

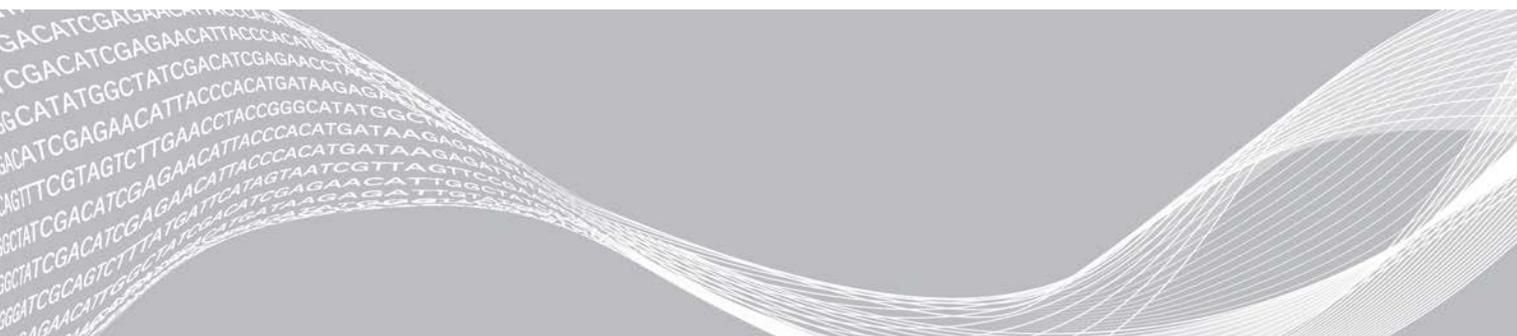
Local Run Manager

Módulo de análise Somatic Variant (Variante Somática)

Guia de fluxo de trabalho do NextSeq 550Dx

PARA UTILIZAÇÃO DE DIAGNÓSTICO IN VITRO

Visão geral	3
Inserir informações de execução	4
Métodos de análise	6
Exibir dados de execução e amostra	7
Relatório de análise	8
Arquivos de resultados de análises	9
Identificação de bases e diversidade de índices	16
Histórico de revisões	17
Assistência técnica	18



Este documento e seu conteúdo são propriedade da Illumina, Inc. e de suas afiliadas (“Illumina”), e destinam-se exclusivamente ao uso contratual de seu cliente com relação ao uso dos produtos descritos neste documento e para nenhuma outra finalidade. Este documento e seu conteúdo não devem ser usados ou distribuídos para nenhuma outra finalidade nem comunicados, divulgados ou reproduzidos de nenhuma forma sem o consentimento prévio por escrito da Illumina. A Illumina não concede nenhuma licença sob seus direitos de patente, marca registrada, direitos autorais ou lei comum, nem direitos semelhantes de terceiros por meio deste documento.

As instruções neste documento devem ser estrita e explicitamente seguidas por pessoal devidamente treinado e qualificado para garantir o uso adequado e seguro dos produtos descritos neste documento. Todo o conteúdo deste documento deve ser inteiramente lido e entendido antes da utilização de tais produtos.

NÃO LER COMPLETAMENTE E NÃO SEGUIR EXPLICITAMENTE TODAS AS INSTRUÇÕES AQUI CONTIDAS PODE RESULTAR EM DANOS AO(S) PRODUTO(S), FERIMENTOS A PESSOAS, INCLUSIVE USUÁRIOS OU OUTROS, E DANOS A OUTROS BENS, ANULANDO TODA GARANTIA APLICÁVEL AO(S) PRODUTO(S).

A ILLUMINA NÃO SE RESPONSABILIZA POR QUALQUER PROBLEMA CAUSADO PELO USO INDEVIDO DO(S) PRODUTO(S) MENCIONADO(S) ACIMA (INCLUINDO PARTES SEPARADAS OU SOFTWARE).

© 2021 Illumina, Inc. Todos os direitos reservados.

Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.

Visão geral

O módulo Somatic Variant (Variante Somática) do Local Run Manager é destinado ao uso com o ensaio do Índice do kit para Amplicon personalizado Illumina TruSeq e o NextSeq 550Dx. Quando usado com o módulo Somatic Variant (Variante Somática), o ensaio é destinado a preparar bibliotecas usadas para o sequenciamento de DNA de tecido fixado em formol e incluído em parafina (FFPE, Formalin-Fixed, Paraffin-Embedded). O ensaio detecta mutações somáticas em baixas frequências de variante.

O módulo de análise avalia pequenas regiões de DNA amplificado, ou Amplicons, para variantes.

O sequenciamento específico de Amplicons possibilita alta cobertura de regiões particulares em um grande número de amostras. O módulo de análise realiza a análise secundária e a geração de relatórios a partir do sequenciamento de execuções usando uma abordagem de quebra dupla que envolve pools de oligo avançado e reverso. Consulte o folheto informativo do *Índice do kit para Amplicon personalizado TruSeq* (documento n.º 1000000029772).

