

Local Run Manager

Соматска варијанта Модул за анализа

Водич за работен тек за NextSeq 550Dx

ЗА ИН ВИТРО ДИЈАГНОСТИЧКА УПОТРЕБА

Преглед	3
Внесете информации за обработка	4
Методи на анализа	6
Прикажи податоци за обработка и примероци	8
Извештај за анализа	8
Излезни датотеки на анализа	9
Одредување на бази и индексна разновидност	18
Претходни ревизии	19
Техничка помош	20



Овој документ и неговата содржина се сопственост на Illumina, Inc. и нејзините филијали („Illumina“) и се наменети исклучиво за договорна употреба од страна на нејзините клиенти во врска со употребата на производите опишани овде и за никаква друга цел. Овој документ и неговата содржина не смеат да се користат или дистрибуираат за никаква друга цел и/или на друг начин да се пренесуваат, обелоденуваат или репродуцираат без претходна писмена согласност од Illumina. Со овој документ Illumina не пренесува никаква лиценца заштитена со патент, трговска марка, авторско право или обичајни и слични права на трети страни.

Квалификуваниот и соодветно обучен персонал мора строго и без никакви отстапки да се придржува до упатствата дадени во овој документ за да се загарантира правилна и безбедна употреба на производите опишани во него. Пред да започнете да ги користите таквите производи, мора да ја прочитате целата содржината на овој документ и да ја разберете.

ДОКОЛКУ НЕ ГИ ПРОЧИТАТЕ ЦЕЛОСНО И НЕ СЕ ПРИДРЖУВАТЕ ДО УПАТСТВАТА ДАДЕНИ ВО ОВОЈ ДОКУМЕНТ, МОЖЕ ДА ДОЈДЕ ДО ОШТЕТУВАЊЕ НА ПРОИЗВОДИТЕ, ПОВРЕДИ НА КОРИСНИЦИТЕ ИЛИ ДРУГИ ЛИЦА И ОШТЕТУВАЊЕ НА ДРУГАТА СОПСТВЕНОСТ, СО ШТО СЕ ПОНИШТУВААТ СИТЕ ГАРАНЦИИ КОИ СЕ ОДНЕСУВААТ НА ПРОИЗВОДИТЕ.

ILLUMINA НЕ ПРЕЗЕМА НИКАКВА ОДГОВОРНОСТ ЗА ШТЕТИТЕ КОИ НАСТАНАЛЕ ПОРАДИ НЕПРАВИЛНА УПОТРЕБА НА ПРОИЗВОДИТЕ КАКО ШТО Е ОПИШАНО ОВДЕ (ВКЛУЧУВАЈЌИ ГИ И НИВНИТЕ ДЕЛОВИ ИЛИ СОФТВЕР).

© 2021 Illumina, Inc. Сите права се задржани.

Сите трговски марки се сопственост на Illumina, Inc. или на нивните соодветни сопственици. За конкретни информации за трговските марки, погледнете на www.illumina.com/company/legal.html.

Преглед

Local Run Manager Модулот за соматска варијанта е за употреба со анализата Illumina TruSeq Custom Amplicon Kit Dx и NextSeq 550Dx. Кога се користи со модулот за соматска варијанта, анализата е наменета за подготовка на библиотечен материјал што се користи за секвенционирање на ДНК од ткиво фиксирано со формалин и вградено во парафин (FFPE). Тестот детектира соматски мутации при ниски варијантни фреквенции.

Модулот за анализа ги проценува кратките региони на амплифицирана ДНК, или ампликони, за варијанти. Фокусираното секвенционирање на ампликони овозможува голема покриеност на одредени региони низ голем број примероци. Модулот за анализа врши секундарна анализа и генерирање извештаи од обработки со секвенционирање користејќи пристап на двоен синџир што вклучува предни и обратни олигогрупи. Погледнете го приложеното упатство *TruSeq Custom Amplicon Kit Dx* (број на документ 1000000029772).

Модулот за анализа на соматски варијанти бара потрошен материјал за секвенционирање од 300 циклуси. За повеќе информации, погледнете го приложеното упатство *NextSeq 550Dx High Output Reagent Kit v2* или *NextSeq 550Dx High Output Reagent Kit v2.5*.

За овој водич

Овој водич дава инструкции за поставување параметри за обработка за секвенционирање и анализа за Соматска варијанта модулот за анализа. За информации за контролната табла и системските поставки на Local Run Manager, погледнете во *Референтен водич за инструментот NextSeq 550Dx* (број на документ 1000000009513).

Прегледување Local Run Manager

Интерфејсот на Local Run Manager се прикажува во Оперативен софтвер за NextSeq 550Dx (NOS) или преку веб-прелистувач. Поддржаниот веб-прелистувач е Chromium.



ЗАБЕЛЕШКА

Ако користите неподдржан прелистувач, преземете го поддржаниот прелистувач кога тоа ќе се побара од вас со пораката Confirm Unsupported Browser (Потврдете неподдржан прегледувач). Изберете „**here**“ (тука) за да ја преземете поддржаната верзија на Chromium.

Приказ на мониторот на инструментот

- 1 За да го видите интерфејсот на Local Run Manager на мониторот на инструментот, изберете една од следниве опции:
 - ▶ Од Home screen (Почетен екран) на NOS, изберете **Local Run Manager**. Како ќе завршите, кликнете X во горниот десен агол за да се вратите во NOS.
 - ▶ Изберете ја NOS иконата Minimize (Минимизирај), отворете го веб-прелистувачот Chromium што се наоѓа на инструментот и внесете **http://localhost** во лентата со адреси. Само администраторите можат да минимизираат NOS.

