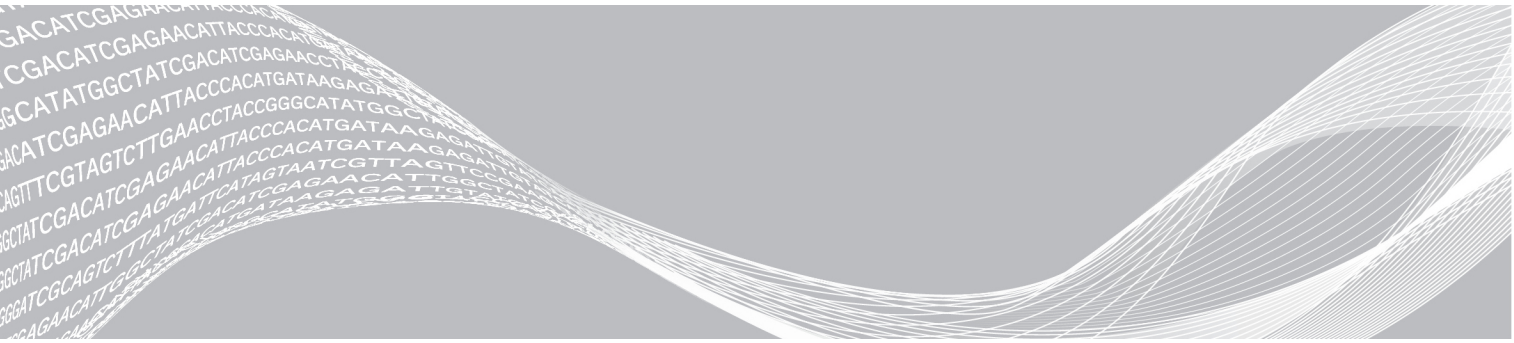
illumina®

Local Run Manager
Mô-đun phân tích Germline Variant

Hướng dẫn về quy trình công việc cho NextSeq 550Dx

DÙNG CHO CHẨN ĐOÁN IN VITRO

Tổng quan	3
Nhập thông tin lần chạy	3
Phương pháp phân tích	5
Xem dữ liệu lần chạy và mẫu	7
Báo cáo phân tích	7
Các tệp đầu ra phân tích	8
Lịch sử sửa đổi	16
Hỗ trợ kỹ thuật	17



Tài liệu số 1000000030329 v04

Tháng 8 năm 2021

TÀI LIỆU ĐỘC QUYỀN CỦA ILLUMINA

DÙNG CHO CHẨN ĐOÁN IN VITRO

Tài liệu này và nội dung trong đó thuộc quyền sở hữu của Illumina, Inc. và các công ty liên kết của Illumina, Inc. ("Illumina") và chỉ dành cho việc sử dụng theo hợp đồng với khách hàng của Illumina liên quan đến việc sử dụng (các) sản phẩm được mô tả trong tài liệu này và không dành cho mục đích nào khác. Tài liệu này và nội dung trong đó sẽ không được sử dụng hay phân phối vì bất kỳ mục đích nào khác và/hoặc không được truyền tải, tiết lộ hay sao chép dưới bất kỳ hình thức nào khác mà không có sự cho phép trước bằng văn bản của Illumina. Illumina không chuyển nhượng bất kỳ giấy phép nào theo các bằng sáng chế, nhãn hiệu, bản quyền hoặc các quyền theo thông luật cũng như các quyền tương tự của bất kỳ bên thứ ba nào thông qua tài liệu này.

Các hướng dẫn nêu trong tài liệu này phải được tuân thủ nghiêm ngặt và rõ ràng bởi cá nhân được đào tạo phù hợp và có đủ trình độ nhằm đảm bảo sử dụng an toàn và đúng cách (các) sản phẩm được mô tả trong tài liệu này. Phải đọc và hiểu hoàn toàn tất cả nội dung của tài liệu này trước khi sử dụng (các) sản phẩm đó.

VIỆC KHÔNG ĐỌC TOÀN BỘ VÀ TUÂN THỦ RÕ RÀNG TẤT CẢ CÁC HƯỚNG DẪN NÊU TRONG TÀI LIỆU NÀY CÓ THỂ DẪN ĐẾN GÂY HƯ HỎNG (CÁC) SẢN PHẨM, GÂY TỔN THƯƠNG CHO CON NGƯỜI, BAO GỒM NGƯỜI DÙNG HOẶC NHỮNG NGƯỜI KHÁC VÀ GÂY THIỆT HẠI TÀI SẢN KHÁC, VÀ SẼ LÀM MẤT HIỆU LỰC BẢO HÀNH ÁP DỤNG CHO (CÁC) SẢN PHẨM ĐÓ.

ILLUMINA KHÔNG CHỊU BẤT KỲ TRÁCH NHIỆM NÀO PHÁT SINH TỪ VIỆC SỬ DỤNG KHÔNG ĐÚNG CÁCH (CÁC) SẢN PHẨM ĐƯỢC MÔ TẢ TRONG TÀI LIỆU NÀY (BAO GỒM CẢ CÁC BỘ PHẬN CỦA SẢN PHẨM HOẶC PHẦN MỀM).

© Năm 2021 Illumina, Inc. Bảo lưu mọi quyền.

Tất cả các nhãn hiệu đều là tài sản của Illumina, Inc. hoặc các chủ sở hữu tương ứng. Để biết thông tin cụ thể về nhãn hiệu, hãy truy cập trang web www.illumina.com/company/legal.html.

Tổng quan

Mô-đun Local Run Manager Germline Variant được sử dụng với xét nghiệm Illumina TruSeq Custom Amplicon Kit Dx và NextSeq 550Dx. Khi được sử dụng với Mô-đun Germline Variant, xét nghiệm được dùng để chuẩn bị thư viện phục vụ cho việc giải trình tự DNA từ mẫu máu toàn phần ngoại vi.

Mô-đun phân tích đánh giá các vùng DNA đã được khuếch đại ngắn, hay còn gọi là amplicon, để tìm biến thể. Giải trình tự tập trung vào các amplicon cho phép đạt độ phủ cao tại các vùng cụ thể. Xem tờ thông tin sản phẩm *TruSeq Custom Amplicon Kit Dx (tài liệu số 1000000029772)*.

