеру, чтобы определить, из какого образца был взят кластер. Сопоставление между кластерами и номером образца записывается в файл демультимплексирования (*.demux) для каждой области сканирования проточной ячейки.

Формат наименования файла демультимплексирования – s_1_X.demux, где X – номер области сканирования.

Файлы демультимплексирования начинаются с заголовка:

- ▶ Версия (4-байтовое целое число), в настоящее время 1
- ▶ Количество кластеров (целое 4-байтовое целое число)

Остальная часть файла состоит из номеров образцов для каждого кластера из области сканирования.

После завершения этапа демультимплексирования программное обеспечение генерирует файл демультимплексирования под названием **DemultiplexSummaryF1L1.txt**.

- ▶ В имени файла **F1** представляет номер проточной ячейки.
- ▶ В имени файла **L1** представляет номер дорожки.
- ▶ Демультимплексирование приводит к созданию таблицы с 1 строкой на область сканирования и 1 столбцом на образец, включая образец 0.
- ▶ Наиболее часто встречающиеся последовательности в индексных прочтениях.

формат файла FASTQ

FASTQ – это текстовый формат файла, который содержит распознавание оснований и значения качества на прочтение. Каждая запись содержит 4 строки:

- ▶ Идентификатор
- ▶ Последовательность
- ▶ Знак плюс (+)
- ▶ Баллы качества Phred в закодированном формате ASCII + 33

Идентификатор отформатирован следующим образом:

@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber

Пример:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

Формат файла BAM

Файл BAM (*.bam) – это сжатая двоичная версия файла SAM, которая используется для представления выровненных последовательностей до 128 Mb. Форматы SAM и BAM подробно описаны на сайте samtools.github.io/hts-specs/SAMv1.pdf.

Файлы BAM используют формат наименования файла **SampleName_S#.bam**, где # – номер образца, определенный по порядку, в котором образцы перечислены для анализа.

Файлы BAM содержат раздел заголовка и раздел выравнивания:

- ▶ **Заголовок** – содержит информацию обо всем файле, такую как имя образца, длина образца и метод выравнивания. Выравнивания в разделе Alignments (Выравнивания) связаны с конкретной информацией в разделе заголовка.
- ▶ **Выравнивания** – содержит имя прочтения, последовательность прочтения, качество прочтения, информацию о выравнивании и пользовательские теги. Имя прочтения включает хромосому, начальную координату, качество выравнивания и строку дескриптора совпадения.

В разделе Alignments (Выравнивания) содержится следующая информация для каждой считываемой или считываемой пары:

- ▶ **AS:** Качество парноконцевого выравнивания
- ▶ **BC:** Тег штрих-кода, указывающий демультимплексный идентификатор образца, связанный с прочтением.
- ▶ **SM:** Качество одноконцевого выравнивания.
- ▶ **XC:** Строка дескриптора совпадения
- ▶ **XN:** Тег имени ампликона, который записывает идентификатор ампликона, связанный с прочтением

Файлы индекса BAM (*.bam.bai) обеспечивают индекс соответствующего файла BAM.

Формат файла VCF

Variant Call Format (VCF) – это общий формат файлов, разработанный научным сообществом геномики. В нем содержится информация о вариантах, обнаруженных в определенных позициях в референсном геноме. Файлы VCF заканчиваются суффиксом .vcf

Заголовок файла VCF содержит версию формата файла VCF и версию ПО для распознавания вариантов, а также содержит аннотации, используемые в оставшейся части файла. Заголовок VCF также содержит референсный файл генома и файл BAM. Последняя строка в заголовке содержит заголовки столбцов для линий данных. Каждая из строк данных файла VCF содержит информацию об одном варианте.

