

DRAGEN for Illumina DNA Prep with Enrichment Dx

Документація на продукт для NovaSeq 6000Dx

ВЛАСНІСТЬ КОМПАНІЇ ILLUMINA

Документ № 200014776, вер. 02

Вересень 2022 р.

ВИКОРИСТОВУВАТИ ДЛЯ ДІАГНОСТИКИ IN VITRO

Історія редакцій

Документ	Дата	Опис зміни
200014776, версія 02	Вересень 2022 р.	Виправлено формат файлу маніфесту з текстового (*.txt) на BED (*.bed) в інструкціях зі створення прогону. Виправлено «консенсусні файли VCF на «файли VCF» у розділі вихідних даних аналізу.
200014776, версія 01	Серпень 2022 р.	Додано: Розділ «Налаштування». Розділ «Фільтрація системного шуму». Оновлено інструкції зі створення прогону, щоб додати докладнішу інформацію. Виправлено друкарські та граматичні помилки. Уточнено, що інструкції призначені для застосунку під час його використання із секвенатором NovaSeq 6000Dx. Оновлено інформацію про вміст вихідного файлу VCF.
200014776, версія 00	Березень 2022 р.	Початкова редакція.

Цей документ і його вміст є власністю компанії Illumina, Inc. та її філій (надалі — Illumina). Він призначений лише для того, щоб користувач використовував вироби тільки за угодою в цілях, описаних у цьому документі. Цей документ і його вміст не слід використовувати або поширювати з будь-якою іншою метою та/або для іншого обговорення, розкриття або відтворення тим або іншим чином без попередньої письмової згоди компанії Illumina. Цим документом компанія Illumina не надає жодного дозволу на свій патент, товарний знак, авторське право або загальноприйняті права, а також на подібні права будь-яких третіх сторін.

Щоб гарантувати правильне та безпечне використання виробів, описаних у цьому документі, кваліфікований і належним чином навчений персонал повинен суворо та чітко дотримуватись інструкцій, описаних у цьому документі. Перед використанням цих виробів потрібно повністю прочитати й зрозуміти весь вміст цього документа.

НЕПОВНЕ ВИВЧЕННЯ ВСІХ ЗАЗНАЧЕНИХ У ЦЬОМУ ДОКУМЕНТІ ВКАЗІВОК І ЇХ НЕЧІТКЕ ДОТРИМАННЯ МОЖЕ ПРИЗВОДИТИ ДО ПОШКОДЖЕННЯ ЦИХ ВИРОБІВ, ТРАВМУВАННЯ ЛЮДЕЙ, ЗОКРЕМА КОРИСТУВАЧІВ АБО ІНШИХ ОСІБ, І ДО ПОШКОДЖЕННЯ ІНШОЇ ВЛАСНОСТІ, А ТАКОЖ ПРИЗВЕДЕ ДО ВТРАТИ БУДЬ-ЯКИХ ГАРАНТІЙНИХ ЗОБОВ'ЯЗАНЬ, ЗАСТОСОВНИХ ДО ЦИХ ВИРОБІВ.

КОМПАНІЯ ILLUMINA НЕ НЕСЕ ЖОДНОЇ ВІДПОВІДАЛЬНОСТІ, ЩО ВИНΙΚАЄ ВНАСЛІДОК НЕНАЛЕЖНОГО ВИКОРИСТАННЯ ВИРОБІВ, ОПИСАНИХ У ЦЬОМУ ДОКУМЕНТІ (ВКЛЮЧНО З ЙОГО ЧАСТИНАМИ АБО ПРОГРАМНИМ ЗАБЕЗПЕЧЕННЯМ).

© 2022 Illumina, Inc. Усі права застережено.

Усі товарні знаки є власністю компанії Illumina, Inc. або їхніх відповідних власників. Конкретну інформацію про товарні знаки зазначено на сторінці www.illumina.com/company/legal.html.

Зміст

Історія редакцій	ii
Огляд	1
Методи аналізу	1
Створення прогону	5
Налаштування	7
Вихідні результати аналізу	8
Файли FASTQ	9
Файли BAM	9
Файли VCF	10
Перегляд результатів аналізу	15
Технічна допомога	16

Огляд

Застосунок DRAGEN™ for Illumina® DNA Prep with Enrichment Dx виконує демультіплексування, генерування FASTQ, зіставлення з еталонним геномом та розпізнавання варіантів, залежно від вибраного робочого процесу.

Методи аналізу

Застосунок DRAGEN for Illumina DNA Prep with Enrichment Dx виконує демультіплексування, генерування FASTQ, зіставлення зчитувань та вирівнювання відповідно до еталонного геному залежно від вибраного робочого процесу:

- генерування файлів FASTQ;
- генерування файлів FASTQ та VCF зародкової лінії;
- генерування файлів соматичних варіантів FASTQ та VCF.