Преглед од мрежен компјутер

- 1 Отворете го веб-прелистувачот Chromium на компјутер со пристап до истата мрежа како инструментот и поврзете ги со IP-адресата или името на инструментот. На пример, **http://myinstrument**.

Внесете информации за обработка

Поставување параметри

- 1 Најавете се на Local Run Manager.
- 2 Изберете **Create Run** (Креирај обработка) и изберете **Somatic Variant** (Соматска варијанта).
- 3 Внесете име на обработката што ќе ја идентификува обработката од секвенционирањето до анализата.
Користете алфанумерички знаци, празни места, долни црти или цртички.
- 4 **[Изборно]** Внесете опис на обработката за полесно идентификување на обработката.
Користете алфанумерички знаци, празни места, долни црти или цртички.
- 5 Изберете го бројот на примероци и комплетот индекси од паѓачката листа.
Земете ги предвид следниве информации кога избирате.
 - ▶ Паѓачката листа содржи броеви на примероци со комплет индекси. На пример, 24-Set 1 (24-Комплет 1) укажува дека ќе се тестираат 24 примероци, со индекси од комплетот индекси 1.
 - ▶ Броевите на комплетот индекси се однесуваат на различни комплети на i5 индекси. Сет 1 и Сет 2 обезбедуваат разновидност на индекси. Понудени се два комплета индекси за да се спречи исцрпувањето на еден комплет.
 - ▶ Изберете го бројот на примероци што е најблиску до бројот на примероци што ги тестирате. Ако точниот број на примероци го нема на списокот, изберете го бројот што е најблиску, но е помал од бројот што го тестирате. На пример, ако сакате да тестирате 18 примероци, изберете 16 примероци.
 - ▶ Примерок бунари и комбинации на индекси што ги исполнуваат условите за разновидност на индекси се нагласени со зелена боја. Ако изберете други бунари и комбинации на индекси, кога ќе ја зачувате обработката, ќе добиете известување ако не се исполнети барањата за разновидност на индекси.

Увезете манифестни датотеки за обработката

- 1 Погрижете се потребните манифести за увезување да бидат достапни на пристапна мрежна локација или на USB диск.
- 2 Изберете **Import Manifests** (Увези манифести).
- 3 Појдете до манифестната датотека и изберете ги манифестите што сакате да ги додадете.



ЗАБЕЛЕШКА

За манифестните датотеки да бидат достапни за сите обработки со модулот за анализа Соматска варијанта, додадете манифести со функцијата Module Settings (Поставки за модулите). За оваа функција се потребни дозволи на ниво на администратор. За повеќе информации, погледнете го *Референтен водич за инструментот NextSeq 550Dx (број на документ 1000000009513)*.

Наведете примероци за обработката

Наведете примероци за обработката со една од опциите и насоките што следуваат.

- ▶ **Рачно внесете примероци** - Користете ја празната табела на екранот Create Run (Креирај обработка).
- ▶ **Увезете примероци** - Одете до надворешна датотека во формат со вредности разделени со запирка (*.csv). На екранот Create Run (Креирај обработка) е достапен шаблон за преземање.

Откако ќе ја пополните табелата со примероци, можете да ги извезете информациите за примерокот во надворешна датотека. Користете ја датотеката како референца кога подготвувате библиотечни материјали или ја увезувате датотеката за друга обработка.

Рачно внесување примероци

- 1 Внесете уникатно име на примерокот во полето Sample Name (Име на примерок). Користете алфанумерички знаци, цртички или долни црти. Името на примерокот автоматски го пополнува соодветниот бунар во другата група.
- 2 **[Изборно]** За позитивните или негативните контролни примероци, кликнете со десното копче и изберете го типот на контрола. Контролата во еден бунар на примероци автоматски го пополнува соодветниот бунар во другата група со истата контрола.
- 3 **[Изборно]** Внесете опис на примерокот во полето Sample Description (Опис на примерокот). Користете алфанумерички знаци, цртички или долни црти. Описот на примерокот автоматски го пополнува соодветниот бунар во другата група. Описите на примероците се поврзани со ИД на примерок. Описите на примероците се препишуваат ако истиот ИД на примерокот се користи повторно во подоцнежна обработка.
- 4 Изберете го адаптерот за индексирање 1 од паѓачката листа Index 1 (i7) (Индекс 1 (i7)). Кога ги користите предложените бунари на примероци, софтверот автоматски ги пополнува индексните адаптери i7 и i5 што ги исполнуваат условите за разновидност на индекси. Ако точниот број на примероци што ги тестираате го нема на списокот, изберете индексни адаптери за дополнителни бунари. Доколку треба да изберете индекси за дополнителни бунари или не ги користите препорачаните комбинации на индексни адаптери, пред да изберете индекси, прочитајте го делот *Одредување на бази и индексна разновидност на страница 18*.
- 5 Изберете го адаптерот за индексирање 2 од паѓачката листа Index 2 (i5) (Индекс 2 (i5)).
- 6 Изберете манифестна датотека од паѓачката листа Manifest (Манифест). За примероците од групата А е потребен различен манифест од тој за примероците во групата Б.
- 7 Изберете опција за прикажување, печатење или зачувување на распоредот на плочата како референца за подготвување на библиотечните материјали:
 - ▶ Изберете ја иконата  **Print** (Печати) за да се прикаже распоредот на плочата. Изберете **Print** (Печати) за да се отпечати распоредот на плочата.
 - ▶ Изберете **Export** (Извези) за да ги извезете информациите за примерокот во надворешна датотека.Проверете дали информациите од манифестот и примерокот се точни. Неточните информации може да влијаат врз резултатите.
- 8 Изберете **Save Run** (Зачувај ја обработката).

