

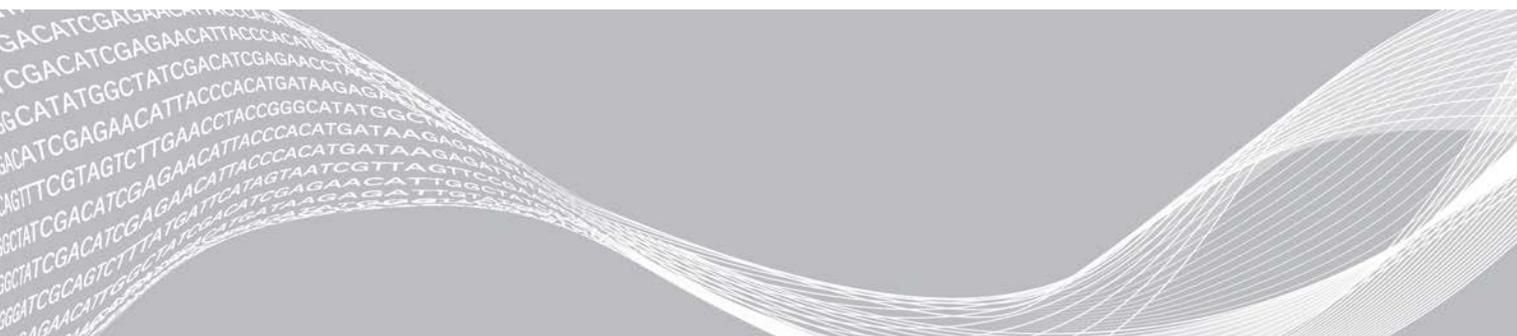
Local Run Manager

Módulo de análise Germline Variant

Manual do fluxo de trabalho do NextSeq 550Dx

PARA UTILIZAÇÃO EM DIAGNÓSTICO IN VITRO

Descrição geral	3
Introduzir informações do ensaio	4
Métodos de análise	6
Ver dados de amostra e ensaio	7
Relatório de análise	8
Ficheiros de saída de análise	9
Histórico de revisão	16
Assistência técnica	17



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 LER INTEGRALMENTE E SEGUIR EXPLICITAMENTE 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).

© 2021 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.

Descrição geral

O módulo Germline Variant do Local Run Manager destina-se a ser utilizado com o ensaio do Kit TruSeq Custom Amplicon Dx e com o NextSeq 550Dx da Illumina. Quando utilizado com o módulo Germline Variant, o ensaio destina-se à preparação das bibliotecas utilizadas para a sequenciação do ADN de exemplares de sangue total periféricos.

O módulo de análise avalia pequenas regiões de ADN amplificado, ou amplicons, de variantes. A sequenciação focada de amplicons permite uma cobertura elevada de determinadas regiões. Consulte o folheto informativo do *Kit TruSeq Custom Amplicon Dx* (documento n.º 1000000029772).

O módulo de análise Germline Variant requer consumíveis de sequenciação de 300 ciclos. Para obter mais informações, consulte o folheto informativo do *Kit de reagentes de saída elevada NextSeq 550Dx v2* ou do *Kit de reagentes de saída elevada NextSeq 550Dx v2.5*.

Acerca deste manual

Este manual fornece instruções para configurar os parâmetros do ensaio para sequenciação e análise no módulo de análise Germline Variant. Para obter informações sobre o painel e as definições do sistema do Local Run Manager, consulte o *Manual de referência do instrumento NextSeq 550Dx* (documento n.º 1000000009513).

Visualizar o Local Run Manager

A interface do Local Run Manager é visualizada no NextSeq 550Dx Operating Software (NOS) ou através de um browser da Internet. O browser suportado é o Chromium.



NOTA

Se utilizar um browser não suportado, transfira o browser suportado quando for solicitado na mensagem "Confirm Unsupported Browser" (Confirmar browser não suportado). Selecione "**here**" (aqui) para transferir a versão suportada do Chromium.

Ver no monitor do instrumento

- 1 Para ver a interface do Local Run Manager no monitor do instrumento, selecione uma das seguintes opções:
 - ▶ No ecrã inicial do NOS, selecione **Local Run Manager**.
Clique no X no canto superior direito para voltar ao NOS quando terminar.
 - ▶ Selecione o ícone de minimização do NOS, abra o browser Chromium no instrumento e introduza **http://localhost** na barra de endereços.
Apenas os utilizadores administrativos podem minimizar o NOS.

Ver a partir de um computador ligado em rede

- 1 Abra um browser do Chromium num computador com acesso à mesma rede que o instrumento e ligue utilizando o endereço IP do instrumento ou nome do instrumento. Por exemplo, **http://myinstrument**.

