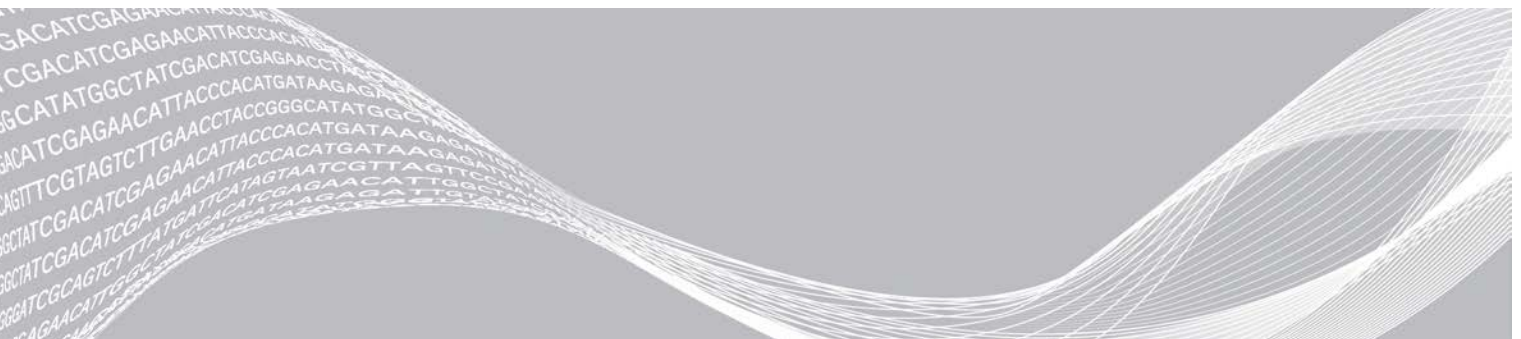


Модул за анализ на вариант за герминативна линия на Local Run Manager

Ръководство за работния процес за NextSeq 550Dx

ЗА ИН ВИТРО ДИАГНОСТИЧНА УПОТРЕБА

Общ преглед	3
Въвеждане на информация за изпълняването	4
Методи за анализ	6
Преглед на данни за изпълняването и пробите	8
Отчет от анализ	8
Изходни файлове на анализ	9
Хронология на редакциите	17
Техническа помощ	18



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 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.

Общ преглед

Модулът за варианти за герминативна линия Local Run Manager е за употреба с анализа за комплект персонализирани ампликони TruSeq Dx и NextSeq 550Dx на Illumina. Когато се използва с модул за вариант за герминативна линия, анализът е предназначен за подготовка на библиотеки, използвани за секвениране на ДНК от проби от периферна цяла кръв.

Модулът за анализ оценява кратки области на амплифицирана ДНК или ампликони за варианти. Фокусираното секвениране на ампликоните позволява високо покритие на определени региони. Вижте листовката за *комплект персонализирани ампликони TruSeq Dx (документ № 1000000029772)*.

Модулът за анализа на вариант за герминативна линия изисква консумативи за секвениране за 300 цикъла. За повече информация вижте листовката на *комплект с реагенти NextSeq 550Dx High Output v2* или *комплект с реагенти NextSeq 550Dx High Output v2.5*.

За това ръководство

Това ръководство предоставя инструкции за настройка на параметри на изпълняване на секвениране и анализ за модула за анализ за вариант на герминативна линия. За информация относно таблото за управление на Local Run Manager и системните настройки 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 поддържания 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 едно 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 от 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**.

Въвеждане на информация за изпълняването

Задаване на параметри

- 1 Влезте в Local Run Manager.
- 2 Изберете **Create Run** (Създаване на изпълняване) и изберете **Germline Variant** (Вариант за герминативна линия).
- 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 от 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 индексния комплект се отнасят до различни комплекти от i5 индекси. Като комплект 1, така и комплект 2 осигуряват индексно разнообразие. Двата индексни комплекта се предлагат, за да подпомогнат в предотвратяването на изчерпването на единичен комплект.
 - ▶ Изберете броя проби, които са най-близо до броя проби, които изследвате. Ако точният брой 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 манифестните файлове, които искате да импортирате, са налични на достъпно мрежово местоположение или на USB устройство.
- 2 Изберете **Import Manifests** (Импортиране на манифестни файлове).
- 3 Навигирайте до манифестния файл и изберете манифестните файлове, които искате да добавите.