Mô-đun phân tích Germline Variant yêu cầu vật tư tiêu hao giải trình tự 300 chu kỳ. Để biết thêm thông tin, xem tờ thông tin sản phẩm của *NextSeq 550Dx High Output Reagent Kit v2* hoặc *NextSeq 550Dx High Output Reagent Kit v2.5*.

Giới thiệu về tài liệu hướng dẫn này

Hướng dẫn này cung cấp các chỉ dẫn để thiết lập các tham số lần chạy cho việc giải trình tự và phân tích trong mô-đun phân tích Germline Variant. Để biết thông tin về trang tổng quan Local Run Manager và cài đặt hệ thống, hãy xem *Hướng dẫn tham khảo về thiết bị NextSeq 550Dx (tài liệu số 1000000009513)*.

Xem Local Run Manager

Bạn có thể xem giao diện Local Run Manager trong NextSeq 550Dx Operating Software (NOS) hoặc thông qua trình duyệt web. Trình duyệt web được hỗ trợ là Chromium.

LƯU Ý

Nếu bạn sử dụng trình duyệt không được hỗ trợ, hãy tải xuống trình duyệt được hỗ trợ khi nhận được thông báo nhắc nhở "Confirm Unsupported Browser" (Xác nhận trình duyệt không được hỗ trợ). Chọn "**here**" (tại đây) để tải xuống phiên bản Chromium được hỗ trợ.

Xem trên màn hình thiết bị

1 Để xem giao diện Local Run Manager trên màn hình thiết bị, hãy chọn một trong các tùy chọn sau:
   - Trên màn hình Home (Trang chủ) của NOS, chọn **Local Run Manager**.
     Nhấp vào dấu X ở góc trên bên phải để quay lại NOS khi hoàn tất.
   - Chọn biểu tượng Minimize (Thu nhỏ) NOS, mở trình duyệt web Chromium trên thiết bị và nhập **http://localhost** vào thanh địa chỉ.
     Chỉ người dùng quản trị mới có thể thu nhỏ NOS.

Xem trên máy tính có kết nối mạng

1 Mở trình duyệt web Chromium trên máy tính có quyền truy cập vào cùng một mạng với thiết bị và kết nối bằng địa chỉ IP của thiết bị hoặc tên thiết bị. Ví dụ: **http://myinstrument**.

Nhập thông tin lần chạy

Thiết lập thông số

1 Đăng nhập vào Local Run Manager.

2 Chọn **Create Run** (Tạo lần chạy) và chọn **Germline Variant** (Biến thể dòng mầm).

3 Nhập tên lượt chạy giúp xác định lượt chạy từ bước giải trình tự đến bước phân tích.
Sử dụng các ký tự chữ và số, dấu cách, dấu gạch dưới hoặc dấu gạch ngang.

4 **[Không bắt buộc]** Nhập thông tin mô tả lần chạy để giúp xác định lần chạy.
Sử dụng các ký tự chữ và số, dấu cách, dấu gạch dưới hoặc dấu gạch ngang.

5 Chọn số mẫu và bộ chỉ thị từ danh sách thả xuống.
Hãy cân nhắc thông tin sau đây khi bạn lựa chọn.

- Danh sách thả xuống chứa số mẫu cùng với một bộ chỉ thị. Ví dụ: 24-Set 1 (24-Bộ 1) thể hiện rằng 24 mẫu sẽ được xét nghiệm, với các chỉ thị từ bộ chỉ thị 1.
- Số bộ chỉ mục đề cập đến các bộ chỉ mục i5 khác nhau. Bộ 1 và Bộ 2 đều cung cấp tính đa dạng chỉ mục. Hai bộ chỉ thị được cung cấp để giúp ngăn tình trạng thiếu hụt nếu chỉ có một bộ duy nhất.
- Chọn số mẫu gần nhất với số mẫu bạn sẽ xét nghiệm. Nếu danh sách không có số mẫu chính xác bạn cần, hãy chọn số mẫu gần nhất, nhưng ít hơn số mẫu bạn sẽ xét nghiệm. Ví dụ: nếu bạn muốn xét nghiệm 18 mẫu, hãy chọn 16 mẫu.
- Các giếng mẫu và tổ hợp chỉ mục đáp ứng yêu cầu về tính đa dạng chỉ mục được nêu bật bằng màu xanh lục. Nếu bạn chọn các giếng và tổ hợp chỉ mục khác, khi lưu lần chạy, bạn sẽ nhận được thông báo nếu các yêu cầu về tính đa dạng chỉ mục không được đáp ứng.

Nhập tệp phiếu kê khai cho lần chạy

1 Đảm bảo rằng các phiếu kê khai bạn muốn nhập có sẵn ở một vị trí mạng có thể truy cập hoặc trên ổ USB.

2 Chọn **Import Manifests** (Nhập phiếu kê khai).

3 Chuyển đến tệp phiếu kê khai và chọn các phiếu kê khai mà bạn muốn thêm.

LƯU Ý

Để chuẩn bị sẵn sàng tệp phiếu kê khai cho tất cả các lần chạy bằng mô-đun phân tích Germline Variant, hãy thêm các phiếu kê khai bằng tính năng Module Settings (Cài đặt mô-đun). Tính năng này yêu cầu bạn phải có quyền cấp người dùng quản trị viên. Để biết thêm thông tin, tham khảo *Hướng dẫn tham khảo về thiết bị NextSeq 550Dx (tài liệu số 1000000009513)*.

Chỉ định mẫu cho lần chạy

Chỉ định mẫu cho lượt chạy bằng một trong các tùy chọn và các hướng dẫn sau đó.

- **Nhập mẫu theo cách thủ công** – Sử dụng bảng trống trên màn hình Create Run (Tạo lần chạy).
- **Nhập mẫu** – Chuyển đến tệp bên ngoài ở định dạng giá trị được phân tách bằng dấu phẩy (*.csv). Trên màn hình Create Run (Tạo lần chạy) sẽ có sẵn một mẫu cho bạn tải xuống.

Sau khi bạn đã điền thông tin vào bảng mẫu, bạn có thể xuất thông tin mẫu ra một tệp bên ngoài. Sử dụng tệp này làm tham chiếu khi chuẩn bị thư viện hoặc nhập tệp cho một lần chạy khác.