Заголовки файлов VCF

Заголовок	Описание
CHROM	Хромосома референсного генома. Хромосомы появляются в том же порядке, что и референсный файл FASTQ.
POS	Однонуклеотидная позиция варианта в референсной хромосоме. Для SNP эта позиция соответствует референсному нуклеотиду, в котором находится вариант; для инделов и делеций или делеций – это референсный нуклеотид, непосредственно предшествующий варианту.
Идентификатор	Номер rs для варианта, полученного из dbSNP.txt, если применимо. Если в этом местоположении несколько rs-чисел, список будет разделен точкой с запятой. Если в этой позиции нет записи dbSNP, используется маркер отсутствующего значения ('.').
REF	Референсный генотип. Например, делеция одного нуклеотида Т представляется как референс ТТ и альтернатива Т. Однонуклеотидная замена А на Т представляется как референс А и альтернатива Т.
ALT	Аллели, отличающиеся от референсного прочтения. Например, вставка одного Т представлена как референс А и альтернатива АТ. Однонуклеотидный вариант замены А на Т представлен как референс А и альтернатива АТ.
QUAL	Оценка качества по шкале Phred, назначенная ПО для распознавания вариантов. Более высокие баллы указывают на более высокую уверенность в варианте и более низкую вероятность ошибок. Для оценки качества Q расчетная вероятность ошибки составляет $10^{-(Q/10)}$. Для набора выявленных вариантов с качеством Q30 вероятность ошибки составляет 0,1 %. Многие ПО для распознавания вариантов присваивают баллы качества на основе своих статистических моделей, при этом эти значения оказываются высокими относительно фактически наблюдаемой частоты ошибок.

Аннотации к файлам VCF

Заголовок	Описание
ФИЛЬТР	<p>Если все фильтры пройдены, в столбце фильтра записывается значение PASS (ПРОЙДЕН).</p> <ul style="list-style-type: none"> • LowDP – применяется к позициям с глубиной покрытия ниже 450x в любом из пулов. Для положений ампликонов, где есть покрытие как прямым, так и обратным прочтением, это эквивалентно 900 перекрывающимся одноконцевым прочтениям. • LowGQ – качество генотипирования (GQ) ниже порогового значения. • q30 – оценка качества < 30. • LowVariantFreq – частота варианта ниже заданного порогового значения. • PB – систематическая ошибка пула зондов. Вариант не найден или обнаружен с низкой частотой в одном или двух пулах зондов. • R3x6 – количество смежных повторов (длиной от 1 до 3 п. о.) для распознаваний вариантов ≥ 6. • SB – смещение по цепи больше заданного порогового значения.
INFO	<p>Возможные записи в столбце INFO:</p> <ul style="list-style-type: none"> • AC – количество аллелей в генотипах для каждого аллеля ALT в том же порядке, в котором они перечислены. • AF – частота аллелей для каждого аллеля ALT в том же порядке, в котором они перечислены. • AN – общее количество аллелей в распознанных генотипах. • CD – флаг, указывающий на то, что SNP находится в пределах кодирующего участка по меньшей мере 1 записи RefGene. • DP – глубина (количество распознаваний оснований, выровненных на позиции и использованных для распознавания вариантов). • Экзон – разделенный запятыми список экзонных областей, считанных из RefGene. • FC – функциональное последствие. • GI – разделенный запятыми список идентификаторов генов, считанных из RefGene. • QD – достоверность/качество варианта по глубине. • TI – разделенный запятыми список идентификаторов транскриптов, считанных из RefGene.
FORMAT	<p>В столбце формата перечислены поля, разделенные двоеточием. Например, GT:GQ. Список полей зависит от используемого ПО для распознавания вариантов. Доступные поля:</p> <ul style="list-style-type: none"> • AD – запись в формате X,Y, где X – число референсных распознаваний, а Y – число альтернативных распознаваний. • DP – приблизительная глубина прочтения; отфильтрованы прочтения с MQ=255 или с некорректными парами прочтений. • GQ – качество генотипа. • GQX – качество генотипа. GQX – это минимальное значение GQ и столбец QUAL. В целом, эти значения схожи; сведение к минимуму делает GQX более консервативным показателем качества генотипа. • GT – генотип. 0 соответствует референсному основанию, 1 соответствует первой записи в столбце ALT и так далее. Косая черта вперед (/) указывает на отсутствие информации о фазировании. • NL – уровень шума; оценка шумов распознавания оснований в этой позиции. • PB – систематическая ошибка пула зондов. Значения, близкие к 0, указывали на большее смещение в отношении одного пула зондов и меньшую уверенность в распознавании варианта. • SB – смещение по цепи в этом положении. Более крупные отрицательные значения указывают на меньшую систематическую ошибку; значения, близкие к 0, указывают на большую систематическую ошибку. • VF – частота варианта; процент прочтений, поддерживающих альтернативный аллель.
ОБРАЗЕЦ	<p>В колонке образца приведены значения, указанные в колонке FORMAT (ФОРМАТ).</p>