O módulo de análise Somatic Variant (Variante Somática) requer consumíveis para sequenciamento em 300 ciclos. Para obter mais informações, consulte o folheto informativo do *Kit de reagentes de alta produção NextSeq 550Dx v2* ou do *Kit de reagentes de alta produção NextSeq 550Dx v2.5*.

Sobre este guia

Este guia dá instruções para configurar os parâmetros de sequenciamento e análise do módulo de análise de Variante Somática. Para obter informações sobre o painel do Local Run Manager e as configurações do sistema, consulte o *Guia de referência do instrumento NextSeq 550Dx* (documento n.º 1000000009513).

Visualização do Local Run Manager

A interface do Local Run Manager é exibida dentro do NextSeq 550Dx Operating Software (NOS) ou por um navegador da Web. O navegador da Web compatível é o Chromium.



OBSERVAÇÃO

Se você usar um navegador incompatível, faça download do navegador compatível quando for avisado pela mensagem "Confirm Unsupported Browser" (Confirme navegador incompatível). Selecione "**here**" (aqui) para fazer download da versão compatível do Chromium.

Visualizar no monitor do instrumento

- 1 Para exibir a interface do Local Run Manager no monitor do instrumento, selecione uma das seguintes opções:
 - ▶ Na tela Home (Página inicial) do NOS, selecione **Local Run Manager**. Clique no X no canto superior direito para retornar ao NOS quando terminar.
 - ▶ Selecione o ícone Minimize NOS (Minimizar NOS), abra o navegador Chromium que está no instrumento e digite **http://localhost** na barra de endereços. Só os usuários Admin podem minimizar o NOS.

Visualizar a partir de um computador em rede

- 1 Abra o navegador Chromium em um computador com acesso à mesma rede do instrumento e conecte usando o endereço IP do instrumento ou o nome deste. Por exemplo, **http://myinstrument**.

Inserir informações de execução

Definir parâmetros

- 1 Faça login no Local Run Manager.
- 2 Selecione **Create Run** (Criar execução) e depois **Somatic Variant** (Variante Somática).
- 3 Digite um nome de execução que identifique a execução do sequenciamento por meio da análise. Use caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 4 **[Opcional]** Digite uma descrição da execução para ajudar a identificá-la. Use caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 5 Selecione o número de amostras e o conjunto do índice a partir da lista suspensa. Leve em consideração as seguintes informações quando fizer uma seleção.
 - ▶ A lista suspensa contém números de amostras com um conjunto do índice. Por exemplo, 24-Set 1 indica 24 amostras a serem testadas com índices do conjunto do índice 1.
 - ▶ Os números do conjunto de índices referem-se a diferentes conjuntos de índices i5. O Conjunto 1 e o Conjunto 2 fornecem, ambos, a diversidade de índice. Dois conjuntos de índices são oferecidos para ajudar a evitar a exaustão de um conjunto único.
 - ▶ Escolha o número de amostras que está mais próximo do número de amostras em teste. Se o número exato de amostras não estiver na lista, selecione o número mais próximo, mas menor que o número que você está testando. Por exemplo, se quiser testar 18 amostras, selecione 16 amostras.
 - ▶ Poços de amostras e combinações de índices que atendem aos requisitos de diversidade de índice são destacados em verde. Se você selecionar outros poços e combinações de índices, quando salvar a execução, você receberá uma notificação se os requisitos de diversidade de índice não forem atendidos.

Importar arquivos de manifesto para a execução

- 1 Verifique se os manifestos que você deseja importar estão disponíveis em um local de rede acessível ou em uma unidade USB.
- 2 Selecione **Import Manifests** (Importar manifestos).
- 3 Navegue até o arquivo do manifesto e selecione o manifesto que você deseja adicionar.



OBSERVAÇÃO

Para disponibilizar os arquivos de manifesto para todas as execuções usando o módulo de análise Somatic Variant (Variante Somática), adicione manifestos usando o recurso Module Settings (Configurações do módulo). Esse recurso requer permissões em nível de usuário admin. Para obter mais informações, consulte o *Guia de referência do instrumento NextSeq 550Dx (documento n.º 1000000009513)*.

Especificar as amostras para a execução

Especifique as amostras para a execução usando uma das opções e as instruções a seguir.

- ▶ **Enter samples manually** (Inserir as amostras manualmente) — Use a tabela em branco na tela Create Run (Criar execução).

- ▶ **Import samples** (Importar amostras) — Navegue até o arquivo externo em um formato de valores separados por vírgulas (*.csv). Está disponível um modelo para fazer download na tela Create Run (Criar execução).

Depois de preencher a tabela de amostras, você pode exportar as informações de amostra para um arquivo externo. Use o arquivo como referência ao preparar bibliotecas ou importar o arquivo para outra execução.

