

# Модул за анализ на соматичен вариант Local Run Manager

Ръководство за работния процес за NextSeq 550Dx

## ЗА ИН ВИТРО ДИАГНОСТИЧНА УПОТРЕБА

Общ преглед	3
Въведете информация за изпълняването	4
Методи за анализ	6
Преглед на данни за изпълняването и пробите	8
Analysis Report (Отчет от анализ)	8
Изходни файлове на анализ	10
Обозначаване на бази и разнообразие на индекси	17
Хронология на редакциите	18
Техническа помощ	19



Настоящият документ и съдържанието му са собственост на Illumina, Inc. и нейните филиали („Illumina“) и са предназначени само за употреба по силата на договор от страна на клиента в във връзка с използването на продукта(ите), описан(и) в настоящия документ, и с никаква друга цел. Този документ и съдържанието му не трябва да се използват или разпространяват за никаква друга цел и/или по друг начин да бъдат съобщавани, разкривани или възпроизведени по какъвто и да е начин без предварителното писмено съгласие от страна на Illumina. Illumina не предоставя посредством този документ никакъв лиценз за свой патент, търговска марка, авторско право или права по силата на общото право, нито подобни права на която и да е трета страна.

Инструкциите в този документ трябва да се следват строго и изрично от страна на квалифициран и правилно обучен персонал, за да се гарантират правилната и безопасната употреба на продукта(ите), описан(и) в настоящия документ. Цялото съдържание на този документ трябва да бъде прочетено и разбрано напълно, преди да се използва(т) такъв(таква) продукт(и).

**АКО ВСИЧКИ ИНСТРУКЦИИ, СЪДЪРЖАЩИ СЕ В НАСТОЯЩИЯ ДОКУМЕНТ, НЕ БЪДАТ НАПЪЛНО ПРОЧЕТИ И ИЗРИЧНО СПАЗВАНИ, ТОВА МОЖЕ ДА ДОВЕДЕ ДО ПОВРЕДА НА ПРОДУКТ(ИТЕ), НАРАНЯВАНЕ НА ЛИЦАТА, ВКЛЮЧИТЕЛНО НА ПОТРЕБИТЕЛИТЕ ИЛИ ДРУГИ ЛИЦА, И УВРЕЖДАНЕ НА ДРУГО ИМУЩЕСТВО, И ЩЕ ОТМЕНИ ВСЯКАКВА ГАРАНЦИЯ, ПРИЛОЖИМА ЗА ПРОДУКТ(ИТЕ).**

**ILLUMINA НЕ ПОЕМА НИКАКВА ОТГОВОРНОСТ В РЕЗУЛТАТ НА НЕПРАВИЛНАТА УПОТРЕБА НА ПРОДУКТА (ИТЕ), ОПИСАН(И) В НАСТОЯЩИЯ ДОКУМЕНТ (ВКЛЮЧИТЕЛНО ТЕХНИ ЧАСТИ ИЛИ СОФТУЕР).**

© 2021 Illumina, Inc. Всички права запазени.

Всички търговски марки са собственост на Illumina, Inc. или съответните им притежатели. За специфична информация относно търговските марки посетете [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

## Общ преглед

Модулът за анализ на соматичен вариант Local Run Manager е за употреба с анализа Комплект за персонализирани ампликони TruSeq и NextSeq 550Dx на Illumina. Когато се използва с модула за соматичен вариант, анализът е предназначен за приготвяне на библиотеки, които се използват за секвенирането на ДНК от фиксирана с формалин и вградена в парафин (FFPE) тъкан. Анализът открива соматични мутации при нисковариантни честоти.

Модулът за анализ оценява кратки области на амплифицирана ДНК или ампликони за варианти. Фокусираното секвениране на ампликоните позволява високо покритие на определени региони в голям брой проби. Модулът за анализ изпълнява вторичен анализ и генерира отчет от изпълняванията за секвениране чрез двойноверижан подход, който включва предни и обратни обединявания на олигонуклеотиди. Вижте листовката за *Комплект за персонализирани ампликони TruSeq Dx (документ № 1000000029772)*.

Модулът за анализ на соматичен вариант изисква консумативи за секвениране за 300 цикъла. За повече информация вижте листовката на *Комплект с реагенти с висок изходен поток NextSeq 550Dx v2* или *Комплект с реагенти с висок изходен поток NextSeq 550Dx v2.5*.

## За това ръководство

Това ръководство предоставя инструкции за настройка на параметри на изпълняване на секвениране и анализ за модула за анализ за соматичен вариант. За информация относно таблото за управление и системните настройки на Local Run Manager вижте *справочното ръководство за инструмента NextSeq 550Dx (документ №1000000009513)*.

## Преглед на Local Run Manager

Интерфейсът на Local Run Manager се преглежда чрез работния софтуер NextSeq 550Dx (NOS) или през уеб браузър. Уеб браузърът, който се поддържа, е Chromium.



### ЗАБЕЛЕЖКА

Ако използвате неподдържан браузър, изтеглете поддържащия браузър, когато бъдете подканени в съобщението „Confirm Unsupported Browser“ („Потвърждаване на неподдържан браузър“). Изберете „**here**“ (тук) за изтегляне на поддържана версия на Chromium.

## Преглеждане на монитора на инструмента

- 1 За да прегледате интерфейса на Local Run Manager на монитора на инструмента, изберете една от следните възможности:
  - ▶ От началния екран на NOS изберете **Local Run Manager**.  
Щракнете върху X в горния десен ъгъл, за да се върнете към NOS, когато сте готови.
  - ▶ Изберете иконата Minimize NOS (Минимизиране на NOS), отворете уеб браузъра Chromium, който е на инструмента, и въведете **http://localhost** в лентата за адреса.  
Само потребителите, които са администратори, могат да минимизират NOS.