Nhập mẫu theo cách thủ công

1 Nhập ID mẫu duy nhất vào trường Sample ID (ID mẫu).
Sử dụng các ký tự chữ và số, dấu gạch ngang hoặc dấu gạch dưới.

2 **[Không bắt buộc]** Đối với các mẫu chứng dương hoặc âm, nhấp chuột phải và chọn loại mẫu chứng.

3 **[Không bắt buộc]** Nhập thông tin mô tả mẫu vào trường Sample Description (Mô tả mẫu).
Sử dụng các ký tự chữ và số, dấu gạch ngang hoặc dấu gạch dưới.
Mô tả mẫu được liên kết với một tên mẫu. Mô tả mẫu sẽ bị ghi đè nếu sử dụng lại cùng một tên mẫu trong một lần chạy khác sau này.

4 Chọn một adapter Chỉ thị 1 từ danh sách thả xuống Index 1 (i7).
Khi bạn sử dụng các giếng mẫu được đề xuất, phần mềm sẽ tự động điền các adapter chỉ thị i7 và i5 đáp ứng các yêu cầu về đa dạng chỉ thị. Nếu danh sách không có số mẫu chính xác mà bạn sẽ xét nghiệm, hãy nhớ chọn adapter chỉ thị cho các giếng trội thêm. Nếu bạn cần chọn chỉ thị cho các giếng bổ sung, hoặc bạn không sử dụng các tổ hợp adapter chỉ thị được đề xuất, hãy đảm bảo đọc *Phát hiện base và tính đa dạng của chỉ thị* trên trang 15 trước khi chọn chỉ thị.

5 Chọn một adapter Chỉ thị 2 từ danh sách thả xuống Index 2 (i5).

6 Chọn tệp phiếu kê khai từ danh sách thả xuống Manifest (Phiếu kê khai).

7 Chọn một tùy chọn để xem, in hoặc lưu cách bố trí khay làm tham chiếu khi chuẩn bị thư viện:
- Chọn biểu tượng **Print** (In) để hiển thị cách bố trí khay. Chọn **Print** (In) để in cách bố trí khay.
- Chọn **Export** (Xuất) để xuất thông tin mẫu ra tệp bên ngoài.

Đảm bảo rằng thông tin phiếu kê khai và mẫu là chính xác. Thông tin không chính xác có thể ảnh hưởng đến kết quả.

8 Chọn **Save Run** (Lưu lần chạy).

Nhập mẫu

1 Chọn **Import Samples** (Nhập mẫu) rồi duyệt đến vị trí của tệp thông tin mẫu. Bạn có thể nhập hai loại tệp.
- Chọn **Template** (Mẫu) trên màn hình Create Run (Tạo lần chạy) để tạo cách bố trí khay mới. Tệp mẫu chứa các tiêu đề cột chính xác để nhập thông tin. Nhập thông tin mẫu vào mỗi cột cho các mẫu trong lần chạy. Xóa thông tin ví dụ trong các ô không sử dụng, sau đó lưu tệp.
- Sử dụng tệp thông tin mẫu đã được xuất từ mô-đun Germline Variant bằng tính năng Export (Xuất).

2 Chọn biểu tượng **Print** (In) để hiển thị cách bố trí khay.

3 Chọn **Print** (In) để in cách bố trí khay làm tham chiếu khi chuẩn bị thư viện.

4 **[Không bắt buộc]** Chọn **Export** (Xuất) để xuất thông tin mẫu ra tệp bên ngoài.
Đảm bảo rằng thông tin phiếu kê khai và mẫu là chính xác. Thông tin không chính xác có thể ảnh hưởng đến kết quả.

5 Chọn **Save Run** (Lưu lần chạy).

Chỉnh sửa lần chạy

Để biết hướng dẫn về cách chỉnh sửa thông tin trong lần chạy của bạn trước khi giải trình tự, hãy xem *Hướng dẫn tham khảo về thiết bị NextSeq 550Dx (tài liệu số 1000000009513)*.

Phương pháp phân tích

Mô-đun phân tích Germline Variant thực hiện các bước phân tích sau đây rồi ghi tệp đầu ra phân tích vào thư mục Alignment (Phân tích so khớp).

- Tách các đoạn đọc chỉ thị

- Tạo tệp FASTQ
- So khớp với tham chiếu
- Xác định biến thể

Tách đoạn

Bước tách đoạn so sánh từng trình tự Đoạn đọc chỉ thị với các trình tự chỉ thị được chỉ định cho lần chạy. Bước này không xem xét giá trị chất lượng nào.

Đoạn đọc chỉ thị được xác định bằng các bước sau:

- Các mẫu được đánh số bắt đầu từ 1 dựa trên thứ tự mẫu được liệt kê cho lần chạy.
- Mẫu số 0 được dành riêng cho các cụm không được chỉ định cho một mẫu.
- Các cụm được chỉ định cho một mẫu khi trình tự chỉ thị khớp chính xác hoặc khi có tối đa một lần không khớp trên mỗi Đoạn đọc chỉ thị.

Tạo tệp FASTQ

Sau khi tách đoạn, phần mềm tạo các tệp phân tích trung gian ở định dạng FASTQ, đây là định dạng văn bản được sử dụng để trình bày các trình tự. Tệp FASTQ chứa các đoạn đọc cho từng mẫu và điểm chất lượng liên quan. Các cụm không đi qua bộ lọc đều bị loại trừ.

Mỗi tệp FASTQ chứa các đoạn đọc cho riêng một mẫu và tên của mẫu đó sẽ có trong tên tệp FASTQ. Tệp FASTQ là đầu vào chính để so khớp. Tám tệp FASTQ được tạo ra cho mỗi mẫu cho, bốn tệp từ Đoạn đọc 1 và bốn tệp từ Đoạn đọc 2.

So khớp

Trong bước so khớp, thuật toán Smith-Waterman dạng dải băng sẽ so khớp các cụm của từng mẫu với các trình tự amplicon trong tệp phiếu kê khai.

Thuật toán Smith-Waterman theo dải băng thực hiện so khớp trình tự cục bộ để xác định các vùng tương tự giữa hai trình tự. Thay vì so sánh tổng trình tự, thuật toán Smith-Waterman so sánh các đoạn thuộc tất cả các độ dài có thể.

