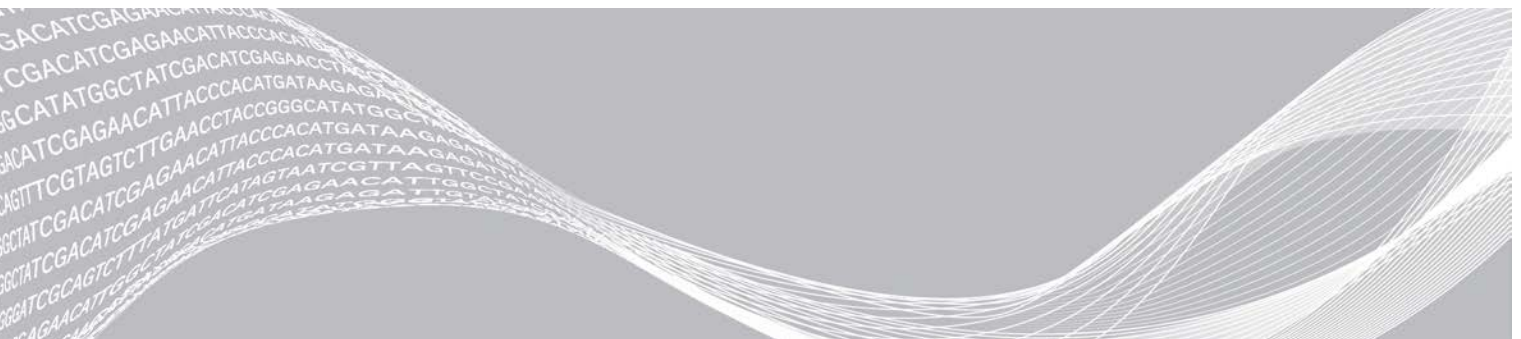


Модуль аналізу Somatic Variant диспетчера Local Run Manager

Посібник із робочого процесу для NextSeq 550Dx

ВИКОРИСТОВУВАТИ ДЛЯ ДІАГНОСТИКИ IN VITRO

Огляд	3
Уведення інформації про прогін	4
Методи аналізу	6
Перегляд даних про прогін і зразок	8
Звіт з аналізу	8
Вихідні файли аналізу	10
Розпізнавання азотистих основ і різноманітність індексів	17
Історія редакцій	18
Технічна допомога	19



Цей документ і його зміст є власністю компанії Illumina, Inc. та її філій (надалі — «Illumina»); він призначений лише для того, щоб користувач використовував вироби виключно за угодою в цілях, описаних у цьому документі. Цей документ і його зміст 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 уміст цього документа.

НЕПОВНЕ ВИВЧЕННЯ ВСІХ ЗАЗНАЧЕНИХ У ЦЬОМУ ДОКУМЕНТІ ВКАЗІВОК І ЇХНЕ НЕЧІТКЕ ДОТРИМАННЯ МОЖЕ ПРИЗВОДИТИ ДО ПОШКОДЖЕННЯ ЦИХ ВИРОБІВ, ТРАВМУВАННЯ ЛЮДЕЙ, ВКЛЮЧНО З КОРИСТУВАЧАМИ АБО ІНШИМИ ОСОБАМИ, І 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).

© Illumina, Inc., 2021. Усі права захищено.

Усі товарні знаки — власність компанії Illumina, Inc. або їхніх відповідних власників. Конкретна інформація про товарні знаки зазначена на сторінці www.illumina.com/company/legal.html.

Огляд

Модуль Somatic Variant (Соматичний варіант) диспетчера Local Run Manager призначено для використання з набором реагентів і витратних матеріалів для аналізу Illumina TruSeq Custom Amplicon Kit Dx і NextSeq 550Dx. У разі використання з модулем Somatic Variant (Соматичний варіант) набір реагентів і витратних матеріалів для аналізу призначено для підготовки бібліотек, використовуваних для секвенування ДНК із зафіксованої формаліном та залитої парафіном (FFPE) тканини. Аналіз виявляє соматичні мутації за низьких частот варіантів.

Цей модуль аналізу оцінює короткі області ампліфікованої ДНК, або амплікони, для розпізнавання варіантів. Сфокусоване секвенування ампліконів забезпечує широке охоплення певних областей на великій кількості зразків. Модуль аналізу виконує вторинний аналіз і генерує звіти з прогонів секвенування з використанням двониткового підходу, який включає в себе прямі та зворотні пули олігонуклеотидів. Див. інструкцію з використання *TruSeq Custom Amplicon Kit Dx* (документ № 1000000029772).

Для роботи модуля аналізу Somatic Variant (Соматичний варіант) потрібні витратні матеріали для 300-циклового секвенування. Щоб отримати додаткову інформацію, див. інструкцію з використання *набору реагентів із високим виходом NextSeq 550Dx версії 2* або *набору реагентів із високим виходом NextSeq 550Dx версії 2.5*.

Про цей посібник

Цей посібник містить інструкції з налаштування параметрів прогону для секвенування й аналізу для модуля аналізу Somatic Variant (Соматичний варіант). Інформацію про панель керування Local Run Manager і системні налаштування див. в *довідковому посібнику до приладу NextSeq 550Dx* (документ № 100000009513).

Перегляд Local Run Manager

Інтерфейс диспетчера Local Run Manager можна переглянути за допомогою системного програмного забезпечення NextSeq 550Dx (NOS) або браузера. Підтримується браузер Chromium.