Увезување примероци

- 1 Изберете **Import Samples** (Увезување примероци) и појдете до локацијата на датотеката со информации за примероците. Постојат два типа датотеки што може да ги увезете.
 - Изберете **Template** (Шаблон) на екранот Create Run (Креирај обработка) за да создадете нов распоред на плоча. Датотеката за шаблонот ги содржи точните наслови на колоните за увезување. Внесете ги информациите за примероците во секоја колона за примероците што се обработуваат. Избришете ги информациите за пример во неискористените ќелии, па потоа зачувајте ја датотеката.
 - Користете ја датотеката со информации за примерок извезена од модулот Соматска варијанта со функцијата Export (Извези).
- 2 Изберете ја иконата  **Print** (Печати) за да се прикаже распоредот на плочата.
- 3 Изберете **Print** (Печати) за да го отпечатите распоредот на плочата како референца за подготвување на библиотечните материјали.
- 4 **[Изборно]** Изберете **Export** (Извези) за да ги извезете информациите за примероците во надворешна датотека.
Проверете дали информациите од манифестот и примерокот се точни. Неточните информации може да влијаат врз резултатите.
- 5 Изберете **Save Run** (Зачувај ја обработката).

Изменете обработка

За упатства за уредување на информациите во вашата обработка пред секвенционирање, погледнете во *Референтен водич за инструментот NextSeq 550Dx (број на документ 1000000009513)*.

Методи на анализа

Соматска варијанта Модулот за анализа ги извршува следните чекори за анализа, а потоа ги запишува излезните датотеки за анализа во папката Alignment (Анализа).

- Демултиплексирани индексни отчитувања
- Генерира FASTQ-датотеки
- Се порамнува со референца
- Идентификува варијанти

Демултиплексирање

Демултиплексирањето ја споредува секоја секвенца на индексно отчитување со секвенците за индексирање наведени за обработката. За овој чекор не се земаат предвид никакви вредности за квалитет.

Индексните отчитувања се идентификуваат со следниве чекори:

- Примероците се нумерирани почнувајќи од 1 според редоследот по којшто се наведени за обработката.
- Примерокот со број 0 е резервиран за кластерите што не се назначени за примерок.
- Кластерите се назначуваат за примерок кога секвенцата за индексирање точно одговара на или кога има најмалку една разлика по индексно отчитување.

Генерирање FASTQ-датотека

По демултиплексирањето, софтверот генерира датотеки за посредна анализа во FASTQ-формат, кој е текстуален формат што се користи за претставување на секвенците. FASTQ-датотеките содржат отчитувања за секој примерок и поврзаните резултати за квалитет. Кластерите што не поминуваат низ филтерот се исклучуваат.

Секоја FASTQ-датотека содржи отчитувања за само еден примерок, а името на тој примерок се вклучува во името на FASTQ-датотеката. FASTQ-датотеките се примарен материјал за порамнување. Осум FASTQ-датотеки се генерираат по примерок по oligo pool, четири од Отчитување 1 и четири од Отчитување 2, што резултира со вкупно 16 FASTQ-датотеки по примерок.

Порамнување

За време на чекорот за порамнување, бандизираниот Smith-Waterman алгоритам ги порамнува кластерите од секој примерок во однос на ампликонските секвенци наведени во манифестната датотека.

Бандизираниот Smith-Waterman алгоритам извршува полуглобални порамнувања на секвенците за да се утврдат слични региони помеѓу двете секвенци. Наместо да ја споредува целата секвенца, Smith-Waterman алгоритмот ги споредува сегментите со сите можни должини.

Секое отчитување со спарен крај се проценува во однос на порамнувањето со релевантните секвенци за сондирање за тоа отчитување.

- ▶ Отчитување 1 се проценува во однос на обратниот дел од Downstream Locus-Specific Oligos (DLSO).
- ▶ Отчитување 2 се проценува во однос на Upstream Locus-Specific Oligos (ULSO).
- ▶ Ако почетокот на отчитувањето одговара на секвенца за сондирање со не повеќе од три разлики (несовпаѓања или поместувања поради водечките индели), целата должина на отчитувањето се порамнува со целта на ампликонот за таа секвенца.
- ▶ Инделите во рамките на DLSO и ULSO не се разгледуваат поради хемиската анализа.

Порамнувањата се филтрираат од резултатите за порамнување врз основа на стапките на несофпаѓање врз регионот на интерес или целиот ампликон, во зависност од должината на ампликонот. Филтрираните порамнувања се впишуваат во датотеките за порамнување како непорамнети и не се користат при одредувањето варијанти.

Одредување варијанти

Развиено од Illumina, одредувачот на Pisces варијанта на идентификува варијанти присутни со ниска фреквенција во примерокот на ДНК.