Генерування файлів FASTQ

Зібрані послідовності записуються у файли FASTQ, по одному файлу на зразок. Файли FASTQ — це текстові файли, які містять дані секвенування та показники якості лише для одного зразка. До кожного зразка створюються окремі файли FASTQ для кожної доріжки проточної кювети та для кожного зчитування секвенування. Назва зразка, вказана під час налаштування прогону, входить до назви файлу FASTQ. Файли FASTQ містять первинні вхідні дані для вирівнювання. Перший етап створення FASTQ — це демультіплексування. Демультіплексування призначає кластери, що пройшли фільтрацію, конкретному зразку шляхом порівняння кожної послідовності зчитування індексу з послідовностями індексів, указаними для прогону. На цьому етапі якісні значення не враховуються. Зчитування індексів ідентифікуються під час описаних далі етапів.

- Зразки нумеруються від 1 в порядку, у якому їх перелічено для прогону.
- Номер зразка 0 резервується для кластерів, які не було присвоєно зразку.
- Кластери присвоюються зразку в разі точної відповідності послідовності індексів або коли є одна розбіжність на зчитування індексу.

У програмне забезпечення входить стискання ORA для стиснення файлів FASTQ. У разі використання формату ORA (*.ora) контрольна сума md5 вмісту FASTQ зберігається після циклу стиснення та декомпресії, що гарантує стиснення без втрат.

Зіставлення та вирівнювання ДНК

Першим етапом зіставлення є створення шаблонів на основі зчитування з подальшим пошуком точних збігів в еталонному геномі. Після цього отримані результати уточнюються шляхом повного вирівнювання за алгоритмом Сміта — Уотермана для ділянок із найвищою щільністю

збігів шаблонів. Цей добре задокументований алгоритм працює шляхом порівняння кожної позиції зчитування з усіма потенційними позиціями еталону. Такі порівняння відповідають матриці можливих варіантів вирівнювання між зчитуванням та еталоном. Для кожної з цих потенційних позицій зіставлення алгоритм Сміта — Уотермана генерує бали, за допомогою яких можна оцінити, яким чином найкраще вирівнювання, що проходить через цю комірку матриці, досягає її: через збіг або розбіжність нуклеотидів (діагональний рух), делецію (горизонтальний рух) або інсерцію (вертикальний рух). Збіг між зчитуванням і еталоном дає бонус до бала, а невідповідність або індел накладає штраф. Шлях із найбільшим загальним балом у матриці обирається як остаточне вирівнювання.

Конкретні значення балів, обрані для цього алгоритму, визначають, як під час вирівнювання з кількома можливими інтерпретаціями збалансувати ймовірність наявності індела порівняно з одним або кількома SNP або надати перевагу вирівнюванню без обрізання. Стандартні значення балів DRAGEN є обґрунтованими для вирівнювання зчитувань середньої довжини щодо цілого еталонного геному людини для завдань розпізнавання варіантів. Будь-який набір параметрів оцінювання Сміта — Уотермана є неточною моделлю геномних мутацій та помилок секвенування. Для певних прикладних завдань доцільнішим може бути використання інакше налаштованих значень балів вирівнювання.

Розпізнавання варіантів зародкової лінії за допомогою DRAGEN

Програма розпізнавання малих варіантів зародкової лінії DRAGEN приймає зіставлені й вирівняні зчитування ДНК як вхідні дані та розпізнає SNP та індели шляхом поєднання постовпчикowego виявлення та локального складання гаплотипів *de novo*.

Спочатку визначаються еталонні ділянки з достатнім покриттям вирівнювання, доступні для розпізнавання. У межах цих еталонних ділянок за допомогою швидкого сканування відсортованих зчитувань ідентифікуються активні ділянки, центровані навколо стовпців накопичення, що свідчать про наявність варіанта. Активні ділянки доповнюються достатнім контекстом, щоб охопити значущий нееталонний вміст поблизу. За наявності ознак інделів активні ділянки отримують додаткове заповнення.

Вирівняні зчитування обрізаються в межах кожної активної ділянки та збираються у граф де Брюїна. Краї обрізаних зчитувань зважуються за кількістю спостережень, при цьому еталонна послідовність виступає основою. Після очищення та спрощення графа всі шляхи від джерела до стоку витягуються як потенційні гаплотипи. Кожен гаплотип вирівнюється відносно еталонного геному за алгоритмом Сміта — Уотермана для ідентифікації варіантів, які він представляє. Цей набір подій може бути доповнений шляхом виявлення на основі позиції. Для кожної пари «зчитування-гаплотип» ймовірність $P(r|H)$ спостереження зчитування за умови, що гаплотип є справжнім вихідним зразком, оцінюється за допомогою парної прихованої марковської моделі.

Шляхом сканування за еталонною позицією в межах активної ділянки формуються потенційні генотипи з диплоїдних комбінацій варіантів (SNP або інделів). Для кожної події (включно з еталоном) умовна ймовірність $P(r|e)$ спостереження кожного зчитування з перекриванням оцінюється як максимальна $P(r|H)$ для гаплотипів, що підтримують цю подію. Вони об'єднуються в

умовну ймовірність $P(r|e1e2)$ для генотипу (пари подій) і множаться для отримання умовної ймовірності $P(R|e1e2)$ спостереження всього накопичення зчитувань. За формулою Байєса обчислюється апостеріорна ймовірність $P(e1e2|R)$ кожного диплоїдного генотипу, після чого обирається найбільш імовірний варіант.