## Преглеждане на мрежови компютър

- 1 Отворете уеб браузъра Chromium на компютър с достъп до същата мрежа като инструмента и се свържете с помощта на IP адреса на инструмента или името на инструмента. Например **http://myinstrument**.

## Въведете информация за изпълняването

### Задаване на параметри

- 1 Влезте в Local Run Manager.
- 2 Изберете **Create Run** (Създаване на изпълняване) и изберете **Somatic Variant** (Соматичен вариант).
- 3 Въведете име на изпълняването, което идентифицира изпълняването от секвенирането до анализа.  
Използвайте буквено-цифрови знаци, интервали, долни черти или тирета.
- 4 **[Незадължително]** Въведете описание на изпълняването, за да подпомогнете идентифицирането на изпълняването.  
Използвайте буквено-цифрови знаци, интервали, долни черти или тирета.
- 5 Изберете броя на проби и индексния комплект от падащия списък.  
Имайте предвид следната информация, когато правите избор.
  - ▶ Падащият списък съдържа броя проби с индексен комплект. Например 24-Set 1 (24 – комплект 1) показва, че ще бъдат изследвани 24 проби с индекси от индексен комплект 1.
  - ▶ Номерата на индексните комплекти се отнасят до различните комплекти от i5 индекси. Като комплект 1, така и комплект 2 осигуряват индексно разнообразие. Двата индексни комплекта се предлагат, за да подпомогнат предотвратяването на изчерпването на единичен комплект.
  - ▶ Изберете броя проби, които са най-близо до броя проби, които изследвате. Ако точният брой на проби не е в списъка, изберете броя, който е най-близо, но не по-малък от броя, който изследвате. Например, ако искате да изследвате 18 проби, изберете 16 проби.
  - ▶ Кладенчетата за проби и индексните комбинации, които отговарят на изискванията за индексно разнообразие, се осветяват в зелено. Ако изберете други кладенчета и индексни комбинации, когато записвате изпълняването, ще получите известие, че изискванията за индексно разнообразие не са спазени.

### Импортиране на манифестни файлове за изпълняването

- 1 Уверете се, че манифестите, които искате да импортирате, са налични на достъпно мрежово местоположение или на USB устройство.
- 2 Изберете **Import Manifests** (Импортиране на манифести).
- 3 Навигирайте до манифестния файл и изберете манифестите, които искате да добавите.



#### **ЗАБЕЛЕЖКА**

За да направите манифестни файлове налични за всички изпълнявания чрез използване на модула за анализ на соматичен вариант, добавете манифестите чрез функцията Module Settings (Настройки на модул). Тази функция изисква разрешения на администраторско ниво. За повече информация вижте *справочното ръководство за инструмента NextSeq 550Dx (документ № 1000000009513)*.


## Посочване на проби за изпълняването

Посочете пробите за изпълняването чрез една от следните опции и указания, които следват.

- ▶ **Enter samples manually** (Ръчно въвеждане на проби) – Използвайте празната таблица на екрана Create Run (Създаване на изпълняване).
- ▶ **Import samples** (Импортиране на проби) – Посочете път до външен файл във формат със стойности, разделени със запетая (\*.csv). На екрана Create Run (Създаване на изпълняване) се предлага шаблон за изтегляне.

След като сте попълнили таблицата за пробите, можете да експортирате информацията за пробите във външен файл. Използвайте файла за справка при приготвяне на библиотеки или импортиране на файла за друго изпълняване.

## Enter Samples Manually (Ръчно въвеждане на проби)

- 1 Въведете уникално име на пробата в полето Sample Name (Име на пробата). Използвайте буквено-цифрови знаци, тирета или долни черти. Името на пробата се попълва автоматично със съответното кладенче в другото обединяване.
- 2 **[Незадължително]** За положителни или негативни контролни проби кликнете с десния бутон на мишката и изберете типа на контролата. Контролата в едно кладенче за проби попълва автоматично съответното кладенче в другото обединяване със същата контрола.
- 3 **[Незадължително]** Въведете описание на пробата в раздела Sample Description (Описание на пробата). Използвайте буквено-цифрови знаци, тирета или долни черти. Описанието на пробата се попълва автоматично със съответното кладенче в другото обединяване. Описанията на пробите са свързани с идентификатор на проба. Описанията на пробите се презаписват, ако същият идентификатор на пробата бъде използван отново в по-късно изпълняване.
- 4 Изберете Index 1 adapter (Индекс 1 адаптер) от падащия списък Index 1 (i7) (Индекс 1 (i7)). Когато използвате предложените кладенчета за проби, софтуерът попълва автоматично индексни адаптери i7 и i5, които отговарят на изискванията на индекса за разнообразие. Ако точният брой проби, които изследвате, не е в списъка, постарайте се да изберете индексни адаптери за допълнителни кладенчета. Ако трябва да изберете индекси за допълнителни кладенчета или не използвате препоръчаните комбинации за индексни адаптери, преди да изберете индекси, се уверете, че сте прочели *Обозначаване на бази и разнообразие на индекси на страница 17*.
- 5 Изберете Index 2 adapter (Индекс 2 адаптер) от падащия списък Index 2 (i5) (Индекс 2 (i5)).
- 6 Изберете манифестен файл от падащия списък Manifest (Манифест). Пробите от Pool A (Обозначаване A) изискват различен манифест от пробите в Pool B (Обозначаване B).
- 7 Изберете опция за преглед, отпечатване или записване на подредбата на плаките като справка за приготвяне на библиотеки:
  - ▶ Изберете иконата  **Print** (Печат) за показване на подредбата на плаката. Изберете **Print** (Печат) за отпечатване на подредбата на плаката.

