

DRAGEN for Illumina DNA Prep with Enrichment Dx trên NextSeq 550Dx

Hướng dẫn sử dụng ứng dụng

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 THIẾT 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).

© 2023 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 tham khảo trang www.illumina.com/company/legal.html.

Lịch sử sửa đổi

Tài liệu	Ngày	Mô tả thay đổi
200025238 v00	Tháng 2 năm 2023	Phát hành lần đầu.

Mục lục

Lịch sử sửa đổi	iii
Tổng quan	1
Phương pháp phân tích	1
Tạo một lần chạy theo kế hoạch	5
Cài đặt	7
Tập bản kê khai	8
Lọc nhiễu (Không bắt buộc)	9
Kết quả phân tích	9
Tập FASTQ	10
Tập BAM	11
Tập VCF	11
Xếp lại hàng đợi phân tích	17
Hỗ trợ kỹ thuật	18

Tổng quan

Ứng dụng DRAGEN for Illumina DNA Prep with Enrichment Dx (DRAGEN for IDPE Dx) được sử dụng để lập kế hoạch và thực hiện phân tích thứ cấp các thư viện IDPE Dx được tạo ra để giải trình tự trên hệ thống NextSeq 550Dx.

DRAGEN for IDPE Dx hỗ trợ từ bước giải trình tự đến phân tích khi được sử dụng cùng với Illumina DNA Prep with Enrichment Dx Library Prep, NextSeq 550Dx và Illumina DRAGEN Server for NextSeq 550Dx.

Phương pháp phân tích

DRAGEN for IDPE Dx thực hiện tách đoạn, tạo FASTQ, ánh xạ đoạn đọc, căn chỉnh với hệ gen tham chiếu và phát hiện biến thể nhỏ tùy thuộc vào quy trình công việc được chọn:

- Tạo FASTQ
- Tạo FASTQ và VCF cho dòng mầm
- Tạo FASTQ và VCF cho soma

LƯU Ý Nén ORA có thể sử dụng cho cả ba quy trình công việc. DRAGEN ORA Compression là phần mềm nén hoàn toàn không mất dữ liệu, tạo tệp với phần mở rộng Original Read Archive (*.ora). Định dạng ORA là định dạng nén dựa trên tham chiếu cho tệp FASTQ, được thiết kế để nén/giải nén rất nhanh và đạt tỷ lệ nén cao.

Tạo FASTQ

Các trình tự đã lắp ráp được ghi vào các tệp FASTQ theo từng mẫu. Tệp FASTQ là tệp văn bản chứa dữ liệu giải trình tự và điểm chất lượng cho một mẫu duy nhất. Với mỗi mẫu, các tệp FASTQ riêng biệt được tạo cho từng làn tế bào dòng chảy và cho từng đoạn đọc giải trình tự. Tên mẫu được chỉ định trong quá trình thiết lập lần chạy sẽ được đưa vào tên tệp FASTQ. Tệp FASTQ là đầu vào chính để căn chỉnh. Bước đầu tiên của quá trình tạo FASTQ là tách đoạn. Quá trình tách đoạn sẽ gán các cụm đi qua bộ lọc cho một mẫu bằng cách so sánh mỗi trình tự đọc chỉ thị với các trình tự chỉ thị đã 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.

Phần mềm bao gồm tính năng nén ORA để nén các tệp FASTQ. Định dạng này có thể được bật không bắt buộc. Khi sử dụng định dạng ORA (*.ora), tổng kiểm md5 của nội dung FASTQ được giữ nguyên sau một chu kỳ nén và giải nén để đảm bảo nén không mất dữ liệu.

Căn chỉnh & ánh xạ DNA

Sau khi tạo tệp FASTQ, các đoạn đọc được ánh xạ và căn chỉnh với hệ gen tham chiếu. Giai đoạn đầu tiên của ánh xạ là tạo các hạt giống từ đoạn đọc, và sau đó tìm các vị trí khớp chính xác trong hệ gen tham chiếu. Các kết quả này được tinh chỉnh bằng cách chạy thuật toán Smith-Waterman đầy đủ tại những vị trí có mật độ hạt giống khớp cao nhất. Thuật toán đã được kiểm chứng này hoạt động bằng cách so sánh từng vị trí của đoạn đọc với tất cả các vị trí ứng viên trên hệ gen tham chiếu. Các phép so sánh này tương ứng với một ma trận các khả năng căn hàng giữa đoạn đọc và hệ gen tham chiếu. Đối với mỗi vị trí căn chỉnh ứng viên này, thuật toán Smith-Waterman sẽ tạo ra các mức điểm được dùng để đánh giá xem đường căn chỉnh tốt nhất đi qua ô ma trận đó đạt được bằng cách: khớp hoặc không khớp nucleotide (di chuyển theo đường chéo), vùng xóa (di chuyển theo chiều ngang), hay vùng chèn (di chuyển theo chiều dọc). Vị trí khớp giữa đoạn đọc và tham chiếu sẽ được cộng điểm thưởng, và vị trí không khớp hoặc có biến đổi indel sẽ bị tính điểm phạt. Đường đi có tổng số điểm cao nhất xuyên qua ma trận chính là phương án căn chỉnh được lựa chọn. Thuật toán này được tăng tốc phần cứng trên các thẻ lập trình mảng cổng logic (FPGA) của hệ thống DRAGEN. Hệ gen tham chiếu được sử dụng trong ứng dụng được tạo từ tệp UCSC hg19 FASTA với tùy chọn DRAGEN để tạo bảng băm có khả năng nhận diện các alen thay thế dựa trên cơ chế chuyển đổi tọa độ hệ gen.

Phát hiện biến thể dòng mầm DRAGEN

DRAGEN Germline Small Variant Caller nhận đầu vào là các đoạn đọc DNA đã được ánh xạ và căn chỉnh và phát hiện các đa hình nucleotide đơn (SNP) cũng như các đột biến chèn hoặc mất đoạn (indel) thông qua sự kết hợp giữa phát hiện theo cột và lắp ráp kiểu đơn bội *de novo* tại chỗ. Để kích hoạt DRAGEN Germline Small Variant Caller, hãy chọn quy trình công việc phân tích biến thể dòng mầm.