Файлы генома VCF

Файлы генома VCF (gVCF) представляют собой файлы VCF v4.1, которые соответствуют набору условных обозначений для представления всех сайтов в геноме в достаточно компактном формате. Файлы gVCF (*.genome.vcf.gz) включают все сайты в пределах области интереса в один файл для каждого образца.

Файл gVCF показывает отсутствие распознавания в положениях, которые не проходят все фильтры. Метка генотипа (GT) ./ указывает на отсутствие распознавания.

Дополнительную информацию см. по ссылке sites.google.com/site/gvcftools/home/about-gvcf.

Файлы VCF каждого пула и консенсуса

Рабочий процесс Соматический вариант создает 2 набора файлов распознавания вариантов.

- ▶ **Файлы VCF для каждого пула** – содержат варианты, распознаваемые в прямом или обратном пуле. Файлы для каждого пула записываются в папку VariantCallingLogs.
- ▶ **Консенсусные файлы VCF** – содержат варианты, распознаваемые из обоих пулов. Файлы консенсуса записываются в папку Выравнивание (Alignment).

Файлы VCF для каждого пула и консенсусные файлы VCF включают файлы VCF (*.vcf) и gVCF (*.genome.vcf) и используют следующий способ наименования, где S# представляет порядок, в котором образец указан для анализа:

- ▶ **Отчеты для всех центров** – SampleName_S#.genome.vcf
- ▶ **Только варианты отчетов** – SampleName_S#.vcf

Программное обеспечение сравнивает файлы VCF для каждого пула и объединяет данные в каждой позиции для создания консенсусного файла VCF для образца.

Распознавания вариантов из различных пулов объединяются в консенсусные файлы VCF с использованием следующих критериев.

Критерии	Результат
Распознавание референса в каждом пуле	Число нераспознаваний референса
Распознавание референса в 1 пуле и распознавание варианта в другом пуле	Отфильтрованное распознавание варианта
Сопоставление распознаваний вариантов с аналогичными частотами в каждом пуле	Распознавание вариантов
Сопоставление распознаваний вариантов со значительно разными частотами в каждом пуле	Отфильтрованное распознавание варианта
Несопоставленные распознавания вариантов в каждом пуле	Отфильтрованное распознавание варианта

Показатели из каждого пула объединяются с использованием следующих значений.

Показатель	Значение
Глубина	Добавление глубины из обоих пулов
Частота варианта	Общее количество вариантов, деленное на общую глубину покрытия
Показатель Q-Score	Минимальное значение обоих пулов

Файл покрытия ампликона

Для каждого файла манифеста создается файл покрытия ампликона. M# в имени файла представляет собой номер манифеста.

Каждый файл содержит строку заголовка, которая содержит идентификаторы образцов, связанных с манифестом. Файл содержит следующую информацию.

- ▶ Идентификатор цели, указанный в манифесте.
- ▶ Глубина охвата прочтений, проходящих фильтр.

Дополнительные файлы выходных данных

Следующие файлы выходных данных предоставляют дополнительную информацию или обобщают результаты запуска и ошибки анализа. Хотя эти файлы не требуются для оценки результатов анализа, их можно использовать для поиска и устранения неисправностей. Все файлы находятся в папке Alignment (Выравнивание), если не указано иное.