- ▶ Изберете **Export** (Експортиране) за експортиране на информация за проба към външен файл.

Уверете се, че манифестът и информацията за пробата са правилни. Неправилната информация може да повлияе на резултатите.

- 8 Изберете **Save Run** (Записване на изпълняване).

## Импортиране на проби

- 1 Изберете **Import Samples** (Импортиране на проби) и преминете към местоположението на файла с информация за пробата. Има два типа файлове, които може да импортирате.
  - ▶ Изберете **Template** (Шаблон) от екрана Create Run (Създаване на изпълняване), за да направите нова подредба на плаките. Файлът с шаблона съдържа правилните заглавия на колони за импортиране. Добавете информация за пробата във всяка колона за всяка проба в изпълняването. Изтрийте примерната информация в неизползвани клетки и след това запишете файла.
  - ▶ Използвайте файл с информация за проба, който е бил експортиран от модула за соматичен вариант чрез функцията Export (Експортиране).
- 2 Изберете иконата  **Print** (Печат) за показване на подредбата на плаката.
- 3 Изберете **Print** (Печат) за отпечатване на подредбата на плаката като справочна информация за приготвяне на библиотеки.
- 4 **[Незадължително]** Изберете **Export** (Експортиране) за експортиране на информация за проба към външен файл.

Уверете се, че манифестът и информацията за пробата са правилни. Неправилната информация може да повлияе на резултатите.
- 5 Изберете **Save Run** (Записване на изпълняване).

## Редактиране на изпълняване

За инструкции за редактиране на информацията във вашето изпълняване преди секвениране вижте *справочното ръководство за инструмента NextSeq 550Dx (документ № 1000000009513)*.

## Методи за анализ

Модулът на анализ на соматичен вариант изпълнява следните стъпки за анализ и след това записва изходните файлове на анализа в папката Alignment (Подравняване).

- ▶ Демултиплексиране на индексни разчитания
- ▶ Генериране на FASTQ файлове
- ▶ Подравняване към препратка
- ▶ Идентифициране на варианти

## Демултиплексиране

Демултиплексирането сравнява всяка последователност от индексни разчитания с индексните последователности, специфични за изпълняването. В тази стъпка не се вземат предвид никакви стойности на качеството.

Индексните разчитания се идентифицират чрез следните стъпки:

- ▶ Пробите се номерират, започвайки от 1, в зависимост от последователността, в която са изброени за изпълняването.
- ▶ Номер на проба 0 е запазен за клъстери, които не са били зададени в проба.
- ▶ Клъстерите се задават към проба, когато индексната последователност съвпада точно или когато има единично несъвпадение на индексно разчитане.

## Генериране на FASTQ файл

След демултиплексиране софтуерът генерира междинни файлове за анализ във формат FASTQ, който е текстов формат, използван за представяне на последователности. FASTQ файловете съдържат разчитания за всяка проба и свързаните с тях резултати за качество. Клъстерите, които не са преминали филтър, са изключени.

Всеки FASTQ файл съдържа разчитания само за една проба и името на тази проба е включено в името на FASTQ файла. FASTQ файловете са основната входна информация за подравняване. Осем FASTQ файлове се генерират на проба на обединяване на олигонуклеотиди, четири от Read 1 (Разчитане 1) и четири от Read 2 (Разчитане 2), което води до общо 16 FASTQ файла на проба.

## Подравняване

По време на стъпката на подравняване лентовият алгоритъм Smith-Waterman подравнява клъстерите от всяка проба спрямо последователности на ампликони, посочени в манифестните файлове.

Лентовият алгоритъм Smith-Waterman изпълнява полуобстойни подравнявания на последователности за определяне на подобни региони между две последователности. Вместо да сравнява цялата последователност, алгоритъмът Smith-Waterman сравнява сегменти с всички възможни дължини.

Всяко разчитане на сдвоени краища се оценява по отношение на неговото подравняване към съответните последователности на пробата за това разчитане.

- ▶ Read 1 (Разчитане 1) се оценява чрез обратно допълнение на локус-специфични олигонуклеотиди след гена (DLSO).
- ▶ Read 2 (Разчитане 2) се оценява чрез локус-специфични олигонуклеотиди преди гена (ULSO).
- ▶ Ако началото на разчитането съвпада с последователността на пробата с повече от три разлики (погрешно сдвояване или измествания поради водещи инсерции и делеции), пълната дължина на разчитането се подравнява спрямо целта на ампликона за тази последователност.
- ▶ Инсерциите и делециите в рамките на DLSO и ULSO не се наблюдават, като се вземе предвид аналитичната химия.

Подравняванията се филтрират от резултатите от подравняването въз основа на проценти на погрешно сдвояване или в региона на интерес, или в пълния ампликон в зависимост от дължината на ампликона. Филтрираните подравнявания се записват във файловете за подравняване като неподравнени и не се използват при обозначаване на варианти.

## Обозначаване на вариант



Разработеният от Illumina инструмент за обозначаване на вариант PISCES идентифицира наличните варианти при ниска честота в ДНК пробата.

Инструментът за обозначаване на вариант PISCES идентифицира SNV, MNV и малки инсерции и делеции в три стъпки:

- ▶ Разглежда всяка позиция в референтния геном поотделно
- ▶ Преброява бази в дадената позиция за подравнени разчитания, които припокриват позицията
- ▶ Изчислява резултат за варианта, който измерва качеството на обозначаването, използвайки модела на Poisson. Варианти с резултат за качество под Q30 не се включват.

Вариантите се обозначават първо за всяко обединяване поотделно. След това вариантите от всяко обединяване се сравняват и комбинират в единичен изходен файл. Ако вариант е наличен и в двете обединявания и преминава всички филтри, посочени в [Анотации на VCF файл на страница 13](#), вариантът се отбелязва като PASS (ПРЕМИНАВА) във файла за обозначаване на вариант (VCF).