Việc phát hiện biến thể dòng mầm thường được sử dụng cho các mẫu dòng mầm có mức bội thể mặc định là hai. Đầu tiên, các vùng tham chiếu có thể phát hiện sẽ được nhận diện dựa trên phạm vi căn chỉnh đầy đủ. Trong các vùng tham chiếu này, một lượt quét nhanh các đoạn đọc đã sắp xếp sẽ giúp xác định các vùng hoạt động, vốn là các vùng tập trung quanh các cột dữ liệu chồng lấp có bằng chứng về biến thể. Các vùng hoạt động này được thêm một khoảng trình tự đủ rộng để bao phủ các nội dung không khớp với tham chiếu quan trọng ở lân cận. Nếu có bằng chứng về indel, các vùng hoạt động sẽ được thêm diện tích.

Các đoạn đọc đã căn chỉnh sẽ được cắt gọn trong phạm vi mỗi vùng hoạt động và lắp ráp thành một đồ thị De Bruijn. Các cạnh của các đoạn đọc được cắt gọn này được gán trọng số dựa trên số lần quan sát được, với trình tự tham chiếu đóng vai trò là khung xương. Sau khi làm sạch và đơn giản hóa đồ thị, tất cả các đường đi từ điểm nguồn đến điểm đích sẽ được trích xuất dưới dạng các kiểu đơn bội ứng viên. Mỗi kiểu đơn bội sau đó được căn chỉnh theo thuật toán Smith-Waterman với hệ gen tham chiếu để xác định các biến thể mà nó đại diện. Tập hợp các sự kiện này có thể được bổ sung bằng phương pháp phát hiện dựa trên vị trí. Đối với mỗi cặp đoạn đọc-kiểu đơn bội, xác suất $P(r|H)$ của quan sát đoạn đọc đó khi giả định kiểu đơn bội là mẫu gốc thực sự sẽ được ước tính bằng mô hình Markov ẩn cặp (HMM).

Bằng cách quét theo vị trí tham chiếu trên vùng hoạt động, các kiểu gen ứng viên được hình thành từ các tổ hợp lưỡng bội của các sự kiện biến thể (SNP hoặc indel). Đối với mỗi sự kiện (bao gồm cả tham chiếu), xác suất có điều kiện $P(r|e)$ của quan sát mỗi đoạn đọc chồng lấn được ước tính là giá trị $P(r|H)$ tối đa cho các kiểu đơn bội hỗ trợ sự kiện đó. Các giá trị này được kết hợp thành xác suất có điều kiện $P(r|e1e2)$ cho một kiểu gen (cặp sự kiện) và nhân lên để tạo ra xác suất có điều kiện $P(R|e1e2)$ của quan sát toàn bộ các đoạn đọc chồng lấn. Sử dụng Công thức Bayes, xác suất hậu nghiệm $P(e1e2|R)$ của mỗi kiểu gen lưỡng bội sẽ được tính toán, và kiểu gen có xác suất cao nhất sẽ được phát hiện.

DRAGEN for IDPE Dx áp dụng cơ chế lọc tự động. Tham khảo phần [Các chú thích trong tệp VCF của quy trình công việc dòng mầm trên trang 13](#) để biết thêm thông tin.

Phát hiện biến thể soma DRAGEN

DRAGEN Somatic Small Variant Caller nhận đầu vào là các đoạn đọc DNA đã được ánh xạ và căn chỉnh và phát hiện các SNV và indel thông qua việc lắp ráp kiểu đơn bội *de novo* tại chỗ trong một vùng hoạt động. Để kích hoạt DRAGEN Somatic Small Variant Caller, hãy chọn ứng dụng biến thể soma.

Phát hiện biến thể soma thường được sử dụng cho các mẫu khối u. Với quy trình công việc này, DRAGEN không đưa ra bất kỳ giả định nào về mức bội thể, điều này cho phép phát hiện các alen có tần suất thấp. Đối với các lô-cut có độ phủ lên tới 100 lần trong mẫu khối u, DRAGEN có ngưỡng phát hiện tại tần suất alen biến thể (VAF) là 5%. Giới hạn này sẽ thay đổi tỉ lệ thuận với độ sâu trình tự tăng dần trên từng lô-cut và giảm đi một nửa mỗi khi độ phủ tăng gấp đôi vượt quá mức 100 lần. Đầu tiên, các vùng tham chiếu có thể phát hiện sẽ được nhận diện dựa trên phạm vi căn chỉnh đầy đủ. Trong các vùng tham chiếu này, một lượt quét các đoạn đọc đã sắp xếp sẽ giúp xác định các vùng hoạt động, vốn là các vùng tập trung quanh các cột dữ liệu chồng lấn có bằng chứng về biến thể trong các đoạn đọc từ khối u. Các vùng hoạt động này được thêm một khoảng trình tự đủ rộng để bao phủ các nội dung không khớp với tham chiếu quan trọng ở lân cận. Nếu có bằng chứng về indel, các vùng hoạt động sẽ được thêm diện tích.

Các đoạn đọc đã căn chỉnh sẽ được cắt gọn trong phạm vi mỗi vùng hoạt động và lắp ráp thành một đồ thị De Bruijn. Các cạnh của các đoạn đọc được cắt gọn này được gán trọng số dựa trên số lần quan sát được, với trình tự tham chiếu đóng vai trò là khung xương. Sau khi làm sạch và đơn giản hóa đồ thị, tất cả các đường đi từ điểm nguồn đến điểm đích sẽ được trích xuất dưới dạng các kiểu đơn bội ứng viên. Mỗi kiểu đơn bội sau đó được căn chỉnh theo thuật toán Smith-Waterman với hệ gen tham chiếu để xác định các biến thể mà nó đại diện. Đối với mỗi cặp đoạn đọc-kiểu đơn bội, xác suất $P(r|H)$ của quan sát đoạn đọc được ước tính bằng mô hình Markov ẩn cặp (HMM), khi giả định kiểu đơn bội là mẫu gốc thực sự.