Introduzir informações do ensaio

Definir parâmetros

- 1 Inicie sessão no Local Run Manager.
- 2 Selecione **Create Run** (Criar ensaio) e selecione **Germline Variant**.
- 3 Introduza um nome para o ensaio que o identifique a partir da sequenciação através da análise. Utilize caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 4 **[Opcional]** Introduza a descrição do ensaio para ajudar a identificar o ensaio. Utilize caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 5 Selecione o número de amostras e o conjunto de índices na lista pendente. Considere as seguintes informações quando selecionar uma opção.
 - ▶ A lista pendente contém números de amostras com um conjunto de índice. Por exemplo, 24-Set 1 indica que existem 24 amostras para testar com índices a partir do conjunto de índices 1.
 - ▶ Os números do conjunto de índices referem-se a conjuntos diferentes de índices i5. O Set 1 (Conjunto 1) e o Set 2 (Conjunto 2) fornecem ambos uma diversidade de índices. São fornecidos dois conjuntos de índices para ajudar a evitar a diminuição de um único conjunto.
 - ▶ Escolha o número de amostras mais aproximado do número de amostras que está a testar. Se o número exato de amostras não estiver na lista, selecione o número mais próximo, mas inferior ao número que está a testar. Por exemplo, se pretende testar 18 amostras, selecione 16 amostras.
 - ▶ Os poços de amostras e as combinações de índices que cumprem os requisitos de diversidade de índices estão realçados a verde. Se selecionar outros poços e combinações de índices, ao guardar o ensaio receberá uma notificação se não forem cumpridos os requisitos da diversidade de índices.

Importar ficheiros de manifesto para o ensaio

- 1 Certifique-se de que os manifestos que pretende importar estão disponíveis numa localização de rede acessível ou numa unidade USB .
- 2 Selecione **Import Manifests** (Importar manifestos).
- 3 Aceda ao ficheiro de manifesto e selecione os manifestos que pretende adicionar.



NOTA

Para tornar os ficheiros de manifesto disponíveis para todos os ensaios utilizando o módulo de análise Germline Variant, adicione os manifestos com a funcionalidade Module Settings (Definições do módulo). Esta funcionalidade requer uma permissão de nível de utilizador administrador. Para obter mais informações, consulte o *Manual de referência do instrumento NextSeq 550Dx (documento n.º 1000000009513)*.

Especificar amostras para o ensaio

Especifique amostras para o ensaio utilizando uma das opções e instruções que se seguem.

- ▶ **Introduza as amostras manualmente**—Utilize a tabela em branco no ecrã Create Run (Criar ensaio).
- ▶ **Importar amostras**—Aceda a um ficheiro externo em formato de valores separados por vírgulas (*.csv). Está disponível um modelo para transferência no ecrã Create Run (Criar ensaio).

Depois de preencher a tabela de amostras, pode exportar as informações da amostra para um ficheiro externo. Utilize o ficheiro como referência quando preparar bibliotecas ou importe o ficheiro para outro ensaio.

Introduzir amostras manualmente

- 1 Introduza uma ID de amostra exclusiva no campo Sample ID (ID da amostra).
Utilize caracteres alfanuméricos, travessões ou sublinhados.
- 2 **[Opcional]** No caso de amostras de controlo positivo ou negativo, clique com o botão direito e selecione o tipo de controlo.
- 3 **[Opcional]** Introduza a descrição da amostra no campo Sample Description (Descrição da amostra).
Utilize caracteres alfanuméricos, travessões ou sublinhados.
As descrições de amostras estão associadas a um nome de amostra. As descrições de amostra são substituídas se o mesmo nome de amostra for novamente utilizado num ensaio posterior.
- 4 Selecione um adaptador de Índice 1 na lista pendente do Índice 1 (i7).
Quando utilizar os poços de amostra sugeridos, o software preenche automaticamente os adaptadores de índice i7 e i5 que cumprem os requisitos de diversidade de índices. Se o número de amostras exato que está a testar não estiver na lista, certifique-se de que seleciona os adaptadores de índice para poços extra. Se for necessário selecionar índices para poços extra ou se não utilizar as combinações recomendadas de adaptador de índices, antes de selecionar os índices, certifique-se de que lê a secção *Identificação de bases e diversidade de índices na página 15*.
- 5 Selecione um adaptador de Índice 2 na lista pendente do Índice 2 (i5).
- 6 Selecione um ficheiro de manifesto na lista pendente do manifesto.
- 7 Selecione uma opção para ver, imprimir ou guardar a disposição das placas, como referência para preparar bibliotecas:
 - ▶ Selecione o ícone  **Print** (Imprimir) para ver a disposição das placas. Selecione **Print** (Imprimir) a disposição das placas.
 - ▶ Selecione **Export** (Exportar) para exportar as informações da amostra para um ficheiro externo. Certifique-se de que as informações da amostra e do manifesto estão corretas. Informações incorretas podem afetar os resultados.
- 8 Selecione **Save Run** (Guardar ensaio).