## Преглед на данни за изпълняването и пробите

- 1 От таблото за управление на Local Run Manager щракнете върху името на изпълняването.
- 2 От раздела Run Overview (Общ преглед на изпълняването) прегледайте показателите за изпълняване на секвенирането.
- 3 **[Незадължително]** Изберете иконата **Copy to Clipboard**  (Копиране в клипборда) за копиране на пътя до папката за изходно изпълняване.
- 4 Щракнете върху раздела Sequencing Information (Информация за секвениране), за да прегледате параметрите на изпълняването и информацията за консумативите.
- 5 Щракнете върху раздела Samples and Results (Проби и резултати), за да видите местоположението на отчета за анализа.
  - ▶ Ако анализът е бил повторен, разгънете падащия списък Select Analysis (Избор на анализ) и изберете подходящия анализ.
- 6 Щракнете върху иконата **Copy to Clipboard**  (Копиране в клипборда) за копиране на пътя до папката Analysis (Анализ).

За повече информация за разделите Run Overview (Общ преглед на изпълняване) и Sequencing Information (Информация за секвениране) и как да поставите анализи на изчакване на опашка вижте *Справочното ръководство за инструмента NextSeq 550Dx (документ №1000000009513)*.

## Analysis Report (Отчет от анализ)

Резултатите от анализа са обобщение в раздела Samples and Results (Проби и резултати) и като обобщен отчет в папката Alignment (Подравняване). Отчет за всяка проба е наличен и във файлов формат PDF за всяка проба.



## Информация за раздели Samples and Results (Проби и резултати)

1 Цракнете върху проба в списъка, за да видите отчета за пробата.

Таблица 1 Информация за изпълняване и проба

Заглавие на колона	Описание
Run Status (Състояние на изпълняване)	Посочва дали изпълняването на секвениране е преминало, или е неуспешно.
Total Yield (GB) (Обща производителност)	Броят бази, които са обозначени в изпълняването на секвениране. Показва прага за преминаване и състояние на преминаване или неуспех.
% ≥ Q30	Процентът на разчитанията в изпълняването на секвениране с резултат за качество 30 (Q30) или по-голям. Показва прага за преминаване и състояние на преминаване или неуспех.
Sample Name (Име на пробата)	Името на пробата, предоставено при създаването на изпълняването.
Total PF Reads (Общо разчитания PF)	Общия брой на разчитания, преминали филтъра.
Read 1% ≥ Q30 (Разчитане 1% ≥ Q30)	Процентът на разчитанията в Разчитане 1 с резултат за качество 30 (Q30) или по-голям за пробата.
Read 2% ≥ Q30 (Разчитане 2% ≥ Q30)	Процентът на разчитанията в Разчитане 2 с резултат за качество 30 (Q30) или по-голям за пробата.
Autosome Call Rate (Честота на обозначаване на автозома)	Броят на геномните позиции в автозомите (хромозоми от 1 до 22), които отговарят на предварително дефиниран праг на стойност на доверителност, разделен на общия брой на автозомни геномни позиции, за които е получена информация. Честотата на обозначаването се описва на основата на проба и се отчита като процент, който се изчислява като 1 минус (брой варианти на автозомни позиции с непълни обозначавания, разделен на общия брой секвенирани автозомни позиции).

Таблица 2 Sample Report Information (Информация за отчет за пробата)

Заглавие на колона	Описание
Sample (Проба)	Името на пробата, предоставено при създаването на изпълняването.
Report Date (Дата на отчета)	Датата, на която отчетът е създаден.
Sample Information (Информация за проба)	Идентификаторът на пробата, който е бил предоставен, когато изпълняването е създадено, общият брой разчитания, които са преминали филтъра в пробата, процентът на разчитания за пробата с резултат за качество 30 (Q30) или по-голям и честотата на автозомно обозначаване.
Amplicon Summary (Обобщение на ампликон)	Общият брой на секвенирани региони на ампликони и общата дължина в двойки бази от секвенирани ампликони в целевите региони, за пробата в Pool A (Обединяване A) и Pool B (Обединяване B) и манифестният файл, използван за всяко обединяване. Манифестният файл определя референтния геном и целевите референтни региони, използвани в стъпката на подравняване.
Read Level Statistics (Статистики за нивото на разчитания)	Броят и процентът на разчитания за пробата, които покриват всяка позиция в справката за Read 1 (Разчитане 1) и Read 2 (Разчитане 2) в Pool A (Обединяване A) и Pool B (Обединяване B).
Variants Summary (Обобщение на вариантите)	Броят на SNV, инсерции и делеции, открити в пробата, които са преминали предложените стойности за определяне дали резултатите за качество са в рамките на приемливата граница.

Заглавие на колона	Описание
Coverage Summary (Обобщение на покриването)	Общият брой на подравнени бази, разделен по размера на целевия регион, и процентът на региони на ампликони с покритие на стойностите, по-големи от прага за ниско покриване от 0,2 * средно покриване на ампликон, за пробата в Pool A (Обединяване A) и Pool B (Обединяване B).
Coverage Plots (Графики за покриване)	Графиката за покриване по регион на ампликон показва покриването сред регионите на ампликони за пробата. Регионите със стойности на покриване, по-ниски от прага на покриване, са маркирани в червено. Средната стойност на всички стойности е обозначена с оранжева линия. Осигурена е графика за покриването на Pool A (Обединяване A) и Pool B (Обединяване B).
Software Versions (Версии на софтуера)	Версиите на софтуера, когато пробата е секвенирана. Включва работен софтуер на NextSeq 550Dx (NOS), софтуер Local Run Manager, софтуер RTA и версията на модула за соматичен вариант.