ЗАБЕЛЕЖКА

За да направите манифестни файлове, налични за всички изпълнявания, чрез използване на модула за анализ на вариант за герминативна линия, добавете манифестните файлове 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 № 100000009513)*.


Посочване на проби за изпълняването

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

- ▶ **Enter samples manually** (Ръчно въвеждане на пробите) – 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, когато подготвяте библиотеки или импортирате файла за друго изпълняване.

Ръчно въвеждане на пробите

- 1 Въведете уникален ИД на пробата в полето Sample ID (ИД на пробата). Използвайте буквено-цифрови знаци, тирета или долни черти.
- 2 **[Незадължително]** За положителни или негативни контролни проби щракнете с десния бутон на мишката и изберете типа на контролата.
- 3 **[Незадължително]** Въведете описание на пробата в раздела Sample Description (Описание на пробата). Използвайте буквено-цифрови знаци, тирета или долни черти. Описанията на пробите са свързани с име на пробата. Описанията на пробите се презаписват, ако същото име на пробата бъде използвано отново в по-късно изпълняване.
- 4 Изберете адаптер на Index 1 (Индекс 1) от падащия списък Index 1 (i7) (Индекс 1 [i7]). Когато използвате предложените кладенчета за проби, софтуерът попълва автоматично адаптерите на индекс 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 адаптери за индекси 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 16*.
- 5 Изберете адаптер Index 2 (Индекс 2) от падащия списък Index 2 (i5) (Индекс 2 [i5]).
- 6 Изберете манифестен файл от падащия списък Manifest (Манифест).
- 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 плаките. Изберете **Print** (Печат) за отпечатване на подредбата на плаките.
 - ▶ Изберете **Export** (Експортиране) за експортиране на информация за проба към външен файл.Уверете се, че информацията за манифеста и пробата е правилна. 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 отидете до местоположението на файла с информацията за проба. 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 на информацията за пробата във всяка колона за всяка проба в изпълняването. Изтрийте примерната информация в неизползвани клетки и след това запишете файла.
 - ▶ Използвайте файл с информацията за проба, който е бил експортиран от модула за вариант за герминативна линия чрез функцията Export (Експортиране).
- 2 Изберете иконата  **Print** (Печат) за показване на подредбата на плаките.
- 3 Изберете **Print** (Печат) за отпечатване на подредбата на плаката като справочна информация за подготовка на библиотеките.
- 4 **[Незадължително]** Изберете **Export** (Експортиране) за експортиране на информацията за проба към външен файл.
Уверете се, че информацията за манифеста и пробата е правилна. 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 вариант за герминативна линия изпълнява следните стъпки за анализ и след това записва изходните файлове на анализа в папката Alignment (Подравняване).

- ▶ Индексни разчитания на демултиплексиране
- ▶ Генериране на FASTQ файлове
- ▶ Подравнява се към препратка
- ▶ Идентифицира варианти

Демултиплексиране

Демултиплексирането сравнява последователността от индексни разчитания към 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 реда, в който са изброени за изпълняването.
- ▶ Номер на проба 0 е запазен за клъстери, които не са били зададени към проба.
- ▶ Клъстерите се задават към проба, когато индексната последователност съвпада точно или когато има до единично съвпадение 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. На проба се генерират по осем FASTQ файла, четири от Разчитане 1 и четири от Разчитане 2.

Подравняване

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

Лентовият алгоритъм Smith-Waterman изпълнява полуглобални подравнявания на последователности за определяне на подобни региони между две последователности. Вместо да сравнява общата 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 разчитане съвпада с последователността на пробата с не повече от три разлики (несъвпадения или промени поради водещи инсерции и делеции), пълната дължина на разчитането се подравнява срещу таргета 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 несъответствие 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

Разработен от Illumina, инструментът за обозначаване на вариант Pisces идентифицира наличните варианти в ДНК проба.