Одредувачот варијанта на Pisces ги идентификува SNV, MNV и малите индели во три чекори:

- ▶ Ја разгледува секоја позиција во референтниот геном одделно
- ▶ Ги брои базите на дадената позиција за порамнети отчитувања што се преклопуваат со позицијата
- ▶ Пресметува варијантен резултат што го мери квалитетот на одредувањето користејќи го моделот на Poisson. Варијантите со оценка за квалитет под Q30 се исклучени.

Варијантите прво се повикуваат за секој на група одделно. Потоа, варијанти од секоја група насе споредени и комбинирани во една излезна датотека. Ако варијантата е присутна во двете групи и ги поминува сите филтри наведени во [Анотации на VCF датотеки на страница 13](#), варијантата е означена како ПОМИНАТО во датотеката за одредување на варијанта (VCF).

Прикажи податоци за обработка и примероци

- 1 Од контролната табла Local Run Manager, кликнете го името на обработката.
- 2 Од картичката Run Overview (Преглед на обработка), прегледајте ги мерните показатели за обработката со секвенционирање.
- 3 **[Изборно]** Кликнете ја иконата **Copy to Clipboard**  (Копирај на таблата) за да ја копирате патеката до датотеката во која се наоѓа излезната папка за обработката.
- 4 Кликнете на картичката Sequencing Information (Информации за секвенционирање) за да ги ревидирате параметрите за обработката и информациите за потрошните материјали.
- 5 Кликнете на картичката Samples and Results (Примероци и резултати) за да ја видите локацијата на извештајот за анализата.
 - Ако била повторена анализата, проширете ја паѓачката листа Select Analysis (Избор на анализа) и изберете ја соодветната анализа.
- 6 Кликнете на иконата **Copy to Clipboard**  (Копирај на таблата) за да ја копирате патеката до папката Analysis (Анализа).

За повеќе информации за картичките Run Overview (Преглед на обработка) и Sequencing Information (Информации за секвенционирање), како и за тоа како повторно да се постави анализата во ред за извршување, погледнете во *Референтен водич за инструментот NextSeq 550Dx (број на документ 1000000009513)*.

Извештај за анализа

Резултатите од анализата се сумирани на картичката Samples and Results (Примероци и резултати), а како збиен извештај во папката Alignment (Порамнување). Извештај за секој примерок е достапен и во PDF формат на датотека за секој примерок.

Информации за картичка Примероци и Резултати

- 1 Кликнете на примерок во листата за да го видите извештајот за примерокот.

Табела 1 Информации за обработката и примерокот

Наслов на колона	Опис
Статус на обработка	Наведува дали секвенционирањето е успешно или неуспешно.
Вкупен принос (GB)	Број на бази одредени во секвенционирањето. Го прикажува прагот на поминување и статусот на поминување или неуспешно поминување.
% ≥ Q30	Процентот на отчитувањата во обработка со секвенционирање со резултат за квалитет од 30 (Q30) или повисоки. Го прикажува прагот на поминување и статусот на поминување или неуспешно поминување.
Име на примерокот	Името на примерокот дадено при креирањето на обработката.
Вкупно PF отчитувања	Вкупниот број отчитувања кои поминуваат низ филтерот.

Наслов на колона	Опис
Отчитување 1 % $\geq$ Q30	Процентот на отчитувања од Отчитување 1 со резултати за квалитет од 30 (Q30) или повисоки.
Отчитување 2 % $\geq$ Q30	Процентот на отчитувања од Отчитување 2 со резултати за квалитет од 30 (Q30) или повисоки.
Стапка на одредување на автозоми	Бројот на геномски позиции низ автозомите (хромозоми од 1 до 22) што исполнуваат претходно дефиниран праг на вредност на доверба, поделен со вкупниот број на испитани автозомски геномски позиции. Стапката на повикување е опишана врз основа на примерок и е прикажана како процент што се пресметува како 1 минус (број на автозомни позиции со нецелосни повикувања поделен со вкупниот број на секвенционирани автозомни позиции).

Табела 2 Извештаи за примероци за Информации

Наслов на колона	Опис
Примерок	Името на примерокот дадено при креирањето на обработката.
Датум на извештајот	Датум кога е генериран извештајот.
Информации за примерокот	ИД на примерокот што беше даден при креирањето на обработката, вкупните отчитувања што го поминале филтерот во примерокот, процентот на отчитувања за примерокот со резултат за квалитет од 30 (Q30) или поголем и автозомната стапка на одредување.
Резиме на ампликон	Вкупен број на секвенционирани ампликонски региони и вкупната должина во базни парови на секвенционирани ампликони во целните региони, за примерокот во Групата А и Групата Б, и манифестната датотека што се користи за секоја група. Манифестната датотека ги специфицира референтниот геном и целните референтни региони што се користат во чекорот на порамнување.
Статистика за ниво на отчитување	Број и процент на отчитувања за примерокот што ја покриваат секоја позиција во референцата, за Отчитување 1 и Отчитување 2 во Група А и Група Б.
Резиме на варијанти	Број на SNV, инсерции и делеции откриени за примерокот што ги поминал предложените вредности за да се утврди дали резултатите за квалитет се во прифатлив опсег,
Резиме на покриеност	Вкупниот број на порамнети бази поделен со големината на целиот регион и процентот на ампликонски региони со вредности на покриеност поголеми од нискиот праг на покриеност од 0,2 * средна покриеност на ампликони, за примерокот во Групата А и Групата Б.
Графикони за покриеност	Графиконот за покриеност по ампликонски регион ја покажува покриеноста низ ампликонските региони за примерокот. Регионите со вредности на покриеност пониски од прагот на покриеност се означени со црвена боја. Просекот од сите вредности е означен со портокалова линија. Обезбеден е графикон за покриеност на Група А и Група Б.
Верзии на софтвер	Верзии на софтверот кога примерокот беше секвенциониран. Вклучува Оперативен софтвер за NextSeq 550Dx (NOS), Софтвер за Local Run Manager, софтвер RTA и верзија на модулот за соматска.