Để xác định điểm số giới hạn phát hiện khối u (TLOD), trước tiên DRAGEN Somatic Small Variant Caller quét theo vị trí tham chiếu cho từng sự kiện soma ứng viên cũng như sự kiện tham chiếu trên khắp vùng hoạt động. Xác suất có điều kiện $P(r|e)$ của quan sát mỗi đoạn đọc chồng lấn được ước tính là giá trị $P(r|H)$ tối đa cho các kiểu đơn bội hỗ trợ sự kiện đó. Các giá trị này được kết hợp thành xác suất có điều kiện $P(r|E)$ cho một giả thuyết sự kiện E, bao gồm một hỗn hợp giữa alen tham chiếu và alen soma ứng

viên trên một dải tần suất alen có thể xảy ra và nhân lên để tạo ra xác suất có điều kiện $P(R|E)$ của quan sát toàn bộ các đoạn đọc chồng lấp. Từ đó, điểm số TLOD được tính toán làm bằng chứng cho thấy sự hiện diện của một alen THAY THẾ tại một lô-cut nhất định trong mẫu khối u.

DRAGEN for IDPE Dx áp dụng cơ chế lọc tự động. Tham khảo phần [Các chú thích trong tệp VCF của quy trình công việc soma trên trang 15](#) để biết thêm thông tin.

Tạo một lần chạy theo kế hoạch

Thực hiện các bước sau để thiết lập một lần chạy trong Illumina Run Manager trên NextSeq 550Dx hoặc bằng cách sử dụng trình duyệt trên máy tính kết nối mạng. Sử dụng trình duyệt trên máy tính kết nối mạng nếu muốn nhập dữ liệu mẫu. Tham khảo Hướng dẫn phần mềm Illumina Run Manager for NextSeq 550Dx (tài liệu số 200025239) để biết hướng dẫn truy cập Illumina Run Manager từ máy tính kết nối mạng.

Có hai cách để tạo một lần chạy theo kế hoạch mới:

- **Nhập lần chạy**—Sử dụng Bảng thông tin mẫu từ một lần chạy hiện có làm mẫu cho lần chạy mới. Tham khảo Hướng dẫn phần mềm Illumina Run Manager for NextSeq 550Dx (tài liệu số 200025239) để biết thông tin về cách nhập lần chạy.
- **Tạo lần chạy**—Nhập thủ công các tham số lần chạy. Các hướng dẫn sau đây mô tả cách tạo một lần chạy.

LƯU Ý Các trường nhập bắt buộc trong giao diện người dùng được đánh dấu hoa thị (*).

Ứng dụng

1. Từ tab Planned (Đã lên kế hoạch) của màn hình Runs (Lần chạy), chọn **Create Run** (Tạo lần chạy).
2. Chọn ứng dụng DRAGEN for Illumina DNA Prep with Enrichment Dx, sau đó chọn **Next** (Tiếp theo).

Các chế độ cài đặt lần chạy

1. Trên màn hình Run Settings (Cài đặt lần chạy), nhập tên lần chạy duy nhất. Tên lần chạy giúp xác định lần chạy từ bước giải trình tự đến bước phân tích.
2. **[Không bắt buộc]** Nhập mô tả cho lần chạy để giúp xác định lần chạy rõ hơn.
3. Chọn (các) bộ kit adapter chỉ thị được sử dụng trong quá trình chuẩn bị thư viện.
4. Xem lại Độ dài đoạn đọc và điều chỉnh nếu cần thiết. Đoạn đọc 1 và Đoạn đọc 2 có giá trị mặc định là 151 chu kỳ. Chỉ thị 1 và Chỉ thị 2 có giá trị cố định là 10 chu kỳ và không thể thay đổi.
5. **[Không bắt buộc]** Nhập ID ống thư viện.
6. Chọn **Next** (Tiếp theo).

Dữ liệu mẫu

Dữ liệu mẫu bao gồm ID mẫu, Vị trí giếng (vị trí giếng trên khay chỉ thị) và Tên thư viện. Khi sử dụng Chỉ thị A&B, Vị trí giếng cũng bao gồm cả mã định danh của Khay.

Có hai cách để nhập dữ liệu mẫu:

- **Nhập mẫu**—Sử dụng tệp mẫu có sẵn để tải xuống trên màn hình Sample Data (Dữ liệu mẫu).

- **Thủ công**—Nhập dữ liệu mẫu trực tiếp vào bảng trên màn hình Sample Data (Dữ liệu mẫu).

Nhập mẫu

Khi lập kế hoạch lần chạy giải trình tự bằng trình duyệt trên máy tính kết nối mạng, một tệp mẫu (*.CSV) có sẵn để tải xuống trên màn hình Sample Data (Dữ liệu mẫu). Tệp mẫu không khả dụng để tải xuống khi truy cập Illumina Run Manager thông qua phần mềm hệ điều hành NextSeq 550Dx. Để nhập dữ liệu mẫu bằng tính năng Import Samples (Nhập mẫu), hãy thực hiện các bước sau.

LƯU Ý Hoàn tất các bước Cài đặt lần chạy trước khi tiếp tục.

1. Chọn **Download Template** (Tải mẫu xuống) để tải xuống một tệp CSV trống.
2. Từ tệp mẫu, nhập dữ liệu mẫu, rồi lưu tệp. Tên thư viện là không bắt buộc.

LƯU Ý Khi sử dụng Chỉ thị A&B, dữ liệu cho cột B phải bao gồm cả vị trí khay và vị trí giếng (vị trí giếng trên khay chỉ thị). Ví dụ: A-A01, A-A02, A-A03.

