

DRAGEN for Illumina DNA Prep with Enrichment Dx no NextSeq 550Dx

Guia do utilizador da aplicação

Este documento e respetivo conteúdo são propriedade da Illumina, Inc. e das suas afiliadas (“Illumina”) e destinam-se unicamente a utilização contratual por parte dos clientes relativamente à utilização dos produtos descritos no presente documento e para nenhum outro fim. Este documento e respetivo conteúdo não podem ser utilizados ou distribuídos para qualquer outro fim e/ou de outra forma transmitidos, divulgados ou reproduzidos por qualquer via, seja de que natureza for, sem a autorização prévia por escrito da Illumina. A Illumina não concede qualquer licença ao abrigo da sua patente, marca comercial, direito de autor ou direitos de jurisprudência nem direitos semelhantes de quaisquer terceiros por via deste documento.

As instruções contidas neste documento têm de ser estrita e explicitamente seguidas por pessoal qualificado e com a devida formação para garantir a utilização adequada e segura dos produtos aqui descritos. Todo o conteúdo deste documento tem de ser integralmente lido e compreendido antes da utilização dos referidos produtos.

A NÃO OBSERVÂNCIA DA RECOMENDAÇÃO PARA LEITURA INTEGRAL E SEGUIMENTO EXPLÍCITO DE TODAS AS INSTRUÇÕES AQUI CONTIDAS PODE RESULTAR EM DANOS NOS PRODUTOS, LESÕES EM PESSOAS, INCLUINDO NOS UTILIZADORES OU OUTROS, E EM DANOS MATERIAIS, E IRÁ ANULAR QUALQUER GARANTIA APLICÁVEL AOS PRODUTOS.

A ILLUMINA NÃO ASSUME QUALQUER RESPONSABILIDADE RESULTANTE DA UTILIZAÇÃO INADEQUADA DOS PRODUTOS AQUI DESCRITOS (INCLUINDO PARTES DOS MESMOS OU DO SOFTWARE).

© 2023 Illumina, Inc. Todos os direitos reservados.

Todas as marcas comerciais são propriedade da Illumina, Inc. ou dos respetivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.

Histórico de revisões

Documento	Data	Descrição da alteração
200025238 v00	Fevereiro de 2023	Edição inicial.

Índice

Histórico de revisões	iii
Descrição geral	1
Métodos de análise	1
Criar um ensaio planeado	5
Definições	8
Ficheiro manifest	8
Filtragem de noise (opcional)	9
Saída de análises	9
Ficheiros FASTQ	11
Ficheiros BAM	11
Ficheiros VCF	12
Recolocar a análise em fila de espera	20
Assistência Técnica	21

Descrição geral

A DRAGEN for Illumina DNA Prep with Enrichment Dx aplicação (DRAGEN for IDPE Dx) é utilizada para planear e realizar uma análise secundária de bancos IDPE Dx gerados para sequenciação no NextSeq 550Dx.

DRAGEN for IDPE Dx suporta a sequenciação para análise, quando utilizado com a Preparação de Banco Illumina DNA Prep With Enrichment Dx e NextSeq 550Dx Illumina DRAGEN Server for NextSeq 550Dx.

Métodos de análise

DRAGEN for IDPE Dx executa a demultiplexação, criação de FASTQ, interpretação de mapeamento, alinhamento com um genoma de referência e identificação de variantes pequenas dependendo dos fluxos de trabalho selecionados:

- Criação de FASTQ
- Criação de Germline FASTQ e VCF
- Criação de Somatic FASTQ e VCF

NOTA A compressão ORA está disponível para utilização com os três fluxos de trabalho. A compressão DRAGEN ORA é um software de compressão totalmente sem perda, que cria um ficheiro com uma extensão Ficheiro de Leitura Original (*.ora). O formato ora é um formato de compressão baseado em referência para ficheiros FASTQ e foi concebido para compressão/descompressão muito rápida e elevado rácio de compressão.

Criação de FASTQ

As sequências unidas são escritas em ficheiros FASTQ por amostra. Os ficheiros FASTQ são ficheiros de texto que contêm dados de sequenciação e pontuações de qualidade, para apenas uma amostra. Para cada amostra, os ficheiros FASTQ são criados por linha de células de fluxo, por leitura de sequenciação. O nome da amostra, especificado durante a configuração do ensaio, é incluído no nome do ficheiro FASTQ. Os ficheiros FASTQ são a principal entrada para o alinhamento. O primeiro passo da criação de um FASTQ é a demultiplexação. A demultiplexação atribui os clusters que passam o filtro de uma amostra, comparando cada sequência de leitura de índice com as sequências de índice especificadas para o ensaio. Não são considerados quaisquer valores de qualidade neste passo. As leituras de índice são identificadas através dos seguintes passos:

- As amostras são numeradas a partir do número 1, na ordem pela qual são listados para o ensaio.
- O número de amostra 0 está reservado para os clusters que não foram atribuídos a uma amostra.
- Os clusters são atribuídos a uma amostra quando a sequência de índice é exatamente correspondente ou quando existe apenas uma única divergência por leitura de índice.

O software inclui a compressão ORA para comprimir os ficheiros FASTQ. Este formato pode ser ativado de forma opcional. Ao utilizar o formato ORA (*.ora), a soma de controlo md5 do conteúdo FASTQ é preservada após o ciclo de compressão e descompressão, para assegurar a compressão sem perdas.