Инструментът за обозначаване на вариант Pisces идентифицира SNV, MNV и малки инсерции и делеции в 3 стъпки:

- ▶ Разглежда всяка позиция в референтния геном поотделно
- ▶ Преброява бази в дадената позиция за подравнени разчитания, които припокриват позицията
- ▶ Изчислява резултат за вариант, който измерва качеството на обозначаването, използвайки модела на Poisson. Варианти с резултат за качество под Q20 не се включват.

Ако даден вариант преминава през всички филтри, вариантът е маркиран като PASS (ПРЕМИНАВА) във VCF.

За повече информация вижте github.com/Illumina/Pisces/wiki.

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 прегледате местоположението на отчета за анализ.
 - ▶ 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 наредите анализи на опашка вижте *справочното ръководство за инструмента NextSeq 550Dx (документ №1000000009513)*.

Отчет от анализ

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

Информация за раздели Samples (Проби) и Results (Резултати)

- 1 Щракнете 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 последователността е преминало, или не.
Total Yield (GB) (Обща производителност)	Броят бази, които са обозначени в изпълняването на секвениране. Показва прага на преминаване и състояние на преминаване или неуспех.
% ≥ Q30	Процентът на разчитанията в изпълняване на секвениране с резултат за качество 30 (Q30) или по-голям. Показва прага на преминаване и състояние на преминаване или неуспех.
Sample ID (ИД на проба)	ИД на пробата, предоставен при създаването на изпълняването.
Total PF Reads (Общо разчитания на PF)	Общият брой на разчитанията, преминаващи филтъра.
Read 1% ≥ Q30 (Разчитане 1% ≥ Q30)	Процентът на разчитанията в Read 1 (Разчитане 1) с резултат за качество 30 (Q30) или по-голям за пробата.

Заглавие на колона	Описание
Read 2% \geq Q30 (Разчитане 2% \geq Q30)	Процентът на разчитанията в Read 2 (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 позиция, за които е получена информация. Честотата на обозначаването се описва на база за извадка и се отчита като процент, който се изчислява като 1 минус (брой варианти на автозомни позиции с непълни обозначавания, разделен на общия брой секвенирани варианти на автозомни позиции).

Таблица 2 Информация за отчет на пробата

Заглавие на колона	Описание
Sample (Проба)	ИД на пробата, предоставен при създаването на изпълняването.
Report Date (Дата на отчета)	Датата, на която отчетът е създаден.
Sample Information (Информация за пробата)	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 и манифестният файл. 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).
Variants Summary (Обобщение на вариантите)	Броят на SNV, инсерции и делеции, открити в пробата, която е преминала предложените стойности за определяне дали резултатите за качество а в рамките на приемливия диапазон.
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 границата за ниско покриване от 0,2 * средно покриване на ампликон за пробата.
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.
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 вариант за герминативна линия.

Изходни файлове на анализ

Следните изходни файлове на анализ се генерират за модула за анализ на вариант за герминативна линия 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 на демултиплексирането.
FASTQ (*.fastq.gz)	Междинни файлове, които съдържат обозначаване на бази с резултат за качество. FASTQ файловете са основната входна информация за стъпката на подравняване.
Файлове за подравняване в BAM формат (*.bam)	Съдържа подравнени разчитания за дадена проба.
Файлове за обозначаване на вариант в генома във VCF формат (*.genome.vcf.gz)	Съдържа генотипа за всяка позиция, независимо дали е обозначена като вариант, или справка.
Файлове за обозначаване на вариант във VCF формат (*.vcf.gz)	Съдържа всички обозначени варианти в целевата област.
AmpliconCoverage_M1.tsv	Съдържа информация относно покриването на ампликон за проба за всеки предоставен манифест. M# представлява манифестния номер.

Формат на файла 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 за 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 със заглавен раздел:

- ▶ 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 за демултиплексиране, наименуван `DemultiplexSummaryF1L1.txt`.