Mỗi đoạn đọc kết đôi được đánh giá về mức độ so khớp với các trình tự đầu dò liên quan cho đoạn đọc đó.

- Đoạn đọc 1 được đánh giá dựa trên phần bổ sung ngược của Oligo dành riêng cho lô-cut xuôi hướng (DLSO).
- Đoạn đọc 2 được đánh giá dựa trên Oligo dành riêng cho lô-cut ngược hướng (ULSO).
- Nếu phần đầu của đoạn đọc khớp với trình tự đầu dò có không quá ba khác biệt (số lần không khớp hoặc thay đổi do indel dẫn đầu) thì toàn bộ độ dài của đoạn đọc được so khớp với mục tiêu amplicon cho trình tự đó.
- Indel trong DLSO và ULSO không được quan sát, xét công nghệ hóa học của xét nghiệm.

Các chuỗi so khớp được lọc từ kết quả so khớp dựa trên tỷ lệ không khớp trên vùng quan tâm hoặc toàn bộ amplicon, tùy thuộc vào độ dài amplicon. Các chuỗi so khớp đã lọc được ghi trong các tệp so khớp là không so khớp và không được sử dụng trong việc phát hiện biến thể.

Phát hiện biến thể

Được phát triển bởi Illumina, Pisces Variant Caller xác định các biến thể có trong mẫu DNA.

Pisces variant caller xác định các SNV, MNV và indel nhỏ theo 3 bước:

- Xem xét riêng từng vị trí trên hệ gen tham chiếu
- Đếm số base tại vị trí đó đối với các đoạn đọc so khớp phủ qua vị trí
- Tính điểm biến thể đo chất lượng phát hiện bằng mô hình Poisson. Các biến thể có điểm chất lượng dưới Q20 sẽ bị loại bỏ.

Nếu vượt qua tất cả các bộ lọc, biến thể được đánh dấu là PASS (ĐẠT) trong VCF.

Để biết thêm thông tin, xem github.com/Illumina/Pisces/wiki.

Xem dữ liệu lần chạy và mẫu

1 Từ trang tổng quan Local Run Manager, chọn tên lần chạy.

2 Từ tab Run Overview (Tổng quan lượt chạy), hãy xem xét các chỉ số của lượt chạy giải trình tự.

3 **[Không bắt buộc]** Nhấp vào biểu tượng **Copy to Clipboard** (Sao chép vào bảng nhớ tạm) để sao chép đường dẫn thư mục lần chạy đầu ra.

4 Nhấp vào tab Sequencing Information (Thông tin giải trình tự) để xem lại các thông số lần chạy và thông tin về vật tư tiêu hao.

5 Nhấp vào tab Samples and Results (Mẫu và kết quả) để xem vị trí báo cáo.
   - Nếu đã lặp lại phân tích, mở rộng danh sách thả xuống Select Analysis (Chọn phân tích) và chọn phân tích thích hợp.

6 Nhấp vào biểu tượng **Copy to Clipboard** (Sao chép vào bảng nhớ tạm) để sao chép đường dẫn thư mục Analysis (Phân tích).

Để biết thêm thông tin về các tab Run Overview (Tổng quan về lần chạy), Sequencing Information (Thông tin giải trình tự) và cách xếp lại hàng đợi phân tích, hãy xem *Hướng dẫn tham khảo về thiết bị NextSeq 550Dx (tài liệu số 1000000009513)*.

Báo cáo phân tích

Kết quả phân tích được tóm tắt trong tab Samples and Results (Mẫu và kết quả), và dưới dạng báo cáo tổng hợp trong thư mục Alignment (So khớp). Báo cáo cho từng mẫu cũng có sẵn ở định dạng tệp PDF cho mỗi mẫu.

Mẫu và Kết quả Tab Thông tin

1 Nhấp vào mẫu trong danh sách để xem báo cáo mẫu.

Bảng 1 Thông tin lần chạy và mẫu

Tiêu đề cột	Mô tả
Run Status (Trạng thái lần chạy)	Cho biết lần chạy giải trình tự đạt hay không.
Total Yield (GB) (Tổng năng suất (GB))	Số lượng base được phát hiện trong lần chạy giải trình tự. Hiển thị ngưỡng đạt chuẩn, và trạng thái đạt hoặc không đạt.
% ≥ Q30	Tỷ lệ phần trăm đoạn đọc trong lần chạy có điểm chất lượng ≥ 30 (Q30). Hiển thị ngưỡng đạt chuẩn, và trạng thái đạt hoặc không đạt.
Sample ID (ID mẫu)	ID mẫu được cung cấp khi quá trình chạy được tạo.

Tiêu đề cột	Mô tả
Total PF Reads (Tổng số đoạn đọc PF)	Tổng số đoạn đọc vượt qua bộ lọc.
Read 1% ≥ Q30 (Đoạn đọc 1% ≥ Q30)	Tỷ lệ đoạn đọc của Đoạn đọc 1 có điểm chất lượng ≥ 30 (Q30) cho mẫu này.
Read 2% ≥ Q30 (Đoạn đọc 2% ≥ Q30)	Tỷ lệ đoạn đọc của Đoạn đọc 2 có điểm chất lượng ≥ 30 (Q30) cho mẫu này.
Autosome Call Rate (Tỷ lệ phát hiện nhiễm sắc thể thường)	Số lượng vị trí hệ gen trên các nhiễm sắc thể thường (nhiễm sắc thể từ 1 đến 22) đạt ngưỡng giá trị độ tin cậy được xác định trước, chia cho tổng số vị trí hệ gen trên nhiễm sắc thể thường được kiểm tra. Tỷ lệ phát hiện được mô tả trên từng mẫu và được báo cáo dưới dạng phần trăm, tính bằng 1 trừ đi (số lượng vị trí trên nhiễm sắc thể thường có kết quả phát hiện không hoàn chỉnh chia cho tổng số vị trí nhiễm sắc thể thường được giải trình tự).