Mapeamento e alinhamento do ADN

Após a geração do FASTQ, as leituras são mapeadas e alinhadas com um genoma de referência. A primeira etapa do mapeamento é gerar sementes da leitura e depois procurar as correspondências exatas no genoma de referência. Estes resultados são então afinados pela execução dos alinhamentos Smith-Waterman nas localizações onde há a maior densidade de correspondência de sementes. Este algoritmo bem documentado funciona pela comparação de cada posição da leitura contra todas as posições candidatas da referência. Estas comparações correspondem a uma matriz de potenciais alinhamentos entre a leitura e a referência. Para cada uma destas posições de alinhamento candidatas, Smith-Waterman criam pontuações que são utilizadas para avaliar se o melhor alinhamento, a passar pela matriz de células, a alcança por uma correspondência ou não-correspondência de nucleótido (movimento diagonal), uma eliminação (movimento horizontal) ou uma inserção (movimento vertical). Uma correspondência entre uma leitura e uma referência proporciona um bónus na pontuação e uma não-correspondência ou indel impõe uma penalidade. A pontuação global mais alta na matriz é o alinhamento escolhido. O algoritmo é acelerado por hardware nas placas de matrizes de portas de campo programáveis (FPGA) DRAGEN. O genoma de referência utilizado na aplicação é criado a partir do FASTA hg19 do UCSC com a opção DRAGEN, para criar uma tabela “alt-aware hash” baseada em liftover.

Identificação de variante de linha germinal DRAGEN

O identificador de pequenas variantes de linha germinal DRAGEN recolhe as leituras de ADN mapeadas e alinhadas, como entrada, e identifica polimorfismos de nucleótido único (SNP) e inserções ou eliminações (indels) através de uma combinação a nível de colunas e de conjunto local *de novo* de haplótipos. Para ativar o identificador de pequenas variantes de linha germinal DRAGEN, selecione o fluxo de trabalho da variante da linha germinal.

A identificação de variantes de linha germinal é tipicamente usada para amostras de linha germinal nas quais se sabe que existem duas ploidias. As regiões referência mobilizáveis são primeiramente identificadas com cobertura suficiente de alinhamento. Nestas regiões referência, uma análise rápida das leituras recolhidas identifica regiões ativas, que são centradas à volta de colunas acumuladoras, com evidência de uma variante. As regiões ativas são preenchidas com contexto suficiente para abranger conteúdo vizinho significativo, não-referência. Se houver evidência de indels, as regiões ativas recebem preenchimento adicional.

As leituras alinhadas são agregadas em cada região ativa e incluídas num gráfico De Bruijn. As margens das leituras agregadas são ponderadas por contagens de observação, com a sequência de referência como espinha dorsal. Após algum trabalho de limpeza e simplificação no gráfico, são extraídas todas as vias como haplótipos candidatos. Cada haplótipo é alinhado, de acordo com Smith-Waterman, com o

genoma referência para identificar as variantes que representam. Este conjunto de eventos pode ser aumentado por uma deteção baseada na posição. Por cada par de leitura de haplótipo, a probabilidade $P(r|H)$ de observar a leitura, assumindo que o haplótipo é a verdadeira amostra na origem, é estimada utilizando um modelo de par oculto de Markov (HMM).

Analisando por posição de referência sobre a região ativa, são formados genótipos candidatos a partir de combinações de diplóides de eventos variantes (SNP ou indels). Para cada evento (incluindo a referência), a probabilidade condicional $P(r|e)$ de observar cada leitura sobreposta é estimada como o valor máximo $P(r|H)$ para os haplótipos que suportam o evento. Estes são combinados em probabilidade condicional $P(R|e1e2)$ para um genótipo (evento par) e multiplicados para produzir a probabilidade condicional $P(R|e1e2)$ de observação de toda a leitura acumulada. Utilizando a fórmula de Bayes, a probabilidade $P(e1e2|R)$ posterior de cada genótipo diplóide é calculado e o vencedor é identificado.

DRAGEN for IDPE Dx aplica filtragem automática. Consulte [Anotações do ficheiro VCF do fluxo de trabalho da linha germinal na página 14](#) para obter mais informações.

Identificação de variante somática DRAGEN

O identificador somático de pequenas variantes DRAGEN recolhe as leituras de ADN mapeadas e alinhadas, como entrada e identificações SNV e indels através de um conjunto local *de novo* de haplótipos, numa região ativa. Para ativar o identificador somático de pequenas variantes DRAGEN, selecione uma aplicação de variantes somáticas.

A identificação somática de variantes é tipicamente usada para amostras de tumores. Com este fluxo de trabalho, a DRAGEN não faz quaisquer pressupostos em termos de ploidia, o que permite a deteção de alelos com baixa frequência. Para os loci com cobertura até 100x na amostra do tumor, o DRAGEN tem um limite de deteção em frequências de alelos de variantes de 5%. O limite aumenta com a profundidade crescente numa base por-locus e reduz para metade, de cada vez que a cobertura duplica para além de 100x. As regiões referência mobilizáveis são primeiramente identificadas com cobertura suficiente de alinhamento. Nestas regiões referência, uma análise das leituras recolhidas identifica regiões ativas, que são centradas em colunas acumuladoras, com evidência de uma variante, nas leituras de um tumor. As regiões ativas são preenchidas com contexto suficiente para abranger conteúdo vizinho significativo, não-referência. Se houver evidência de indels, as regiões ativas recebem preenchimento adicional.

As leituras alinhadas são agregadas em cada região ativa e incluídas num gráfico De Bruijn. As margens das leituras agregadas são ponderadas por contagens de observação, com a sequência de referência como espinha dorsal. Após algum trabalho de limpeza e simplificação no gráfico, são extraídas todas as vias como haplótipos candidatos. Cada haplótipo é alinhado, de acordo com Smith-Waterman, com o genoma referência para identificar as variantes que representam. Por cada par de leitura de haplótipo, a probabilidade $P(r|H)$ de observar a leitura é estimada utilizando um modelo de par oculto de Markov (HMM), assumindo que o haplótipo é a verdadeira amostra na origem.