3. Chọn **Import Samples** (Nhập mẫu) và duyệt đến tệp mẫu chứa thông tin dữ liệu mẫu từ bước trước.
4. Chọn **Open** (Mở), **Proceed** (Tiếp tục) rồi chọn **Next** (Tiếp theo).

LƯU Ý Thay đổi ID mẫu trước khi chọn Next (Tiếp theo) có thể dẫn đến lỗi. Hãy hoàn tất việc thiết lập lần chạy trước khi thực hiện thay đổi để tránh lỗi.

Nhập mẫu theo cách thủ công

Sử dụng bảng trên màn hình Dữ liệu mẫu (Sample Data) để nhập dữ liệu mẫu theo cách thủ công.

1. Nhập ID mẫu duy nhất vào trường Sample ID (ID mẫu).
2. Sử dụng **Well Position** (Vị trí giếng) (Chỉ thị A hoặc Chỉ thị B) hoặc **Plate - Well Position** (Khay - Vị trí giếng) (Chỉ thị A&B) để chọn chỉ thị liên kết cho các mẫu.
Các trường i7 Index (Chỉ thị i7), Index 1 (Chỉ thị 1), i5 Index (Chỉ thị i5) và Index 2 (Chỉ thị 2) sẽ tự động được điền.
3. **[Không bắt buộc]** Nhập tên thư viện.
4. Thêm các hàng và lặp lại bước 1–3 khi cần cho đến khi tất cả mẫu đã được thêm vào bảng. Bạn có thể thêm nhiều hàng cùng lúc bằng cách nhập trước số lượng hàng cần thêm, sau đó chọn biểu tượng +. Bạn cũng có thể xóa các hàng bằng cách chọn ô bên cạnh số hàng, sau đó nhấp vào biểu tượng thùng rác.
5. Chọn **Next** (Tiếp theo).

Cài đặt phân tích

1. Chọn quy trình công việc phân tích mong muốn:

- Tạo FASTQ
 - Tạo FASTQ và VCF cho quy trình công việc dòng mầm (yêu cầu Tập bản kê khai)
 - Tạo FASTQ và VCF cho quy trình công việc soma (yêu cầu Tập bản kê khai)
2. **[Không bắt buộc] Generate ORA compressed FASTQs** (Tạo FASTQ nén bằng ORA) được bật theo mặc định. Nén FASTQ bằng ORA giúp nén không mất dữ liệu các tệp FASTQ lên đến 5 lần so với fastq.gz. Bỏ chọn **Generate ORA compressed FASTQs** (Tạo FASTQ nén bằng ORA) nếu bạn muốn dữ liệu không nén (fastq.gz).
 3. Đối với các quy trình công việc dòng mầm và soma, cần có tệp bản kê khai. Sử dụng menu thả xuống **Manifest File Selection** (Chọn tệp bản kê khai) để chọn tệp bản kê khai. Tệp bản kê khai là tệp BED (*.bed) được phân tách bằng tab, dùng để chỉ định tên và vị trí của các vùng tham chiếu được nhắm mục tiêu. Tham khảo phần [Tệp bản kê khai trên trang 8](#) để biết thêm thông tin.
 4. **[Không bắt buộc]** Đối với các quy trình công việc Soma, sử dụng menu thả xuống **Noise File Selection** (Chọn tệp nhiễu) để chọn tệp nhiễu hệ thống.
Có thể chỉ định một tệp BED (*.bed.gz) với mức độ nhiễu theo vị trí để lọc bỏ nhiễu hệ thống. Để biết thêm thông tin, tham khảo phần [Lọc nhiễu \(Không bắt buộc\) trên trang 9](#).
 5. Chọn **Next** (Tiếp theo).

Xem lại lần chạy

1. Trên màn hình Review (Xem lại), hãy kiểm tra lại thông tin cho mục Run Settings (Cài đặt lần chạy), Sample Data (Dữ liệu mẫu) và Analysis Settings (Cài đặt phân tích).
2. Chọn **Save** (Lưu).
Lần chạy được lưu trên tab Planned (Được lên kế hoạch) trên màn hình Runs (Lần chạy).

Cài đặt

Để xem hoặc thay đổi cài đặt ứng dụng DRAGEN for IDPE Dx, trước tiên hãy chọn biểu tượng Ứng dụng từ màn hình chính. Sau đó chọn ứng dụng bạn muốn xem hoặc thay đổi. Cần có tài khoản Quản trị viên để thay đổi cài đặt.

Cấu hình

Màn hình cấu hình hiển thị các cài đặt ứng dụng sau:

- **Library Prep Kits** (Bộ kit chuẩn bị thư viện)—Hiển thị bộ kit chuẩn bị thư viện mặc định cho ứng dụng. Cài đặt này không thể thay đổi.
- **Index Adapter Kits** (Bộ kit adapter chỉ thị)—Hiển thị bộ kit adapter chỉ thị mặc định cho ứng dụng. Cài đặt này không thể thay đổi.
- **Read lengths** (Độ dài đoạn đọc)—Độ dài đoạn đọc được đặt mặc định là 151 cho ứng dụng, nhưng có thể thay đổi trong quá trình tạo lần chạy.

- **Manifest and Noise Files** (Tập bản kê khai và nhiễu)—Tải lên và thay đổi cài đặt cho các tập bản kê khai và nhiễu.
 - Chọn **Upload File** (Tải lên tệp) để tải lên tệp sử dụng trong phân tích.
 - Chọn nút radio **Default** (Mặc định) để đặt tệp làm tập bản kê khai hoặc nhiễu mặc định được chọn trong quá trình tạo lần chạy khi chọn ứng dụng.
 - Chọn hộp kiểm **Enabled** (Đã bật) để hiển thị tệp trong menu thả xuống trong quá trình tạo lần chạy.

Các quyền

Sử dụng các hộp kiểm trên màn hình Permissions (Quyền) để quản lý quyền truy cập của người dùng cho ứng dụng.