Inserir as amostras manualmente

- 1 Informar um nome exclusivo da amostra no campo Sample Name (Nome da amostra).
Use caracteres alfanuméricos, sublinhados ou travessões.
O nome da amostra preenche automaticamente o poço correspondente no outro pool.
- 2 **[Opcional]** Para amostras de controle positivas ou negativas, clique com o botão direito e selecione o tipo de controle. O controle em um poço de amostra preenche automaticamente o poço correspondente no outro pool com o mesmo controle.
- 3 **[Opcional]** Informe uma descrição para a amostra no campo Sample Description (Descrição da amostra).
Use caracteres alfanuméricos, sublinhados ou travessões.
A descrição da amostra preenche automaticamente o poço correspondente no outro pool.
As descrições das amostras são associadas ao ID das amostras. As descrições das amostras são substituídas se o mesmo ID de amostra é usado em uma execução posterior.
- 4 Selecione um adaptador de Índice 1 na lista suspensa Index 1 (i7) (Índice 1 (i7)).
Quando você usar os poços sugeridos de amostras, o software preencherá automaticamente os adaptadores de índice i7 e i5 que atenderem aos requisitos de índice de diversidade. Se o número exato de amostras que você está testando não estiver na lista, certifique-se de selecionar adaptadores de índice para poços extras. Se precisar selecionar índices para poços extras, ou se você não usa as combinações recomendadas de adaptadores de índice, antes de escolher os índices, certifique-se de ler *Identificação de bases e diversidade de índices na página 16*.
- 5 Selecione um adaptador de Índice 2 na lista suspensa Index 2 (i5) (Índice 2 (i5)).
- 6 Selecione um arquivo de manifesto na lista suspensa Manifest (Manifesto).
As amostras do Pool A exigem um manifesto diferente das amostras no Pool B.
- 7 Escolha uma opção para exibir, imprimir ou salvar o layout da placa como referência para preparar bibliotecas:
 - ▶ Selecione o ícone  **Print** (Imprimir) para exibir o layout da placa. Selecione **Print** (Imprimir) para imprimir o layout da placa.
 - ▶ Selecione **Export** (Exportar) para exportar as informações da amostra para um arquivo externo. Certifique-se de que as informações das amostras e do manifesto estejam corretas. Informações incorretas podem afetar os resultados.
- 8 Selecione **Save Run** (Salvar execução).

Importar amostras

- 1 Selecione **Import Samples** (Importar amostras) e procure o local do arquivo de informações da amostra. Há dois tipos de arquivos que você pode importar.
 - ▶ Selecione **Template** (Modelo) na tela Create Run (Criar execução) para fazer um novo layout da placa. O arquivo do modelo contém os cabeçalhos corretos da coluna para importação. Insira as informações da amostra em cada coluna para as amostras da execução. Exclua as informações de exemplo nas células não utilizadas e depois salve o arquivo.

- ▶ Use um arquivo de informações da amostra que foi exportado do módulo de Variante Somática com o recurso Export (Exportar).
- 2 Selecione o ícone  **Print** (Imprimir) para exibir o layout da placa.
- 3 Selecione **Print** (Imprimir) para imprimir o layout da placa como referência para preparar as bibliotecas.
- 4 **[Opcional]** Selecione **Export** (Exportar) para exportar informações das amostras para um arquivo externo. Certifique-se de que as informações das amostras e do manifesto estejam corretas. Informações incorretas podem afetar os resultados.
- 5 Selecione **Save Run** (Salvar execução).

Editar uma execução

Para obter instruções sobre a edição das informações em sua execução antes do sequenciamento, consulte o *Guia de referência do instrumento NextSeq 550Dx (documento nº 1000000009513)*.

Métodos de análise

O módulo de análise de Variante Somática realiza as seguintes etapas de análise e depois grava os arquivos de resultados de análise na pasta Alignment (Alinhamento).

- ▶ Demultiplexa leituras de índices
- ▶ Gera arquivos FASTQ
- ▶ Alinha a uma referência
- ▶ Identifica variantes

Demultiplexação

A demultiplexação compara cada sequência de leitura de índice com as sequências de índice especificadas para a execução. Não são considerados valores de qualidade nesta etapa.

As leituras de índice são identificadas com o uso das seguintes etapas:

- ▶ As amostras são numeradas começando por 1 com base na ordem em que estão relacionadas para a execução.
- ▶ O número de amostras 0 é reservado para clusters que não foram atribuídos a uma amostra.
- ▶ Os clusters são atribuídos a uma amostra quando a sequência de índice corresponder exatamente ou quando houver até uma única discrepância por leitura de índice.

Geração de arquivos FASTQ

Depois da demultiplexação, o software gera arquivos intermediários de análise no formato FASTQ, que é um formato de texto usado para representar sequências. Os arquivos FASTQ contêm leituras de cada amostra e das pontuações de qualidade associadas. Os clusters que não passaram pelo filtro são excluídos.