## Изходни файлове на анализ

Следните изходни файлове на анализ се генерират за модула за анализ на соматичен вариант и осигуряват резултати от анализа за подравняване и обозначаване на вариант. Изходните файлове на анализ се намират в папката Alignment (Подравняване).

Име на файла	Описание
Демултиплексиране (*.txt)	Междинни файлове, които съдържат обобщителни резултати за демултиплексирането.
FASTQ (*.fastq.gz)	Междинни файлове, които съдържат обозначавания на бази с резултат за качество. FASTQ файловете са основната входна информация за стъпката на подравняване.
Файловете за подравняване в BAM формат (*.bam)	Съдържа подравнени разчитания за дадена проба.
Файлове за обозначаване на вариант по обединяване във VCF формат (*.vcf)	Съдържа обозначените варианти за всяка позиция за предно обединяване или обратно обединяване.
Файлове за обозначаване на вариант в генома в VCF формат (*.genome.vcf.gz)	Съдържа генотипа за всяка позиция, независимо дали е обозначена като вариант, или справка.
Файлове за обозначаване на консенсусен вариант във VCF формат (*.vcf.gz)	Съдържа варианти, обозначени за всяка позиция от двете обединявания.
AmpliconCoverage_M1.tsv	Съдържа информация относно покриването на ампликон за проба за всеки предоставен манифест. M# представлява манифестния номер.

## Файлов формат за демултиплексиране

Процесът на демултиплексиране разчита индексната последователност, прикрепена към всеки клъстер, за да се определи от коя проба произхожда клъстерът. Съпоставянето между клъстери и номер на проба се записва във файл на демултиплексиране (\*.demux) за всяка плочка на поточната клетка.

Форматът за името на файла за демултиплексиране е **s\_1\_X.demux**, където X е номерът на плочката.

Файловете за демултиплексиране започват с горен колонтитул:

- Версия (4 байта целочислени типове данни), в момента 1

- ▶ Брой клъстери (4 байта целочислени типове данни)

Останалата част от файла се състои от номера на проби за всеки клъстер от плочката.

Когато стъпката за демултиплексиране е завършена, софтуерът ще създаде файл на демултиплексиране, наименуван **DemultiplexSummaryF1L1.txt**.

- ▶ **F1** представлява номера на поточната клетка в името на файла.
- ▶ **L1** представлява номера на линията в името на файла.
- ▶ Демултиплексирането завършва с таблица с 1 ред на плочка и 1 колона на проба, включително проба 0.
- ▶ Най-често срещаните последователности в индексни разчитания.

## Файлов формат FASTQ

FASTQ е текстово-базиран файлов формат, който съдържа обозначавания на бази и стойности на качеството на едно разчитане. Всеки запис съдържа 4 линии:

- ▶ Идентификаторът
- ▶ Последователността
- ▶ Знак плюс (+)
- ▶ Резултатът за качество Phred се оценява в кодиран формат ASCII + 33

Идентификаторът е представен във формат:

**@Инструмент:Идентификатор на изпълняването:Идентификатор на поточна клетка:Линия:Плочка:X:Y Номер разчитане:Флаг за филтър:0:Номер на проба**

Пример:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

## Файлов формат BAM

BAM файлът (\*.bam) е компресирана двоична версия на SAM файл, който се използва за представяне на подравнени последователности до 128 Mb. Форматите SAM и BAM са описани подробно в [samtools.github.io/hts-specs/SAMv1.pdf](https://samtools.github.io/hts-specs/SAMv1.pdf).

BAM файловете използват формат за именуване на файла **SampleName\_S#.bam**, където # е номерът на пробата, определен от последователността, в която пробите са посочени за изпълняването.

BAM файловете съдържат раздел за горен колонтитул и раздел за подравняване:

- ▶ **Горен колонтитул** – Съдържа информация за целия файл, като име на пробата, дължина на пробата и метод на подравняване. Подравняванията в раздела за подравнявания са свързани със специфична информация в раздела за горен колонтитул.
- ▶ **Подравнявания** – Съдържа име на разчитането, последователност на разчитането, качество на разчитането, информация за подравняване и персонализирани тагове. Името на разчитането включва хромозомата, началната координата, качеството на подравняване и низа на дескриптора за съвпадение.

Разделът за подравнявания включва следната информация за всяко разчитане или двойка разчитания:

- ▶ **AS:** Качество на подравняване на сдвоени краища
- ▶ **BC:** Баркод таг, който показва идентификатора на демултиплексираната проба, свързан с разчитането.
- ▶ **SM:** Качество на подравняване на единични краища.
- ▶ **XC:** Низ на дескриптора за съвпадение
- ▶ **XN:** Таг за името на ампликон, който записва идентификатора на ампликона, свързан с разчитането

Индексните файлове BAM (\*.bam.bai) предоставят индекс на съответния BAM файл.

## Файлов формат VCF

Формат за обозначаване на вариант (VCF) е често срещан файлов формат, разработен от научната общност на геномиката. Той съдържа информация за варианти, открити на конкретни позиции в референтен геном. VCF файловете завършват с наставка .vcf.

Горният колонтитул на VCF файла включва версията на файловия формат VCF и версията на инструмента за обозначаване на вариант и посочва анотациите, използвани в останалата част на файла. Горният колонтитул на VCF също включва файла на референтния геном и BAM файла. Последният ред в горния колонтитул съдържа заглавията на колоните за редовете с данни. Всеки от редовете за данни на VCF файла съдържа информация за един вариант.