ПРИМІТКА

У разі використання браузера, який не підтримується, завантажте браузер, що підтримується, коли з'явиться повідомлення Confirm Unsupported Browser (Підтвердити браузер, який не підтримується). Натисніть **тут**, щоб завантажити версію Chromium, яка підтримується.

Перегляд на моніторі приладу

- 1 Для перегляду інтерфейсу диспетчера Local Run Manager на моніторі приладу виберіть один із наведених нижче параметрів.
 - ▶ На головному екрані NOS виберіть **Local Run Manager**. Натисніть на X у правому верхньому кутку, щоб повернутися до NOS, коли все буде готово.
 - ▶ Виберіть значок Minimize NOS (Згорнути NOS), відкрийте браузер Chromium, що встановлений на приладі, і введіть **http://localhost** у рядок адреси. Тільки користувачі з правами адміністратора можуть згорнути NOS.

Перегляд із мережевого комп'ютера

- 1 Відкрийте браузер Chromium на комп'ютері, що має доступ до тієї ж мережі, до якої належить прилад, і підключіться, використовуючи IP-адресу або ім'я приладу. Приклад: **http://myinstrument**.

Уведення інформації про прогін

Налаштування параметрів

- 1 Увійдіть до Local Run Manager.
- 2 Виберіть **Create Run** (Створити прогін) і виберіть **Somatic Variant** (Соматичний варіант).
- 3 Уведіть ім'я прогону, яке ідентифікує прогін від секвенування до аналізу. Використовуйте літери та цифри, пробіл, підкреслювання або дефіс.
- 4 **[Додатково]** Уведіть опис прогону, який допоможе його ідентифікувати. Використовуйте літери та цифри, пробіл, підкреслювання або дефіс.
- 5 Виберіть кількість зразків і набори індексів у розкритому списку. Під час вибору враховуйте наведену далі інформацію.
 - ▶ Розкритий список містить номери зразків зі встановленими індексами. Наприклад, 24-Set 1 показує 24 зразки для тестування, з індексами з набору індексів 1.
 - ▶ Номер наборів індексів належать до різних наборів індексів i5. Обидва набори, i Set 1, i Set 2, забезпечують різноманітність індексів. Для запобігання виснаженню одного набору пропонуються два набори індексів.
 - ▶ Виберіть кількість зразків, яка є найближчою до кількості зразків, що ви тестуєте. Якщо точної кількості зразків немає в списку, виберіть найближчу, але меншу, ніж кількість, що ви тестуєте. Наприклад, якщо ви хочете протестувати 18 зразків, виберіть 16 зразків.
 - ▶ Зеленим кольором виділені лунки й комбінації індексів зразків, які відповідають вимогам щодо різноманітності індексів. Якщо ви виберете інші лунки та комбінації індексів, то під час збереження прогону ви отримаєте повідомлення про те, що вимоги щодо різноманітності індексів не виконуються.

Імпорт файлів маніфесту для прогону

- 1 Переконайтеся, що маніфести, які вам потрібно імпортувати, доступні в мережі або на USB-накопичувачі.
- 2 Натисніть **Import Manifests** (Імпорт маніфестів).
- 3 Перейдіть до файлу маніфесту та виберіть маніфести, які хочете додати.



ПРИМІТКА

Щоб зробити маніфести доступними для всіх прогонів за допомогою модуля аналізу Somatic Variant (Соматичний варіант), додайте маніфести, використовуючи функцію налаштування модуля. Ця функція вимагає дозволів на рівні адміністратора. Додаткову інформацію див. в довідковому посібнику до приладу NextSeq 550Dx (документ № 100000009513).


Зазначення зразків для прогону

Укажіть зразки для прогону, використовуючи один із варіантів і наведені далі інструкції.


- ▶ **Enter samples manually** (Увести зразки вручну) — скористайтеся порожньою таблицею на екрані Create Run (Створити прогін).
- ▶ **Import samples** (Імпорт зразків) — перейдіть до зовнішнього файлу у форматі розділених комами значень (*.csv). Шаблон можна завантажити на екрані Create Run (Створити прогін).

Після заповнення таблиці зразків можна експортувати інформацію про зразки в зовнішній файл. Використовуйте цей файл як довідковий під час підготовки бібліотек або імпорту файлу для іншого прогону.