Para determinar a classificação do limite de deteção do tumor (TLOD), o identificador somático de pequenas variantes DRAGEN analisa primeiro por posição de referência para cada evento somático candidato, bem como o evento de referência sobre a região ativa. A probabilidade condicional $P(r|e)$ de observar cada leitura sobreposta é estimada como o valor máximo $P(r|H)$ para os haplótipos, que suportam o evento. Estes são combinados em probabilidade condicional $P(r|E)$ para um evento hipotético, E , envolvendo uma mistura de um alelo somático de referência e candidato de um intervalo de frequências possíveis de alelos e multiplicados para produzir a probabilidade condicional $P(R|E)$ de observação de toda a leitura acumulada. A partir daí, é calculada uma pontuação TLOD como prova de que está presente um alelo ALT na amostra de tumor, num determinado locus.

DRAGEN for IDPE Dx aplica filtragem automática. Consulte [Anotações de ficheiro VCF do fluxo de trabalho somático na página 17](#), para obter mais informações.

Criar um ensaio planeado

Utilize as seguintes etapas para configurar um ensaio no Illumina Run Manager ou no NextSeq 550Dx, utilizando um navegador num computador ligado à rede. Utilize um navegador num computador em rede, se pretender importar dados de amostra. Consulte o Illumina Run Manager para o guia de software NextSeq 550Dx (documento n.º 200025239) para obter instruções sobre como aceder a Illumina Run Manager, a partir de um computador em rede.

Existem duas formas diferentes de criar um novo ensaio planeado:

- **Import Run** (Importar ensaio) — Utilize uma folha de amostra de um ensaio existente como modelo para um novo ensaio. Consulte o Illumina Run Manager para o guia de software NextSeq 550Dx (documento n.º 200025239) para obter informações sobre como importar um ensaio.
- **Create Run** (Criar ensaio) — Insira manualmente os parâmetros do ensaio. As instruções seguintes descrevem como criar um ensaio.

NOTA Os campos de entrada obrigatórios na interface do utilizador estão marcados com um símbolo de asterisco (*).

Aplicação

1. No separador Planned (Planeados) do ecrã Runs (Ensaio), selecione **Create Runs** (Criar ensaios).
2. Selecione a aplicação DRAGEN for Illumina DNA Prep with Enrichment Dx e, em seguida, selecione **Next** (Seguinte).

Definições do ensaio

1. No ecrã Run Settings (Definições do ensaio), introduza um nome exclusivo de ensaio. O nome do ensaio é o nome que o identifica desde a sequenciação à análise.
2. **[Opcional]** Introduza uma descrição do ensaio, para ajudar a identificá-lo.
3. Selecione o(s) kit(s) do adaptador de índice utilizado(s) durante a preparação do banco.
4. Reveja o comprimento de leitura e altere, se necessário. A Leitura 1 e a Leitura 2 têm um valor predefinido de 151 ciclos. O índice 1 e o índice 2 têm um valor predefinido de 10 ciclos, que não pode ser modificado.
5. **[Opcional]** Introduza o ID do tubo para o banco.
6. Selecione **Next** (Seguinte).

Dados de amostra

Os dados da amostra incluem o Sample ID (ID da amostra), a Well Position (Posição do poço) (posição do poço da placa de índice) e o Library Name (nome do banco). Ao utilizar o Índice A&B, a Posição do Poço também inclui o identificador da Placa.

Existem duas formas de introduzir os dados de amostra:

- **Import Samples** (Importar amostras) — Use um ficheiro de modelo disponível para descarregamento no ecrã Sample Data (Dados de amostra).
- **Manually** (Manualmente) — Insira os dados de amostra diretamente na tabela no ecrã Sample Data (Dados de amostra).

Importar amostras

Ao planear um ensaio de sequenciação utilizando um navegador num computador com rede, está disponível um ficheiro de modelo (*.csv) para descarregamento no ecrã Sample Data (Dados de amostra). O ficheiro de modelo não está disponível para descarregamento ao aceder a Illumina Run Manager, através do software do sistema operativo NextSeq 550Dx. Para introduzir os dados de amostra utilizando a funcionalidade Importar Amostras, execute os seguintes passos.

NOTA Conclua os passos de Run Settings (Definições do ensaio), antes de prosseguir.

1. Selecione **Download Template** (Descarregar modelo) para descarregar um ficheiro CSV em branco.
2. No ficheiro de modelo, introduza os dados de amostra e, em seguida, guarde-o. O Nome do banco é opcional.

NOTA Ao usar o Índice A&B, os dados para a coluna B devem incluir a posição da placa e do poço (posição do poço da placa de índice). Exemplo: A-A01, A-A02, A-A03.

3. Selecione **Import Samples** (Importar amostras) e procure o ficheiro modelo, que contém as informações dos dados de amostra do passo anterior.
4. Selecione **Open** (Abrir), **Proceed** (Avançar), e em seguida, **Next** (Próximo).

NOTA Alterar Sample ID (ID da amostra) antes de selecionar Next (Seguinte) pode resultar num erro. Termine de configurar o ensaio, antes de fazer alterações para evitar erros.

Introduzir amostras manualmente

Utilize a tabela no ecrã Sample Data (Dados de amostra), para introduzir manualmente os dados de amostra.