Bảng 2 Mẫu Báo cáo Thông tin

Tiêu đề cột	Mô tả
Sample (Mẫu)	ID mẫu được cung cấp khi quá trình chạy được tạo.
Report Date (Ngày báo cáo)	Ngày báo cáo được tạo.
Sample Information (Thông tin mẫu)	Mã định danh mẫu được cung cấp khi tạo lần chạy, tổng số đoạn đọc vượt qua bộ lọc trong mẫu, tỷ lệ phần trăm đoạn đọc của mẫu có điểm chất lượng từ 30 (Q30) trở lên, và tỷ lệ phát hiện nhiễm sắc thể thường.
Amplicon Summary (Tóm tắt amplicon)	Tổng số vùng amplicon được giải trình tự và tổng chiều dài tính bằng cặp base của các amplicon được giải trình tự trong các vùng mục tiêu, đối với mẫu và tệp phiếu kê khai. Tệp phiếu kê khai chỉ định hệ gen tham chiếu và các vùng tham chiếu mục tiêu được sử dụng trong bước so khớp.
Read Level Statistics (Thống kê cấp độ đoạn đọc)	Số lượng và tỷ lệ đoạn đọc cho mẫu bao phủ từng vị trí trên tham chiếu, cho Đoạn đọc 1 và Đoạn đọc 2.
Variants Summary (Tóm tắt biến thể)	Số lượng SNV, đột biến thêm đoạn và đột biến mất đoạn được phát hiện cho mẫu đạt các giá trị khuyến nghị để xác định kết quả chất lượng có nằm trong phạm vi chấp nhận được hay không,
Coverage Summary (Tóm tắt độ bao phủ)	Tổng số base đã so khớp chia cho kích thước vùng mục tiêu, và tỷ lệ phần trăm các vùng amplicon có độ bao phủ lớn hơn ngưỡng độ bao phủ thấp (0,2 * độ bao phủ trung bình của amplicon), đối với mẫu.
Coverage Plots (Biểu đồ độ bao phủ)	Biểu đồ độ bao phủ theo Vùng Amplicon thể hiện độ bao phủ trên toàn bộ các vùng amplicon cho mẫu. Các vùng có giá trị độ bao phủ thấp hơn ngưỡng độ bao phủ được tô màu đỏ. Giá trị trung bình của tất cả các giá trị được thể hiện bằng đường màu cam.
Software Versions (Phiên bản phần mềm)	Các phiên bản phần mềm khi mẫu được giải trình tự. Bao gồm NextSeq 550Dx Operating Software (NOS), Phần mềm Local Run Manager, Phần mềm RTA, và phiên bản mô-đun Germline Variant.

Các tệp đầu ra phân tích

Các tệp đầu ra phân tích sau đây được tạo cho mô-đun phân tích Germline Variant và cung cấp kết quả phân tích cho quá trình so khớp và phát hiện biến thể. Các tệp đầu ra phân tích nằm trong thư mục Alignment (So khớp).

Tên tệp	Mô tả
Demultiplexing (*.txt)	Các tệp trung gian chứa kết quả tóm tắt tách đoạn.
FASTQ (*.fastq.gz)	Tệp trung gian chứa các phát hiện base đã được chấm điểm chất lượng. Các tệp FASTQ là dữ liệu đầu vào chính cho bước so khớp.
Các tệp so khớp ở định dạng BAM (*.bam)	Chứa các đoạn đọc đã được so khớp cho một mẫu cụ thể.
Các tệp phát hiện biến thể ở định dạng genome VCF (*.genome.vcf.gz)	Chứa kiểu gen cho từng vị trí, dù được phát hiện là biến thể hay tham chiếu.
Các tệp phát hiện biến thể ở định dạng VCF (*.vcf.gz)	Chứa tất cả các biến thể được phát hiện trên toàn bộ vùng mục tiêu.
AmpliconCoverage_M1.tsv	Chứa thông tin về độ bao phủ trên từng amplicon cho từng mẫu đối với mỗi phiếu kê khai được cung cấp. M# đại diện cho số phiếu kê khai.

Định dạng tệp tách đoạn

Quá trình tách đoạn đọc trình tự chỉ thị gắn với mỗi cụm để xác định cụm đó bắt nguồn từ mẫu nào. Ánh xạ giữa các cụm và số mẫu được ghi vào một tệp tách đoạn (*.demux) cho mỗi ô lưới của tế bào dòng chảy.

Định dạng tên tệp tách đoạn là s_1_X.demux, trong đó X là số ô lưới.

Các tệp tách đoạn bắt đầu với phần đầu:

- Phiên bản (số nguyên 4 byte), hiện tại là 1
- Số lượng cụm (số nguyên 4 byte)

Phần còn lại của tệp bao gồm số mẫu cho mỗi cụm từ ô lưới.

Khi bước tách đoạn hoàn tất, phần mềm tạo ra tệp tách đoạn tên là DemultiplexSummaryF1L1.txt.

- Trong tên tệp, **F1** đại diện cho số tế bào dòng chảy.
- Trong tên tệp, **L1** đại diện cho số làn.
- Việc tách đoạn tạo ra bảng có 1 hàng cho mỗi ô lưới và 1 cột cho mỗi mẫu, bao gồm cả mẫu 0.
- Ngoài ra, tệp còn ghi lại các chuỗi thường gặp nhất trong đoạn đọc chỉ thị.

Định dạng tệp FASTQ

FASTQ là một định dạng tệp văn bản chứa các phát hiện base và giá trị chất lượng cho mỗi đoạn đọc. Mỗi bản ghi có 4 dòng:

- Mã nhận dạng
- Trình tự
- Dấu cộng (+)
- Điểm chất lượng Phred được mã hóa theo định dạng ASCII + 33

Mã nhận dạng được định dạng là:

@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber

Ví dụ:

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

Định dạng tệp BAM

Tệp BAM (*.bam) là phiên bản nén nhị phân của tệp SAM, được dùng để biểu diễn các trình tự đã so khớp lên đến 128 Mb. Các định dạng SAM và BAM được mô tả chi tiết tại samtools.github.io/hts-specs/SAMv1.pdf.

Tệp BAM dùng định dạng tên SampleName_S#.bam, trong đó # là số mẫu được xác định bằng thứ tự mà các mẫu được liệt kê cho lần chạy.