## Заглавия на VCF файл

Заглавие	Описание
<b>CHROM</b>	Хромозомата от референтния геном. Хромозомите се появяват в същата последователност както в референтния FASTQ файл.
<b>POS</b>	Еднобазова позиция на варианта в референтната хромозома. За SNP тази позиция е референтната база с варианта; за инсерции и делеции или делеции тази позиция е референтната база веднага преди варианта.
<b>ID</b>	RS номерът за варианта, получен от dbSNP.txt, ако е приложимо. Ако има няколко rs номера на това местоположение, списъкът е разделен с двоеточие. Ако не съществува dbSNP запис в тази позиция, се използва маркер за липсваща стойност ('.').
<b>REF</b>	Референтният генотип. Например делеция на единична T е представено като референтна TT и алтернативна T. А към T единичен нуклеотиден вариант е представен като референтна A и алтернативна T.
<b>ALT</b>	Алелите, които се различават от референтното разчитане. Например инсерция на единична T е представена като референтна A и алтернативна AT. А към T единичен нуклеотиден вариант е представен като референтна A и алтернативна T.
<b>QUAL</b>	Резултат за качество, определен чрез скала Phred, назначен от инструмента за обозначаване на варианти. По-високият резултат показва по-висока доверителност във варианта и по-малка възможност за грешки. За резултат за качество на Q изчислената възможност за грешка е $10^{-(Q/10)}$ . Например комплект от Q30 обозначава има честота на грешки 0,1%. Много инструменти за обозначаване на варианти назначават резултати за качество на основата на техните статистически модели, които са високи във връзка с наблюдаваната честота на грешки.

## Анотации на VCF файл

Заглавие	Описание
<b>FILTER (ФИЛТЪР)</b>	<p>Ако всички филтри са преминали, в колоната за филтъра се изписва <b>PASS</b> (ПРЕМИНАВА).</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b> (Ниска дълбочина) – Прилага се за места с дълбочина на покриване под 450x във всяко обединяване. За позиции на ампликони, покрити както от предно, така и обратно разчитане, това е еквивалентно на покритие 900x при едно разчитане.</li> <li>• <b>LowGQ</b> (Ниско качество на генотипиране) – Качеството на генотипиране (GT) е под праговата стойност.</li> <li>• <b>q30</b> – Резултат за качество &lt; 30.</li> <li>• <b>LowVariantFreq</b> (Ниско честота на вариант) – Честотата на варианта е по-ниска от зададения праг.</li> <li>• <b>PB</b> – Отклонение на обединението на пробата. Вариантът не е открит или е намерен с ниска честота в едно или две обединения на пробата.</li> <li>• <b>R3x6</b> – Броят на съседни повторения (с дължина 1 до 3 bp) при обозначавания на варианти ≥ 6.</li> <li>• <b>SB</b> – Отклонението на веригата е по-голямо от зададения праг.</li> </ul>
<b>INFO (ИНФОРМАЦИЯ)</b>	<p>Възможните въвеждания в колоната INFO включват:</p> <ul style="list-style-type: none"> <li>• <b>AC</b> – Броят алели в генотиповете за всеки ALT алел в същата последователност, както е посочено.</li> <li>• <b>AF</b> – Честотата на алелите за всеки ALT алел в същата последователност, както е посочено.</li> <li>• <b>AN</b> – Общият брой алели в посочения генотип.</li> <li>• <b>CD</b> – Означение, показващо, че SNP се е появило в рамките на кодиращия регион в най-малко 1 запис на RefGene.</li> <li>• <b>DP</b> – Дълбочината (броят обозначавания на бази, подравнени в позиция и използвани в обозначаването на бази).</li> <li>• <b>Exon</b> – Списък, разделен със запетаи, за регионите на екзони, разчетени от RefGene.</li> <li>• <b>FC</b> – Функционални последици.</li> <li>• <b>GI</b> – Списък, разделен със запетаи, за идентификаторите на гените, разчетени от RefGene.</li> <li>• <b>QD</b> – Доверителност/качество на варианта по дълбочина.</li> <li>• <b>TI</b> – Списък, разделен със запетаи, за идентификаторите на транскриптите, разчетени от RefGene.</li> </ul>
<b>FORMAT (ФОРМАТ)</b>	<p>В колоната за формат са изброени полета, разделени с двоеточие. Например GT:GQ. Предоставеният списък с полета зависи от използвания инструмент за обозначаване на вариант. Наличните полета включват:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> – Запис на формата X, Y, където X е броят на референтни обозначавания и Y е броят на редуващи се обозначавания.</li> <li>• <b>DP</b> – Приблизителна дълбочина на разчитане; разчитания с MQ=255 или лоши свързвания са филтрирани.</li> <li>• <b>GQ</b> – Качество на генотипа.</li> <li>• <b>GQX</b> – Качество на генотипа. GQX е минималната стойност на GQ и колоната QUAL. Като цяло тези стойности са сходни; като се има предвид, че най-малкото прави GQX по-умерената мярка за качество на генотипа.</li> <li>• <b>GT</b> – Генотип. 0 отговаря на референтната база, 1 отговаря на първия запис в колоната ALT и така нататък. Наклонената черта (/) показва, че не е налична информация за фазиране.</li> <li>• <b>NL</b> – Ниво на шум; изчисление на шума на обозначаване на бази на тази позиция.</li> <li>• <b>PB</b> – Отклонение на обединението на пробата. Стойности, близки до 0, показали по-голямо отклонение към едно обединяване на проба и по-малка доверителност в обозначаването на вариант.</li> <li>• <b>SB</b> – Отклонение на верига в тази позиция. По-големи отрицателни стойности показват по-малко отклонение; стойности, близки до 0, показват по-голямо отклонение.</li> <li>• <b>VF</b> – Честота на вариант; процентът на разчитания, поддържащи алтернативен алел.</li> </ul>
<b>SAMPLE (ПРОБА)</b>	Колоната за проба дава стойностите, специфични за колоната FORMAT (ФОРМАТ).