Имя файла	Описание
AnalysisLog.txt	Журнал обработки, в котором описывается каждый шаг, который произошел во время анализа текущей папки запуска. Этот файл не содержит сообщений об ошибках. Находится в папке Alignment (Выравнивание).
AnalysisError.txt	Обработка журнала, в котором перечислены все ошибки, возникшие во время анализа. Этот файл будет пустым, если ошибок не произошло. Находится в папке Alignment (Выравнивание).
DemultiplexSummaryF1L1#.txt	Результаты демultipлексирования отображаются в таблице, где каждая ячейка соответствует 1 строке на область сканирования и 1 столбцу на образец. # представляет дорожку 1, 2, 3 или 4 проточной ячейки. Находится в папке Alignment (Выравнивание).
AmpliconRunStatistics.xml	Содержит сводную статистику, относящуюся к запуску. Находится в папке Alignment (Выравнивание).

Папка Analysis (Анализ)

В папке анализа хранятся файлы, созданные программным обеспечением Local Run Manager.

Взаимосвязь между папкой выходных данных и папкой анализа обобщена следующим образом:


- ▶ Во время секвенирования Анализ в режиме реального времени (Real-Time Analysis) (RTA) заполняет папку файлами выходных данных, созданными во время анализа изображений, распознавания оснований и оценки качества.
- ▶ RTA копирует файлы в папку анализа в режиме реального времени. После того как RTA присваивает оценку качества каждой базе для каждого цикла, программное обеспечение записывает файл RTAComplete.txt в обе папки.
- ▶ При наличии файла RTAComplete.txt начинается анализ.

- ▶ По мере продолжения анализа Local Run Manager записывает файлы выходных данных в папку анализа, а затем копирует файлы обратно в папку выходных данных.

Папки выравнивания

Каждый раз, когда анализ перезапускается, Local Run Manager создает папку Alignment (Выравнивание) под названием **Alignment_N**, где N – это порядковый номер.

Структура папок

 **Выравнивание** – содержит файлы *.bam, *.vcf, FASTQ и файлы, относящиеся к модулю анализа.

 **Метка даты и времени** – метка даты_времени анализа как ГГГГММДД_ЧЧММСС

 AnalysisError.txt

 AnalysisLog.txt

 aggregate.report.html

 aggregate.report.pdf

 aggregate.summary.csv

 AmpliconCoverage_M#.tsv

 AmpliconRunStatistics.xml

 Sample1.genome.vcf.gz

 Sample1.coverage.csv

 Sample1.report.pdf

 Sample1.summary.csv

 Sample1.vcf.gz

 Sample1.bam

 **FASTQ**

 **Sample1**

 Образец1_L001_R1_001_fastq.gz

 **Stats**

 DemuxSummaryF1L1.txt

 FastqSummaryF1L1.txt

 **Data**


 **Intensities**


 **BaseCalls**

 **L001** – содержит файлы *.bcl.

 **L001** – содержит файлы *.locs.

 **Журналы RTA** – содержит файлы журналов из анализа программного обеспечения RTA.

 **InterOp** – содержит двоичные файлы, используемые для создания отчетов о показателях запуска секвенирования.

 **Журналы** – содержит файлы журналов, описывающие шаги, выполняемые во время секвенирования.

 RTAComplete.txt

 RunInfo.xml

 RunParameters.xml

Определение последовательности нуклеотидов и различие индексов

При секвенировании образцов с помощью прибора NextSeq 550Dx определение последовательности нуклеотидов позволяет установить основание (А, С, G или Т) для каждого кластера указанной области сканирования или зоны визуализации на проточной ячейке в определенном цикле. В приборе NextSeq 550Dx используется двухканальное секвенирование, при котором нужны только два изображения для кодирования всех четырех оснований ДНК: одно изображение из красного канала и одно из зеленого.

Процесс распознавания оснований в ходе прочтения индекса отличается от распознавания оснований в ходе других прочтений.

Прочтения индекса должны начинаться в случае, когда в первых двух циклах обнаружено хотя бы одно основание, отличное от G. Если прочтение индекса начинается с двух оснований, распознанных как G, интенсивность сигнала не генерируется. Чтобы обеспечить выполнение демультимплексирования, сигнал должен присутствовать хотя бы в одном из двух первых циклов.