Tệp BAM gồm có phần đầu và phần so khớp:

- **Phần đầu**–Chứa thông tin về toàn bộ tệp, chẳng hạn như tên mẫu, độ dài mẫu, và phương pháp so khớp. Các so khớp trong phần so khớp sẽ được liên kết với thông tin cụ thể trong phần đầu.
- **So khớp**–Chứa tên đoạn đọc, trình tự đoạn đọc, chất lượng đọc, thông tin so khớp và thẻ tùy chỉnh. Tên đoạn đọc bao gồm nhiễm sắc thể, tọa độ bắt đầu, chất lượng so khớp, và chuỗi mô tả khớp.

Phần so khớp bao gồm các thông tin sau cho mỗi đoạn đọc hoặc cặp đoạn đọc:

- **AS:** Chất lượng so khớp của đoạn đọc kết đôi
- **RG:** Nhóm đoạn đọc, cho biết số lượng đoạn đọc của một mẫu cụ thể.
- **BC:** Thẻ mã vạch, dùng để chỉ định ID mẫu đã được tách đoạn gắn với đoạn đọc.
- **SM:** Chất lượng so khớp của đoạn đọc đơn.
- **XC:** Chuỗi mô tả khớp
- **XN:** Thẻ tên amplicon, ghi lại ID amplicon gắn với đoạn đọc

Các tệp chỉ thị BAM (*.bam.bai) cung cấp một chỉ thị cho tệp BAM tương ứng.

Định dạng tệp VCF

Variant Call Format (VCF) là định dạng tệp chung do cộng đồng khoa học di truyền học phát triển. Nó chứa thông tin về các biến thể được tìm thấy tại các vị trí cụ thể trên hệ gen tham chiếu. Các tệp VCF kết thúc với hậu tố .vcf

Phần đầu tệp VCF bao gồm phiên bản định dạng tệp VCF, phiên bản công cụ phát hiện biến thể và liệt kê các chú giải được sử dụng trong phần còn lại của tệp. Phần đầu tệp VCF cũng bao gồm tệp hệ gen tham chiếu và tệp BAM. Dòng cuối cùng trong phần đầu chứa tiêu đề cột cho các dòng dữ liệu. Mỗi dòng dữ liệu trong tệp VCF chứa thông tin về một biến thể.

Các tiêu đề trong tệp VCF

Tiêu đề	Mô tả
CHROM	Nhiễm sắc thể của hệ gen tham chiếu. Các nhiễm sắc thể xuất hiện theo cùng thứ tự như trong tệp FASTQ tham chiếu.
POS	Vị trí một base duy nhất của biến thể trên nhiễm sắc thể tham chiếu. Đối với SNP, vị trí này là base tham chiếu cùng với biến thể; đối với các indel hoặc vùng xóa, vị trí này là base tham chiếu ngay trước biến thể.
ID	Số rs cho biến thể được lấy từ dbSNP.txt, nếu có. Nếu tại vị trí này có nhiều số rs, danh sách sẽ được phân tách bằng dấu chấm phẩy. Nếu không có mục nhập dbSNP tại vị trí này, sẽ sử dụng ký hiệu giá trị thiếu ('.').
REF	Hệ gen tham chiếu. Ví dụ: một đột biến mất một nucleotide T được biểu diễn là kiểu gen tham chiếu TT và kiểu gen thay thế T. Một biến thể đơn nucleotide từ A sang T được biểu diễn là kiểu gen tham chiếu A và kiểu gen thay thế T.
ALT	Các alen khác với đoạn đọc tham chiếu. Ví dụ: một vùng chèn thêm một T duy nhất được biểu diễn là kiểu gen tham chiếu A và kiểu gen thay thế AT. Một biến thể đơn nucleotide từ A sang T được biểu diễn là kiểu gen tham chiếu A và kiểu gen thay thế T.
QUAL	Điểm chất lượng theo thang Phred do trình phát hiện biến thể gán. Điểm càng cao thì độ tin cậy về biến thể càng lớn và xác suất lỗi càng thấp. Với điểm chất lượng Q, xác suất lỗi ước tính là $10^{-(Q/10)}$. Ví dụ: bộ phát hiện Q30 có tỷ lệ lỗi 0,1%. Nhiều trình phát hiện biến thể gán điểm chất lượng dựa trên mô hình thống kê, thường cao hơn so với tỷ lệ lỗi quan sát được.

Các chú thích trong tệp VCF

Tiêu đề	Mô tả
FILTER	Nếu tất cả các bộ lọc đều đạt, giá trị **PASS** sẽ được ghi trong cột bộ lọc. • **LowDP**–Áp dụng cho các vị trí có độ sâu bao phủ dưới 150x. Đối với các vị trí amplicon được bao phủ bởi cả đoạn đọc xuôi và đoạn đọc ngược, điều này tương đương với 300 đoạn đọc kết đôi chồng lấp. • **q20**–Điểm chất lượng <20. • **MultiAllelicSite**–Biến thể không tuân theo mô hình lưỡng bội. • **R5x9**–Số lượng chuỗi lặp lại liền kề (độ dài 1 đến 5 bp) so với kết quả phát hiện biến thể ≥ 9. • **SB**–Độ lệch theo chiều mạch lớn hơn ngưỡng quy định.
INFO	Các mục nhập có thể có trong cột INFO bao gồm: • **AC**–Số lượng alen trong kiểu gen đối với mỗi alen THAY THẾ (ALT), theo đúng thứ tự được liệt kê. • **AF**–Tần suất alen đối với mỗi alen THAY THẾ, theo đúng thứ tự được liệt kê. • **AN**–Tổng số lượng alen trong các kiểu gen đã được phát hiện. • **CD**–Cờ cho biết SNP xảy ra trong vùng mã hóa của ít nhất một mục nhập RefGene. • **DP**–Độ sâu (số lượng các lần phát hiện base được so khớp tại một vị trí và được sử dụng trong việc phát hiện biến thể). • **Exon**–Danh sách các vùng exon, được phân tách bằng dấu phẩy, đọc từ RefGene. • **FC**–Hậu quả chức năng. • **GI**–Danh sách các ID gen, được phân tách bằng dấu phẩy, đọc từ RefGene. • **QD**–Độ tin cậy/Chất lượng biến thể theo độ sâu đọc. • **TI**–Danh sách các ID bản phiên mã, được phân tách bằng dấu phẩy, đọc từ RefGene.
FORMAT	Cột định dạng liệt kê các trường được phân tách bằng dấu hai chấm. Ví dụ: GT:GQ. Danh sách các trường được cung cấp phụ thuộc vào công cụ phát hiện biến thể được sử dụng. Các trường có sẵn bao gồm: • **AD**–Mục nhập có dạng X,Y, trong đó X là số lần phát hiện alen tham chiếu, và Y là số lần phát hiện alen thay thế. • **DP**–Độ sâu đoạn đọc xấp xỉ; các đoạn đọc có MQ=255 hoặc có cặp lỗi sẽ bị lọc bỏ. • **GQ**–Chất lượng kiểu gen. • **GT**–Kiểu gen. 0 tương ứng với base tham chiếu. 1 tương ứng với mục nhập đầu tiên trong cột ALT. Và tiếp tục như vậy cho các mục nhập kế tiếp. Dấu gạch chéo (/) cho biết không có thông tin định pha khả dụng. • **NC**–Tỷ lệ các base không được phát hiện, hoặc có chất lượng phát hiện base dưới ngưỡng tối thiểu. • **NL**–Mức nhiễu; ước lượng mức độ nhiễu trong việc phát hiện base tại vị trí này. • **SB**–Sai lệch theo chiều sợi tại vị trí này. Giá trị âm càng lớn cho thấy càng ít sai lệch hơn; các giá trị gần 0 cho thấy sai lệch nhiều hơn. • **VF**–Tần suất biến thể; phần trăm số đoạn đọc hỗ trợ cho alen thay thế.
SAMPLE	Cột mẫu cung cấp các giá trị được chỉ định trong cột FORMAT.