У режимі gVCF, що використовується для масштабованого розпізнавання варіантів у багатьох зразках, програму розпізнавання малих варіантів зародкової лінії DRAGEN можна запускати для кожного зразка окремо для створення проміжного файлу розпізнавання геномних варіантів (gVCF). Після цього gVCF можна використовувати для ефективного спільного генотипування кількох зразків, що дозволяє здійснювати швидку інкрементну обробку зразків та масштабування до великих розмірів когорт.

Оскільки програма розпізнавання малих варіантів зародкової лінії DRAGEN має алгоритми, що дають змогу ефективно відрізнити корельовані помилки від справжніх варіантів, правила фільтрації дуже прості.

Розпізнавання соматичних варіантів за допомогою DRAGEN

Програма розпізнавання малих соматичних варіантів DRAGEN приймає зіставлені й вирівняні зчитування ДНК як вхідні дані та розпізнає SNV та індели шляхом локального складання гаплотипів *de novo* в активній ділянці.

Спочатку визначаються еталонні ділянки з достатнім покриттям вирівнювання, доступні для розпізнавання. У межах цих еталонних ділянок шляхом сканування відсортованих зчитувань ідентифікуються активні ділянки, центровані навколо стовпців накопичення, що свідчать про наявність варіанта у зчитуваннях пухлини. Активні ділянки доповнюються достатнім контекстом, щоб охопити значущий нееталонний вміст поблизу. За наявності ознак інделів активні ділянки отримують додаткове заповнення.

Вирівняні зчитування обрізаються в межах кожної активної ділянки та збираються у граф де Брюїна. Краї обрізаних зчитувань зважаються за кількістю спостережень, при цьому еталонна послідовність виступає основою. Після очищення та спрощення графа всі шляхи від джерела до стоку витягуються як потенційні гаплотипи. Кожен гаплотип вирівнюється відносно еталонного геному за алгоритмом Сміта — Уотермана для ідентифікації варіантів, які він представляє. Для кожної пари «зчитування-гаплотип» ймовірність $P(r|H)$ спостереження зчитування оцінюється за допомогою парної прихованої марковської моделі за припущення, що гаплотип є справжнім вихідним зразком.

Для визначення бала TLOD програма розпізнавання малих соматичних варіантів DRAGEN спочатку виконує сканування за еталонною позицією для кожної потенційної соматичної події, а також для еталонної події в межах активної ділянки. Умовна ймовірність $P(r|e)$ спостереження кожного зчитування з перекриванням оцінюється як максимальна $P(r|H)$ для гаплотипів, що підтримують цю подію. Вони об'єднуються в умовну ймовірність $P(r|E)$ для гіпотези події (E), яка включає суміш еталонних та потенційних соматичних алелів у діапазоні можливих частот алелів, і

множаться для отримання умовної ймовірності $P(R|E)$ спостереження всього накопичення зчитувань. На основі цього розраховується бал TLOD як доказ того, що алель ALT присутній у зразку пухлини в певному локусі.

Створення прогону

Виконуйте наведені нижче дії, щоб налаштувати прогін через Illumina Run Manager на секвенаторі NovaSeq 6000Dx або через браузер на підключеному до мережі комп'ютері. Дані зразка можна вводити вручну або шляхом імпорту протоколу аналізу.

Налаштування застосунку і прогону

1. На екрані Runs (Прогони) виберіть **Create Run** (Створити прогін).
2. Виберіть застосунок DRAGEN for Illumina DNA Prep with Enrichment Dx, а потім натисніть **Next** (Далі).
3. На екрані Run Settings (Параметри прогону) введіть ім'я прогону. Ім'я прогону ідентифікує прогін на всіх етапах від секвенування до аналізу.
4. **[Необов'язково]** Введіть опис прогону, який допоможе додатково його ідентифікувати.
5. Переконайтеся, що вибрано набір для підготовки бібліотек Illumina DNA Prep with Enrichment Dx.
6. Виберіть потрібний набір індексних адаптерів.
7. Введіть тривалість зчитування.
Зчитування 1 і зчитування 2 має значення за замовчуванням 151 цикл.
Індекс 1 та індекс 2 мають фіксоване значення 10 циклів.
8. **[Необов'язково]** Введіть ідентифікатор пробірки з бібліотекою.
9. Виберіть **Next** (Далі).

Дані зразка

Використовуйте таблицю на екрані Sample Data (Дані зразка) для введення інформації про зразок вручну. Або виберіть **Import Samples** (Імпортувати зразки), щоб завантажити інформацію про зразки. Інформацію щодо імпорту інформації про зразок див. у розділі [Імпорт зразків на стор. 6](#).