Tập bản kê khai

Khi sử dụng DRAGEN for IDPE Dx, một tập bản kê khai là đầu vào bắt buộc cho các quy trình công việc sau:

- Tạo FASTQ và VCF cho quy trình công việc dòng mầm
- Tạo FASTQ và VCF cho quy trình công việc soma

Tập bản kê khai là tệp văn bản được phân cách bằng tab sử dụng định dạng BED (*.bed), trong đó chỉ định tên và vị trí của các vùng tham chiếu được nhắm mục tiêu. Phần chính của tập bản kê khai là phần Vùng và phải chứa các cột dữ liệu sau:

Cột	Mô tả
Name (Tên)	Tên duy nhất do người dùng chỉ định cho vùng mục tiêu
Chromosome (Nhiễm sắc thể)	Vị trí nhiễm sắc thể (ví dụ: chr10, chr5, ...)
Start (Bắt đầu)	Chỉ thị bắt đầu của vùng mục tiêu tính theo hệ 1 base
Stop (Dừng)	Chỉ thị kết thúc của vùng mục tiêu tính theo hệ 1 base
Upstream Probe Length (Độ dài mẫu dò phía ngược dòng)	Độ dài của mẫu dò phía ngược dòng. Đối với ứng dụng DRAGEN for IDPE Dx, giá trị này phải đặt là 0.
Downstream Probe Length (Độ dài của mẫu dò phía xuôi dòng)	Độ dài của mẫu dò phía xuôi dòng. Đối với ứng dụng DRAGEN for IDPE Dx, giá trị này phải đặt là 0.

LƯU Ý Định dạng tập bản kê khai hợp lệ là bắt buộc cho phân tích. DRAGEN sẽ dừng phân tích nếu tập bản kê khai không hợp lệ.

Lọc nhiễu (Không bắt buộc)

Bộ lọc nhiễu hệ thống có sẵn cho quá trình phát hiện biến thể soma, và có thể được sử dụng để giảm các phát hiện dương tính giả bằng cách tính đến nhiễu đặc hiệu theo vị trí. Tập nhiễu hệ thống được tạo ra bằng cách trước tiên thu thập khoảng 50 mẫu bình thường (tốt nhất là đặc hiệu cho bảng, quy trình chuẩn bị thư viện, và máy giải trình tự), sau đó tổng tần suất alen dưới 30% tại mỗi vị trí có độ phủ đủ sẽ được chia cho tổng số mẫu (các tần suất alen trên 30% được giả định là biến thể dòng mầm và không được coi là nhiễu). Khi các giá trị nhiễu đã được tạo ra, các biến thể soma được phát hiện tại vị trí đó sẽ được lọc bỏ.

Bộ lọc có thể được sử dụng trong chế độ Khối u-Bình thường, nhưng đặc biệt hữu ích cho các lần chạy Chỉ-Khối u khi không có mẫu bình thường khớp. Tập nhiễu hệ thống phải sử dụng tệp BED với phần mở rộng tệp (*.bed.gz) và phải bao gồm bốn cột: Nhiễm sắc thể, Vị trí bắt đầu, Vị trí kết thúc và mức nhiễu đặc hiệu theo vị trí cho mỗi hàng. Không bắt buộc phải áp dụng lọc nhiễu hệ thống.

Kết quả phân tích

Các lần chạy đang diễn ra được hiển thị trong tab Active (Đang hoạt động). Các lần chạy đã hoàn thành được hiển thị trong tab Completed (Đã hoàn tất). DRAGEN for IDPE Dx tạo một thư mục phân tích với tên duy nhất cho mỗi lần chạy, tách biệt với thư mục chứa dữ liệu giải trình tự. Thư mục phân tích bao gồm thông tin sau:

- Tệp bản kê khai được sử dụng
- Phiên bản phần mềm
- ID mẫu
- Tổng số đoạn đọc đã căn chỉnh
- Tỷ lệ phần trăm đoạn đọc đã căn chỉnh trên mỗi mẫu
- Số lượng SNV được phát hiện trên mỗi mẫu
- Số lượng indel được phát hiện trên mỗi mẫu
- Thống kê độ phủ

Tệp đầu ra phân tích

Vị trí thư mục phân tích được xác định bởi thiết lập Lưu trữ bên ngoài cho kết quả phân tích. Tham khảo Hướng dẫn phần mềm Illumina Run Manager for NextSeq 550Dx (tài liệu số 200025239) để biết thêm thông tin về thiết lập Lưu trữ bên ngoài để cho kết quả phân tích.

Trên màn hình Run Details (Chi tiết lần chạy), trường External Location (Vị trí bên ngoài) cung cấp đường dẫn đến dữ liệu giải trình tự. Trường Analysis Output Folder (Thư mục đầu ra phân tích) cung cấp tên thư mục phân tích duy nhất trên màn hình Run Details (Chi tiết lần chạy). Các tệp chính xác được tạo ra phụ thuộc vào quy trình công việc phân tích được sử dụng. Ứng dụng tạo ra các tệp đầu ra phân tích sau đây.

LƯU Ý Nếu gặp lỗi giới hạn độ dài đường dẫn tệp tối đa khi truy cập tệp đầu ra phân tích, hãy thử di chuyển tệp đến vị trí có đường dẫn ngắn hơn hoặc dùng phương pháp khác để mở tệp.