Излезни датотеки на анализа

Следните излезни датотеки на анализата се генерирани за Соматска варијанта модул за анализа и да се обезбедат резултати од анализата за порамнување и одредување на варијанти. Излезните датотеки на анализата се наоѓаат во папката Alignment (Порамнување).

Назив на датотека:	Опис
Демултиплексирање (*.txt)	Меѓудатотеки што содржат сумарни резултати од демултиплексирање.
FASTQ (*.fastq.gz)	Меѓудатотеки што содржат основни одредувања со оценка на квалитет. FASTQ-датотеките се примарен материјал за чекорот на порамнување.
Датотеки за порамнување во BAM формат (*.bam)	Содржи порамнети отчитувања за даден примерок.
Датотеки за одредување на варијанти по базен во VCF формат (*.vcf)	Содржи варијанти одредени на секоја позиција од предната или обратната група.
Датотеки со одредување варијанти во геномот VCF формат (*.genome.vcf.gz)	Го содржи го генотипот за секоја позиција, без разлика дали е одреден како варијанта или како референца.
Датотеки за консензусни одредувања варијанти во VCF формат (*.vcf.gz)	Содржи варијанти одредени на секоја позиција од двете групи.
AmpliconCoverage_M1.tsv	Содржи информации за покриеноста по ампликон по примерок за секој обезбеден манифест. M# го претставува манифестниот број.

Формат на датотека за демултиплексирање

Процесот на демултиплексирање ја отчитуваа индексната секвенца прикачена на секој кластер за да утврди од кој примерок потекнува кластерот. Мапирањето помеѓу кластерите и бројот на примерокот се запишува во демултиплексирана (*.demux) датотека за секој сликоред од проточната ќелија.

Форматот за именување на датотеките за демултиплексирање е **s_1_X.demux**, каде што X е бројот на сликоред.

Датотеките за демултиплексирање започнуваат со заглавие:

- ▶ Верзија (цел број од 4 бајти), моментално 1
- ▶ Број на кластери (цел број од 4 бајти)

Остатокот од датотеката се состои од примерочни броеви за секој кластер од сликоредот.

Кога чекорот на демултиплексирање е завршен, софтверот генерира датотека за демултиплексирање наречена **DemultiplexSummaryF1L1.txt**.

- ▶ Во името на датотеката, **F1** го претставува бројот на проточната ќелија.
- ▶ Во името на датотеката, **L1** го претставува бројот на лентата.
- ▶ Демултиплексирањето резултира со табела со 1 ред по сликоред и 1 колона по примерок, вклучувајќи го и примерокот 0.
- ▶ Најчесто појавени секвенци при индексни отчитувања.

Формат на датотека FASTQ

FASTQ е текстуален формат на датотека што содржи одредувања на бази и вредности за квалитет по отчитување. Секој запис содржи 4 реда:

- ▶ Идентификаторот
- ▶ Секвенцата
- ▶ Знак плус (+)

- Резултатите за квалитет на Phred во ASCII + 33 кодиран формат

Идентификаторот е форматиран како:

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

BAM формат на датотека

BAM-датотеката (*.bam) е компресирана бинарна верзија на SAM-датотека што се користи за претставување на порамнети секвенци до 128 Mb. SAM и BAM форматите се детално опишани на samtools.github.io/hts-specs/SAMv1.pdf.

BAM-датотеките го користат форматот за именување на датотеки **SampleName_S#.bam**, каде што # е бројот на примерокот определен според редоследот по кој примероците се наведени за извршувањето.

BAM датотеките содржат дел за заглавие и дел за порамнување:

- **Заглавие** - Содржи информации за целата датотека, како што се име на примерокот, должина на примерокот и метод на порамнување. Порамнувањата во делот за порамнувања се поврзани со специфични информации во делот за заглавие.
- **Порамнувања** - Содржи име на отчитување, секвенца на отчитување, квалитет на отчитување, информации за порамнување и приспособени ознаки. Името на отчитување ги вклучува хромозомот, почетната координата, квалитетот на порамнувањето и низата за опишување на совпаѓањето.

Делот за порамнувања ги содржи следните информации за секое отчитување или пар отчитувања:

- **AS:** Квалитет на порамнување на спарени краеве
- **BC:** Ознака за баркод, која го означува демултиплексираниот ИД на примерокот поврзан со отчитувањето.
- **SM:** Квалитет на порамнување на еден крај.
- **XC:** Низа за опис на совпаѓање
- **XN:** Ознака за име на ампликон, која го евидентира ИД на ампликонот поврзан со отчитувањето

BAM индексните датотеки (*.bam.bai) обезбедуваат индекс на соодветната BAM датотека.