Введення зразків вручну

1. Введіть унікальний ідентифікатор зразка в поле Sample ID (Ідентифікатор зразка).
2. Щоб вибрати положення лунки, використовуйте **Plate - Well Position** (Планшет – положення лунки).
Поля i7 Index (Індекс i7), Index 1 (Індекс 1), i5 Index (Індекс i5) та Index 2 (Індекс 2) заповнюються автоматично.
3. **[Необов'язково]** Введіть ім'я бібліотеки.
4. Додайте рядки та за потреби повторіть дії [1–3](#), доки всі зразки не будуть додані до таблиці.
5. Виберіть **Next** (Далі).

Імпорт зразків

Шаблон (*.csv) можна завантажити на екрані Sample Data (Дані зразка) під час планування прогону в Illumina Run Manager через браузер на підключеному до мережі комп'ютері.

1. Щоб завантажити порожній файл CSV, натисніть **Download Template** (Завантажити шаблон).
2. Із файлу CSV введіть дані зразка і збережіть файл.
Файл протоколу аналізу CSV містить такі стовпці даних: Sample ID (Ідентифікатор зразка), Plate - Well Position (Планшет – положення лунки), **[Необов'язково]** Library Name (Ім'я бібліотеки).
3. Виберіть **Import Samples** (Імпорт зразків) і перейдіть до розташування інформаційного файлу CSV.
4. Виберіть **Next** (Далі).

Параметри аналізу

1. Виберіть потрібний робочий процес аналізу:
 - Генерування файлів FASTQ
 - Генерування файлів FASTQ та VCF зародкової лінії для робочого процесу із зародковою лінією
 - Генерування файлів Somatic FASTQ та VCF робочого процесу для соматичних варіантів
2. **[Необов'язково]** За бажанням установіть прапорець **Generate ORA compressed FASTQs** (Створити стиснені FASTQ ORA), щоб увімкнути стиснення FASTQ ORA.
3. **[Робочі процеси генерування VCF]** Використовуйте розкритне меню **Manifest File Selection** (Вибір файлу маніфесту), щоб вибрати файл маніфесту.
Файл маніфесту є обов'язковим вхідним файлом для роботи застосунку DRAGEN for Illumina DNA Prep with Enrichment Dx. Маніфест — це файл BED (*.bed), розділений знаками табуляції, який визначає назви та розташування цільових еталонних ділянок.
4. **[Робочий процес генерування файлів Somatic FASTQ та VCF]** Скористайтеся розкритним меню **Noise File Selection** (Вибір файлу шуму), щоб вибрати файл шуму.
Для фільтрації системного шуму можна вказати файл BED із рівнем шуму для конкретної ділянки. Додаткову інформацію див. у розділі [Фільтрування шуму на стор. 7](#).
5. Виберіть **Next** (Далі).

Перегляд параметрів прогону

1. На екрані Review (Огляд) перегляньте інформацію, введenu на екранах Run Settings (Параметри прогону), Sample Data (Дані зразка) та Analysis Settings (Параметри аналізу).
2. Виберіть **Save** (Зберегти).
Прогін зберігається у вкладці Planned (Заплановані) на екрані Runs (Прогони).

Налаштування

Щоб переглянути й змінити поточні налаштування, виберіть застосунок на екрані Applications (Застосунки).

Конфігурація

На екрані конфігурації відображаються такі налаштування застосунку:

- **Library Prep Kits** (Набори для підготовки бібліотек) — відображає набори для підготовки бібліотек, які використовуються застосунком за замовчуванням. Це налаштування не можна змінити.
- **Набори індексних адаптерів** — відображає набір індексних адаптерів за замовчуванням для програми. Це налаштування не можна змінити.
- **Довжина зчитування** — для застосунку за замовчуванням встановлено довжину зчитування 151 п. о., проте її можна змінити під час створення прогону.
- **Файли маніфесту та шуму** — завантаження та зміни налаштувань файлів маніфесту й системного шуму.
 - Виберіть **Upload File** (Завантажити файл), щоб завантажити файли для використання в аналізі.
 - Виберіть перемикач **Default** (За замовчуванням), щоб файл маніфеста або шуму автоматично обирався під час створення запуску через цей застосунок.
 - Встановіть прапорець **Enabled** (Увімкнено), щоб файл відображався у розкритому меню під час створення прогону.

Дозволи

Використовуйте прапорці на екрані Permissions (Дозволи), щоб керувати доступом користувачів до застосунку.

Фільтрування шуму

Фільтрація системного шуму доступна під час використання робочого процесу для соматичних варіантів. Фільтр можна використовувати в режимі Tumor-Normal (Пухлина — норма), але він особливо корисний для прогонів у режимі Tumor-Only (Лише пухлина), коли відповідний нормальний зразок відсутній.

Файл BED із системним шумом слід генерувати нормальних зразків. Рекоменується створювати файли системного шуму, специфічні для підготовки бібліотек, системи секвенування та панелі. Для створення файлу шуму рекомендовано використовувати приблизно 50 нормальних зразків.

Вихідні результати аналізу

Застосунок DRAGEN for Illumina DNA Prep with Enrichment Dx зберігає наведені нижче відомості в папці аналізу. Лише робочі процеси для гермінальних і соматичних варіантів формують PDF-файл.