1. Introduza um ID único de amostra no campo Sample ID (ID da amostra).

2. Use **Well Position** (Posição do Poço) (Índice A ou Índice B) ou **Plate – Well Position** (Placa – Posição do Poço) (Índice A&B), para selecionar o índice associado para as amostras.
Os campos dos i7 Index, Index 1, i5 Index, and Index 2 (índices i7, 1, i5 e 2) são preenchidos automaticamente.
3. **[Opcional]** Introduza um nome do banco.
4. Adicione linhas e repita os passos 1–3, conforme necessário, até que todas as amostras tenham sido adicionadas. Pode adicionar várias linhas de uma só vez, introduzindo primeiro o número de linhas a adicionar e, em seguida, selecionando o ícone +. Também pode remover linhas selecionando a caixa junto ao número da linha e, em seguida, clicando no ícone de lixo.
5. Selecione **Next** (Seguinte).

Definições das análises

1. Selecione o fluxo de trabalho desejado das análises:
 - Criação de FASTQ
 - Criação de FASTQ e VCF para um fluxo de trabalho de linha germinal (ficheiro manifest necessário)
 - Criação de FASTQ e VCF para um fluxo de trabalho somático (ficheiro manifest necessário)
2. **[Opcional] Generate ORA compressed FASTQs** (Gerar FASTQ comprimidos com o ORA) está ativado por padrão. A compressão FASTQ ORA comprime sem perda ficheiros FASTQ até 5x, em comparação com fastq.gz. Desmarque **Generate ORA compressed FASTQs** (Gerar FASTQ comprimidos com o ORA) se forem preferidos dados não comprimidos (fastq.gz).
3. Para os fluxos de trabalho de linha germinal e somático, é necessário um ficheiro manifest. Utilize o menu pendente **Manifest File Selection** (seleção de ficheiro manifest), para selecionar um ficheiro manifest. O manifest é um ficheiro de texto delimitado por tabulações BED (*.bed), que especifica os nomes e localizações de regiões referência visadas. Consulte [Ficheiro manifest na página 8](#), para obter mais informações.
4. **[Opcional]** Para fluxos de trabalho somáticos, use o menu suspenso **Noise File Selection** (Seleção de ficheiro Noise), para selecionar um ficheiro Noise sistemático.
Pode ser especificado um ficheiro BED (*.bed.gz) com um determinado nível de noise, para filtrar noise sistemático. Consulte [Filtragem de noise \(opcional\) na página 9](#), para obter mais informações.
5. Selecione **Next** (Seguinte).

Rever ensaio

1. No ecrã Review (Rever), reveja a informação introduzida nos ecrãs Run Settings (Definições do ensaio), Sample Data (Dados de amostra) e Analysis Settings (Definições das análises).
2. Selecione **Save** (Guardar).
O ensaio é guardado no separador Planned (Planeado) no ecrã Runs (ensaios).

Definições

Para ver ou alterar as definições da aplicação DRAGEN for IDPE Dx, selecione primeiro o ícone Applications (Aplicações) no ecrã principal. Em seguida, selecione a aplicação que pretende ver ou alterar. É necessária uma conta de administrador para alterar as definições.

Configuração

O ecrã de configuração mostra as seguintes definições da aplicação:

- **Library Prep Kits** (Kits de preparação de bancos) — Mostra o kit de preparação do banco predefinido para a aplicação. Esta predefinição não pode ser alterada.
- **Index Adapter Kits** (Kits do adaptador de índice) — Mostra o kit adaptador de índice predefinido para a aplicação. Esta predefinição não pode ser alterada.
- **Read lengths** (Comprimentos de leitura) — Os comprimentos de leitura são definidos para 151 na aplicação por defeito, mas pode ser alterado durante a criação do ensaio.
- **Manifest and Noise Files** (Ficheiros Manifest e Noise) — Definições de carregamento e alteração de ficheiros Manifest e Noise.
 - Selecione **Upload File** (Carregar ficheiro) para carregar ficheiros para utilização em análises.
 - Selecione o botão de opção **Default** (Predefinição) para definir o ficheiro como o ficheiro Manifest ou Noise selecionado, durante a criação do ensaio, quando a aplicação for selecionada.
 - Selecione a caixa de seleção **Enabled** (Ativado) para definir o ficheiro para mostrar no menu pendente, durante a criação do ensaio.

Autorizações

Utilize as caixas de seleção no ecrã Permissions (Autorizações), para gerir o acesso do utilizador à aplicação.

Ficheiro manifest

Ao usar DRAGEN for IDPE Dx, é necessário a entrada de um ficheiro manifest, para os seguintes fluxos de trabalho:

- Criação de FASTQ e VCF para um fluxo de trabalho de linha germinal
- Criação de FASTQ e VCF, para um fluxo de trabalho somático

O ficheiro manifest é um ficheiro de texto delimitado por tabulações BED (*.bed), que especifica os nomes e localizações de regiões referência visadas. A secção principal do ficheiro manifest é a secção Regions (Regiões) e deve conter as seguintes colunas de dados:

Coluna	Descrição
Nome	Nome exclusivo especificado pelo utilizador para o alvo
Cromossoma	Localização do cromossoma (por exemplo, chr10, chr5, etc.)
Iniciar	Índice de base 1 para a posição de início do alvo
Parar	Índice de base 1 para a posição de paragem do alvo
Comprimento da sonda a montante	O comprimento da sonda a montante. Para a aplicação DRAGEN for IDPE Dx, o comprimento deve ser definido para 0.
Comprimento da sonda a jusante	O comprimento da sonda a jusante. Para a aplicação DRAGEN for IDPE Dx, o comprimento deve ser definido para 0.