VCF формат на датотека

Форматот за одредување варијанти (VCF) е вообичаен формат на датотека развиен од научната заедница за геномика. Содржат информации за варијанти пронајдени на специфични позиции во референтниот геном. VCF датотеките завршуваат со наставката .vcf

Заглавието на VCF-датотеката ја вклучува верзијата на форматот на VCF-датотеката и верзијата на одредувачот на варијанти и ги наведува анотациите што се користат во остатокот од датотеката. Заглавието на VCF исто така ја вклучува референтната датотека со геном и

BAM датотеката. Последниот ред во заглавието ги содржи насловите на колоните за редовите со податоци. Секоја од линиите со податоци на датотеката VCF содржи информации за една варијанта.

Наслови на VCF датотеки

Наслов	Опис
CHROM	Хромозомот на референтниот геном. Хромозомите се појавуваат по истиот редослед како и референтната FASTQ-датотека.
POS	Позицијата на една база на варијантата во референтниот хромозом. За SNP, оваа позиција е референтната база со варијантата; за индели или делеции, оваа позиција е референтната база веднаш пред варијантата.
ID	rs бројот за варијантата добиена од dbSNP.txt, доколку е применливо. Ако на оваа локација има повеќе rs броеви, листата е разграничена со точка-запирка. Ако на оваа позиција не постои dbSNP запис, се користи маркер за недостасувачка вредност ('.').
REF	Референтен генотип. На пример, делецијата на еден T е претставена како референтен TT и алтернативен T. Варијанта од A до T со еден нуклеотид е претставена како референтен A и алтернативен T.
ALT	Алелите што се разликуваат од референтното отчитување. На пример, инсерција на едно T е претставено како референтно A и алтернативно AT. Варијанта од еден нуклеотид од A до T е претставена како референтен A и алтернативен T.
QUAL	Phred-скалирана оценка за квалитет доделена од одредувачот на варијантата. Повисоките резултати укажуваат на поголема доверба во варијантата и помала веројатност за грешки. За оценка за квалитет Q, проценетата веројатност за грешка е $10^{-(Q/10)}$. На пример, множеството на одредувања Q30 има стапка на грешка од 0,1%. Многу одредувачи на варијанти доделуваат оценки за квалитет врз основа на нивните статистички модели, кои се високи во однос на забележаната стапка на грешки.

Анотации на VCF датотеки

Наслов	Опис
ФИЛТЕР	Ако сите филтри се поминати, PASS (ПОМИНАТО) се запишува во колоната за филтрирање. • LowDP - Се применува на локации со длабочина на покриеност под 450x во која било група. За позициите на ампликоните опфатени и со напредно и со обратно отчитување, ова е еквивалентно на покриеност од 900x единечно отчитување. • LowGQ - Квалитетот на генотипизацијата (GQ) е под граничната вредност. • q30 - Оценка за квалитет < 30. • Ниска фреквенција на варијанта - Фреквенцијата на варијантата е помала од дадениот праг. • PB - Пристрасност на групата на сонди. Варијантата не е пронајдена или е пронајдена со ниска фреквенција во една или две групи со сонди. • R3x6 - Бројот на соседни повторувања (со должина од 1 до 3 bp) на одредувањата на варијантните е ≥ 6. • SB - Пристрасноста на нишката е поголема од дадениот праг.
ИНФО	Можни записи во колоната INFO вклучуваат: • AC - Број на алели во генотипите за секој ALT алел, по истиот редослед како што е наведен. • AF - Фреквенција на алели за секој алел на ALT, по истиот редослед како што е наведен. • AN - Вкупен број на алели во веќе одредени генотипи. • CD - Знаменце што означува дека SNP се јавува во рамките на кодирачкиот регион од најмалку 1 RefGene запис. • DP - Длабочината (број на одредувања на бази порамнети со позиција и користени при одредување на варијанта). • Егзон - Список на егзонски региони одделен со запирки отчитан од RefGene. • FC - Функционална последица. • GI - список на генски идентификатори одделен со запирки отчитан од RefGene. • QD - Доверба/Квалитет на варијанта според длабочината. • TI - Список на идентификатори на транскрипти одделен со запирки отчитан од RefGene.

Наслов	Опис
ФОРМАТ	Колоната за формат ги наведува полињата одделени со две точки. На пример, GT:GQ. Листата на дадени полиња зависи од користениот одредувач на варијанта. Достапните полиња вклучуваат: • AD - Внес во форма X,Y, каде што X е бројот на референтни одредувања, а Y е бројот на алтернативни одредувања. • DP - Приближна длабочина на отчитување; отчитувањата со MQ=255 или со лоши партнери се филтрираат. • GQ - Квалитет на генотипот. • GQX - Квалитет на генотипот. GQX е минимумот од вредноста GQ и колоната QUAL. Генерално, овие вредности се слични; земањето на минимумот го прави GQX поконзервативна мерка за квалитетот на генотипот. • GT - Генотип. 0 одговара на референтната база, 1 одговара на првиот запис во колоната ALT и така натаму. Косата црта (/) означува дека не се достапни информации за фазирање. • NL - Ниво на бучава; проценка на бучавата при одредување на базата на оваа позиција. • PB - Пристрасност на групата на сонди. Вредностите поблиски до 0 укажуваат на поголема пристрасност кон една група сонди и помала доверба во одредување на варијанта. • SB - Пристрасност на нишката на оваа позиција. Поголемите негативни вредности укажуваат на помала пристрасност; вредностите близу 0 укажуваат на поголема пристрасност. • VF - Варијантна фреквенција; процентот на отчитувања што го поддржуваат алтернативниот алел.
ПРИМЕРОК	Колоната со примерок ги дава вредностите наведени во колоната FORMAT (ФОРМАТ).