Importar amostras

- 1 Selecione **Import Samples** (Importar amostras) e aceda à localização do ficheiro de informações da amostra. Existem dois tipos de ficheiros que pode importar.
 - ▶ Selecione **Template** (Modelo) no ecrã Create Run (Criar ensaio) para criar uma nova disposição de placas. O ficheiro modelo contém os cabeçalhos de coluna corretos para importação. Introduza informações em cada coluna das amostras do ensaio. Elimine as informações de exemplo das células não utilizadas e, em seguida, guarde o ficheiro.
 - ▶ Utilize um ficheiro com informações da amostra que tenha sido exportado do módulo Germline Variant através da função Export (Exportar).
- 2 Selecione o ícone  **Print** (Imprimir) para ver a disposição das placas.
- 3 Selecione **Print** (Imprimir) para imprimir a disposição das placas como referência para preparar bibliotecas.

- 4 **[Opcional]** Selecione **Export** (Exportar) para exportar as informações da amostra para um ficheiro externo.
Certifique-se de que as informações da amostra e do manifesto estão corretas. Informações incorretas podem afetar os resultados.
- 5 Selecione **Save Run** (Guardar ensaio).

Editar um ensaio

Para obter instruções para editar as informações de um ensaio antes da sequenciação, consulte o *Manual de referência do instrumento NextSeq 550Dx (documento n.º 1000000009513)*.

Métodos de análise

O módulo de análise Germline Variant executa os seguintes passos de análise e, em seguida, grava os ficheiros de saída de análise na pasta Alignment (Alinhamento).

- ▶ Desmultiplexagem de leituras de índice
- ▶ Gera ficheiros FASTQ
- ▶ É alinhado a uma referência
- ▶ Identifica variantes

Desmultiplexagem

A desmultiplexagem compara 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 aos clusters que não foram atribuídos a uma amostra.
- ▶ Os clusters são atribuídos a uma amostra quando a sequência de índice corresponde exatamente ou quando existe pelo menos uma única divergência por Leitura de índice.

Geração de ficheiros FASTQ

Após a desmultiplexagem, o software gera ficheiros de análise intermédia em formato FASTQ, que é um formato de texto usado para representar sequências. Os ficheiros FASTQ contêm leituras de cada amostra e as pontuações de qualidade associadas. Os clusters que não passarem no filtro serão excluídos.

Cada ficheiro FASTQ contém leituras de apenas uma amostra e o nome dessa amostra está incluído no nome de ficheiro FASTQ. Os ficheiros FASTQ são a principal entrada para alinhamento. São gerados oito ficheiros FASTQ por amostra; quatro a partir da Leitura 1 e quatro a partir da Leitura 2.

Alinhamento

Durante o passo de alinhamento, o algoritmo de Smith-Waterman com faixas alinha os clusters de cada amostra com as sequências de amplicons especificadas no ficheiro de manifesto.

O algoritmo de Smith-Waterman com faixas executa alinhamentos de sequências semiglobais para determinar regiões semelhantes entre duas sequências. Em vez de comparar a totalidade da sequência, o algoritmo de Smith-Waterman compara segmentos de todos os tamanhos possíveis.

Cada leitura de extremidade emparelhada é avaliada em termos de alinhamento em relação às sequências de investigação relevantes dessa leitura.

- ▶ A Leitura 1 é avaliada em relação ao complemento inverso de DLSO (Downstream Locus-Specific Oligos).
- ▶ A Leitura 2 é avaliada em relação ao ULSO (Upstream Locus-Specific Oligos).
- ▶ Se o início de uma leitura corresponder a uma sequência de investigação com até três diferenças (divergências ou mudanças devido aos primeiros Indels), o tamanho total da leitura é alinhado em relação ao alvo amplicon dessa sequência.
- ▶ Os Indels do DLSO e do ULSO não são observados devido à química do ensaio.

Os alinhamentos são filtrados a partir dos resultados de alinhamento com base nas taxas de divergência na região de interesse ou de todo o amplicon, consoante o tamanho do amplicon. Os alinhamentos filtrados são gravados em ficheiros de alinhamento como não alinhados e não são utilizados na identificação de variantes.

Identificação de variantes

Desenvolvido pela Illumina, o identificador de variantes Pisces identifica as variantes presentes na amostra de ADN.