- Використаний файл маніфесту
- Версія ПЗ
- Ідентифікатори зразків
- Загальна кількість вирівняних зчитувань
- Відсоток вирівняних зчитувань на зразок
- Кількість SNV, визначених для кожного зразка
- Кількість інделів, визначених для кожного зразка
- Статистичні дані покриття

Вихідні файли аналізу

Застосунок формус наведені нижче вихідні файли. Точний перелік сформованих файлів залежить від того, який робочий процес аналізу використовується. Вихідні файли розташовано в папці аналізу.

Вихідний файл	Опис
FASTQ (*.fastq.gz або *.fastq.ora)	Проміжні файли, що містять розпізнавання азотистих основ із відповідним показником якості. Файли FASTQ містять первинні вхідні дані для етапу вирівнювання. Якщо вибрано формат ORA, це буде відображено в імені файлу.
BAM-файли вирівнювання (*.bam)	Містять вирівняні зчитування для конкретного зразка.
VCF-файли геному (*.gvcf.gz)	Містять генотип для кожної позиції незалежно від того, чи її визначено як варіант, чи як еталон.
Файли VCF (*.vcf.gz)	Містять варіанти, визначені в кожній позиції.
Звіт про показники прогону (*.csv)	Містить показники якості прогону, зокрема загальний вихід і показник Q30.

Файли FASTQ

FASTQ (*.fastq.gz, *.fastq.ora) — це текстовий формат файлів, що містить визначення основ і значення якості для кожного зчитування. Кожен файл містить таку інформацію:

- ідентифікатор зразка;
- послідовність;
- знак плюс (+);
- показники якості Phred у кодованому форматі ASCII + 33.

Ідентифікатор зразка має такий формат:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
Приклад:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

Файли BAM

Файл BAM (*.bam) — це стиснена бінарна версія файлу SAM (зіставлення вирівнювання послідовностей), яка використовується для подання вирівняних послідовностей розміром до 128 Мб. У файлах BAM використовується формат іменування файлів `SampleName_S#.bam`, де # — це номер зразка, який визначається порядком перерахування зразків для прогону. У багатовузловому режимі для S# завжди встановлюється значення S1, незалежно від черговості зразка.

Файли BAM містять розділ заголовка й розділ вирівнювання.

- Header (Заголовок) — містить інформацію про весь файл, таку як ім'я зразка, довжина зразка та метод вирівнювання. Вирівнювання в розділі вирівнювання пов'язане з конкретною інформацією в розділі заголовка.
- Alignments (Вирівнювання) — містить назву зчитування, послідовність зчитування, якість зчитування, інформацію про вирівнювання та користувацькі мітки. Назва зчитування включає в себе хромосому, початкову координату, якість вирівнювання і рядок характеристики відповідності.

Розділ вирівнювання містить таку інформацію для кожного зчитування або пари зчитувань:

- AS: якість вирівнювання парних кінцевих фрагментів;
- RG: групу зчитувань, яка вказує кількість зчитувань для конкретного зразка;

- BC: тег із штрихкодом, який вказує на демультіплексований ідентифікатор зразка, пов'язаний зі зчитуванням;
- SM: якість вирівнювання одиничних кінцевих фрагментів;
- XC: рядок характеристики відповідності;
- XN: мітку з назвою амплікона, яка записує ідентифікатор амплікона, пов'язаний зі зчитуванням. Індексні файли BAM (*.bam.bai) забезпечують індексацію відповідного файлу BAM.

Файли VCF

Файли формату розпізнавання варіантів (*.vcf) містять інформацію про варіанти, виявлені в конкретних позиціях еталонного геному.

Заголовок файлу VCF включає в себе версію формату файлу VCF і версію засобу розпізнавання варіанта, а також список анотацій, що використовуються в іншій частині файлу. У заголовок VCF також включений файл еталонного геному й файл BAM. Останній рядок у заголовку містить заголовки стовпчиків для рядків даних. Кожний із рядків даних файлу VCF містить інформацію про один варіант.

Таблиця 1. Заголовки файлу VCF

Заголовок	Опис
CHROM	Хромосома еталонного геному. Хромосоми відображаються в тому ж порядку, що й довідковий файл FASTA.
POS	Одноосновні позиції варіантів у еталонній хромосомі. Для однонуклеотидних варіантів (SNV) ця позиція відповідає еталонній основі з варіантом. Для інделів ця позиція відповідає еталонній основі, що безпосередньо передуює варіанту.
ID	Номер rs (еталонний SNP) для SNP, отриманий з <i>dbSNP.txt</i> , якщо застосовно. Якщо в цьому місці існує кілька номерів rs, то список розділяється крапками з комою. Якщо запис <i>dbSNP</i> у цій позиції відсутній, використовується маркер пропущеного значення ('.').
REF	Еталонний генотип. Наприклад, делеція окремого T представлена як еталонний TT і альтернативний T. Однонуклеотидний варіант від A до T представлений як еталонний A й альтернативний T.
ALT	Алелі, які відрізняються від зчитаного еталонного геному. Наприклад, інсерція окремого T представлена як еталонний A й альтернативний AT. Однонуклеотидний варіант від A до T представлений як еталонний A й альтернативний T.
QUAL	Показник якості Phred, визначений засобом розпізнавання варіантів. Більш високі показники якості варіанта свідчать про те, що він має більш високий ступінь вірогідності, а також про меншу ймовірність помилок. Для показника якості Q оцінна вірогідність помилки становить $10^{-Q/10}$. Наприклад, набір із розпізнаваннями Q30 має коефіцієнт помилок 0,1%. Багато засобів розпізнавання варіантів визначають показники якості на основі своїх статистичних моделей, які є високими порівняно зі спостережуваним коефіцієнтом помилок.