Уведення зразків уручну

- 1 Уведіть унікальне ім'я зразка в поле Sample Name (Ім'я зразка). Використовуйте літери та цифри, дефіс або підкреслювання. Ім'я зразка автоматично додається у відповідну лунку в іншому пулі.
- 2 **[Додатково]** Для зразків позитивного та негативного контролю клацніть правою клавшею миші й виберіть тип контролю. Контроль у лунці одного зразка автоматично додається у відповідну лунку в іншому пулі з таким самим контролем.
- 3 **[Додатково]** Уведіть опис зразка в поле Sample Description (Опис зразка). Використовуйте літери та цифри, дефіс або підкреслювання. Опис зразка автоматично додається у відповідну лунку в іншому пулі. Описи зразків пов'язані з ідентифікатором зразка. Описи зразків перезаписуються, якщо один і той самий ідентифікатор зразка повторюється в подальшому прогоні.
- 4 Виберіть адаптер індексу 1 у розкритому списку Index 1 (Індекс 1) (i7). У разі використання запропонованих лунок зразка програмне забезпечення автоматично заповнює індексні адаптери i7 і i5, які відповідають вимогам щодо різноманітності індексів. Якщо точної кількості зразків для тестування немає в списку, обов'язково виберіть індексні адаптери для додаткових лунок. Якщо вам потрібно вибрати індекси для додаткових лунок або ви не використовуєте рекомендовані комбінації індексних адаптерів, перш ніж вибрати індекси, обов'язково прочитайте розділи *Розпізнавання азотистих основ і різноманітність індексів* на стор. 17.
- 5 Виберіть адаптер індексу 2 у розкритому списку Index 2 (Індекс 2) (i5).
- 6 Виберіть файл маніфесту з розкритого списку Manifest (Маніфест). Зразкам у Pool A (Пул А) потрібен інший маніфест, ніж зразкам у Pool B (Пул В).
- 7 Оберіть варіант для перегляду, друку або збереження макету планшета як довідкового матеріалу для приготування бібліотек.
 - ▶ Натисніть значок  **Print** (Друк), щоб відобразити макет планшета. Виберіть **Print** (Друк), щоб роздрукувати макет планшета.
 - ▶ Виберіть **Export** (Експорт), щоб експортувати інформацію про зразки до зовнішнього файлу.Переконайтеся, що інформація стосовно маніфесту й зразка правильна. Неправильна інформація може вплинути на результати.
- 8 Виберіть **Save Run** (Зберегти прогін).

Імпорт зразків

- 1 Виберіть **Import Samples** (Імпорт зразків) і перейдіть до розташування інформаційного файлу зразка. Ви можете імпортувати два типи файлів.
 - ▶ Виберіть **Template** (Шаблон) на екрані Create Run (Створити прогін), щоб створити новий макет планшета. Файл шаблону містить правильні заголовки стовпчиків для імпорту. Уведіть інформацію про зразок у кожний стовпчик для зразків у прогоні. Видаліть приклад даних із невикористаних клітинок і збережіть файл.
 - ▶ Використайте файл з інформацією про зразок, який було експортовано з модуля Somatic Variant (Соматичний варіант) за допомогою функції Export (Експорт).
- 2 Натисніть значок  **Print** (Друк), щоб відобразити макет планшета.
- 3 Виберіть **Print** (Друк), щоб роздрукувати макет планшета як еталон під час підготовки бібліотек.
- 4 **[Додатково]** Виберіть **Export** (Експорт), щоб експортувати інформацію про зразки до зовнішнього файлу. Переконайтеся, що інформація стосовно маніфесту й зразка правильна. Неправильна інформація може вплинути на результати.
- 5 Виберіть **Save Run** (Зберегти прогін).

Редагування прогону

Інструкції з редагування інформації в прогоні перед секвенуванням див. в *довідковому посібнику до приладу NextSeq 550Dx (документ № 1000000009513)*.

Методи аналізу

Модуль аналізу Somatic Variant (Соматичний варіант) виконує наведені нижче кроки аналізу, а потім записує вихідні файли аналізу в папку Alignment (Вирівнювання).

- ▶ Демультіплексує зчитування індексів.
- ▶ Генерує файли FASTQ.
- ▶ Виконує вирівнювання відповідно до еталонного геному.
- ▶ Визначає варіанти.

Демультіплексування

Під час демультіплексування виконується порівняння кожної послідовності зчитування індексу з послідовностями індексів, зазначеними для кожного прогону. На цьому етапі якісні значення не враховуються.

Зчитування індексів ідентифікуються під час описаних далі етапів.

- ▶ Зразки нумеруються від 1 залежно від 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.

Генерування файлів FASTQ

Після демультіплексування програмне забезпечення генерує проміжні файли аналізу у форматі FASTQ, який є текстовим форматом для представлення послідовностей. Файли FASTQ містять зчитування для кожного зразка та пов'язані оцінки якості. Кластери, які не пройшли крізь фільтри, виключено.

Кожен файл FASTQ містить зчитування лише для одного зразка, а ім'я цього зразка включається в ім'я файлу FASTQ. Файли FASTQ містять первинні вхідні дані для вирівнювання. На кожен зразок генерується вісім файлів FASTQ на кожен пул олігонуклеотидів, чотири з Read 1 (Зчитування 1) і чотири з Read 2 (Зчитування 2), у результаті чого на кожен зразок припадає загалом 16 файлів FASTQ.

Вирівнювання

Під час етапу вирівнювання попарний алгоритм Сміта — Уотермана вирівнює кластери з одного зразка відносно послідовностей ампліконів, заданих у файлі маніфесту.

Попарний алгоритм Сміта — Уотермана виконує напівглобальне вирівнювання послідовностей для визначення схожих областей між двома послідовностями. Замість порівняння всієї послідовності алгоритм Сміта — Уотермана порівнює сегменти всіх можливих довжин.

Кожне зчитування парних кінців оцінюється з точки зору його вирівнювання відносно відповідних послідовностей зондів для цього зчитування.

- ▶ Зчитування 1 оцінюється відносно зворотно комплементарної ділянки низхідного локус-специфічного олігонуклеотиду (Downstream Locus-Specific Oligos, DLSO).
- ▶ Зчитування 2 оцінюється відносно висхідного локус-специфічного олігонуклеотиду (Upstream Locus-Specific Oligos, ULSO).
- ▶ Якщо початок зчитування збігається з послідовністю зонда не більше ніж із трьома відмінностями (розбіжності або зсуви через переважання інделів), повна довжина зчитування вирівнюється відносно мішені амплікона для цієї послідовності.
- ▶ Індели для DLSO й ULSO не спостерігалися з огляду на хімічні властивості аналізу.