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

Файлов формат FASTQ

FASTQ е текстовобазирани файлове формат, който съдържа обозначаване на бази и стойности на качеството на едно разчитане. Всеки запис съдържа 4 линии:

- ▶ Идентификаторът
- ▶ Последователността
- ▶ Знак плюс (+)
- ▶ Качеството на PhRed се оценява в кодиран формат ASCII + 33

Идентификаторът е представен във формат:

@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=AAAAAAAAAAA9#:<#<;<<<????#=#
```

Файлов формат BAM

BAM файл (*.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 на 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 включени в списъка за изпълнение.

BAM файловете съдържат заглавен раздел и секция за подравняване:

- ▶ **Заглавен раздел** – Съдържа информация за целия файл, като име на проба, дължина на пробата и метод на подравняване. Подравняванията в раздела за подравняване са свързани със специфична информация в заглавния раздел.
- ▶ **Подравнявания** – Съдържа име за четене, последователност на четене, качество на четене, информация за подравняване и персонализирани тагове. Прочетеното наименование включва хромозомата, началната координатна система, качеството на подравняване и дескриптора на съвпадение.

Разделът за подравнявания включва следната информация за всяка двойка четене или четене:

- ▶ **AS:** Качество на сдвоено подравняване.
- ▶ **RG:** Група за разчитане, която посочва броя разчитания за конкретна проба.
- ▶ **BC:** Баркод таг, който показва демултиплексиращия идентификатор на пробата, свързан с разчитането.
- ▶ **SM:** Качество на подравняване в един край.
- ▶ **XC:** Отговаря на низа на дескриптора.
- ▶ **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

Заглавният ред на VCF файла включва версията на VCF файловия формат и версията на инструмента за обозначаване на варианта и посочва анотациите, използвани в останалата част на файла. Заглавният ред на VCF също включва файла на референтния геном и BAM файла. Последният ред в заглавката 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
CHROM	Хромозомата от референтния геном. Хромозомите се появяват в същата последователност както в референтния FASTQ файл.
POS	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.
ID	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 ('.').
REF	Референтният генотип. Например делеция на единична T е представено като референтна TT и алтернативна T. A към T единичен нуклеотиден вариант е представен като референтна A и алтернативна T.
ALT	Алелите, които се различават от референтното разчитане. Например инсерция на единична T е представена като референтна A и алтернативна AT. A към T единичен нуклеотиден вариант е представен като референтна A и алтернативна T.
QUAL	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 Q изчислената възможност за грешка е $10^{-(Q/10)}$. Например комплект от Q30 обозначавания 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
FILTER (ФИЛТЪР)	Ако всички филтри са преминали, в колоната за филтъра се изписва PASS (ПРЕМИНАВА). <ul style="list-style-type: none"> • LowDP (Ниска дълбочина) – Прилага се за места с дълбочина на покриване под 150x. За позиции на ампликони, покрити както от предно, така и обратно разчитане, това е еквивалентно на 300 покриващи се разчитания със съдвоени краища. • q20 – Резултат за качество < 20. • MultiAllelicSite – Вариантът не отговаря на диплоидния модел. • R5x9 – Броят на съседни повторения (с дължина 1 до 5 bp) при обозначавания на варианти ≥ 9. • SB – Отклонението на веригата е по-голямо от зададения праг.
INFO (ИНФОРМАЦИЯ)	Възможните въвеждания в колоната INFO (ИНФОРМАЦИЯ) включват: <ul style="list-style-type: none"> • AC – Броят алели в генотиповете за всеки ALT алел в същата последователност, както е посочено. • AF – Честота на алели за всеки ALT алел в същата последователност, както е посочено. • AN – Общият брой алели в посочения генотип. • CD – Означение, показващо, че SNP се е появило в рамките на кодиращия регион в най-малко 1 запис на RefGene. • DP – Дълбочината (броят на обозначаване на бази, подравнени в позиция и използвани в обозначаването на бази). • Exon – Списък, разделен със запетаи, за регионите на екзони, разчетени от RefGene. • FC – функционални последици. • GI – Списък, разделен със запетаи, за идентификаторите на гените, разчетени от RefGene. • QD – Доверителност/качество на варианта по дълбочина. • TI – Списък, разделен със запетаи, за идентификаторите на транскриптите, разчетени от RefGene.