Cada arquivo FASTQ contém leituras de apenas uma amostra, e o nome dessa amostra é incluído no nome do arquivo FASTQ. Os arquivos FASTQ são a entrada primária do alinhamento. Oito arquivos FASTQ são gerados por amostra por pool de oligo, quatro da Leitura 1 e quatro da Leitura 2, resultando em um total de 16 arquivos FASTQ por amostra.

Alinhamento

Durante a etapa de alinhamento, o algoritmo de Smith-Waterman com faixas alinha clusters de cada amostra em relação a sequências de amplicon especificadas no arquivo de manifesto.

O algoritmo de Smith-Waterman com faixas realiza alinhamentos de sequência semi-globais para determinar regiões similares entre duas sequências. Em vez de comparar a sequência total, o algoritmo de Smith-Waterman compara segmentos de todas as durações possíveis.

Cada leitura do tipo paired-end é avaliada em termos de alinhamento com as sequências relevantes da sonda daquela leitura.

- ▶ A Leitura 1 é avaliada em relação ao complemento inverso do Downstream Locus-Specific Oligos (DLSO).
- ▶ A Leitura 2 é avaliada em relação aos oligos específicos de lugar a montante (ULSO, Upstream Locus-Specific Oligos).
- ▶ Se o início de uma leitura corresponde a uma sequência de sondagem com no máximo três diferenças (disparidades e alterações devidas a indels de alinhamento), o tamanho total da leitura é alinhado em relação ao destino do Amplicon para essa sequência.
- ▶ Os indels dentro do DLSO e do ULSO não são observados devido à química do ensaio.

Os alinhamentos são filtrados dos resultados do alinhamento com base nas taxas de discrepância sobre a região de interesse ou do amplicon total, dependendo da duração do amplicon. Os alinhamentos filtrados são gravados nos arquivos de alinhamento como não alinhados e não são usados na identificação de variantes.

Chamada de variante

Desenvolvido pela Illumina, o chamador de variante Pisces identifica variantes presentes em baixa frequência na amostra de DNA.

O chamador de variante Pisces identifica variantes de um único nucleotídeo, variantes em múltiplos nucleotídeos e pequenos indels em três etapas:

- ▶ Considera cada posição no genoma de referência separadamente
- ▶ Conta bases em uma determinada posição para leituras alinhadas que coincidem com a posição
- ▶ Computa uma pontuação de variante que mede a qualidade da chamada usando o modelo Poisson. Variantes com uma pontuação de qualidade abaixo de Q30 são excluídas.

As variantes primeiro são chamadas para cada pool separadamente. Depois, as variantes de cada pool são comparadas e combinadas em um só arquivo de resultados. Se houver uma variante presente em ambos os pools e ela passar em todos os filtros listados em *Anotações de arquivos VCF na página 12*, a variante será marcada como PASS (APROVADA) no arquivo de chamada de variante (VCF, Variant Call File).

Exibir dados de execução e amostra

- 1 No painel de controle do Local Run Manager, clique no nome de uma execução.
- 2 Na guia Run Overview (Visão geral da execução), analise as medidas de execução de sequenciamento.
- 3 **[Opcional]** Clique no ícone **Copy to Clipboard**  (Copiar para área de transferência) para copiar o caminho da pasta com os resultados da execução.
- 4 Clique na guia Sequencing Information (Informações de sequenciamento) para analisar os parâmetros de execução e as informações dos consumíveis.

- 5 Clique na guia Samples and Results (Amostras e resultados) para exibir o local do relatório de análise.
 - ▶ Se a análise tiver sido repetida, expanda a lista suspensa Select Analysis (Selecionar análise) e selecione a análise apropriada.
- 6 Clique no ícone **Copy to Clipboard**  (Copiar para área de transferência) para copiar o caminho da pasta Analysis (Análise).

Para obter mais informações sobre as guias Run Overview (Visão geral da execução) e Sequencing Information (Informações de sequenciamento) e sobre como enfileirar novamente a análise, consulte o *Guia de referência do instrumento NextSeq 550Dx (documento n.º 1000000009513)*.

Relatório de análise

Os resultados de análise são resumidos na guia Samples and Results (Amostras e resultados) e como um relatório agregado na pasta Alignment (Alinhamento). Um relatório para cada amostra também está disponível em um arquivo em formato PDF para cada amostra.

Informações na guia Samples and Results (Amostras e resultados)

- 1 Clique em uma amostra na lista para ver o relatório da amostra.