Tệp đầu ra	Mô tả
Báo cáo tóm tắt biến thể (* .pdf)	Chứa bản tóm tắt về thông tin tệp, phiên bản phần mềm, thông tin mẫu, thống kê ở mức đọc, và các tóm tắt về SNV, chèn, mất đoạn cũng như độ phủ. Chỉ có quy trình công việc dòng mầm và soma mới tạo ra báo cáo biến thể.
FASTQ (* .fastq.gz hoặc * .fastq.ora)	Tệp trung gian chứa các phát hiện base đã được chấm điểm chất lượng. Tệp FASTQ là đầu vào chính cho bước căn chỉnh. Nếu chọn nén ORA, phần mở rộng sẽ là * .fastq.ora.
Tệp Alignment BAM (* .bam)	Chứa các đoạn đọc đã được căn chỉnh cho một mẫu cụ thể.
Tệp Genome VCF (* .gvcf.gz)	Chứa kiểu gen cho từng vị trí, dù được phát hiện là biến thể hay tham chiếu.
Tệp VCF (* .vcf.gz)	Chứa các biến thể được phát hiện tại mỗi vị trí.
Báo cáo số liệu lần chạy (* .csv)	Chứa các số liệu chất lượng của lần chạy, bao gồm tổng hiệu suất dữ liệu thu được không có chỉ thị và điểm Q30.

Tệp FASTQ

FASTQ (* .fastq.gz, * .fastq.ora) là một định dạng tệp văn bản chứa các phát hiện base và giá trị chất lượng cho mỗi đoạn đọc. Mỗi tệp chứa những thông tin sau:

- Mã nhận dạng mẫu
- Trình tự
- Điểm chất lượng Phred được mã hóa theo định dạng ASCII + 33

Mã nhận dạng mẫu được định dạng như sau:

```
@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=AAAAAAAAA9#:<#<;<<<????#=#
```

Tập BAM

Tập BAM (*.bam) là phiên bản nhị phân nén của tập SAM (ánh xạ căn chỉnh giải trình tự), được dùng để biểu diễn các trình tự đã căn chỉnh có độ dài lên đến 128 Mb. Tập BAM sử dụng định dạng tên tập `SampleName_S#.bam`. # là số mẫu, được xác định theo thứ tự các mẫu được liệt kê trong lần chạy. Ở chế độ đa nút, ký hiệu S# luôn được đặt là S1, bất kể thứ tự của mẫu.

Tập BAM gồm có phần đầu và phần căn chỉnh:

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

Phần căn chỉnh bao gồm các thông tin sau cho mỗi đoạn đọc hoặc cặp đoạn đọc:

- AS: Chất lượng căn chỉnh 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 căn chỉnh của đoạn đọc đơn.
- XC: Chuỗi mô tả khớp.
- XN: Thẻ tên amplicon, ghi lại ID amplicon gắn với đoạn đọc

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

Tập VCF

Tập Variant call format (*.vcf) chứa thông tin về các biến thể được tìm thấy ở các vị trí cụ thể trong hệ gen tham chiếu.

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

Bảng 1 Các tiêu đề trong tập 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 tập FASTA tham chiếu.

Tiêu đề	Mô tả
POS	Vị trí một base duy nhất của biến thể trên nhiễm sắc thể tham chiếu. Đối với biến thể đơn nucleotide (SNV), vị trí này là base tham chiếu có biến thể. Đối với indel, vị trí này là base tham chiếu ngay trước biến thể.
ID	Số rs (SNP tham chiếu) cho SNP lấy từ <code>dbSNP.txt</code> , nếu có. Nếu có nhiều số rs tại cùng vị trí, 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.

Bảng 2 Các chú thích trong tệp VCF của quy trình công việc dòng mầm

Tiêu đề	Mô tả
FILTER	Nếu tất cả các bộ lọc đều đạt, giá trị PASS sẽ được ghi trong cột bộ lọc. Các mục nhập FILTER có thể bao gồm: • DRAGENSnpHardQUAL—Áp dụng nếu điểm QUAL của biến thể SNP không đạt ngưỡng • DRAGENIndelHardQUAL—Áp dụng nếu điểm QUAL của biến thể indel không đạt ngưỡng • LowDepth—Vị trí bị lọc vì độ sâu phạm vi không đạt ngưỡng • LowGQ—Vị trí bị lọc vì chất lượng kiểu gen không đạt ngưỡng • PloidyConflict—Kiểu gen từ trình phát hiện biến thể không phù hợp với mức bội thể của nhiễm sắc thể • base_quality—Vị trí bị lọc vì chất lượng base trung vị của các đoạn đọc thay thế tại lô-cut này không đạt ngưỡng • filtered_reads—Vị trí bị lọc vì tỷ lệ quá lớn các đoạn đọc đã bị lọc bỏ • fragment_length—Vị trí bị lọc vì chênh lệch tuyệt đối giữa độ dài phân đoạn trung vị của các đoạn đọc thay thế và tham chiếu tại lô-cut này vượt ngưỡng • low_depth—Vị trí bị lọc vì độ sâu đoạn đọc quá thấp • low_frac_info_reads—Vị trí bị lọc vì tỷ lệ đoạn đọc thông tin thấp hơn ngưỡng • low_normal_depth—Vị trí bị lọc vì độ sâu đoạn đọc của mẫu bình thường quá thấp • long_indel—Vị trí bị lọc vì chiều dài indel quá lớn • mapping_quality—Vị trí bị lọc vì chất lượng ánh xạ trung vị của các đoạn đọc thay thế tại lô-cut này không đạt ngưỡng • multiallelic—Vị trí bị lọc vì có hơn hai alen thay thế đi qua ngưỡng LOD của khối u • non_homref_normal—Vị trí bị lọc vì kiểu gen của mẫu bình thường không phải là tham chiếu đồng hợp tử • no_reliable_supporting_read—Vị trí bị lọc vì không có đoạn đọc soma hỗ trợ đáng tin cậy • panel_of_normals—Được ghi nhận trong ít nhất một mẫu trong tệp VCF của bảng nhập giá trị bình thường • read_position—Vị trí bị lọc vì trung vị khoảng cách giữa đầu/cuối đoạn đọc và lô-cut này thấp hơn ngưỡng • RMxNRepeatRegion—Vị trí bị lọc vì toàn bộ hoặc một phần alen biến thể là lặp lại của tham chiếu • strand_artifact—Vị trí bị lọc vì sai lệch theo chiều sợi nghiêm trọng • str_contraction—Vị trí bị lọc do nghi ngờ lỗi PCR, trong đó alen thay thế ít hơn một đơn vị lặp so với tham chiếu • too_few_supporting_reads—Vị trí bị lọc vì có quá ít đoạn đọc hỗ trợ trong mẫu khối u • weak_evidence—Điểm biến thể soma không đạt ngưỡng