Таблиця 2. Анотації до файлів VCF

Заголовок	Опис
FILTER (ФІЛЬТР)	<p>Якщо всі фільтри пройдено, то в стовпчику фільтра записується PASS (Пройдено). Можливі записи у стовпчику FILTER (ФІЛЬТР) робочого процесу для зародкової лінії:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL — застосовується, якщо показник QUAL для варіанта SNP не відповідає пороговому значенню. • DRAGENIndelHardQUAL — застосовується, якщо показник QUAL для варіанта інделу не відповідає пороговому значенню. • LowDepth — позицію відфільтровано, оскільки глибина покриття не відповідає пороговому значенню. • LowGQ — позицію відфільтровано, оскільки якість генотипу не відповідає пороговому значенню. • PloidyConflict — результат визначення генотипу, отриманий модулем визначення варіантів, не узгоджується з плоїдністю хромосоми. • base_quality — позицію відфільтровано, оскільки медіанна якість основи альтернативних зчитувань у цьому локусі не відповідає пороговому значенню. • filtered_reads — позицію відфільтровано, оскільки відфільтровано надто велику частку зчитувань. • fragment_length — позицію відфільтровано, оскільки абсолютна різниця між медіанною довжиною фрагмента альтернативних зчитувань і медіанною довжиною фрагмента еталонних зчитувань у цьому локусі перевищує порогове значення. • low_depth — позицію відфільтровано, оскільки глибина зчитування надто низька. • low_frac_info_reads — позицію відфільтровано, оскільки частка інформативних зчитувань нижча за порогове значення. • low_normal_depth — позицію відфільтровано, оскільки глибина зчитування нормального зразка надто низька. • long_indel — позицію відфільтровано, оскільки довжина інделу надто велика. • mapping_quality — позицію відфільтровано, оскільки медіанна якість вирівнювання альтернативних зчитувань у цьому локусі не відповідає пороговому значенню. • multiallelic — позицію відфільтровано, оскільки поріг LOD пухлини проходять більше двох альтернативних алелів. • non_homref_normal — позицію відфільтровано, оскільки генотип нормального зразка не є гомозиготним за референтним алелем. • no_reliable_supporting_read — позицію відфільтровано, оскільки немає жодного надійного соматичного зчитування, що підтверджує варіант. • panel_of_normals — виявлено принаймні в одному зразку у VCF-файлі панелі нормальних зразків. • read_position — позицію відфільтровано, оскільки медіана відстаней між початком/кінцем зчитування та цим локусом нижча за порогове значення. • RMxNRepeatRegion — позицію відфільтровано, оскільки весь варіантний алель або його частина є повтором еталонної послідовності. • strand_artifact — позицію відфільтровано через виражений зсув за ланцюгами. • str_contraction — позицію відфільтровано через підозру на помилку ПЛР, за якої альтернативний алель на одну одиницю повтору коротший за еталонний. • too_few_supporting_reads — позицію відфільтровано, оскільки в пухлинному зразку надто мало зчитувань, що підтверджують варіант. • weak_evidence — оцінка соматичного варіанта не відповідає пороговому значенню.

Заголовок	Опис
FILTER (ФІЛЬТР) (продовження)	<p>Можливі записи у стовпчику FILTER (ФІЛЬТР) робочого процесу для соматичних варіантів:</p> <ul style="list-style-type: none"> • base_quality — позицію відфільтровано, оскільки медіанна якість основи альтернативних зчитувань у цьому локусі не відповідає пороговому значенню. • filtered_reads — позицію відфільтровано, оскільки відфільтровано надто велику частку зчитувань. • fragment_length — позицію відфільтровано, оскільки абсолютна різниця між медіанною довжиною фрагмента альтернативних зчитувань і медіанною довжиною фрагмента еталонних зчитувань у цьому локусі перевищує порогове значення. • low_depth — позицію відфільтровано, оскільки глибина зчитування надто низька. • low_frac_info_reads — позицію відфільтровано, оскільки частка інформативних зчитувань нижча за порогове значення. • low_normal_depth — позицію відфільтровано, оскільки глибина зчитування нормального зразка надто низька. • long_indel — позицію відфільтровано, оскільки довжина інделу надто велика. • mapping_quality — позицію відфільтровано, оскільки медіанна якість вирівнювання альтернативних зчитувань у цьому локусі не відповідає пороговому значенню. • multiallelic — позицію відфільтровано, оскільки поріг LOD пухлини проходять більше двох альтернативних алелів. • non_homref_normal — позицію відфільтровано, оскільки генотип нормального зразка не є гомозиготним за референтним алелем. • no_reliable_supporting_read — позицію відфільтровано, оскільки немає жодного надійного соматичного зчитування, що підтверджує варіант. • panel_of_normals — виявлено принаймні в одному зразку у VCF-файлі панелі нормальних зразків. • read_position — позицію відфільтровано, оскільки медіана відстаней між початком/кінцем зчитування та цим локусом нижча за порогове значення. • RMxNRepeatRegion — позицію відфільтровано, оскільки весь варіантний алель або його частина є повтором еталонної послідовності. • strand_artifact — позицію відфільтровано через виражений зсув за ланцюгами. • str_contraction — позицію відфільтровано через підозру на помилку ПЛР, за якої альтернативний алель на одну одиницю повтору коротший за еталонний. • too_few_supporting_reads — позицію відфільтровано, оскільки в пухлинному зразку надто мало зчитувань, що підтверджують варіант. • weak_evidence — оцінка соматичного варіанта не відповідає пороговому значенню. • systematic_noise — позицію відфільтровано на підставі ознак систематичного шуму в нормальних зразках.