O identificador de variantes Pisces identifica SNV (variantes de nucleótido único), MNV (variantes de nucleótido múltiplo) e pequenos Indels em três passos:

- ▶ Considera cada posição no genoma de referência em separado.
- ▶ Conta as bases na posição específica para leituras alinhadas que se sobrepõem à posição.
- ▶ Calcula uma pontuação de variante que mede a qualidade da identificação utilizando o modelo Poisson. São excluídas as variantes com uma pontuação de qualidade inferior a Q20.

Se uma variante passar em todos os filtros, é assinalada como PASS (Aprovada) no VCF.

Para obter mais informações, consulte github.com/Illumina/Pisces/wiki.

Ver dados de amostra e ensaio

- 1 No painel do Local Run Manager, clique no nome do ensaio.
- 2 No separador Run Overview (Descrição geral do ensaio), reveja as métricas do ensaio de sequenciação.
- 3 **[Opcional]** Clique no ícone **Copy to Clipboard**  (Copiar para a área de transferência) para copiar o caminho da pasta de saída do ensaio.
- 4 Clique no separador Sequencing Information (Informações de sequenciação) para rever os parâmetros do ensaio e as informações dos consumíveis.
- 5 Clique no separador Samples and Results (Amostras e resultados) para ver a localização do relatório de análise.
 - ▶ Se a análise tiver sido repetida, expanda a lista pendente Select Analysis (Selecionar análise) e selecione a análise apropriada.
- 6 Clique no ícone **Copy to Clipboard**  (Copiar para a área de transferência) para copiar o caminho da pasta Analysis (Análise).

Para obter mais informações sobre os separadores Run Overview (Descrição geral do ensaio) e Sequencing Information (Informações de sequenciação) e sobre como realinhar análises, consulte o *Manual de referência do instrumento NextSeq 550Dx (documento n.º 1000000009513)*.

Relatório de análise

Os resultados da análise são resumidos no separador Samples and Results (Amostras e resultados) e como relatório agregado na pasta Alignment (Alinhamento). Também está disponível um relatório em formato PDF para cada amostra.

Informações do separador Samples and Results (Amostras e resultados)

- 1 Clique numa amostra da lista para ver o relatório de amostras.

Tabela 1 Informações de ensaios e de amostras

Cabeçalho da coluna	Descrição
Estado do ensaio	Indica se o ensaio de sequenciação foi aprovado ou reprovado.
Produto Total (GB)	Número de bases identificadas num ensaio de sequenciação. Mostra o limiar de passagem e o estado de aprovação ou reprovação.
% ≥ Q30	A percentagem de leituras do ensaio de sequenciação com uma pontuação de qualidade de 30(Q30) ou superior. Mostra o limiar de passagem e o estado de aprovação ou reprovação.
ID da amostra	A ID de amostra fornecida quando o ensaio foi criado.
Total de leituras PF	O número total de leituras aprovadas no filtro.
Leitura 1% ≥ Q30	A percentagem de leituras na Leitura 1 com uma pontuação de qualidade de 30 (Q30) ou superior para a amostra.
Leitura 2% ≥ Q30	A percentagem de leituras na Leitura 2 com uma pontuação de qualidade de 30 (Q30) ou superior para a amostra.
Taxa de identificação de autossomas	O número de posições genómicas dos autossomas (cromossomas de 1 a 22) que cumprem o limiar predefinido do valor de confiança, dividido pelo número total de posições genómicas autossómicas interrogadas. A taxa de identificação é descrita numa base por amostra e reportada como percentagem que é calculada como 1 menos o número das posições autossómicas com identificações incompletas dividido pelo número total de posições autossómicas sequenciadas.

Tabela 2 Informações do relatório de amostras

Cabeçalho da coluna	Descrição
Amostra	A ID de amostra fornecida quando o ensaio foi criado.
Data do relatório	A data em que o relatório foi gerado.
Informações da amostra	O ID da amostra que foi fornecido quando o ensaio foi criado, total de leituras que passaram no filtro na amostra, a percentagem de leituras da amostra com uma pontuação de qualidade de 30(Q30) ou superior e a taxa de identificação de autossomas.
Resumo de amplicons	Número total de regiões de amplicons e tamanho total em pares de bases dos amplicons sequenciados nas regiões alvo, para a amostra e o ficheiro de manifesto. O ficheiro de manifesto especifica o genoma de referência e as regiões de referência alvo utilizadas no passo de alinhamento.
Estatísticas do nível de leitura	Número e percentagem de leituras da amostra que cobrem cada posição na referência, para a Leitura 1 e Leitura 2.