## Геномни VCF файлове

Геномните VCF (gVCF) файлове са VCF v4.1 файлове, които следват набор от конвенции за представяне на всички места в генома в приемливо компактен формат. gVCF (\*.genome.vcf.gz) файловете включват всички места в рамките на района на интерес в един файл за всяка проба. gVCF файлът показва липсата на обозначавания на позиции, които не преминават всички филтри. Таг за генотип (GT) ./. показва, че няма обозначаване.

За повече информация вижте [sites.google.com/site/gvcftools/home/about-gvcf](https://sites.google.com/site/gvcftools/home/about-gvcf).

## VCF файлове по обединяване и консенсусни VCF файлове.

Работният процес за соматичен вариант генерира 2 комплекта от файлове за обозначаване на вариант.

- ▶ **Per-pool VCF files** (VCF файлове по обединяване) – Съдържа обозначените варианти за предно обединяване или обратно обединяване. Файловете по обединяване са записани в папката VariantCallingLogs (Хронологии на обозначаване на варианти).
- ▶ **Consensus VCF files** (Консенсусни VCF файлове) – Съдържат варианти от двете обединявания. Консенсусните файлове са записани в папката Alignment (Подравняване).

VCF файлове по обединяване и консенсусни VCF файлове включват както VCF (\*.vcf), така и gVCF (\*.genome.vcf) файлове и използват следните конвенции за именуване, където S# представлява последователността, по която пробата е посочена за изпълняване.

- ▶ **Reports for all sites** (Отчети за всички места) – SampleName\_S#.genome.vcf
- ▶ **Reports variants only** (Отчети само за варианти) – SampleName\_S#.vcf

Софтуерът сравнява VCF файловете по обединяване и комбинира данни за всяка позиция, за да създаде консенсусен VCF файл за пробата.

Обозначаванията за варианти за всяко обединяване са слети в консенсусни VCF файлове по следните критерии.

Критерии	Резултат
Референтно обозначаване за всяко обединяване	Reference call (Референтно обозначаване)
Референтно обозначаване в 1 обединяване и обозначаване на вариант в другото обединяване	Filtered variant call (Филтрирано обозначаване на вариант)
Съвпадащи обозначавания на варианти с подобни честоти във всяко обединяване	Variant call (Обозначаване на вариант)
Съвпадащи обозначавания на варианти със значително различни честоти във всяко обединяване	Filtered variant call (Филтрирано обозначаване на вариант)
Несъвпадащи обозначавания на варианти във всяко обединяване	Filtered variant call (Филтрирано обозначаване на вариант)

Измерванията от всяко обединяване са слети чрез следните стойности.

Метричен	Стойност
Depth (Дълбочина)	Сборът на дълбочините от двете обединявания
Variant Frequency (Честотата на вариант)	Общите стойности на вариантите, разделени от общата дълбочина на покриването

Метричен	Стойност
Q-Score (Резултат за качество)	Минималната стойност на двете обединявания

## Файл за покритие на ампликон

Файл за покритие на ампликон се генерира за всеки манифестен файл. M# в името на файла представлява манифестния номер.

Всеки файл включва ред за главен колонтитул, който съдържа идентификатора на пробата, свързан с манифеста. Файлът съдържа следната информация.

- ▶ Идентификаторът на целта, както е посочен в манифеста.
- ▶ Дълбочината на покритието на филтъра за преминаване на разчитанията.

## Допълнителни изходни файлове

Следните изходни файлове осигуряват допълнителна информация или обобщават резултатите от изпълняванията и грешки в анализа. Въпреки че тези файлове не са необходими за оценяване на резултатите от анализа, те могат да се използват за целите на отстраняване на неизправности. Всички файлове се намират в папката Alignment (Подравняване), освен ако не е посочено друго.

Име на файла	Описание
<b>AnalysisLog.txt</b>	Хронология на обработването, която описва всяка стъпка, която се е случила по време на анализа на сегашната папка за изпълнявания. Файлът не съдържа съобщения за грешки. Намира се в папката Alignment (Подравняване).
<b>AnalysisError.txt</b>	Хронология на обработването, която изброява всякакви грешки, които са се случили по време на анализа. Този файл ще бъде празен, ако не са настъпили никакви грешки. Намира се в папката Alignment (Подравняване).
<b>DemultiplexSummaryF1L1#.txt</b>	Отчита резултати от демултиплексирането в таблица с 1 ред на плочка и 1 колона на проба. # представлява линия 1, 2, 3 или 4 от поточната клетка. Намира се в папката Alignment (Подравняване).
<b>AmpliconRunStatistics.xml</b>	Съдържа обобщаващи статистики, специфични за изпълняването. Намира се в папката Alignment (Подравняване).

## Папка Analysis (Анализ)

Папката Analysis (Анализ) съдържа файловете, генерирани от софтуера Local Run Manager.


Връзката между изходната папка и папката Analysis (Анализ) е обобщена, както следва:


- ▶ По време на секвенирането анализът в реално време (RTA) попълва изходна папка с файловете, генерирани по време на анализа на изображения, обозначаване на бази и резултати за качество.
- ▶ RTA копира файлове в папката Analysis (Анализ) в реално време. След като RTA назначи резултат за качеството за всяка база за всеки цикъл, софтуерът записва файла RTAComplete.txt и в двете папки.
- ▶ Когато файлът RTAComplete.txt е наличен, анализът започва.
- ▶ Докато анализът продължава, Local Run Manager записва изходните файлове в папката Analysis (Анализ) и след това копира файловете обратно в изходната папка.













## Папки Alignment (Подравняване)

Всеки път, когато анализът се нарежда отново на опашката, Local Run Manager създава папка Alignment (Подравняване) с име **Alignment\_N**, където N е пореден номер.