Геномски VCF датотеки

VCF датотеките на геномот (gVCF) се датотеки VCF v4.1 кои следат збир на конвенции за претставување на сите места во геномот во разумно компактен формат. Датотеките gVCF (*.genome.vcf.gz) ги вклучуваат сите локации во регионот од интерес во една датотека за секој примерок.

Датотеката gVCF прикажува no-calls (неодредување) на позиции кои не ги поминуваат сите филтри. Ознаката за генотип (GT) од ./. означува дека нема одредување.

За повеќе информации, погледнете на sites.google.com/site/gvcftools/home/about-gvcf.

VCF датотеки по група и консензус

Работниот тек на Соматска варијанта генерира 2 сета датотеки за одредување варијанти.

- ▶ **VCF датотеки по група** - Содржи варијанти одредени или во напредната група или во обратната група. Датотеките по група се запишуваат во папката VariantCallingLogs.
- ▶ **Консензусни VCF датотеки** - Содржи варијанти одредени од двете групи. Датотеките со консензус се запишуваат во папката Alignment (Порамнување).

VCF датотеките по група и консензусните датотеки вклучуваат VCF (*.vcf) и gVCF (*.genome.vcf) датотеки и ја користат следната конвенција за именување, каде што S# го претставува редоследот по кој примерокот е наведен за обработката:

- ▶ **Извештаи за сите локации** - SampleName_S#.genome.vcf
- ▶ **Пријавува само варијанти** - SampleName_S#.vcf

Софтверот ги споредува VCF датотеките по група и ги комбинира податоците на секоја позиција за да создаде консензус VCF датотека за примерокот.

Одредувањата на варијанти од секоја група се спојуваат во консензусни VCF датотеки користејќи ги следните критериуми.

Критериуми	Резултат
Референтно одредување во секоја група	Референтно одредување
Референтно одредување во 1 група и одредување на варијанта во другата група	Филтрирано одредување на варијанта
Совпаѓање на одредување варијанти со слични фреквенции во секоја група	Одредување варијанта
Совпаѓање на одредување варијанти со значително различни фреквенции во секоја група	Филтрирано одредување на варијанта
Неспоредливо одредување на варијанти во секоја група	Филтрирано одредување на варијанта

Мерните показатели од секоја група се спојуваат со користење на следните вредности.

Мерен показател	Вредност
Длабочина	Додавање на длабочини од двете групи
Фреквенција на варијанта	Вкупен број на варијанти поделени со вкупната длабочина на покриеност
Q-Оценка	Минимална вредност на двете групи

Датотека со покриеност на ампликон

За секоја манифестна датотека се генерира датотека со покриеност на ампликони. M# во името на датотеката го претставува бројот на манифестот.

Секоја датотека вклучува ред во заглавието што ги содржи ИД-те на примероците поврзани со манифестот. Датотеката ги содржи следниве информации.

- ИД на целта како што е наведен во манифестот.
- Длабочината на покриеност на отчитувањата што минуваат низ филтерот.

Дополнителни излезни датотеки

Следните излезни датотеки даваат дополнителни информации или ги сумираат резултатите од обработката и грешките во анализата. Иако овие датотеки не се потребни за проценка на резултатите од анализата, тие можат да се користат за решавање проблеми. Сите датотеки се наоѓаат во папката Alignment (Порамнување) освен ако не е поинаку наведено.

Назив на датотека:	Опис
AnalysisLog.txt	Дневник на обработка што го опишува секој чекор што се случил за време на анализата на тековната папка за обработка. Оваа датотека не содржи пораки за грешка. Се наоѓа во папката Alignment (Порамнување).
AnalysisError.txt	Дневник за обработка што ги наведува сите грешки што се случиле за време на анализата. Оваа датотека ќе биде празна ако не се појават грешки. Се наоѓа во папката Alignment (Порамнување).
DemultiplexSummaryF1L1#.txt	Ги прикажува резултатите од демултипликсањето во табела со 1 ред по сликоред и 1 колона по примерок. # ја претставува лентата 1, 2, 3 или 4 од проточната ќелија. Се наоѓа во папката Alignment (Порамнување).
AmpliconRunStatistics.xml	Содржи резиме на статистика специфична за обработката. Се наоѓа во папката Alignment (Порамнување).

Папка Analysis (анализа)

Папката за анализи ги содржи датотеките генерирани од Local Run Manager софтвер.

Врската помеѓу излезната папка и папката за анализа е сумирана на следниов начин:

- ▶ За време на секвенционирањето, Real-Time Analysis (RTA) ја пополнува излезната папка со датотеки генерирани за време на анализата на сликата, одредувањето на базата и оценувањето на квалитетот.
- ▶ RTA ги копира датотеките во папката за анализа во реално време. Откако RTA ќе додели оценка за квалитет на секоја база за секој циклус, софтверот ја запишува датотеката RTAComplete.txt во обете папки.
- ▶ Кога датотеката RTAComplete.txt е присутна, анализата започнува.
- ▶ Како што продолжува анализата, Local Run Manager ги запишува излезните датотеки во папката за анализа, а потоа ги копира датотеките назад во излезната папка.