Tabela 1 Informações das execuções e amostras

Cabeçalho da coluna	Descrição
Status de execução	Declara se a execução do sequenciamento foi aprovada ou reprovada.
Rendimento total (GB)	Número de bases chamadas na execução do sequenciamento. Mostra o limite de aprovação e o status aprovado ou reprovado.
% ≥ Q30	O percentual de leituras na execução do sequenciamento com uma pontuação de qualidade de 30(Q30) ou superior. Mostra o limite de aprovação e o status aprovado ou reprovado.
Sample Name (Nome da amostra)	O nome da amostra fornecido quando a execução foi criada.
Total PF Reads (Total de leituras aprovadas no filtro)	O número total de leituras que são aprovadas no filtro.
Read 1% ≥ Q30 (Leitura 1% ≥ Q30)	O percentual de leituras na Leitura 1 com uma pontuação de qualidade de 30 (Q30) ou superior para a amostra.
Read 2% ≥ Q30 (Leitura 1% ≥ Q30)	O percentual de leituras na Leitura 2 com uma pontuação de qualidade de 30 (Q30) ou superior para a amostra.
Autosome Call Rate (taxa de chamada de autossomo)	O número de posições genômicas nos autossomos (cromossomos 1 a 22) que atendem a um limite de valor de segurança predefinido, dividido pelo número total de posições genômicas autossômicas interrogadas. A taxa de chamada é descrita por amostra e informada como um percentual calculado como 1 menos o número de posições autossômicas com chamadas incompletas dividido pelo número total de posições autossômicas sequenciadas.

Tabela 2 Informações do relatório de amostras

Cabeçalho da coluna	Descrição
Sample (Amostra)	O nome da amostra fornecido quando a execução foi criada.
Report Date (Data do relatório)	Data em que o relatório foi gerado.

Cabeçalho da coluna	Descrição
Sample Information (Informações da amostra)	O ID da amostra que foi fornecido quando a execução foi criada, total de leituras que passaram no filtro na amostra, o percentual de leituras para a amostra com uma pontuação de qualidade de 30(Q30) ou mais e a taxa de chamada autossômica.
Amplicon Summary (Resumo do Amplicon)	Número total de regiões sequenciadas do Amplicon e o comprimento total em pares de base de Amplicons sequenciados nas regiões de destino, para a amostra no Pool A e no Pool B, bem como o arquivo de manifesto usado para cada pool. O arquivo de manifesto especifica o genoma de referência e as regiões de referência de destino usadas na etapa de alinhamento.
Read Level Statistics (Estatísticas de nível de leitura)	Número e percentual de leituras para a amostra que cobrem cada posição na referência, para a Leitura 1 e a Leitura 2 no Pool A e no Pool B.
Variants Summary (Resumo de variantes)	Número de variantes de um único nucleotídeo, inserções e deleções detectadas para a amostra aprovada nos valores sugeridos para determinar se os resultados de qualidade estão dentro de uma faixa aceitável
Coverage Summary (Resumo de cobertura)	O número total de bases alinhadas dividido pelo tamanho da região de destino, além da porcentagem de regiões do Amplicon com valores de cobertura superiores ao limite de baixa cobertura de cobertura média do Amplicon de 0,2 *, para a amostra no Pool A e no Pool B.
Coverage Plots (Gráficos de cobertura)	O gráfico de Cobertura por região do Amplicon mostra a cobertura em regiões do Amplicon para a amostra. Regiões com valores de cobertura inferiores ao limite são destacadas em vermelho. A média de todos os valores é indicada por uma linha laranja. Um gráfico é fornecido para a cobertura do Pool A e do Pool B.
Software Versions (Versões do software)	Versões do software quando a amostra foi sequenciada. Inclui versão do NextSeq 550Dx Operating Software (NOS), Software Local Run Manager, Software RTA e do módulo Somatic Variant (Variante Somática).

Arquivos de resultados de análises

Os seguintes arquivos de resultados de análise são gerados para o módulo de análise Somatic Variant (Variante Somática) e fornecem resultados de análise para alinhamento e chamada de variante. Os arquivos de resultados de análise estão localizados na pasta Alignment (Alinhamento).

Nome do arquivo	Descrição
Demultiplexação (*.txt)	Arquivos intermediários que contêm resultados resumidos da demultiplexação.
FASTQ (*.fastq.gz)	Arquivos intermediários que contêm identificações de bases com pontuações de qualidade. Os arquivos FASTQ são a entrada primária para a etapa de alinhamento.
Arquivos de alinhamento no formato BAM (*.bam)	Contém leituras alinhadas para uma determinada amostra.
Arquivos de chamada de variante por pool no formato VCF (*.vcf)	Contém variantes chamadas em cada posição do pool avançado ou do pool reverso.
Arquivos de chamada de variante no formato VCF de genoma (*.genome.vcf.gz)	Contém o genótipo para cada posição, seja chamada como variante ou como referência.
Arquivos consensuais da chamada de variante no formato VCF (*.vcf.gz)	Contém variantes chamadas em cada posição de ambos os pools.
AmpliconCoverage_M1.tsv	Contém informações sobre cobertura por Amplicon por amostra para cada manifesto fornecido. M# representa o número do manifesto.

Formato dos arquivos de demultiplexação

O processo de demultiplexação lê a sequência de índice conectada a cada cluster para determinar a amostra de origem do cluster. O mapeamento entre os clusters e o número de amostras é gravado em um arquivo de demultiplexação (*.demux) de cada bloco na lâmina de fluxo.

O formato de nome do arquivo de demultiplexação é **s_1_X.demux**, em que X é o número do bloco.