NOTA É necessário um formato de ficheiro manifest válido para análise. O DRAGEN irá parar a análise, se o ficheiro manifest for inválido.

Filtragem de ruído (opcional)

O filtro de ruído sistemático está disponível para identificação de variantes somáticas e pode ser utilizado para reduzir as identificações de falsos positivos, tendo em conta o ruído específico do local. O ficheiro de ruído sistemático é gerado pela recolha inicial de aproximadamente 50 amostras normais (de preferência específicas para o painel, preparação do banco e sequenciador) e, em seguida, a soma das frequências de alelos abaixo de 30% em cada local com cobertura suficiente é dividida pelo número total de amostras (as frequências de alelos acima de 30% são assumidas como sendo variantes da linha germinal e não ruído). Assim que os valores de ruído forem gerados, as variantes somáticas detetadas nesse local serão filtradas.

O filtro pode ser utilizado no modo Tumor-Normal, mas é especialmente útil nos ensaios Tumor-Only, onde não está disponível um normal com correspondência. O ficheiro de ruído sistemático deve usar um ficheiro BED que tenha uma extensão (*.bed.gz) e deve incluir quatro colunas: Cromossoma, Início, Fim e níveis de ruído específicos do local para cada linha. A filtragem de ruído sistemático é opcional.

Saída de análises

Os ensaios em curso atualmente são mostrados no separador Active (Ativo). Os ensaios concluídos são apresentados no separador Completed (Concluídos). DRAGEN for IDPE Dx cria uma pasta de análise com nome único para cada análise, separada da pasta que contém dados de sequenciação. A pasta de análise inclui a seguinte informação:

- Ficheiro “manifest” utilizado
- Versão do software
- ID da amostra

- Total de leituras alinhadas
- Percentagem de leituras alinhadas por amostra
- Número de SNV identificados por amostra
- Número de indels identificados por amostra
- Cobertura de estatística

Ficheiros de saída da análise

A localização da pasta de análise é especificada pelo Armazenamento Externo para definição dos Resultados de Análise. Consulte o Illumina Run Manager para o guia de software NextSeq 550Dx (documento n.º 200025239) para obter mais informações sobre o Armazenamento Externo para definição dos Resultados de Análise.

No ecrã Run details (Detalhes do ensaio), o campo External Location (Local externo) indica o caminho para dados de sequenciação. O nome único da pasta de análise é indicado no campo Analysis Output Folder (Pasta de saída de análise) no ecrã Run Details (Detalhes do ensaio). Os ficheiros exatos gerados dependem do fluxo de trabalho usado. A aplicação gera os seguintes ficheiros de resultados de análises.

NOTA Se ocorrer um erro de limitação máxima de comprimento do caminho do ficheiro, ao aceder aos ficheiros de saída de análise, tente mover o ficheiro para um caminho mais curto de localização ou utilize um método diferente para abrir o ficheiro.

Ficheiro de saída	Descrição
Relatório de resumo de variantes (*.pdf)	Contém um resumo das informações do ficheiro, versões do software, informações da amostra, estatísticas do nível de leitura e SNV, inserções, eliminações e resumos da cobertura. Apenas os fluxos de trabalho da linha germinal e somáticos produzem um relatório de variante.
FASTQ (*.fastq.gz ou *.fastq.ora)	Ficheiros intermédios que contêm identificações de base com pontuação de qualidade. Os ficheiros FASTQ são a principal entrada para o passo de alinhamento. Quando a compressão ORA é selecionada, é utilizada a extensão de ficheiro *.fastq.ora.
Ficheiros de alinhamento BAM (*.bam)	Contém leituras alinhadas de uma determinada amostra.
Ficheiros genoma VCF (*.gvcf.gz)	Contém o genótipo de cada posição, seja denominada como variante ou como referência.
Ficheiros VCF (*.vcf.gz)	Contém variantes denominadas em cada posição.

Ficheiro de saída	Descrição
Relatório de métrica do ensaio (*.csv)	Contém métricas de qualidade sobre o ensaio, incluindo a produtividade total não indexada e a pontuação Q30.

Ficheiros FASTQ

FASTQ (*.fastq.gz, *.fastq.ora) é um formato de ficheiro baseado em texto, que contém identificações de base e valores de qualidade por leitura. Cada ficheiro contém a seguinte informação:

- O identificador da amostra
- A sequência
- As pontuações de qualidade Phred em formato com codificação ASCII + 33