Cabeçalho da coluna	Descrição
Resumo de variantes	Número de SNVs, inserções e eliminações detetados para a amostra que passaram os valores sugeridos para determinar se os resultados de qualidade estão dentro de um intervalo aceitável.
Resumo de cobertura	Número total de bases alinhadas a dividir pelo tamanho da região alvo, e a percentagem de regiões de amplicons com valores de cobertura superiores ao limiar de cobertura inferior de 0,2 * de cobertura média de amplicons, para a amostra.
Traçados de cobertura	O traçado de cobertura de regiões de amplicons mostra a cobertura nas regiões de amplicons da amostra. As regiões com valores de cobertura inferiores ao limiar de cobertura são realçadas a vermelho. A média de todos os valores é indicada por uma linha laranja.
Versões de software	Versões de software quando a amostra foi sequenciada. Inclui o NextSeq 550Dx Operating Software (NOS), o Software Local Run Manager, o Software RTA e a versão do módulo Germline Variant.

Ficheiros de saída de análise

Os seguintes ficheiros de saída de análise são gerados para o módulo de análise Germline Variant e fornecem resultados de análise para alinhamento e identificação de variantes. Os ficheiros de saída de análise estão localizados na pasta Alignment (Alinhamento).

Nome de ficheiro	Descrição
Desmultiplexagem (*.txt)	Ficheiros intermédios que contêm o resumo dos resultados de desmultiplexagem.
FASTQ (*.fastq.gz)	Ficheiros intermédios que contêm identificações de base de pontuação de qualidade. Os ficheiros FASTQ são a principal entrada para o passo de alinhamento.
Ficheiros de alinhamento em formato BAM (*.bam)	Contêm leituras alinhadas de uma determinada amostra.
Ficheiros de identificação de variantes no formato VCF genómico (*.genome.vcf.gz)	Contêm o genótipo de cada posição, que tenha sido identificada como variante ou como referência.
Ficheiros de identificação de variantes no formato VCF (*.vcf.gz)	Contêm todas as variantes identificadas na região alvo.
AmpliconCoverage_M1.tsv	Contêm informações sobre a cobertura por amplicon, por amostra, de cada manifesto fornecido. M# representa o número do manifesto.

Formato de ficheiro de desmultiplexagem

O processo de desmultiplexagem faz a leitura da sequência de índice ligada a cada cluster para determinar qual é a amostra de origem do cluster. O mapeamento entre clusters e número de amostra é gravado num ficheiro de desmultiplexagem (*.demux) de cada secção da célula de fluxo.

O formato de nomes do ficheiro de desmultiplexagem é **s_1_X.demux**, em que X corresponde ao número da secção.

Os ficheiros de desmultiplexagem começam por um cabeçalho:

- ▶ Versão (inteiro de 4 bytes), atualmente 1
- ▶ Contagem de clusters (inteiro de 4 bytes)

O resto do ficheiro é composto por números de amostra de cada cluster da secção.

Quando o passo de desmultiplexagem estiver concluído, o software gera um ficheiro de desmultiplexagem chamado **DemultiplexSummaryF1L1.txt**.

- ▶ No nome do ficheiro, **F1** representa o número da célula de fluxo.
- ▶ No nome do ficheiro, **L1** representa o número da faixa.
- ▶ A desmultiplexagem cria uma tabela com 1 linha por secção e 1 coluna por amostra, incluindo a amostra 0.
- ▶ As sequências que ocorrem mais frequentemente nas leituras de índice.

Formato de ficheiro FASTQ

FASTQ é um formato de ficheiro baseado em texto, que contém identificações de base e valores de qualidade por leitura. Cada registo contém 4 linhas:

- ▶ O identificador
- ▶ A sequência
- ▶ Um sinal de adição (+)
- ▶ As pontuações de qualidade Phred em formato com codificação ASCII + 33

O identificador está formatado como:

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

Exemplo:

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

Formato de ficheiro BAM

Um ficheiro BAM (*.bam) é a versão binária comprimida de um ficheiro SAM que é utilizada para representar sequências alinhadas até 128 Mb. Os formatos SAM e BAM são descritos em detalhe no samtools.github.io/hts-specs/SAMv1.pdf.

Os ficheiros BAM utilizam o formato de nome de ficheiro **SampleName_S#.bam**, em que # é o número da amostra determinado pela ordem em que as amostras são apresentadas para o ensaio.

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

- ▶ **Cabeçalho**—Contém informações sobre todo o ficheiro, como 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.
- ▶ **Alinhamentos**—Contém o nome da leitura, a sequência da leitura, a qualidade da leitura, informações de alinhamento e tags personalizadas. O nome da leitura inclui o cromossoma, as coordenadas iniciais, a qualidade do alinhamento e a cadeia do descritor de correspondência.

A secção de alinhamentos inclui as seguintes informações de cada leitura ou par de leituras:

- ▶ **AS:** qualidade de alinhamento de extremidade emparelhada.
- ▶ **RG:** grupo de leitura, que indica o número de leituras relativas a uma amostra específica.
- ▶ **BC:** tag de código de barras, que indica o ID de amostra desmultiplexada associado à leitura.
- ▶ **SM:** qualidade de alinhamento de extremidade individual.