Os arquivos de demultiplexação começam com um cabeçalho:

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

O restante do arquivo consiste em números de amostra para cada cluster do bloco.

Quando a etapa de demultiplexação está concluída, o software gera um arquivo de demultiplexação chamado **DemultiplexSummaryF1L1.txt**.

- ▶ No nome do arquivo, **F1** representa o número da lâmina de fluxo.
- ▶ No nome do arquivo, **L1** representa o número da cavidade.
- ▶ Resultados da demultiplexação em uma tabela com 1 linha por bloco e 1 coluna por amostra, incluindo a amostra 0.
- ▶ As sequências que ocorrem com mais frequência em leituras de índice.

Formato do arquivo FASTQ

FASTQ é um formato de arquivo baseado em texto que contém identificações de bases e valores de qualidade por leitura. Cada registro contém 4 linhas:

- ▶ O identificador
- ▶ A sequência
- ▶ Um sinal de mais (+)
- ▶ As pontuações de qualidade Phred em um formato codificado ASCII + 33

O identificador tem formato 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=AAAAAAAAA9#:<#<;<<<????#=#
```

Formato de arquivo BAM

Um arquivo BAM (*.bam) é a versão binária compactada de um arquivo SAM usado para representar sequências alinhadas até 128 Mb. Os formatos SAM e BAM são descritos em detalhes em samtools.github.io/hts-specs/SAMv1.pdf.

Os arquivos BAM usam o formato de nome de arquivo **SampleName_S#.bam**, em que # é o número da amostra determinado pela ordem em que as amostras são listadas para a execução.

Os arquivos BAM contêm uma seção de cabeçalho e uma seção de alinhamento:

- ▶ **Header** (Cabeçalho) — contém informações sobre todo o arquivo, como nome da amostra, tamanho da amostra e método de alinhamento. Os alinhamentos na seção Alignments (Alinhamentos) são associados a informações específicas na seção de cabeçalho.
- ▶ **Alignments** (Alinhamentos) — contém o nome, a sequência e a qualidade da leitura, as informações de alinhamento e os identificadores personalizados. O nome da leitura inclui o cromossomo, a coordenada inicial, a qualidade de alinhamento e a sequência descritora correspondente.

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

- ▶ **AS:** qualidade de alinhamento tipo paired-end
- ▶ **BC:** identificador de código de barras, que indica o ID da amostra demultiplexada associado à leitura.
- ▶ **SM:** qualidade de alinhamento tipo single-end.
- ▶ **XC:** sequência descritora correspondente
- ▶ **XN:** identificador de nome do Amplicon, que registra o ID do Amplicon associado à leitura

Os arquivos de índice BAM (*.bam.bai) fornecem um índice do arquivo BAM correspondente.

Formato de arquivo VCF

O formato de chamada de variante (VCF, Variant Call Format) é um formato de arquivo comum desenvolvido pela comunidade científica genômica. Ele contém informações sobre variantes encontradas em posições específicas em um genoma de referência. Os arquivos VCF terminam com o sufixo .vcf

O cabeçalho do arquivo VCF inclui a versão do formato de arquivo VCF e a versão do chamador de variante, além de listar as anotações usadas no restante do arquivo. O cabeçalho VCF também inclui o arquivo do genoma de referência e o arquivo BAM. A última linha no cabeçalho contém os cabeçalhos da coluna para as linhas de dados. Cada uma das linhas de dados do arquivo VCF contém informações sobre uma variante.

Cabeçalhos de arquivo VCF

Cabeçalho	Descrição
CHROM	O cromossomo do genoma de referência. Os cromossomos aparecem na mesma ordem do arquivo FASTQ de referência.
POS	A posição de base única da variante no cromossomo de referência. Para SNPs, essa posição é a base de referência com a variante; para indels ou deleções, essa posição é a base de referência imediatamente antes da variante.
ID	O número de rs para a variante obtida a partir de dbSNP.txt, se aplicável. Se houver múltiplos números de rs nesse local, a lista será delimitada por ponto e vírgula. Se não existir nenhuma entrada dbSNP nessa posição, um marcador (!) de valor ausente será usado.
REF	O genótipo de referência. Por exemplo, uma deleção de um só T é representada como um TT de referência e um T alternativo. Uma variante de nucleotídeo único 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 A de referência e AT alternativo. Uma variante de nucleotídeo único A a T é representada como A de referência e T alternativo.
QUAL	Uma pontuação de qualidade com escala Phred atribuída pelo chamador de variante. Pontuações mais altas indicam maior confiança na variante e 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 Q30 chamadas tem uma taxa de erro de 0.1%. Muitos chamadores de variante atribuem pontuações de qualidade baseadas em seus modelos estatísticos, que são altos em relação à taxa de erro observada.

Anotações de arquivos VCF

Cabeçalho	Descrição