Заголовок	Опис
INFO (ІНФОРМАЦІЯ)	<p>Можливі записи у стовпчику INFO (ІНФОРМАЦІЯ) робочого процесу для зародкової лінії:</p> <ul style="list-style-type: none"> • AC — підрахунок алелів у генотипах для кожного алеля ALT, у тому ж порядку, що й у списку. • AF — частота алелів для кожного алеля ALT в тому ж порядку, що й у списку. • AN — загальна кількість алелів у розпізнаних генотипах. • DB — присутність у базі даних DbSNP. • FS — р-значення за шкалою Phred, отримане за допомогою точного тесту Фішера для виявлення зміщення ланцюга. • QD — вірогідність варіанта / якість за глибиною. • R2_5P_bias — показник, що базується на зміщенні спарених зчитувань та відстані від 5'-кінця. • SOR — симетричне відношення шансів таблиці спряженості 2x2 для виявлення зміщення ланцюга. • DP — приблизна глибина зчитування (інформативні та неінформативні зчитування); деякі зчитування могли бути відфільтровані на підставі mapq тощо. • END — кінцева позиція інтервалу. • FractionInformativeReads — частка інформативних зчитувань від загальної кількості зчитувань. • MQ — середньоквадратична (RMS) якість вирівнювання. • MQRankSum — Z-оцінка за критерієм рангових сум Вілкоксона для якості вирівнювання Alt порівняно з Ref. • ReadPosRankSum — Z-оцінка за критерієм рангових сум Вілкоксона для зсуву позиції зчитування Alt порівняно з Ref. • SOMATIC — принаймні один варіант у цій позиції є соматичним. <p>Можливі записи у стовпчику INFO (ІНФОРМАЦІЯ) робочого процесу для соматичних варіантів:</p> <ul style="list-style-type: none"> • DP — приблизна глибина зчитування (інформативні та неінформативні зчитування); деякі зчитування могли бути відфільтровані на підставі mapq тощо. • END — кінцева позиція інтервалу. • FractionInformativeReads — частка інформативних зчитувань від загальної кількості зчитувань. • MQ — середньоквадратична (RMS) якість вирівнювання. • MQRankSum — Z-оцінка за критерієм рангових сум Вілкоксона для якості вирівнювання Alt порівняно з Ref. • ReadPosRankSum — Z-оцінка за критерієм рангових сум Вілкоксона для зсуву позиції зчитування Alt порівняно з Ref. • AQ — оцінка систематичного шуму. • hotspot — відома соматична позиція, яку використовують для підвищення впевненості у визначенні. • SOMATIC — принаймні один варіант у цій позиції є соматичним.