Tệp Genome VCF

Tệp Genome VCF (gVCF) là tệp VCF v4.1 tuân theo bộ quy ước biểu diễn mọi vị trí trong hệ gen dưới định dạng nhỏ gọn hợp lý. Tệp gVCF (*.genome.vcf.gz) bao gồm tất cả vị trí trong vùng quan tâm trong một tệp duy nhất cho từng mẫu.

Tệp gVCF hiển thị không phát hiện tại các vị trí không đi qua tất cả bộ lọc. Thẻ kiểu gen (GT) **./.** biểu thị trạng thái không phát hiện.

Để biết thêm thông tin, xem sites.google.com/site/gvcftools/home/about-gvcf.

Tệp độ bao phủ amplicon

Một tệp độ bao phủ amplicon được tạo cho từng tệp phiếu kê khai. M# trong tên tệp biểu thị số phiếu kê khai.

Mỗi tệp gồm một dòng tiêu đề chứa các ID mẫu liên kết với phiếu kê khai. Tệp chứa các thông tin sau.

- ID đích như được liệt kê trong phiếu kê khai.
- Độ sâu bao phủ của các đoạn đọc vượt qua bộ lọc.

Tệp đầu ra bổ sung

Các tệp đầu ra sau đây cung cấp thông tin bổ sung hoặc tóm tắt kết quả chạy và lỗi phân tích. Mặc dù các tệp này không bắt buộc để đánh giá kết quả phân tích, nhưng chúng có thể được sử dụng cho mục đích xử lý sự cố. Tất cả các tệp đều nằm trong thư mục Alignment (So khớp), trừ khi có chỉ định khác.

Tên tệp	Mô tả
AnalysisLog.txt	Nhật ký xử lý mô tả từng bước đã diễn ra trong quá trình phân tích thư mục chạy hiện tại. Tệp này không chứa thông báo lỗi. Nằm ở thư mục Alignment (So khớp).
AnalysisError.txt	Nhật ký xử lý liệt kê bất kỳ lỗi nào xảy ra trong quá trình phân tích. Tệp này sẽ trống nếu không có lỗi nào xảy ra. Nằm ở thư mục Alignment (So khớp).
DemultiplexSummaryF1L1#.txt	Báo cáo kết quả tách đoạn trong một bảng, với 1 hàng cho mỗi ô lưới và 1 cột cho mỗi mẫu. # đại diện cho làn 1, 2, 3, hoặc 4 của tế bào dòng chảy. Nằm ở thư mục Alignment (So khớp).
AmpliconRunStatistics.xml	Chứa thống kê tóm tắt cụ thể cho lần chạy. Nằm ở thư mục Alignment (So khớp).

Analysis Folder (Thư mục phân tích)

Thư mục phân tích chứa các tệp được phần mềm Local Run Manager tạo ra.

Mối quan hệ giữa thư mục đầu ra và thư mục phân tích được tóm tắt như sau:

- Trong quá trình giải trình tự, Real-Time Analysis (RTA) sẽ điền vào thư mục đầu ra các tệp được tạo ra trong quá trình phân tích hình ảnh, phát hiện base và chấm điểm chất lượng.
- RTA sao chép các tệp vào thư mục phân tích trong thời gian thực. Sau khi RTA gán điểm chất lượng cho từng base ở mỗi chu kỳ, phần mềm sẽ ghi tệp RTAComplete.txt vào cả hai thư mục.
- Khi tệp RTAComplete.txt xuất hiện, quá trình phân tích bắt đầu.
- Khi quá trình phân tích tiếp tục, Local Run Manager sẽ ghi các tệp đầu ra vào thư mục phân tích, sau đó sao chép các tệp này trở lại thư mục đầu ra.

Các thư mục so khớp

Mỗi lần phân tích được xếp lại hàng đợi, Local Run Manager sẽ tạo một thư mục Alignment (So khớp) có tên là **Alignment_N**, trong đó N là một số thứ tự tăng dần.

Cấu trúc thư mục

Alignment (So khớp)–Chứa các tệp *.bam, *.vcf, FASTQ và các tệp dành riêng cho mô-đun phân tích.