Папки Alignment (порамнување)

Секој пат кога таа анализа е повторно ставена во ред за извршување, Local Run Manager креира папка Alignment (Порамнување) со име **Alignment_N**, каде што N е секвенцијален број.

Структура на папки

 **Alignment** - Содржи *.bam, *.vcf, FASTQ-датотеки и датотеки специфични за модулот за анализа.

 **Date and Time Stamp** - Ознака за датум_време на анализата како 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**

 Sample1_L001_R1_001_fastq.gz

 **Stats**

 DemuxSummaryF1L1.txt

 FastqSummaryF1L1.txt

 **Data**

 **Intensities**

 **BaseCalls**

 **L001** - Содржи *.bcl датотеки.

 **L001** - Содржи *.locs датотеки.

 **RTA Logs** - Содржи дневници од анализа на RTA софтверот.

 **InterOp** - Содржи бинарни датотеки што се користат за пријавување на мерни показатели за обработка со секвенционирање.

 **Logs** - Содржи датотеки за евиденција што ги опишуваат чекорите извршени за време на секвенционирањето.

 RTAComplete.txt

 RunInfo.xml



RunParameters.xml

Одредување на бази и индексна разновидност

Кога примероците се секвенционираат на NextSeq 550Dx инструментот, одредувањето на базата одредува база (A, C, G или T) за секој кластер на даден сликоред или област за снимање на проточната ќелија, во одреден циклус. Инструментот NextSeq 550Dx користи двоканално секвенционирање, за што се потребни само две слики за да се шифрираат податоците за четири ДНК бази, една од црвениот канал и една од зелениот.

Процесот за одредување на индексни отчитувања на база се разликува од одредувањето на база за време на други отчитувања.

Индексните отчитувања мора да започнуваат со најмалку една база која не е G во кој било од првите два циклуси. Ако индексното отчитување започнува со две одредувања бази G, нема да се генерира интензитет на сигналот. Мора да има сигнал во кој било од првите два циклуси за да се осигури перформанс на демултиплексирање.

При избирање на индекси за време на креирањето на обработка, ќе се појави предупредување за ниска разновидност ако индексите не ги исполнуваат барањата за разновидност. За да го спречите предупредувањето за ниска разновидност, изберете индексни секвенци што обезбедуваат сигнал во двата канали за секој циклус.

- ▶ Црвен канал - A или C
- ▶ Зелен канал - A или T

Процесот за одредување бази осигурува точност при анализирањето примероци со низок плекс. За повеќе информации за секвенците на вашите индекси, видете го приложеното упатство *TruSeq Custom Amplicon Kit Dx (број на документ 1000000029772)*.

За време на креирањето на обработката во Local Run Manager ќе го изберете бројот на примероци што ќе се тестираат. Предложените комбинации на индекси што ги исполнуваат барањата за разновидност на индекси автоматски се пополнуваат од софтверот. Иако не сте обврзани да ги користите предложените комбинации на индекси, тоа е препорачливо.

Претходни ревизии

Документ	Датум	Опис на измената
Бр. на документ 1000000030330 v04	Август 2021 г.	Ажурирана е адресата на овластениот застапник за ЕУ.
Бр. на документ 1000000030330 v03	Декември 2019 г.	Ажурирана е адресата на овластениот застапник за ЕУ. Ажурирана е адресата на австралискиот спонзор.
Бр. на документ 1000000030330 v02	Јануари 2019 г.	Додадени се информации за комплетите со реагенси v2.5.
Бр. на документ 1000000030330 v01	Август 2018 г.	Ажурирани се регулаторните ознаки.
Бр. на документ 1000000030330 v00	Ноември 2017 г.	Првично издание.

Техничка помош

За техничка помош, контактирајте со техничката поддршка на Illumina.

Веб-локација: www.illumina.com
Е-пошта: techsupport@illumina.com

Illumina Телефонски броеви на поддршката за корисници

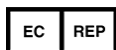
Регион	Бесплатен телефонски број	Регионално
Северна Америка	+1.800.809.4566	
Австралија	+1.800.775.688	
Австрија	+43 800006249	+43 19286540
Белгија	+32 80077160	+32 34002973
Кина	400.066.5835	
Данска	+45 80820183	+45 89871156
Финска	+358 800918363	+358 974790110
Франција	+33 805102193	+33 170770446
Германија	+49 8001014940	+49 8938035677
Хонг Конг	800960230	
Ирска	+353 1800936608	+353 016950506
Италија	+39 800985513	+39 236003759
Јапонија	0800.111.5011	
Холандија	+31 8000222493	+31 207132960
Нов Зеланд	0800.451.650	
Норвешка	+47 800 16836	+47 21939693
Сингапур	+1.800.579.2745	
Шпанија	+34 911899417	+34 800300143
Шведска	+46 850619671	+46 200883979
Швајцарија	+41 565800000	+41 800200442
Тајван	00806651752	
Обединето Кралство	+44 8000126019	+44 2073057197
Останати земји	+44.1799.534000	

Безбедносни листови (SDS) - Достапни на веб-локацијата на Illumina на support.illumina.com/sds.html.

Документација за производот - Достапна за преземање во PDF од веб-страницата на Illumina. Одете на support.illumina.com, изберете производ, а потоа изберете **Documentation & Literature** (Документација и литература).



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



Illumina Netherlands B. V.
Steenoven 19
5626 DK Ајндховен
Холандија

Австралиски спонзор
Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
Melbourne, VIC 3000
Австралија