Tiêu đề	Mô tả
INFO	Các mục nhập INFO có thể 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. • DB—Thành viên của dbSNP. • FS—Giá trị p theo thang Phred sử dụng kiểm định chính xác Fisher để phát hiện sai lệch theo chiều sợi. • QD—Độ tin cậy/Chất lượng biến thể theo độ sâu đọc. • R2_5P_bias—Điểm dựa trên sai lệch cặp đọc và khoảng cách từ đầu mồi 5. • SOR—Tỷ số chênh lệch đối xứng của bảng liên hợp 2x2 để phát hiện sai lệch theo chiều sợi. • DP—Độ sâu đoạn đọc xấp xỉ (thông tin và không thông tin); một số đoạn đọc có thể đã bị lọc dựa trên chất lượng ánh xạ, v.v. • END—Vị trí kết thúc của khoảng. • FractionInformativeReads—Tỷ lệ đọc thông tin trên tổng số đọc. • MQ—Chất lượng ánh xạ RMS. • MQRankSum—Giá trị Z-score từ phép kiểm định tổng hạng Wilcoxon, so sánh chất lượng ánh xạ của các đoạn đọc Thay thế so với Tham chiếu. • ReadPosRankSum—Giá trị Z-score từ phép kiểm định tổng hạng Wilcoxon, so sánh sai lệch vị trí của các đoạn đọc Thay thế so với Tham chiếu. • SOMATIC—Có ít nhất một biến thể tại vị trí này là soma.
FORMAT	Cột định dạng liệt kê các trường được phân tách bằng dấu hai chấm. Ví dụ: GT:GQ. Các trường có sẵn bao gồm: • AD—Độ sâu alen (chỉ tính các đoạn đọc thông tin trong tổng số đoạn đọc) cho alen tham chiếu và alen biến thể theo thứ tự liệt kê. • AF—Tỷ lệ alen cho các alen biến thể theo thứ tự liệt kê. • DP—Độ sâu đoạn đọc xấp xỉ (các đoạn đọc có MQ=255 hoặc cặp lỗi sẽ bị lọc bỏ). • F1R2—Số lượng đoạn đọc trong hướng cặp F1R2 hỗ trợ mỗi alen. • F2R1—Số lượng đoạn đọc trong hướng cặp F2R1 hỗ trợ mỗi alen. • 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. • MB—Thống kê thành phần theo mẫu để phát hiện sai lệch cặp đọc. • PS—Thông tin ID định pha vật lý, mỗi ID duy nhất trong một mẫu (chứ không áp dụng giữa các mẫu) kết nối các bản ghi trong một nhóm định pha. • SB—Thống kê thành phần theo mẫu, bao gồm Kiểm định chính xác Fisher để phát hiện sai lệch theo chiều sợi. • SQ—Chất lượng soma.
SAMPLE	Cột mẫu cung cấp các giá trị được chỉ định trong cột FORMAT.

Bảng 3 Các chú thích trong tệp VCF của quy trình công việc soma

Tiêu đề	Mô tả
FILTER	Nếu tất cả các bộ lọc đều đạt, giá trị PASS sẽ được ghi trong cột bộ lọc. Các mục nhập FILTER có thể bao gồm: • base_quality—Vị trí bị lọc vì chất lượng base trung vị của các đoạn đọc thay thế tại lô-cut này không đạt ngưỡng • filtered_reads—Vị trí bị lọc vì tỷ lệ quá lớn các đoạn đọc đã bị lọc bỏ • fragment_length—Vị trí bị lọc vì chênh lệch tuyệt đối giữa độ dài phân đoạn trung vị của các đoạn đọc thay thế và tham chiếu tại lô-cut này vượt ngưỡng • low_depth—Vị trí bị lọc vì độ sâu đoạn đọc quá thấp • low_frac_info_reads—Vị trí bị lọc vì tỷ lệ đoạn đọc thông tin thấp hơn ngưỡng • low_normal_depth—Vị trí bị lọc vì độ sâu đoạn đọc của mẫu bình thường quá thấp • long_indel—Vị trí bị lọc vì chiều dài indel quá lớn • mapping_quality—Vị trí bị lọc vì chất lượng ánh xạ trung vị của các đoạn đọc thay thế tại lô-cut này không đạt ngưỡng • multiallelic—Vị trí bị lọc vì có hơn hai alen thay thế đi qua ngưỡng LOD của khối u • non_homref_normal—Vị trí bị lọc vì kiểu gen của mẫu bình thường không phải là tham chiếu đồng hợp tử • no_reliable_supporting_read—Vị trí bị lọc vì không có đoạn đọc soma hỗ trợ đáng tin cậy • panel_of_normals—Được ghi nhận trong ít nhất một mẫu trong tệp VCF của bảng nhập giá trị bình thường • read_position—Vị trí bị lọc vì trung vị khoảng cách giữa đầu/cuối đoạn đọc và lô-cut này thấp hơn ngưỡng • RMxNRepeatRegion—Vị trí bị lọc vì toàn bộ hoặc một phần alen biến thể là lặp lại của tham chiếu • strand_artifact—Vị trí bị lọc vì sai lệch theo chiều sợi nghiêm trọng • str_contraction—Vị trí bị lọc do nghi ngờ lỗi PCR, trong đó alen thay thế ít hơn một đơn vị lặp so với tham chiếu • too_few_supporting_reads—Vị trí bị lọc vì có quá ít đoạn đọc hỗ trợ trong mẫu khối u • weak_evidence—Điểm biến thể soma không đạt ngưỡng • systematic_noise—Vị trí bị lọc dựa trên bằng chứng về nhiễu hệ thống trong các mẫu bình thường