FORMAT (ФОРМАТ)	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: <ul style="list-style-type: none"> • AD – Запис на формата X,Y, където X е броят на референтни обозначавания и Y е броят на редуващи обозначавания. • DP – Приблизителна дълбочина на разчитане, разчитания с MQ=255 или с лоши свързвания са филтрирани. • GQ – Качество на генотипа. • GT – 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. • NC – Част от необозначени бази или с качество на обозначаване на бази под минималния праг. • NL – Ниво на шум; изчисление на шума на обозначаване на бази на тази позиция. • SB – Отклонение на верига в тази позиция. По-големи отрицателни стойности показват по-малко отклонение; стойности, близки до 0 показват по-голямо отклонение. • VF – Честота на вариант; процентът на разчитания, поддържащи алтернативен алел.
SAMPLE (ПРОБА)	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 разумно компактен формат. gVCF (*.genome.vcf.gz) файловете включват всички места в района на интерес в един файл за всяка проба.

Файлът gVCF показва липса на обаждания на позиции, които не преминават всички филтри. Таг за генотип (GT) ./. показва, че няма съвпадение.

За повече информация вижте sites.google.com/site/gvcftools/home/about-gvcf.

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 заглавен ред, който съдържа идентификатори на пробата, свързани 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 оценка на резултатите от анализа, те могат да се използват за отстраняване на неизправности. Всички файлове се намират в папката Alignment (Подравняване), освен ако не е посочено друго.

Име на файла	Описание
AnalysisLog.txt	Хронология на обработването, която описва всяка стъпка, която се е случила по време на анализа на папката за сегашното изпълняване. Файлът не съдържа съобщения за грешки. Намира се в папката Alignment (Подравняване).
AnalysisError.txt	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).
DemultiplexSummaryF1L1#.txt	Отчита резултати от демултиплексирането в таблица с 1 ред на плочка и 1 колона на проба. # представлява пътека 1, 2, 3 или 4 от поточната клетка. Намира се в папката Alignment (Подравняване).
AmpliconRunStatistics.xml	Съдържа обобщаващи статистики, специфични за изпълняването. Намира се в папката Alignment (Подравняване).

Папка Analysis (Анализ)

Папка 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 за анализ в реално време. След като RTA назначи оценка на 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 към 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 за 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 анализа.

 **Date and Time Stamp** (Клеймо за дата и час) – Клеймо за 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**

 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 една от зеления канал.

Процесът за индексни разчитания с 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 двата първи цикъла. Ако индексното разчитане започне с две обозначавания на бази от G, не се генерира сигнал за 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 изпълнение, предупреждение за ниско разнообразие ще се появи, ако индексите не срещат 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 проби с малък брой. За повече информация относно последователностите на вашите индекси вижте листовката на комплект персонализирани ампликони TruSeq (документ № 1000000029772).

По време на създаването на изпълнение 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 № 1000000030329 v04	Август 2021 г.	Актуализиран адрес на упълномощен представител на ЕС.
Документ № 1000000030329 v03	Декември 2019 г.	Актуализиран адрес на упълномощен представител на ЕС. Актуализиран адрес на спонсора за Австралия.
Документ № 1000000030329 v02	Януари 2019 г.	Добавена е информацията относно комплекти с реагенти v2.5.
Документ № 1000000030329 v01	Август 2018 г.	Актуализирана регулаторна маркировка.
Документ № 1000000030329 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
Имейл: 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	

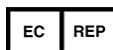
Информационни листове за безопасност (ИЛБ) – налични на уеб сайта на Illumina на адрес support.illumina.com/sds.html.

Документация на продукта – налична за изтегляне в PDF формат на уеб сайта на Illumina. Посетете 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[®]