- ▶ **XC:** cadeia do descritor de correspondência.
- ▶ **XN:** o tag do nome do amplicon, que regista o ID do amplicon associado à leitura.

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

Formato de ficheiro VCF

O VCF (Variant Call Format) é um formato de ficheiro comum desenvolvido pela comunidade científica de genómica. Contém informações sobre variantes encontradas em posições específicas num genoma de referência. Os ficheiros VCF acabam com o sufixo .vcf

O cabeçalho do ficheiro VCF inclui a versão do formato de ficheiro VCF e a versão do identificador de variantes e apresenta as anotações utilizadas no resto do ficheiro. O cabeçalho do VCF também inclui o ficheiro de genoma de referência e o ficheiro BAM. A última linha do cabeçalho contém os cabeçalhos de coluna para as linhas de dados. Cada uma das linhas de dados do ficheiro VCF contém informações sobre uma variante.

Cabeçalhos do ficheiro VCF

Cabeçalho	Descrição
CHROM	O cromossoma do genoma de referência. Os cromossomas são apresentados pela mesma ordem que o ficheiro de referência FASTQ.
POS	A posição de base única da variante no cromossoma de referência. No caso dos SNPs, esta posição é a base de referência com a variante; no caso de Indels ou eliminações, esta posição é a base de referência imediatamente anterior à variante.
ID	O número rs da variante obtida do dbSNP.txt, se aplicável. Se houver múltiplos números rs nesta localização, a lista é delimitada por ponto e vírgula. Se não houver uma entrada dbSNP nesta posição, é utilizado um marcador de valor em falta ('.') .
REF	O genótipo de referência. Por exemplo, a eliminação de um T único é representada como TT de referência e um T alternativo. Uma variante única de nucleótido de A a T é representada como A de referência e T alternativo.
ALT	Os alelos que diferem da leitura de referência. Por exemplo, a inserção de um T único é representada como um A de referência e um AT alternativo. Uma variante única de nucleótido de A a T é representada como A de referência e T alternativo.
QUAL	Uma pontuação de qualidade com escala Phred atribuída pelo identificador da variante. As pontuações mais altas indicam uma maior confiança na variante e uma menor probabilidade de erros. Para uma pontuação de qualidade de Q, a probabilidade estimada de um erro é $10^{-(Q/10)}$. 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 seus modelos estatísticos, que são elevados em relação à taxa de erro observada.

Anotações do ficheiro VCF

Cabeçalho	Descrição
FILTER (FILTRO)	<p>Se todos os filtros passarem, será apresentada a indicação PASS (APROVADO) na coluna Filter (Filtro).</p> <ul style="list-style-type: none"> • LowDP—Aplicado aos locais com uma profundidade de cobertura inferior a 150x. Nas posições de amplicons abrangidas por leituras avançadas e inversas, isto equivale a 300 leituras de extremidade emparelhada sobrepostas. • q20—Pontuação de qualidade < 20. • MultiAllelicSite—A variante não está em conformidade com o modelo diploide. • R5x9—O número de repetições adjacentes (com tamanho 1 a 5 bp) às identificações de variantes ≥ 9. • SB—A tendência de cadeia é superior ao limiar especificado.
INFO (INFORMAÇÕES)	<p>As possíveis entradas na coluna INFO (Informações) incluem:</p> <ul style="list-style-type: none"> • AC—Contagem de alelos nos genótipos para cada alelo ALT, pela mesma ordem que são listados. • AF—Frequência de alelos para cada alelo ALT, pela mesma ordem que são listados. • AN—O número total de alelos nos genótipos identificados. • CD—Um sinalizador indica que o SNP ocorre na região de codificação de pelo menos 1 entrada RefGene. • DP—A profundidade (número de identificações de bases alinhadas numa posição e utilizadas na identificação de variantes). • Exon—Uma lista separada por vírgulas da leitura das regiões exon do RefGene. • FC—Consequência funcional. • GI—Uma lista separada por vírgulas das leituras dos IDs de genes do RefGene. • QD—Confiança da variante/Qualidade por profundidade. • TI—Uma lista separada por vírgulas da leitura dos IDs de transcrição do RefGene.
FORMAT (FORMATO)	<p>A coluna Format (Formato) apresenta uma lista dos campos separados por dois pontos. Por exemplo, GT:GQ. A lista dos campos fornecidos depende do identificador de variantes utilizado. Os campos disponíveis incluem:</p> <ul style="list-style-type: none"> • AD—Entrada do formulário X,Y, em que X corresponde ao número de identificações de referência e Y corresponde ao número de identificações alternativas. • DP—Profundidade de leitura aproximada; são filtradas leituras com MQ=255 ou parceiros incorretos. • GQ—Qualidade do genótipo. • GT—Genótipo. 0 corresponde à base de referência, 1 corresponde à primeira entrada da coluna ALT, e assim por diante. A barra (/) indica que não estão disponíveis informações de faseamento. • NC—Fração de bases que não foram identificadas ou com qualidade de identificação de bases inferior ao limite mínimo. • NL—Nível de ruído; uma estimativa do ruído da identificação de bases nesta posição. • SB—Tendência de cadeia nesta posição. Valores negativos maiores indicam menos tendência; valores próximos de 0 indicam mais tendência. • VF—Frequência de variantes; a percentagem de leituras que suportam o alelo alternativo.
SAMPLE (AMOSTRA)	A coluna Sample (Amostra) inclui os valores especificados na coluna FORMAT (Formato).