O identificador da amostra está formatado como segue:

```
@Instrumento:IDEnsaio:IDCélulaDeFluxo:Faixa:Bloco:X:Y
NúmLeitura:IdentificaçãoFiltro:0:NúmeroAmostra
Exemplo:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

Ficheiros BAM

Um ficheiro BAM (*.bam) é a versão binária comprimida de um ficheiro SAM (mapa de sequência de alinhamento) utilizado para representar sequências alinhadas até 128 Mb. Os ficheiros BAM utilizam o formato de nome de `SampleName_S#.bam`. O # corresponde ao número da amostra determinado pela ordem em que as amostras se encontram enumeradas no ensaio. No modo multinódulo, o n.º do S é definido para S1, independentemente da ordem da amostra.

Os ficheiros BAM contêm uma secção de cabeçalho e uma secção de alinhamento:

- **Header** (Cabeçalho) — Contém informações sobre todo o ficheiro, como, por exemplo, o nome da amostra, o tamanho da amostra e o método de alinhamento. Os alinhamentos da secção de alinhamentos estão associados a informações específicas da secção de cabeçalho.
- **Alignments** (Alinhamentos) — Contém o nome da leitura, a sequência da leitura, a qualidade da leitura, informações de alinhamento e etiquetas personalizadas. O nome da leitura inclui o cromossoma, a coordenada de início, qualidade do alinhamento e o descritor de correspondências.

A secção de alinhamentos inclui a seguinte informação para cada leitura ou par de leituras:

- AS: Extremidades emparelhadas com qualidade de alinhamento.
- RG: Grupo de leitura, que indica o número de leituras numa determinada amostra.

- BC: Etiqueta de codificação de barras, que indica o ID da amostra da demultiplexagem, associado à leitura.
- SM: Extremidade única com qualidade de alinhamento.
- XC: Descritor de correspondências.
- XN: O nome da etiqueta do amplicão, que regista o ID do amplicão associado à leitura

Ficheiros de índice BAM (*.bam.bai) fornecem um índice do ficheiro BAM correspondente.

Ficheiros VCF

Os ficheiros em formatos de identificação de variantes (*.vcf) contêm informações sobre variantes encontrados em posições específicas num genoma de referência.

O cabeçalho do ficheiro VCF inclui a versão do formato do ficheiro VCF, a versão identificadora da variante e lista as anotações na parte restante do ficheiro. O cabeçalho VCF inclui igualmente o ficheiro do genoma de referência e o ficheiro BAM. A última linha no cabeçalho contém os títulos das colunas, para as linhas de dados. Cada linha do ficheiro VCF contém informação sobre uma variante única.

Tabela 1 Títulos do ficheiro VCF

Cabeçalho	Descrição
CHROM	O cromossoma do genoma de referência. Os cromossomas aparecem pela mesma ordem que os do ficheiro de referência FASTA.
POS	A posição de base única da variante no cromossoma de referência. Para variantes de nucleótido único (SNV), esta posição é a base de referência com a variante. Para os indels, esta posição é a base de referência imediatamente precedente da variante.
ID	O número rs (referência SNP) para o SNP obtido do <code>dbSNP.txt</code> , se aplicável. Se existirem vários números rs nesta localização, a lista é delimitada por ponto e vírgula. Caso não exista uma entrada dbSNP nesta posição, é utilizado um marcador de valor em falta ('.').
REF	O genótipo de referência. Por exemplo, uma eliminação de um único T é representada como referência TT e T alternativo. Um A para variante de nucleótido de T único é representado como referência A e T alternativo.
ALT	Os alelos que diferem da leitura de referência. Por exemplo, uma inserção de um único T é representada como referência A e AT alternativa. Um A para variante de nucleótido de T único é representado como referência A e T alternativa.

Cabeçalho	Descrição
QUAL	Uma pontuação de qualidade com escala Phred atribuída pelo identificador de variante. Pontuações mais altas indicam um grau de confiança maior na variante e menor probabilidade de erros. Para uma pontuação de qualidade de Q, a probabilidade estimada de ocorrerem erros é de $10^{-(Q/10)}$. Por exemplo, o conjunto de identificações Q30 tem uma taxa de erro de 0,1%. Muitos identificadores de variantes atribuem pontuações de qualidade com base nos respetivos modelos estatísticos, que são elevados em relação à taxa de erros observada.

Tabela 2 Anotações do ficheiro VCF do fluxo de trabalho da linha germinal

Cabeçalho	Descrição
<p>FILTER (FILTRO)</p>	<p>Se todos os filtros passarem, PASS (APROVADO) encontra-se escrito na coluna dos filtros. As entradas possíveis do FILTRO incluem:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL — Aplicado se a pontuação QUAL da variante SNP não atingir o limiar • DRAGENIndelHardQUAL — Aplicado se a pontuação QUAL da variante Indel não atingir o limiar • LowDepth — Filtrado no local porque a profundidade da cobertura não atinge o limiar • LowGQ — Filtrado no local porque a qualidade do genótipo não atinge o limiar • PloidyConflict — Identificação do genótipo pelo identificador de variantes não consistente com a ploidia de cromossomas • base_quality — Filtrado no local porque a qualidade mediana base das leituras ALT neste locus não atinge o limiar • filtered_reads — Filtrado no local porque uma fração de leituras demasiado grande foi excluída pela filtragem • fragment_length — Filtrado no local porque a diferença absoluta entre o comprimento mediano do fragmento das leituras ALT e REF, neste locus, excede o limiar • low_depth — Filtrado no local porque a profundidade da leitura é demasiado baixa • low_frac_info_reads — Filtrado no local porque a fração de leituras informativas se encontra abaixo do limiar • low_normal_depth — Filtrado no local porque a profundidade da leitura normal da amostra é demasiado baixa • long_indel — Filtrado no local porque o comprimento do Indel é demasiado longo • mapping_quality — Filtrado no local porque a qualidade mediana do mapeamento das leituras ALT neste locus não atinge o limiar • multiallelic — Filtrado no local porque mais de dois alelos ALT passam o LoD do tumor • non_homref_normal — Filtrado no local porque o genótipo normal da amostra não é referência homocigótica • no_reliable_supporting_read — Filtrado no local porque não existe uma leitura somática de apoio fiável • panel_of_normals — Visto, pelo menos, numa amostra no painel dos VCF normais • read_position — Filtrado no local porque a mediana de distâncias entre o início/fim da leitura e este locus encontra-se abaixo do limiar • RMxNRepeatRegion — Filtrado no local porque todos ou parte dos alelos da variante são uma repetição da referência