Вирівнювання фільтруються з результатів вирівнювання на підставі частоти розбіжності для досліджуваної області або повного амплікона, залежно від довжини амплікона. Відфільтровані вирівнювання записуються у файли вирівнювання як невирівняні та не використовуються в розпізнаванні варіантів.

Розпізнавання варіантів



Програма розпізнавання варіантів Pisces, розроблена компанією Illumina, ідентифікує варіанти, які з низькою частотою присутні в зразку ДНК.

Програма розпізнавання варіантів Pisces ідентифікує SNV, MNV та малі індели в три етапи:

- ▶ розгляд кожної позиції в еталонному геномі окремо;
- ▶ підрахунок основ у заданій позиції для вирівняних зчитувань, які перекривають позиції;
- ▶ розрахунок показника якості варіанта, який визначає якість розпізнавання за допомогою моделі Пуассона. Варіанти з показником якості нижче Q30 виключаються.

Варіанти спочатку розпізнаються окремо для кожного пулу. Потім варіанти з кожного пулу порівнюються й об'єднуються в один вихідний файл. Якщо варіант присутній в обох пулах і проходить крізь усі фільтри, перераховані в розділі *Анотації до файлів VCF на стор. 13*, то у файлі розпізнавання варіантів (VCF) варіант позначається як PASS.

Перегляд даних про прогін і зразок

- 1 На панелі інструментів Local Run Manager натисніть на назву прогону.
- 2 На вкладці Run Overview (Огляд прогону) перегляньте показники прогону секвенування.
- 3 **[Додатково]** Натисніть **Copy to clipboard**  (Копіювати до буфера обміну), щоб скопіювати шлях до папки вихідних даних прогону.
- 4 Перейдіть на вкладку Sequencing Information (Інформація про секвенування), щоб переглянути інформацію про параметри прогону й витратні матеріали.
- 5 Перейдіть на вкладку Samples and Results (Зразки й результати), щоб переглянути розташування звіту про аналіз.
 - ▶ Якщо аналіз був повторений, відкрийте розкривний список Select Analysis (Вибір аналізу) і виберіть відповідний аналіз.
- 6 Натисніть на значок **Copy to clipboard**  (Копіювати до буфера обміну), щоб скопіювати шлях до папки Analysis (Аналіз).

Додаткові відомості про вкладки Run Overview (Огляд прогону) і Sequencing Information (Інформація про секвенування), а також про те, як повернути аналіз у чергу, див. в довідковому посібнику до приладу *NextSeq 550Dx (документ № 1000000009513)*.

Звіт з аналізу

Результати аналізу зібрано на вкладці Samples and Results (Зразки й результати), а також у вигляді зведеного звіту в папці Alignment (Вирівнювання). Звіт за кожним зразком також доступний у форматі PDF для кожного зразка.

Інформація на вкладці Samples and Results

- 1 Натисніть на зразок у списку, щоб побачити звіт про зразок.

Таблиця 1 Інформація про прогін і зразок

Заголовок стовпчика	Опис
Run Status (Стан прогону)	Визначає, пройшов прогін секвенування чи ні.
Total Yield (GB) (Загальний вихід (ГБ))	Кількість розпізнаних азотистих основ у прогоні секвенування. Показує прохідний поріг, а також стан «пройдено» або «не пройдено».
% ≥ Q30	Відсоток зчитування під час прогону секвенування з показником якості 30 (Q30) або вище. Показує прохідний поріг, а також стан «пройдено» або «не пройдено».
Sample Name (Назва зразка)	Назва зразка, указана під час створення прогону.
Total PF Reads (Усього зчитувань ПФ)	Загальна кількість зчитувань, що проходять крізь фільтр.

Заголовок стовпчика	Опис
Read 1% \geq Q30 (Зчитування 1, % \geq Q30)	Відсоток зчитувань у Read 1 (Зчитування 1) із показником якості 30 (Q30) або вище для зразка.
Read 2% \geq Q30 (Зчитування 2, % \geq Q30)	Відсоток зчитувань у Read 2 (Зчитування 2) із показником якості 30 (Q30) або вище для зразка.
Autosome Call Rate (Частота розпізнавання аутосом)	Кількість геномних позицій серед аутосом (хромосоми з 1 по 22), які задовольняють визначеному порогу довірчої ймовірності, поділена на загальну кількість досліджених позицій аутосом. Частота розпізнавань описується на один зразок і визначається величиною у відсотках, яка розраховується як 1 мінус кількість аутосомних позицій із незавершеними розпізнаваннями, поділена на загальну кількість секвенованих аутосомних позицій.