### Структура на папката

 **Alignment** (Подравняване) – Съдържа \*.bam, \*.vcf, FASTQ файлове и файлове, специфични за модула на анализ.

 **Date and Time Stamp** (Клеймо с дата и час) – Клеймо с дата и час на анализа, като YYYYMMDD\_HHMMSS



-  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**

-  Sample1\_L001\_R1\_001\_fastq.gz

##### **Stats (Статистика)**

-  DemuxSummaryF1L1.txt
-  FastqSummaryF1L1.txt


#### **Data (Данни)**

##### **Intensities (Интензитети)**


###### **BaseCalls**


 **L001** – съдържа \*.bcl файлове.

 **L001** – съдържа \*.locs файлове.

 **RTA Logs** (Хронологии за RTA) – Съдържа файлове за хронологии от анализа на софтуера RTA.



 **InterOp** – Съдържа двоични файлове, използвани за отчитане на измерванията при изпълняване на секвениране.

 **Logs** (Хронологии) – Съдържа файлове с хронологии, описващи изпълнените стъпки по време на секвенирането.

 RTAComplete.txt

 RunInfo.xml

 RunParameters.xml

## Обозначаване на бази и разнообразие на индекси

Когато пробите се секвенират на инструмента NextSeq 550Dx, обозначаването на базите определя база (A, C, G или T) за всеки клъстър на дадена плочка или изобразявана област на поточната клетка в специфичен цикъл. Инструментът NextSeq 550Dx използва двуканално секвениране, което изисква само две изображения за кодиране на данните за четири ДНК бази, едно от червения канал и едно от зеления канал.

Процесът за индексни разчитания на обозначаване с база се различава от обозначаването на бази по време на други разчитания.

Индексните разчитания започват с най-малко една база, различна от G, в който и да е от първите два цикъла. Ако индексното разчитане започне с две обозначавания на бази от G, не се генерира сигнален интензитет. За да се гарантира демултиплексна функционалност, сигналът трябва да е налице в някой от първите два цикъла.

Когато се избират индекси по време на създаването на изпълняване, предупреждение за ниско разнообразие ще се появи, ако индексите не отговарят на изискванията за разнообразие. За да предотвратите предупреждението за ниско разнообразие, изберете индексни последователности, които осигуряват сигнал и в двата канала за всеки цикъл.

- ▶ Червен канал – A или C
- ▶ Зелен канал – A или T

Този процес на обозначаване на бази обезпечава точността, когато се анализират малкоплексни проби. За повече информация относно последователностите на вашите индекси вижте листовката на комплект за персонализирани ампликони *TruSeq Dx* (документ № 1000000029772).

По време на създаването на изпълняване в Local Run Manager ще изберете броя на пробите, които ще бъдат изследвани. Предложените индексни комбинации, които отговарят на изискванията за индексно разнообразие, се попълват автоматично от софтуера. Въпреки че не е задължително да използвате предложените индексни комбинации, това е препоръчително.

## Хронология на редакциите

Документ	Дата	Описание на промяната
Документ № 1000000030330 v04	Август 2021 г.	Актуализиран адрес на упълномощен представител на ЕС.
Документ № 1000000030330 v03	Април 2020 г.	Актуализиран адрес на упълномощен представител на ЕС. Актуализиран адрес на спонсора за Австралия.
Документ № 1000000030330 v02	Януари 2019 г.	Добавена е информацията относно комплекти с реагенти v2.5.
Документ № 1000000030330 v01	Август 2018 г.	Актуализирана регулаторна маркировка.
Документ № 1000000030330 v00	Ноември 2017 г.	Първоначална версия.

## Техническа помощ

За техническа помощ се свържете с отдела за техническа поддръжка на Illumina.

Уеб сайт: [www.illumina.com](http://www.illumina.com)  
Имейл: [techsupport@illumina.com](mailto:techsupport@illumina.com)

## Телефонни номера на отдела за техническа поддръжка на Illumina

Регион	Безплатен	Регионален
Северна Америка	+1.800.809.4566	
Австралия	+1.800.775.688	
Австрия	+43 800006249	+43 19286540
Белгия	+32 80077160	+32 34002973
Германия	+49 8001014940	+49 8938035677
Дания	+45 80820183	+45 89871156
Ирландия	+353 1800936608	+353 016950506
Испания	+34 911899417	+34 800300143
Италия	+39 800985513	+39 236003759
Китай	400.066.5835	
Нидерландия	+31 8000222493	+31 207132960
Нова Зеландия	0800.451.650	
Норвегия	+47 800 16836	+47 21939693
Обединеното кралство	+44 8000126019	+44 2073057197
Сингапур	+1.800.579.2745	
Тайван	00806651752	
Финландия	+358 800918363	+358 974790110
Франция	+33 805102193	+33 170770446
Хонконг	800960230	
Швейцария	+41 565800000	+41 800200442
Швеция	+46 850619671	+46 200883979
Япония	0800.111.5011	
Други държави	+44.1799.534000	

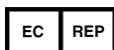
**Информационни листове за безопасност (ИЛБ)** – налични на уеб сайта на Illumina на адрес [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Документация на продукта** – налична за изтегляне в PDF формат на уебсайта на Illumina. Посетете [support.illumina.com](http://support.illumina.com), изберете продукт, след което изберете **Documentation & Literature** (Документация и литература).



Illumina  
5200 Illumina Way  
San Diego, California 92122, САЩ  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (извън Северна Америка)  
techsupport@illumina.com  
www.illumina.com

CE



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
Нидерландия

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

**ЗА ИН ВИТРО ДИАГНОСТИЧНА УПОТРЕБА**

© 2021 Illumina, Inc. Всички права запазени.

**illumina®**