- **Date and Time Stamp** (Dấu ngày và giờ)–Dấu ngày giờ phân tích dạng YYYYMMDD_HHMMSS
  - AnalysisError.txt
  - AnalysisLog.txt
  - aggregate.report.html
  - aggregate.report.pdf
  - aggregate.summary.csv
  - AmpliconCoverage_M#.tsv
  - AmpliconRunStatistics.xml
  - Sample1.genome.vcf.gz
  - Sample1.coverage.csv
  - Sample1.report.pdf
  - Sample1.summary.csv
  - Sample1.vcf.gz
  - Sample1.bam
  - **FASTQ**
    - **Sample1**
      - Sample1_L001_R1_001_fastq.gz
    - **Stats**
      - DemuxSummaryF1L1.txt
      - FastqSummaryF1L1.txt
- **Data**
  - **Intensities**
    - **BaseCalls**
      - **L001**–Chứa các tệp *.bcl.
    - **L001**–Chứa các tệp *.locs.
  - **RTA Logs**–Chứa các tệp nhật ký từ phân tích phần mềm RTA.
- **InterOp**–Chứa các tệp nhị phân dùng để báo cáo thông số của lần chạy giải trình tự.
- **Logs**–Chứa các tệp nhật ký mô tả các bước được thực hiện trong khi giải trình tự.
- RTAComplete.txt
- RunInfo.xml
- RunParameters.xml

Phát hiện base và tính đa dạng của chỉ thị

Khi các mẫu được giải trình tự trên thiết bị NextSeq 550Dx, quá trình phát hiện base sẽ xác định một base (A, C, G hoặc T) cho mỗi cụm của một ô cho trước hoặc vùng chụp ảnh trên tế bào dòng chảy ở một chu kỳ cụ thể. Thiết bị NextSeq 550Dx sử dụng phương pháp giải trình tự hai kênh, chỉ yêu cầu hai hình ảnh để mã hóa dữ liệu cho bốn base DNA, một từ kênh đỏ và một từ kênh xanh lục.

Quy trình dành cho các đoạn đọc chỉ thị phát hiện base khác với quy trình phát hiện base trong các đoạn đọc khác.

Đoạn đọc chỉ thị phải bắt đầu với ít nhất một base không phải loại G ở một trong hai chu kỳ đầu. Nếu Đoạn đọc chỉ thị bắt đầu với hai phát hiện base loại G thì sẽ không có cường độ tín hiệu nào được tạo ra. Tín hiệu phải xuất hiện ở một trong hai chu kỳ đầu để đảm bảo hiệu suất tách đoạn.

Khi lựa chọn các chỉ thị trong quá trình tạo lần chạy, cảnh báo tính đa dạng thấp sẽ xuất hiện nếu các chỉ thị không đáp ứng yêu cầu về tính đa dạng. Để ngăn xuất hiện cảnh báo tính đa dạng thấp, hãy chọn trình tự chỉ thị cung cấp tín hiệu trong cả hai kênh cho mọi chu kỳ.

- Kênh đỏ: A hoặc C
- Kênh xanh lục: A hoặc T

Quy trình phát hiện base này đảm bảo độ chính xác khi phân tích các mẫu ít kênh. Để biết thêm thông tin về trình tự của các chỉ thị của bạn, xem tờ hướng dẫn sử dụng sản phẩm *TruSeq Custom Amplicon Kit Dx (tài liệu số 1000000029772)*.

Trong quá trình tạo lần chạy trong Local Run Manager, bạn sẽ chọn số lượng mẫu cần được xét nghiệm. Các tổ hợp chỉ thị được đề xuất đáp ứng các yêu cầu về tính đa dạng chỉ thị sẽ được phần mềm tự động điền. Dù không bắt buộc sử dụng các tổ hợp chỉ thị được đề xuất, bạn vẫn nên áp dụng các tổ hợp này.

Lịch sử sửa đổi

Tài liệu	Ngày	Mô tả thay đổi
Tài liệu số 1000000030329 v04	Tháng 8 năm 2021	Đã cập nhật địa chỉ của Đại diện được ủy quyền tại Liên minh châu Âu.
Tài liệu số 1000000030329 v03	Tháng 12 năm 2019	Đã cập nhật địa chỉ của Đại diện được ủy quyền tại Liên minh châu Âu. Cập nhật địa chỉ của Nhà bảo trợ tại Úc.
Tài liệu số 1000000030329 v02	Tháng 1 năm 2019	Đã thêm thông tin về bộ kit thuốc thử v2.5.
Tài liệu số 1000000030329 v01	Tháng 8 năm 2018	Cập nhật nhãn công bố hợp quy.
Tài liệu số 1000000030329 v00	Tháng 11 năm 2017	Phát hành lần đầu.

Hỗ trợ kỹ thuật

Để được hỗ trợ kỹ thuật, hãy liên hệ với bộ phận Hỗ trợ kỹ thuật của Illumina.

Trang web: www.illumina.com
Email: techsupport@illumina.com

Các số điện thoại hỗ trợ khách hàng của Illumina

Khu vực	Số miễn cước	Khu vực
Bắc Mỹ	+1.800.809.4566	
Úc	+1.800.775.688	
Áo	+43 800006249	+43 19286540
Bỉ	+32 80077160	+32 34002973
Trung Quốc	400.066.5835	
Đan Mạch	+45 80820183	+45 89871156
Phần Lan	+358 800918363	+358 974790110
Pháp	+33 805102193	+33 170770446
Đức	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Ý	+39 800985513	+39 236003759
Nhật Bản	0800.111.5011	
Hà Lan	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Na Uy	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Tây Ban Nha	+34 911899417	+34 800300143
Thụy Điển	+46 850619671	+46 200883979
Thụy Sĩ	+41 565800000	+41 800200442
Đài Loan	00806651752	
Vương quốc Anh	+44 8000126019	+44 2073057197
Các quốc gia khác	+44.1799.534000	

Các bảng dữ liệu an toàn (SDS) – Có trên trang web của Illumina tại địa chỉ support.illumina.com/sds.html.

Tài liệu hướng dẫn về sản phẩm – Có thể tải xuống bản PDF từ trang web của Illumina. Truy cập support.illumina.com, chọn một sản phẩm, sau đó chọn **Documentation &Literature** (Tài liệu hướng dẫn & Y văn).

Illumina
5200 Illumina Way
San Diego, California 92122
U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (ngoài khu vực Bắc Mỹ)
techsupport@illumina.com
www.illumina.com

IVD

EC REP

Illumina Netherlands B. V.
Steenoven 19
5626 DK Eindhoven
Hà Lan

Nhà bảo trợ tại Úc

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

DÙNG CHO CHẨN ĐOÁN IN VITRO

© Năm 2021 Illumina, Inc. Bảo lưu mọi quyền.

illumina®