Ficheiros VCF genómico

Os ficheiros VCF genómico (gVCF) são ficheiros VCF v4.1 que seguem um conjunto de convenções para representar todos os locais no genoma num formato razoavelmente compacto. Os ficheiros gVCF (*.genome.vcf.gz) incluem todos os locais na região de interesse num único ficheiro para cada amostra.

Os ficheiros gVCF mostram não-identificações nas posições que não passam todos os filtros. Uma identificação de genótipo (GT) de ./ indica uma não-identificação.

Para obter mais informações, consulte sites.google.com/site/gvcftools/home/about-gvcf.

Ficheiro de cobertura amplicon

É gerado um ficheiro de cobertura amplicon por cada ficheiro de manifesto. O M# no nome do ficheiro representa o número do manifesto.

Cada ficheiro inclui uma linha de cabeçalho que contém os IDs da amostra associados ao manifesto. O ficheiro contém as seguintes informações.

- ▶ O ID alvo conforme apresentado no manifesto.
- ▶ A profundidade de cobertura das leituras aprovadas no filtro.

Ficheiros de saída complementares

Os seguintes ficheiros de saída fornecem informações complementares ou um resumo dos resultados e erros de análise. Embora estes ficheiros não sejam necessários para avaliar os resultados da análise, estes ficheiros podem ser utilizados para efeitos de resolução de problemas. Todos os ficheiros estão localizados na pasta Alignment (Alinhamento) a menos que especificado em contrário.

Nome de ficheiro	Descrição
AnalysisLog.txt	Registo de processamento que descreve cada passo que ocorreu durante a análise da pasta do ensaio atual. Este ficheiro não contém mensagens de erro. Localizado na pasta Alignment (Alinhamento).
AnalysisError.txt	Registo de processamento que indica todos os erros que ocorreram durante a análise. Este ficheiro estará vazio se não tiverem ocorrido erros. Localizado na pasta Alignment (Alinhamento).
DemultiplexSummaryF1L1#.txt	Reporta os resultados de desmultiplexagem numa tabela com 1 linha por secção e 1 coluna por amostra. O # representa as faixas 1, 2, 3 ou 4 da célula de fluxo. Localizado na pasta Alignment (Alinhamento).
AmpliconRunStatistics.xml	Contém as estatísticas de resumo específicas do ensaio. Localizado na pasta Alignment (Alinhamento).

Pasta Analysis (Análise)

A pasta de análise inclui os ficheiros gerados pelo software Local Run Manager.

A relação entre a pasta de saída e a pasta de análise é resumida da seguinte forma:

- ▶ Durante a sequenciação, o RTA (Real-Time Analysis) coloca na pasta de saída os ficheiros gerados durante a análise de imagens, a identificação de bases e a pontuação de qualidade.
- ▶ O RTA copia os ficheiros para a pasta de análise em tempo real. Depois de o RTA atribuir uma pontuação de qualidade a cada base para cada ciclo, o software grava o ficheiro RTAComplete.txt em ambas as pastas.
- ▶ Quando o ficheiro RTAComplete.txt está presente, a análise é iniciada.
- ▶ À medida que a análise continua, o Local Run Manager grava os ficheiros de saída na pasta de análise e, em seguida, copia novamente os ficheiros para a pasta de saída.

Pasta Alignment (Alinhamento)

Sempre que as filas de análise forem realinhadas, o Local Run Manager cria uma pasta de alinhamento com o nome **Alignment_N**, em que N corresponde a um número sequencial.