Таблиця 2 Інформація про зразок у звітах

Заголовок стовпчика	Опис
Sample (Зразок)	Назва зразка, указана під час створення прогону.
Report Date (Дата звіту)	Дата, коли було згенеровано звіт.
Sample Information (Інформація про зразок)	Ідентифікатор зразка, який був наданий під час створення прогону, загальна кількість зчитувань, що пройшли крізь фільтр у зразку, відсоток зчитувань для зразка із показником якості 30 (Q30) або вище, а також частота розпізнавання аутосом.
Amplicon Summary (Зведена інформація про амплікони)	Загальна кількість секвенованих областей ампліконів і загальна довжина в парах основ секвенованих ампліконів у цільових областях для зразка в Pool A (Пул A) і Pool B (Пул B) і файлі маніфесту, що використовується для кожного пулу. Файл маніфесту визначає еталонний геном і цільові еталонні області, що використовуються на етапі вирівнювання.
Read Level Statistics (Статистичні дані рівня зчитування)	Кількість і відсоток зчитувань для зразка, що охоплює кожен позицію в еталонному геномі, для Read 1 (Зчитування 1) і Read 2 (Зчитування 2) у Pool A (Пул A) і Pool B (Пул B).
Variants Summary (Зведена інформація про варіанти)	Кількість SNV, інсерцій і делецій, виявлених для зразка, що пройшов запропоновані значення, щоб визначити, чи знаходяться в прийнятних межах результати за показником якості.
Coverage Summary (Зведена інформація про охоплення)	Загальна кількість вирівняних азотистих основ, поділена на розмір цільової області, і відсоток областей ампліконів зі значеннями охоплення, які перевищують нижній поріг охоплення, що дорівнює $0,2 * \text{середнє охоплення ампліконів}$ для зразка в Pool A (Пул A) і Pool B (Пул B).
Coverage Plots (Діаграми охоплення)	На діаграмі «Охоплення областями ампліконів» показано охоплення областями ампліконів для зразка. Области зі значеннями охоплення нижче порога охоплення виділено червоним кольором. Середнє значення всіх величин позначено помаранчевою лінією. Діаграма надається для покриття Pool A (Пул A) і Pool B (Пул B).
Software Versions (Версії програмного забезпечення)	Версії програмного забезпечення, коли зразок було секвеновано. Включає системне програмне забезпечення NextSeq 550Dx (NOS), програмне забезпечення Local Run Manager, програмне забезпечення RTA і версію модуля Somatic Variant (Соматичний варіант).

Вихідні файли аналізу

Указані далі вихідні файли аналізу генеруються для модуля аналізу Somatic Variant (Соматичний варіант) і надають результати аналізу для вирівнювання й розпізнавання варіантів. Вихідні файли аналізу знаходяться в папці Alignment (Вирівнювання).

Ім'я файлу	Опис
Demultiplexing (*.txt)	Проміжні файли, що містять результати демультіплексування.
FASTQ (*.fastq.gz)	Проміжні файли, що містять розпізнавання азотистих основ із відповідним показником якості. Файли FASTQ містять первинні вхідні дані для етапу вирівнювання.
Файли вирівнювання у форматі BAM (*.bam)	Містять вирівняні зчитування для заданого зразка.
Файли розпізнавання варіантів на пул у форматі VCF (*.vcf)	Містять варіанти, розпізнавані в кожній позиції з прямого або зворотного пулу.
Файли розпізнавання варіантів у форматі геному VCF (*.genome.vcf.gz)	Містять генотип для кожної позиції, незалежно від того, розпізнаний він як варіант або як еталонний генотип.
Файли розпізнавання варіантів консенсусних послідовностей у форматі VCF (*.vcf.gz)	Містять варіанти, розпізнавані в кожній позиції з обох пулів.
AmpliconCoverage_M1.tsv	Містить інформацію про охоплення на амплікон на зразок для кожного представленого маніфесту. M# — це номер маніфесту.

Демультіплексований формат файлів

У процесі демультіплексування зчитується індексна послідовність, прикріплена до кожного кластера, щоб визначити, з якого зразка утворився кластер. Для кожного сегмента проточної кювети в демультіплексований (*.demux) файл записується відповідність між кластерами й номером зразка.

Формат іменування файлів демультіплексування — **s_1_X.demux**, де X — номер сегмента.

Файли демультіплексування починаються із заголовка:

- ▶ версія (4-байтове ціле число), поточна: 1;
- ▶ лічильник кластерів (4-байтове ціле число);

інша частина файлу складається з номерів зразків для кожного кластера із сегмента.

Після завершення етапу демультіплексування програма генерує файл демультіплексування з назвою **DemultiplexSummaryF1L1.txt**.

- ▶ В імені файлу **F1** позначає номер проточної кювети.
- ▶ В імені файлу **L1** позначає номер доріжки.
- ▶ Результатом демультіплексування є таблиця з одним рядком для сегмента й одним стовпчиком для зразка, включно зі зразком 0.
- ▶ Зчитуються послідовності в індексі, що трапляються найчастіше.

Формат файлів FASTQ

FASTQ — це текстовий формат файлу, який містить розпізнавання азотистих основ і значення якості за кожним зчитуванням. Кожен запис містить 4 рядки:

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

Ідентифікатор відформатований як:

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

Формат файлів BAM

Файл BAM (*.bam) — це стиснена бінарна версія файлу SAM, яка використовується для подання вирівняних послідовностей, розміром до 128 Мб. Формати SAM і BAM докладно описано в samtools.github.io/hts-specs/SAMv1.pdf.

У файлах BAM використовується формат іменування файлів **SampleName_S#.bam**, де # — це номер зразка, який визначається порядком перерахування зразків для прогону.

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;
- ▶ **BC:** тег із бар-кодом, який вказує на демультіплексований ідентифікатор зразка, пов'язаний зі зчитуванням;
- ▶ **SM:** якість вирівнювання одиничних кінцевих фрагментів;
- ▶ **XC:** рядок характеристики відповідності;
- ▶ **XN:** мітка з назвою амплікона, яка містить ідентифікатор амплікона, пов'язаний зі зчитуванням.