Cabeçalho	Descrição
FILTER (FILTRO)	<ul style="list-style-type: none"> • strand_artifact — Filtrado no local pela tendência grave da cadeia • str_contraction — Filtrado no local devido a suspeita de erro de PCR, onde o alelo ALT é uma unidade de repetição inferior à referência • too_few_supporting_reads — Filtrado no local porque existem poucas leituras de apoio na amostra do tumor • weak_evidence — Pontuação somática da variante não atinge o limite
INFO	<p>As possíveis entradas de INFO incluem:</p> <ul style="list-style-type: none"> • AC — Contagem de alelos em genótipos por cada alelo ALT, na mesma ordem listada. • AF — Frequência de alelos por cada alelo ALT, na mesma ordem listada. • AN — O número total de alelos em genótipos identificados. • DB — Adesão dbSNP. • FS — Valor de p com escala Phred, utilizando o teste exato de Fisher para detetar tendências da cadeia. • QD — Fiabilidade da variante/qualidade por profundidade. • R2_5P_bias — Pontuação com base em tendência mate e distância da extremidade principal 5. • SOR — Taxa de probabilidade simétrica de tabela de contingência 2x2, para detetar tendências da cadeia. • DP — Profundidade aproximada de leitura (informativa e não-informativa); algumas leituras podem ter sido filtradas com base no MAPQ, etc. • END — Posição stop do intervalo. • FractionInformativeReads — A fração de leituras informativas do total de leituras. • MQ — Qualidade RMS do mapeamento. • MQRankSum — Pontuação Z do teste de soma de ranks de Wilcoxon de ALT vs. qualidades de mapeamento de leitura REF. • ReadPosRankSum — Pontuação Z do teste de soma de ranks de Wilcoxon de ALT vs. tendências de posição de leitura REF. • SOMATIC — Pelo menos uma variante nesta posição é somática.

Cabeçalho	Descrição
<p>FORMAT (FORMATO)</p>	<p>O formato da coluna indica os campos separados por dois pontos. Por exemplo, GT:GQ. Os campos disponíveis incluem:</p> <ul style="list-style-type: none"> • AD — Profundidades dos alelos (contando apenas leituras informativas do total de leituras) para os alelos REF e ALT, na ordem listada. • AF — Frações de alelos para os alelos ALT, na ordem listada. • DP — Profundidade aproximada de leitura (leituras com MQ=255 ou com maus mates são filtradas). • F1R2 — Contagem de leituras em orientação de par F1R2 de apoio a cada alelo. • F2R1 — Contagem de leituras em orientação de par F2R1 de apoio a cada alelo. • GT — Genótipo. 0 corresponde à base de referência, 1 corresponde à primeira entrada na coluna ALT e assim por diante. A barra inclinada (/) indica que não está disponível informação faseada. • MB — Estatísticas de componente por amostra, para detetar a tendência mate. • PS — ID de informação física faseada, onde cada ID único, numa determinada amostra (mas não através das amostras), se liga aos registos dentro de um grupo faseado. • SB — Estatísticas de componente por amostra, que abrange o Teste Exato de Fisher, para detetar as tendências da cadeia. • SQ — Qualidade somática.
<p>SAMPLE (AMOSTRA)</p>	<p>A coluna amostra dá-lhe os valores especificados na coluna FORMATO.</p>

Tabela 3 Anotações de ficheiro VCF do fluxo de trabalho somático

Cabeçalho	Descrição
<p>FILTER (FILTRO)</p>	<p>Se todos os filtros passarem, PASS (APROVADO) encontra-se escrito na coluna dos filtros. As entradas possíveis do FILTRO incluem:</p> <ul style="list-style-type: none"> • base_quality — Filtrado no local porque a qualidade mediana base das leituras ALT neste lócus não atinge o limiar • filtered_reads — Filtrado no local porque uma fração de leituras demasiado grande foi excluída pela filtragem • fragment_length — Filtrado no local porque a diferença absoluta entre o comprimento mediano do fragmento das leituras ALT e REF, neste lócus, excede o limiar • low_depth — Filtrado no local porque a profundidade da leitura é demasiado baixa • low_frac_info_reads — Filtrado no local porque a fração de leituras informativas se encontra abaixo do limiar • low_normal_depth — Filtrado no local porque a profundidade da leitura normal da amostra é demasiado baixa • long_indel — Filtrado no local porque o comprimento do Indel é demasiado longo • mapping_quality — Filtrado no local porque a qualidade mediana do mapeamento das leituras ALT neste lócus não atinge o limiar • multiallelic — Filtrado no local porque mais de dois alelos ALT passam o LoD do tumor • non_homref_normal — Filtrado no local porque o genótipo normal da amostra não é referência homozigótica • no_reliable_supporting_read — Filtrado no local porque não existe uma leitura somática de apoio fiável • panel_of_normals — Visto, pelo menos, numa amostra no painel dos VCF normais • read_position — Filtrado no local porque a mediana de distâncias entre o início/fim da leitura e este lócus encontra-se abaixo do limiar • RMxNRepeatRegion — Filtrado no local porque todos ou parte dos alelos da variante são uma repetição da referência • strand_artifact — Filtrado no local pela tendência grave da cadeia • str_contraction — Filtrado no local devido a suspeita de erro de PCR, onde o alelo ALT é uma unidade de repetição inferior à referência • too_few_supporting_reads — Filtrado no local porque existem poucas leituras de apoio na amostra do tumor • weak_evidence — Pontuação somática da variante não atinge o limite • systematic_noise — Filtrado no local com base em evidência de noise sistemático nos normais