При выборе индексов в ходе создания запуска отобразится предупреждение о низком различии, если индексы не соответствуют требованиям к различию. Чтобы избежать отображения предупреждения о низком различии, выберите индексные последовательности, которые обеспечивают сигнал в обоих каналах для каждого цикла.

- ▶ Красный канал – А или С
- ▶ Зеленый канал – А или Т

Такой процесс распознавания оснований обеспечивает точность при анализе малоплексных образцов. Чтобы узнать подробнее о последовательностях индексов вашего анализа, см. листок-вкладыш *Комплект TruSeq Custom Amplicon Kit Dx (документ No 1000000029772)*.

Во время создания запуска с помощью программного обеспечения Local Run Manager пользователь сможет выбрать количество образцов для анализа. Данное программное обеспечение автоматически заполняет предлагаемые сочетания индексов, соответствующие требованиям к различию индексов. Предлагаемые сочетания индексов являются не обязательными, а рекомендованными.

История изменений

Документ	Дата	Описание изменений
Документ № 1000000030330, версия 04	Август 2021 г.	Обновлен адрес уполномоченного представителя в ЕС.
Документ № 1000000030330, версия 03	Декабрь 2019 г.	Обновлен адрес уполномоченного представителя в ЕС. Обновлен адрес спонсора в Австралии.
Документ № 1000000030330, версия 02	Январь 2019 г.	Добавлена информация о наборах реагентов версии 2.5.
Документ № 1000000030330, версия 01	Август 2018 г.	Обновлена нормативно-правовая маркировка.
Документ № 1000000030330, версия 00	Ноябрь 2017 г.	Первый выпуск.

Техническая помощь

Для получения технической помощи свяжитесь со службой технической поддержки Illumina.

Веб-сайт: www.illumina.com
Электронная почта: techsupport@illumina.com

Номера телефонов службы поддержки клиентов Illumina

Регион	Бесплатный звонок	Региональные отделения
Северная Америка	+1 800 809 4566	
Австралия	+1 800 775 688	
Австрия	+43 800006249	+43 19286540
Бельгия	+32 80077160	+32 34002973
Китай	400 066 5835	
Дания	+45 80820183	+45 89871156
Финляндия	+358 800918363	+358 974790110
Франция	+33 805102193	+33 170770446
Германия	+49 8001014940	+49 8938035677
Гонконг	800960230	
Ирландия	+353 1800936608	+353 016950506
Италия	+39 800985513	+39 236003759
Япония	0800 111 5011	
Нидерланды	+31 8000222493	+31 207132960
Новая Зеландия	0800 451 650	
Норвегия	+47 800 16836	+47 21939693
Сингапур	+1 800 579 2745	
Испания	+34 911899417	+34 800300143
Швеция	+46 850619671	+46 200883979
Швейцария	+41 565800000	+41 800200442
Тайвань	00806651752	
Великобритания	+44 8000126019	+44 2073057197
Другие страны	+44 1799 534000	

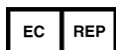
Паспорта безопасности веществ (SDS) – см. на веб-сайте компании Illumina по адресу support.illumina.com/sds.html.

Документация на изделие – можно загрузить с веб-сайта компании Illumina в формате PDF. Перейдите на сайт support.illumina.com, выберите продукт, затем выберите **Documentation & Literature** (Документация и литература).



Illumina
5200 Illumina Way
San Diego, California 92122
США
+1-800-809-ILMN (4566)
+1-858-202-4566 (за пределами
Северной Америки)
techsupport@illumina.com
www.illumina.com

CE



Illumina Netherlands B. V.
Steenoven 19
5626 DK Eindhoven
The Netherlands (Нидерланды)

Спонсор в Австралии
Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
Melbourne, VIC 3000
Австралия

ДЛЯ ДИАГНОСТИКИ IN VITRO

© 2021 г. Illumina, Inc. Все права защищены.

illumina®