Tiêu đề	Mô tả
INFO	Các mục nhập INFO có thể bao gồm: • DP—Độ sâu đoạn đọc xấp xỉ (thông tin và không thông tin); một số đoạn đọc có thể đã bị lọc dựa trên chất lượng ánh xạ, v.v. • END—Vị trí kết thúc của khoảng. • FractionInformativeReads—Tỷ lệ đọc thông tin trên tổng số đọc. • MQ—Chất lượng ánh xạ RMS. • MQRankSum—Giá trị Z-score từ phép kiểm định tổng hạng Wilcoxon, so sánh chất lượng ánh xạ của các đoạn đọc Thay thế so với Tham chiếu. • ReadPosRankSum—Giá trị Z-score từ phép kiểm định tổng hạng Wilcoxon, so sánh sai lệch vị trí của các đoạn đọc Thay thế so với Tham chiếu. • AQ—Điểm nhiễu hệ thống. • hotspot—Vị trí soma đã biết, được sử dụng để tăng độ tin cậy trong việc phát hiện. • SOMATIC—Có ít nhất một biến thể tại vị trí này là soma.
FORMAT	Cột định dạng liệt kê các trường được phân tách bằng dấu hai chấm. Ví dụ: GT:GQ. Các trường có sẵn bao gồm: • AD—Độ sâu alen (chỉ tính các đoạn đọc thông tin trong tổng số đoạn đọc) cho alen tham chiếu và alen biến thể theo thứ tự liệt kê. • AF—Tỷ lệ alen cho các alen biến thể theo thứ tự liệt kê. • DP—Độ sâu đoạn đọc xấp xỉ (các đoạn đọc có MQ=255 hoặc cặp lỗi sẽ bị lọc bỏ). • F1R2—Số lượng đoạn đọc trong hướng cặp F1R2 hỗ trợ mỗi alen. • F1R2—Số lượng đoạn đọc trong hướng cặp F2R1 hỗ trợ mỗi alen. • GP—Xác suất hậu nghiệm theo thang Phred cho kiểu gen, như được định nghĩa trong thông số kỹ thuật VCF. • 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. • MB—Thống kê thành phần theo mẫu để phát hiện sai lệch cặp đọc. • PL—Xác suất hợp lý theo thang Phred đã chuẩn hóa cho kiểu gen, như được định nghĩa trong thông số kỹ thuật VCF. • PRI—Xác suất tiên nghiệm theo thang Phred cho kiểu gen. • PS—Thông tin ID định pha vật lý, mỗi ID duy nhất trong một mẫu (chứ không áp dụng giữa các mẫu) kết nối các bản ghi trong một nhóm định pha. • SB—Thống kê thành phần theo mẫu, bao gồm Kiểm định chính xác Fisher để phát hiện sai lệch theo chiều sợi. • SQ—Chất lượng soma.
SAMPLE	Cột mẫu cung cấp các giá trị được chỉ định trong cột FORMAT.

Tập Genome VCF

Tập Genome VCF (*.gvcf.gz) tuân theo một tập hợp quy ước để biểu diễn tất cả các vị trí trong hệ gen theo định dạng gọn nhẹ hợp lý. Các tập gVCF bao gồm toàn bộ các vị trí trong vùng quan tâm trong một tệp duy nhất cho mỗi mẫu. Tập gVCF hiển thị không phát hiện tại các vị trí không đi qua tất cả bộ lọc. Một thẻ kiểu gen (GT) có giá trị ./, biểu thị trạng thái không phát hiện.

Xếp lại hàng đợi phân tích

Bạn có thể xếp lại hàng đợi phân tích nếu quá trình phân tích bị ngừng lại, không thành công hoặc nếu bạn muốn phân tích lại lần chạy bằng các chế độ cài đặt khác. Để xếp lại hàng đợi phân tích, hãy làm theo các bước sau:

1. Từ màn hình Run (Chạy), chọn tab Completed (Đã hoàn tất), sau đó chọn tên lần chạy cần phân tích lại.
Nếu trước đó đã thực hiện Xếp lại hàng đợi phân tích, hãy chọn tên lần chạy của Parent Run (Lần chạy gốc).
2. Từ màn hình Run Details (Chi tiết lần chạy), sau phần Sequencing Information (Thông tin giải trình tự), chọn **Requeue Analysis** (Xếp lại hàng đợi phân tích).
3. Chọn một tùy chọn:
 - Xếp lại hàng đợi phân tích mà không thay đổi gì
 - Chỉnh sửa cài đặt lần chạy và xếp lại hàng đợi phân tích
 - Xếp lại hàng đợi phân tích với ứng dụng khác
4. Xác nhận rằng vị trí dữ liệu giải trình tự hiện tại đã được cung cấp trong trường **Sequencing data file path** (Đường dẫn tệp dữ liệu giải trình tự).

LƯU Ý Đường dẫn đến dữ liệu giải trình tự phải khớp với đường dẫn trong cài đặt External Storage for Analysis Results (Lưu trữ bên ngoài cho kết quả phân tích). Tham khảo Hướng dẫn phần mềm Illumina Run Manager for NextSeq 550Dx (tài liệu số 200025239) để biết thông tin về cách thay đổi đường dẫn lưu trữ bên ngoài.

5. Nhập Lý do phân tích lại.
6. Chọn **Requeue Analysis** (Xếp lại hàng đợi phân tích).
7. Chỉnh sửa các thay đổi mong muốn trong Run Settings (Cài đặt lần chạy), Sample Data (Dữ liệu mẫu) và Analysis Settings (Cài đặt phân tích).
8. Chọn **Save** (Lưu). Quá trình phân tích sẽ bắt đầu sử dụng các thông số phân tích hiện tại.

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 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 từ support.illumina.com.



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



Illumina Netherlands B.V.
Steenoven 19
5626 DK Eindhoven
The Netherlands

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.

© 2023 Illumina, Inc. Bảo lưu mọi quyền.

illumina®