Cabeçalho	Descrição
INFO	<p>As possíveis entradas de INFO incluem:</p> <ul style="list-style-type: none"> • DP — Profundidade aproximada de leitura (informativa e não-informativa); algumas leituras podem ter sido filtradas com base no MAPQ, etc. • END — Posição stop do intervalo. • FractionInformativeReads — A fração de leituras informativas do total de leituras. • MQ — Qualidade RMS do mapeamento. • MQRankSum — Pontuação Z do teste de soma de ranks de Wilcoxon de ALT vs. qualidades de mapeamento de leitura REF. • ReadPosRankSum — Pontuação Z do teste de soma de ranks de Wilcoxon de ALT vs. tendências de posição de leitura REF. • AQ — Pontuação de noise sistemático. • hotspot — Local somático conhecido, utilizado para aumentar a fiabilidade na identificação. • SOMATIC — Pelo menos uma variante nesta posição é somática.

Cabeçalho	Descrição
FORMAT (FORMATO)	<p>O formato da coluna indica os campos separados por dois pontos. Por exemplo, GT:GQ. Os campos disponíveis incluem:</p> <ul style="list-style-type: none"> • AD — Profundidades dos alelos (contando apenas leituras informativas do total de leituras) para os alelos REF e ALT, na ordem listada. • AF — Frações de alelos para os alelos ALT, na ordem listada. • DP — Profundidade aproximada de leitura (leituras com MQ=255 ou com maus mates são filtradas). • F1R2 — Contagem de leituras em orientação de par F1R2 de apoio a cada alelo. • F2R1 — Contagem de leituras em orientação de par F2R1 de apoio a cada alelo. • GP — Probabilidades com escala Phred posteriores para genótipos, como definido na especificação do VCF. • GQ — Qualidade de genótipo. • GT — Genótipo. 0 corresponde à base de referência, 1 corresponde à primeira entrada na coluna ALT e assim por diante. A barra inclinada (/) indica que não está disponível informação faseada. • MB — Estatísticas de componente por amostra, para detetar a tendência mate. • PL — Probabilidades normalizadas com escala Phred para genótipos, como definido na especificação do VCF. • PRI — Probabilidades anteriores com escala Phred para genótipos. • PS — ID de informação física faseada, onde cada ID único, numa determinada amostra (mas não através das amostras), se liga aos registos dentro de um grupo faseado. • SB — Estatísticas de componente por amostra, que abrange o Teste Exato de Fisher, para detetar as tendências da cadeia. • SQ — Qualidade somática.
SAMPLE (AMOSTRA)	A coluna amostra dá-lhe os valores especificados na coluna FORMATO.

Ficheiros genoma VCF

Os ficheiros VCF (*.gvcf.gz) sobre genomas seguem um conjunto de convenções para representar todos os locais relacionados com o genoma, num formato razoavelmente compacto. Os ficheiros gVCF incluem todos os centros na região de interesse num único ficheiro, para cada amostra. O ficheiro gVCF não mostra identificações em posições que não passam todos os filtros. Uma etiqueta de genótipo (GT) de ./ indica uma não-identificação.

Recolocar a análise em fila de espera

Pode recolocar a análise em fila de espera se a análise tiver sido interrompida, se a análise tiver sido malsucedida ou se pretender voltar a analisar um ensaio com definições diferentes. Para recolocar a análise, execute os seguintes passos:

1. No ecrã Run (Ensaio), selecione o separador Completed (Concluídos) e, em seguida, selecione o nome do ensaio a reanalisar.
Se Requeue Analysis (Recolocar a análise em fila de espera) tiver sido realizada anteriormente, selecione o nome de ensaio de Parent Run (Ensaio principal).
2. No ecrã Run Details (Detalhes do ensaio), depois de Sequencing Information (Informação de sequenciação), selecione **Requeue Analysis** (Recolocar a análise em fila de espera).
3. Selecione uma opção:
 - Requeue analysis with no changes (Recoloque as análises em fila de espera, sem qualquer alteração)
 - Edit run settings and requeue analysis (Editar definições do ensaio e de recolocação de análise em fila de espera)
 - Requeue analysis with a different application (Recolocação de análise com uma aplicação diferente)
4. Confirme se o local onde os dados de sequenciação residem atualmente é fornecido no campo **Sequencing data file path** (Caminho do ficheiro de dados de sequenciação).

NOTA O caminho para os dados de sequenciação deve corresponder ao caminho na definição External Storage for Analysis Results (Armazenamento externo para resultados de análise). Consulte o Illumina Run Manager para o guia de software NextSeq 550Dx (documento n.º 200025239) para obter informações sobre como alterar o caminho de armazenamento externo.

5. Introduza um motivo de reanálise.
6. Selecione **Requeue Analysis** (Recolocar análise em fila de espera).
7. Edite as alterações pretendidas em Run Settings (Definições do ensaio), Sample Data (Dados de amostra) e Analysis Settings (Definições de análise).
8. Selecione **Save** (Guardar). A análise é iniciada utilizando os parâmetros de análise atuais.

Assistência Técnica

Para obter assistência técnica, contacte o Suporte Técnico da Illumina.

Sítio Web: www.illumina.com

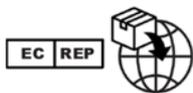
E-mail: techsupport@illumina.com

Fichas de dados de segurança (SDS) — Disponíveis no sítio Web da Illumina em support.illumina.com/sds.html.

Documentação do produto — Disponível para descarregamento em support.illumina.com.



Illumina
5200 Illumina Way
San Diego, Califórnia 92122 EUA
+1 800 809 ILMN (4566)
+1 858 202 4566 (fora da América do Norte)
techsupport@illumina.com
www.illumina.com



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

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

PARA UTILIZAÇÃO EM DIAGNÓSTICO IN VITRO.

© 2023 Illumina, Inc. Todos os direitos reservados.

illumina[®]