FILTRO	<p>Se todos os filtros forem aprovados, PASS (APROVADO) será registrado na coluna de filtros.</p> <ul style="list-style-type: none"> • LowDP — aplicado a locais com profundidade de cobertura abaixo de 450x qualquer um dos pools. Para posições do Amplicon cobertas pela leitura avançada e reversa, é equivalente à cobertura de leitura única de 900x. • LowGQ — a qualidade de genotipagem (GQ, Genotyping Quality) está abaixo de um corte. • q30 — pontuação de qualidade < 30. • LowVariantFreq — a frequência da variante é menor do que o limite determinado. • PB — tendência do pool de sondagem. Variante não encontrada ou encontrada com baixa frequência em um ou dois pools de sondagem. • R3x6 — o número de repetições adjacentes (de 1 a 3 bp de comprimento) para as chamadas de variante ≥ 6. • SB — a tendência de quebra é maior que o limite determinado.
INFO (INFORMAÇÕES)	<p>Possíveis entradas na coluna INFO incluem:</p> <ul style="list-style-type: none"> • AC — contagem de alelos em genótipos para cada alelo ALT na mesma ordem listada. • AF — frequência de alelos para cada alelo ALT na mesma ordem listada. • AN — o número total de alelos em genótipos chamados. • CD — um marcador indicando que o SNP ocorre dentro da região de codificação de no mínimo 1 entrada RefGene. • DP — a profundidade (número de identificações de bases alinhadas a uma posição e usadas na chamada de variante). • Exon — uma lista separada por vírgulas de regiões exon lidas do RefGene. • FC — consequência funcional. • GI — uma lista separada por vírgulas de IDs de gene lidos do RefGene. • QD — confiança da variante/qualidade por profundidade. • TI — uma lista separada por vírgulas de IDs de transcrições lidos do RefGene.
FORMAT (FORMATO)	<p>A coluna de formatos lista campos separados por dois pontos. Por exemplo, GT:GQ. A lista de campos fornecidos depende do chamador de variante usado. Entre os campos disponíveis estão:</p> <ul style="list-style-type: none"> • AD — entrada do formulário X,Y, em que X é o número de chamadas de referência e Y é o número de chamadas alternativas. • DP — profundidade de leitura aproximada; leituras com MQ=255 ou com pares incorretos são filtradas. • GQ — qualidade do genótipo. • GQX — qualidade do genótipo. GQX é o mínimo do valor de GQ e da coluna QUAL. Em geral, esses valores são semelhantes; usar o mínimo torna o GQX a medida mais conservadora de 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 há informações de fases disponíveis. • NL — nível de ruído; uma estimativa do ruído de identificação de bases nesta posição. • PB — tendência do pool de sondagem. Valores próximos de 0 indicaram mais tendência em relação a um pool de sondagem e menos confiança em uma chamada de variante. • SB — tendência de quebra nesta posição. Valores negativos maiores indicam menos tendência; valores próximos de 0 indicam mais tendência. • VF — frequência da variante; a porcentagem de leituras que dão suporte ao alelo alternativo.
SAMPLE (AMOSTRA)	<p>A coluna de amostras dá os valores especificados na coluna FORMAT (FORMATO).</p>

Arquivos VCF de genoma

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

O arquivo gVCF mostra não-codificantes em posições que não passam em todos os filtros. Um marcador de genótipo (GT) de ./ indica uma não-codificante.

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

Arquivos VCF consensuais e por pool

O fluxo de trabalho Somatic Variant (Variante Somática) gera 2 conjuntos de arquivos de chamada de variante.

- ▶ **Per-pool VCF files** (Arquivos VCF por pool) — contém variantes chamadas no pool avançado ou no pool reverso. Os arquivos por pool são gravados na pasta VariantCallingLogs.
- ▶ **Consensus VCF files** (Arquivos VCF consensuais) — contém variantes chamadas de ambos os pools. Os arquivos consensuais são gravados na pasta Alignment (Alinhamento).

Arquivos VCF consensuais e por pool incluem tanto arquivos VCF (*.vcf) quanto gVCF (*.genome.vcf), e usam a seguinte convenção de nomenclatura, onde S# representa a ordem em que a amostra é listada para a execução:

- ▶ **Reports for all sites** (Relatórios para todos os locais) — SampleName_S#.genome.vcf
- ▶ **Reports variants only** (Relatórios apenas de variantes) — SampleName_S#.vcf

O software compara os arquivos VCF por pool e combina os dados em cada posição para criar um arquivo VCF consensual para a amostra.

As chamadas de variante de cada pool são combinadas em arquivos VCF consensuais usando os seguintes critérios.

Critérios	Resultado
Uma chamada de referência em cada pool	Chamada de referência
Uma chamada de referência em um pool e uma chamada de variante no outro pool	Chamada de variante filtrada
Chamadas de variante correspondentes com frequências similares em cada pool	Chamada de variante
Chamadas de variante correspondentes com frequências significativamente diferentes em cada pool	Chamada de variante filtrada
Chamadas de variante não correspondentes em cada pool	Chamada de variante filtrada

As medidas de cada pool são combinadas usando os seguintes valores.

Medida	Valor