Файли індексів BAM (*.bam.bai) надають індекс відповідного файлу BAM.

Формат файлу VCF

Формат розпізнавання варіантів (Variant Call Format, VCF) — це загальний формат файлів, розроблений науковим співтовариством із геноміки. Містить інформацію про варіанти, знайдені в конкретних позиціях еталонного геному. Файли VCF закінчуються суфіксом .vcf.

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 заголовку 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.

Заголовки файлу VCF

Заголовок	Опис
CHROM	Хромосома еталонного геному. Хромосоми відображаються в тому ж порядку, що й довідковий файл FASTQ.
POS	Одноосновні позиції варіантів у еталонній хромосомі. Для SNP ця позиція є еталонною основою з варіантом; для інделів або делецій ця позиція є еталонною основою безпосередньо перед варіантом.
ID	Номер rs для варіанта, отриманого з dbSNP.txt, якщо є. Якщо в цьому місці є кілька номерів rs, то список ділиться крапкою з комою. Якщо в цій позиції немає запису dbSNP, то використовується маркер відсутнього значення ('.').
REF	Еталонний генотип. Наприклад, делеція окремого T представляється як еталонний TT і альтернативний T. Однонуклеотидний варіант від A до T представляється як еталонний A й альтернативний T.
ALT	Алелі, які відрізняються від зчитаного еталонного геному. Наприклад, інсерція окремого T представляється 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.

Анотації до файлів VCF

Заголовок	Опис
FILTER (ФІЛЬТР)	<p>Якщо всі фільтри пройдено, то в стовпчику фільтра записується PASS (Пройдено).</p> <ul style="list-style-type: none"> • LowDP: застосовується на ділянках із глибиною охоплення нижче 450 зчитувань у кожному пулі. Для позицій ампліконів, охоплених як прямим, так і зворотним зчитуванням, це еквівалентно охопленню в 900 зчитувань одиничних кінцевих фрагментів. • LowGQ: якість генотипування (GQ) нижче граничного значення. • q30: показник якості < 30. • LowVariantFreq: частота варіантів менше заданого порога. • PB: зміщення пулу зондів. Варіант не знайдено чи знайдено з низькою частотою в одному або двох пулах зондів. • R3x6: кількість сусідніх повторів (довжиною від 1 до 3 п. о.) до розпізнавань варіантів ≥ 6. • SB: зміщення ниток більше за заданий поріг.
INFO (ІНФОРМАЦІЯ)	<p>Можливі записи в стовпчику INFO (ІНФОРМАЦІЯ) уключають:</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. • CD: мітка, яка вказує, що SNP відбувається в області кодування щонайменше в одному записі RefGene. • DP: глибина (кількість розпізнавань основ, вирівняних за позицією і використовуваних у розпізнаванні варіантів). • Exon: розділений комами список областей екзонів, зчитаних із RefGene. • FC: функціональні наслідки. • GI: розділений комами список ідентифікаторів генів, зчитаних із RefGene. • QD: вірогідність варіанта / якість за глибиною. • TI: розділений комами список ідентифікаторів транскриптів, зчитаних із RefGene.
FORMAT (ФОРМАТ)	<p>У стовпчику формату перераховано поля, розділені двокрапкою. Наприклад, GT:GQ. Список наданих полів залежить від використовуваного засобу розпізнавання варіантів. Доступні поля містять:</p> <ul style="list-style-type: none"> • AD: запис форми X, Y, де X — кількість еталонних розпізнавань, а Y — кількість альтернативних розпізнавань. • DP: приблизна глибина зчитування; зчитування з MQ = 255 або поганими парами фільтруються. • GQ: якість генотипу. • GQX: якість генотипу. GQX: це мінімальне значення параметра GQ та стовпця QUAL. Загалом ці значення схожі, але в разі використання мінімального значення GQX дає більш занижену оцінку якості генотипу. • GT: генотип. 0 відповідає еталонній основі, 1 відповідає першому запису в стовпчику ALT і так далі. Скісна риска (/) вказує на відсутність інформації про фазування. • NL: рівень шуму; оцінка шуму розпізнавання основ у цій позиції. • PB: зміщення пулу зондів. Значення ближче до 0 вказують на більше зміщення в бік одного пулу зондів і менший ступінь вірогідності в розпізнаванні варіанта. • SB: зміщення ниток у цьому позиції. Великі негативні значення вказують на менше зміщення; значення близько 0 вказують на більше зміщення. • VF: частота варіанта; відсоток зчитувань, які підтримують альтернативний алель.
SAMPLE (ЗРАЗОК)	Стовпчик зразка містить значення, указані в стовпчику FORMAT (ФОРМАТ).

Файли геному VCF

Файли геному VCF (gVCF) — це файли VCF v4.1, які відповідають набору правил для представлення всіх ділянок геному в досить компактному форматі. Файли gVCF (*.genome.vcf.gz) включають усі ділянки в досліджуваній області в один файл для кожного зразка.

Файл gVCF показує нерозпізнавання в позиціях, які не проходять крізь усі фільтри. Тег генотипу (GT) *.I.* указує на нерозпізнавання.

Для отримання додаткової інформації див. sites.google.com/site/gvcftools/home/about-gvcf.