Estrutura de pastas

- 📁 **Alignment**—Contém ficheiros *.bam, *.vcf, FASTQ e ficheiros específicos do módulo de análise.
 - 📁 **Date and Time Stamp**—Carimbo de data e hora da análise, como AAAAMMDD_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**—Contém ficheiros *.bcl.
 - 📁 **L001**—Contém ficheiros *.locs.
 - 📁 **RTA Logs**—Contém ficheiros de registo do software de análise RTA.
- 📁 **InterOp**—Contém ficheiros binários utilizados para reportar métricas de ensaios de sequenciação.
- 📁 **Logs**—Contém ficheiros de registo com a descrição dos passos executados durante a sequenciação.
 - 📄 RTAComplete.txt
 - 📄 RunInfo.xml
 - 📄 RunParameters.xml

Identificação de bases e diversidade de índices

Quando as amostras são sequenciadas no instrumento NextSeq 550Dx, a identificação de bases determina a base (A, C, G ou T) de cada cluster de uma determinada secção ou área de imagiologia na célula de fluxo num ciclo específico. O instrumento NextSeq 550Dx utiliza a sequenciação de dois canais, que requer apenas duas imagens para codificar os dados de quatro bases de ADN, uma do canal vermelho e outra do canal verde.

O processo das leituras de índice de identificação de bases é diferente da identificação de bases durante outras leituras.

As leituras de índice têm de começar por pelo menos uma base diferente de G, num dos primeiros dois ciclos. Se uma leitura de índice começar por duas identificações de base de G, não será gerada qualquer intensidade de sinal. O sinal tem de estar presente num dos primeiros dois ciclos para garantir o desempenho da desmultiplexagem.

Ao seleccionar os índices durante a criação de ensaios, será apresentado um aviso de baixa diversidade se os índices não cumprirem os requisitos de diversidade. Para evitar um aviso de baixa diversidade, seleccione as sequências de índice que fornecem o sinal em ambos os canais de cada ciclo.

- ▶ Canal vermelho—A ou C
- ▶ Canal verde—A ou T

Este processo de identificação de bases assegura a precisão ao analisar amostras de baixa-plexagem. Para obter mais informações sobre as sequências dos seus índices, consulte o folheto informativo do *Kit TruSeq Custom Amplicon Dx* (documento n.º 1000000029772).

Durante a criação de ensaios no Local Run Manager irá escolher o número de amostras para testar. As sugestões de combinação de índices que cumprem os requisitos de diversidade de índice são preenchidas automaticamente pelo software. Embora não seja necessário utilizar as sugestões de combinação de índices, recomenda-se que o faça.

Histórico de revisão

Documento	Data	Descrição da alteração
Documento n.º 1000000030329 v04	Agosto de 2021	Foi atualizada a morada do Representante autorizado na UE.
Documento n.º 1000000030329 v03	Abril de 2020	Foi atualizada a morada do Representante autorizado na UE. Foi atualizada a morada do Patrocinador australiano.
Documento n.º 1000000030329 v02	Janeiro de 2019	Foram adicionadas informações sobre os kits de reagentes v2.5.
Documento n.º 1000000030329 v01	Agosto de 2018	Foram atualizadas as marcações regulamentares.
Documento n.º 1000000030329 v00	Novembro de 2017	Edição inicial.

Assistência técnica

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

Sítio Web: www.illumina.com
E-mail: techsupport@illumina.com

Números de telefone do Apoio ao Cliente da Illumina

Região	Número gratuito	Regional
América do Norte	+1.800.809.4566	
Alemanha	+49 8001014940	+49 8938035677
Austrália	+1.800.775.688	
Áustria	+43 800006249	+43 19286540
Bélgica	+32 80077160	+32 34002973
China	400.066.5835	
Dinamarca	+45 80820183	+45 89871156
Espanha	+34 911899417	+34 800300143
Finlândia	+358 800918363	+358 974790110
França	+33 805102193	+33 170770446
Hong Kong	800960230	
Irlanda	+353 1800936608	+353 016950506
Itália	+39 800985513	+39 236003759
Japão	0800.111.5011	
Noruega	+47 800 16836	+47 21939693
Nova Zelândia	0800.451.650	
Países Baixos	+31 8000222493	+31 207132960
Reino Unido	+44 8000126019	+44 2073057197
Singapura	+1.800.579.2745	
Suécia	+46 850619671	+46 200883979
Suíça	+41 565800000	+41 800200442
Taiwan	00806651752	
Outros países	+44.1799.534000	

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

Documentação do produto — Disponível para transferência em PDF a partir do sítio Web da Illumina. Aceda a support.illumina.com, selecione um produto e, em seguida, selecione **Documentation & Literature** (Documentação e literatura).



Illumina
5200 Illumina Way
San Diego, California 92122 EUA
+1.800.809.ILMN (4566)
+1.858.202.4566 (fora da América do Norte)
techsupport@illumina.com
www.illumina.com

CE



Illumina Netherlands B.V.
Steenoven 19
5626 DK Eindhoven
Países Baixos

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

PARA UTILIZAÇÃO EM DIAGNÓSTICO IN VITRO

© 2021 Illumina, Inc. Todos os direitos reservados.

illumina®