Заголовок	Опис
FORMAT (ФОРМАТ)	<p>У стовпчику формату перераховано поля, розділені двокрапкою. Наприклад, GT:GQ. Доступні поля робочого процесу для зародкової лінії:</p> <ul style="list-style-type: none"> • AD — алельні глибини (з урахуванням лише інформативних зчитувань із загальної кількості зчитувань) для еталонного та альтернативних алелів у наведеному порядку. • AF — частки альтернативних алелів у наведеному порядку. • DP — приблизна глибина зчитування (зчитування з MQ=255 або з некоректними парами відфільтровано). • F1R2 — кількість зчитувань в орієнтації пари F1R2, що підтримують кожний алель. • F2R1 — кількість зчитувань в орієнтації пари F2R1, що підтримують кожний алель. • GP — апостеріорні ймовірності генотипів, масштабовані за шкалою Фреда, як визначено у нормативах VCF. • GQ — якість генотипу. • GT — генотип. 0 відповідає еталонній основі, 1 відповідає першому запису в стовпчику ALT і так далі. Скисна риска (/) указує на відсутність інформації про фазування. • MB — покомпонентна статистика на рівні зразка для виявлення зсуву за mate-парами. • PL — нормалізовані ймовірності генотипів, масштабовані за шкалою Фреда, як визначено у нормативах VCF. • PRI — апіорні ймовірності генотипів, масштабовані за шкалою Фреда. • PS — інформація про ідентифікатор фізичного фазування, де кожний унікальний ідентифікатор у межах певного зразка (але не між різними зразками) пов'язує записи в межах однієї групи фазування. • SB — статистика компонентів для кожного зразка, що використовується в точному тесті Фішера для виявлення зміщення ланцюга. • SQ — соматична якість. <p>Доступні поля робочого процесу для соматичних варіантів:</p> <ul style="list-style-type: none"> • AD — алельні глибини (з урахуванням лише інформативних зчитувань із загальної кількості зчитувань) для еталонного та альтернативних алелів у наведеному порядку. • AF — частки альтернативних алелів у наведеному порядку. • DP — приблизна глибина зчитування (зчитування з MQ=255 або з некоректними парами відфільтровано). • F1R2 — кількість зчитувань в орієнтації пари F1R2, що підтримують кожний алель. • F2R1 — кількість зчитувань в орієнтації пари F2R1, що підтримують кожний алель. • GT — генотип. 0 відповідає еталонній основі, 1 відповідає першому запису в стовпчику ALT і так далі. Скисна риска (/) указує на відсутність інформації про фазування. • MB — покомпонентна статистика на рівні зразка для виявлення зсуву за mate-парами. • PS — інформація про ідентифікатор фізичного фазування, де кожний унікальний ідентифікатор у межах певного зразка (але не між різними зразками) пов'язує записи в межах однієї групи фазування. • SB — статистика компонентів для кожного зразка, що використовується в точному тесті Фішера для виявлення зміщення ланцюга. • SQ — соматична якість.

Заголовок	Опис
SAMPLE (ЗРАЗОК)	Стовпчик зразка містить значення, наведені у стовпчику FORMAT (ФОРМАТ).

VCF-файли геному

VCF-файли геному (*.gvcf.gz) відповідають набору правил для представлення всіх ділянок геному в досить компактному форматі. Файли gVCF включають усі ділянки в досліджуваній ділянці в один файл для кожного зразка. Файл gVCF відображає невизначення в позиціях, які не проходять усі фільтри. Тег генотипу (GT) ./. означає невизначення.

Перегляд результатів аналізу

Прогони в процесі виконання відображаються у вкладці Active (Активні). Завершені прогони відображаються у вкладці Completed (Завершені). Докладнішу інформацію про перегляд результатів див. у розділі [Документація на продукт NovaSeq 6000Dx \(документ № 200010105\)](#).

Технічна допомога

Для отримання технічної допомоги зв'яжіться зі службою технічної підтримки компанії Illumina.

Вебсайт: www.illumina.com
 Електронна пошта: techsupport@illumina.com

Номери телефонів служби технічної підтримки компанії Illumina

Регіон	Безкоштовний номер	Номер у міжнародному форматі
Австралія	+61 1800 775 688	
Австрія	+43 800 006249	+43 1 9286540
Бельгія	+32 800 77 160	+32 3 400 29 73
Канада	+1 800 809 4566	
Китай		+86 400 066 5835
Данія	+45 80 82 01 83	+45 89 87 11 56
Фінляндія	+358 800 918 363	+358 9 7479 0110
Франція	+33 8 05 10 21 93	+33 1 70 77 04 46
Німеччина	+49 800 101 4940	+49 89 3803 5677
Гонконг, Китай	+852 800 960 230	
Індія	+91 8006500375	
Індонезія		0078 036 510 048
Ірландія	+353 1800 936608	+353 1 695 0506
Італія	+39 800 985513	+39 236003759
Японія	+81 0800 111 5011	
Малайзія	+60 1800 80 6789	
Нідерланди	+31 800 022 2493	+31 20 713 2960
Нова Зеландія	+64 800 451 650	
Норвегія	+47 800 16 836	+47 21 93 96 93
Філіппіни	+63 180016510798	
Сінгапур	1 800 5792 745	

Регіон	Безкоштовний номер	Номер у міжнародному форматі
Південна Корея	+82 80 234 5300	
Іспанія	+34 800 300 143	+34 911 899 417
Швеція	+46 2 00883979	+46 8 50619671
Швейцарія	+41 800 200 442	+41 56 580 00 00
Тайвань, Китай	+886 8 06651752	
Таїланд	+66 1800 011 304	
Сполучене Королівство	+44 800 012 6019	+44 20 7305 7197
Сполучені Штати Америки	+1 800 809 4566	+1 858 202 4566
В'єтнам	+84 1206 5263	

Паспорти безпеки продукції (SDS) доступні на вебсайті Illumina за адресою support.illumina.com/sds.html.

Документація на продукт доступна для завантаження на вебсайті support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A. (США)
+1 800 809.ILMN (4566)
+1 858 202 4566 (за межами Північної
Америци)
techsupport@illumina.com
www.illumina.com

CE



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

Австралійський спонсор
Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
Melbourne, VIC 3000
Австралія

ВИКОРИСТОВУВАТИ ДЛЯ ДІАГНОСТИКИ IN VITRO

© 2022 Illumina, Inc. Усі права застережено.

illumina®