Файли розпізнавання варіантів на пул і файли розпізнавання консенсусних послідовностей у форматі VCF

Під час робочого процесу Somatic Variant (Соматичний варіант) генерується два набори файлів розпізнавання варіантів.

- ▶ **Per-pool VCF files** (Файли розпізнавання варіантів на пул у форматі VCF): містять варіанти, розпізнавані з прямого або зворотного пулу. Файли розпізнавання варіантів на пул записуються в папку VariantCallingLogs.
- ▶ **Consensus VCF files** (Файли розпізнавання консенсусних послідовностей): містять варіанти, розпізнавані з обох пулів. Файли розпізнавання консенсусних послідовностей записуються в папку Alignment (Вирівнювання).

Файли розпізнавання варіантів на пул і файли розпізнавання консенсусних послідовностей у форматі VCF включають у себе як файли VCF (*.vcf), так і файли gVCF (*.genome.vcf) і мають указане далі умовне позначення, де S# — це порядок, у якому цей зразок розміщено в списку для прогону.

- ▶ **Reports for all sites** (Звіт про всі ділянки) — SampleName_S#.genome.vcf
- ▶ **Reports variants only** (Звіт тільки про варіанти) — SampleName_S#.vcf

Програмне забезпечення порівнює файли розпізнавання варіантів на пул у форматі VCF і об'єднує дані в кожній позиції для створення файлу розпізнавання консенсусних послідовностей у форматі VCF для зразка.

Розпізнавання варіантів із кожного пулу об'єднуються в файли розпізнавання консенсусних послідовностей у форматі VCF за наведеними нижче критеріями.

Критерій	Результат
Еталонне розпізнавання в кожному пулі	Еталонне розпізнавання
Еталонне розпізнавання в пулі 1 і розпізнавання варіанта в іншому пулі	Фільтроване розпізнавання варіанта
Розпізнавання варіантів, які збігаються, зі схожими частотами в кожному пулі	Розпізнавання варіантів
Розпізнавання варіантів, які збігаються, з істотно різними частотами в кожному пулі	Фільтроване розпізнавання варіанта
Розпізнавання варіантів, які не збігаються, у кожному пулі	Фільтроване розпізнавання варіанта

Показники з кожного пулу об'єднуються за допомогою вказаних далі значень.

Показник	Значення
Глибина	Збільшення глибини в обох пулах
Частота варіантів	Загальна кількість варіантів, поділена на загальну глибину покриття
Показник Q	Мінімальне значення обох пулів

Файл охоплення ампліконів

Для кожного файлу маніфесту створюється файл охоплення ампліконів. M# в імені файлу — це номер маніфесту.

Кожен файл включає рядок заголовка, що містить ідентифікатори зразків, пов'язані з відповідним маніфестом. Цей файл містить таку інформацію.

- ▶ Target ID (ідентифікатор мішені) у тому вигляді, в якому він вказаний у маніфесті.
- ▶ Глибина охоплення зчитувань, що проходять крізь фільтр.

Додаткові вихідні файли

Указані далі вихідні файли надають додаткову інформацію або підсумовують результати прогону й помилки аналізу. Ці файли не потрібні для оцінки результатів аналізу, але їх можна використовувати для пошуку й усунення несправностей. Усі файли знаходяться в папці Alignment (Вирівнювання), якщо не вказано інше.

Ім'я файлу	Опис
AnalysisLog.txt	Журнал обробки, в якому описано кожен крок, пройдений під час аналізу поточної папки прогону. Цей файл не містить повідомлень про помилки. Знаходиться в папці Alignment (Вирівнювання).
AnalysisError.txt	Журнал обробки, в якому перераховано всі помилки, що виникли під час аналізу. Цей файл буде порожнім, якщо помилки не виникли. Знаходиться в папці Alignment (Вирівнювання).
DemultiplexSummaryF1L1#.txt	Подає результати демультіплексування у вигляді таблиці з одним рядком для сегмента й одним стовпчиком для зразка. # вказує номер доріжки проточної кювети: 1, 2, 3 або 4. Знаходиться в папці Alignment (Вирівнювання).
AmpliconRunStatistics.xml	Містить зведені статистичні дані для відповідного прогону. Знаходиться в папці Alignment (Вирівнювання).

Папка аналізу

У папці аналізу зберігаються файли, згенеровані програмним забезпеченням Local Run Manager.

Взаємозв'язок між папкою вихідних даних і папкою аналізу можна викласти так.

- ▶ Під час секвенування програмне забезпечення Real-Time Analysis (RTA) заповнює папку вихідних даних файлами, що генеруються під час аналізу зображень, розпізнавання азотистих основ і оцінки якості.
- ▶ RTA копіює файли в папку аналізу в режимі реального часу. Після того як RTA присвоює показник якості кожній азотистій основі для кожного циклу, програма записує файл RTAComplete.txt в обидві папки.
- ▶ Аналіз починається за наявності файлу RTAComplete.txt.
- ▶ Під час проведення аналізу Local Run Manager записує вихідні файли в папку аналізу, а потім копіює їх назад у папку вихідних даних.

Папки вирівнювання

Кожного разу, коли аналіз повертається в чергу, Local Run Manager створює папку вирівнювання з назвою **Alignment_N**, де N — порядковий номер.

Структура папок

Alignment — містить файли *.bam, *.vcf, FASTQ, а також файли, специфічні для модуля аналізу.

Date and Time Stamp — дата й час аналізу у форматі PPPPMMДД_ГГХХСС

- 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 — містить файли *.bcl.

L001 — містить файли *.locs.

RTA Logs — містить файли журналу аналізів програмного забезпечення RTA.

InterOp — містить бінарні файли, які використовуються для звіту за показниками прогону секвенування.

Logs — містить файли журналу з описом кроків, виконуваних під час секвенування.

- RTAComplete.txt
- RunInfo.xml
- RunParameters.xml

Розпізнавання азотистих основ і різноманітність індексів

Під час секвенування зразків на приладі NextSeq 550Dx виконується розпізнавання азотистих основ, яке визначає основи (A, C, G або T) для кожного кластера заданого сегмента або області зображення на проточній кюветі в певному циклі. У приладі NextSeq 550Dx використовується двоканальне секвенування, яке вимагає тільки двох зображень для кодування даних для чотирьох основ ДНК, одне з яких із червоного, а інше із зеленого каналу.

Процес зчитування індексу для розпізнавання азотистих основ відрізняється від розпізнавання азотистих основ під час інших зчитувань.

Зчитування індексу мають починатися хоча б з однієї основи, відмінної від G, у будь-якому з перших двох циклів. Якщо зчитування індексу починається з двох розпізнаних азотистих основ G, інтенсивність сигналу не генерується. Щоб забезпечити продуктивність демультимплексування, сигнал має бути наявний у будь-якому з перших двох циклів.

Якщо індекси не відповідають вимогам щодо різноманітності, під час вибирання індексів у процесі створення прогону з'явиться попередження про низький ступінь різноманітності. Щоб запобігти появі попередження про низький ступінь різноманітності, вибирайте індексні послідовності, які забезпечують сигнал в обох каналах для кожного циклу.

- ▶ Червоний канал — A чи C
- ▶ Зелений канал — A чи T

Цей процес розпізнавання азотистих основ забезпечує точність під час аналізу зразків із низькою щільністю. Додаткову інформацію про секвенування ваших індексів див. в інструкції з використання *TruSeq Custom Amplicon Kit Dx* (документ № 1000000029772).

Під час створення прогону в Local Run Manager ви вибираєте кількість зразків для тестування. Пропоновані комбінації індексів, які відповідають вимогам щодо різноманітності індексів, автоматично заповнюються програмою. Незважаючи на те що від вас не вимагається використання запропонованих комбінацій індексів, рекомендовано їх використовувати.

Історія редакцій

Документ	Дата	Опис зміни
Документ № 1000000030330, версія 03	Грудень 2019 р.	Оновлено адресу уповноваженого представника в Європейському Союзі. Оновлено адресу австралійського спонсора.
Документ № 1000000030330, версія 02	Січень 2019 р.	Додано інформацію про набори реагентів версії 2.5.
Документ № 1000000030330, версія 01	Серпень 2018 р.	Оновлено маркування відповідності нормативно-правовим вимогам.
Документ № 1000000030330, версія 00	Листопад 2017 р.	Початкова редакція.

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 підтримки користувачів компанії Illumina

Регіон	Безкоштовний	Регіональний
Північна Америка	+1 800 809 4566	
Австралія	+1 800 775 688	
Австрія	+43 800006249	+43 19286540
Бельгія	+32 80077160	+32 34002973
Велика Британія	+44 8000126019	+44 2073057197
Гонконг	800960230	
Данія	+45 80820183	+45 89871156
Ірландія	+353 1800936608	+353 016950506
Іспанія	+34 911899417	+34 800300143
Італія	+39 800985513	+39 236003759
Китай	400 066 5835	
Нідерланди	+31 8000222493	+31 207132960
Німеччина	+49 8001014940	+49 8938035677
Нова Зеландія	0800 451 650	
Норвегія	+47 800 16836	+47 21939693
Сингапур	+1 800 579 2745	
Тайвань	00806651752	
Фінляндія	+358 800918363	+358 974790110
Франція	+33 805102193	+33 170770446
Швейцарія	+41 565800000	+41 800200442
Швеція	+46 850619671	+46 200883979
Японія	0800 111 5011	
Інші країни	+44 1799 534000	

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

Документація продукції доступна для завантаження на вебсайті Illumina у форматі PDF. Перейдіть за адресою support.illumina.com і виберіть спочатку продукт, а потім — сторінку **Documentation & Literature** (Документація та література).

Уповноважений представник в Україні (Authorized Representative in Ukraine):



ТОВ «БІОЛАБТЕХ ЛТД»

проспект Героїв Сталінграда буд.42-А, кв.45
м. Київ, 04213, Україна

Тел.: +38 044 492 81 88

Електронна адреса: info@biolabtech.com.ua, ЄДРПОУ
34891619



Illumina

5200 Illumina Way
San Diego, California 92122
U.S.A. (США)

+1 800 809.ILMN (4566)

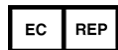
+1 858 202 4566 (за межами

Північної Америки)

techsupport@illumina.com

www.illumina.com

CE



Illumina Netherlands B. V.
Freddy van Riemsdijkweg 15
5657 EE Eindhoven
The Netherlands (Нідерланди)

Австралійський спонсор

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

ВИКОРИСТОВУВАТИ ДЛЯ ДІАГНОСТИКИ IN VITRO

© 2021 Illumina, Inc. Усі права захищено.

illumina®