Profundidade	Adição de profundidades de ambos os pools
Frequência de variante	Contagens de variantes totais divididas pela profundidade total de cobertura
Q-Score	Valor mínimo de ambos os pools

Arquivo de cobertura do Amplicon

Um arquivo de cobertura do Amplicon é gerado para cada arquivo de manifesto. O M# no nome do arquivo representa o número do manifesto.

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

- ▶ O ID de destino conforme listado no manifesto.

- ▶ A profundidade de cobertura das leituras que passam pelo filtro.

Arquivos de resultados complementares

Os arquivos de resultados a seguir fornecem informações complementares ou resumem os resultados da execução e os erros de análise. Embora esses arquivos não sejam necessários para avaliar os resultados de análise, eles podem ser usados para fins de solução de problemas. Todos os arquivos estão localizados na pasta Alignment (Alinhamento), salvo se especificado de outra forma.

Nome do arquivo	Descrição
AnalysisLog.txt	Registro de processamento que descreve todas as etapas que ocorreram durante a análise da pasta de execução atual. Esse arquivo não contém mensagens de erro. Localizado na pasta Alignment (Alinhamento).
AnalysisError.txt	Registro de processamento que lista todos os erros que ocorreram durante a análise. Este arquivo estará vazio se não forem encontrados erros. Localizado na pasta Alignment (Alinhamento).
DemultiplexSummaryF1L1#.txt	Informa resultados de demultiplexação em uma tabela com 1 linha por bloco e 1 coluna por amostra. O # representa a cavidade 1, 2, 3 ou 4 da lâmina de fluxo. Localizado na pasta Alignment (Alinhamento).
AmpliconRunStatistics.xml	Contém estatísticas resumidas específicas da execução. Localizado na pasta Alignment (Alinhamento).

Pasta de análise

A pasta de análise contém os arquivos gerados pelo software Local Run Manager.

A relação entre a pasta de resultados e a pasta de análise é resumida conforme segue:

- ▶ Durante o sequenciamento, o Real-Time Analysis (RTA) preenche a pasta de resultados com arquivos gerados durante a análise de imagens, a identificação de bases e a pontuação de qualidade.
- ▶ O RTA copia arquivos para a pasta de análise em tempo real. Depois que o RTA atribui uma pontuação de qualidade a cada base para cada ciclo, o software grava o arquivo RTAComplete.txt em ambas as pastas.
- ▶ Quando o arquivo RTAComplete.txt está presente, a análise começa.
- ▶ À medida que a análise continua, o Local Run Manager grava os arquivos de resultados na pasta de análise e depois copia os arquivos de volta para a pasta de resultados.

Pastas de alinhamento

Toda vez que a análise é enfileirada novamente, o Local Run Manager cria uma pasta de alinhamento chamada de **Alignment_N**, em que N é um número sequencial.

Estrutura de pastas

 **Alignment** (Alinhamento) — contém arquivos *.bam, *.vcf, arquivos FASTQ e arquivos específicos do módulo de análise.

 **Date and Time Stamp** (Carimbo de data e hora) — carimbo de data_hora da análise como AAAMMDD_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** (Dados)

 **Intensities** (Intensidades)

 **BaseCalls** (Identificações de bases)

 **L001** — contém arquivos *.bcl.

 **L001** — contém arquivos *.locs.

 **RTA Logs** (Registros do RTA) — contém arquivos de registro da análise do software RTA.

 **InterOp** (Interoperabilidade) — contém arquivos binários usados para informar medidas de execução de sequenciamento.

 **Logs** (Registros) — contém arquivos de registro descrevendo as etapas realizadas durante o sequenciamento.

 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 uma base (A, C, G ou T) para todos os clusters de um determinado bloco ou área de imagem na lâmina de fluxo, em um ciclo específico. O instrumento NextSeq 550Dx utiliza o sequenciamento de dois canais, o que requer apenas duas imagens para codificar os dados de quatro bases de DNA, uma imagem do canal vermelho e outra do canal verde.

O processo para leituras de índice de identificação de bases difere da identificação de bases durante outras leituras.

As leituras de índice devem começar com pelo menos uma base diferente de G em qualquer um dos primeiros dois ciclos. Se uma leitura de índice inicia com duas identificações de bases de G, nenhuma intensidade de sinal é gerada. O sinal deve estar presente em um dos dois primeiros ciclos para garantir o desempenho da demultiplexação.

Ao selecionar índices durante a criação da execução, um aviso de baixa diversidade será exibido se os índices não atenderem aos requisitos de diversidade. Para evitar o aviso de baixa diversidade, selecione as sequências de índice que fornecem sinal em ambos os canais de cada ciclo.

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

Esse processo de identificação de bases garante a precisão na análise de amostras de baixo plex. Para obter mais informações sobre as sequências de seus índices, consulte o folheto informativo do *Índice do kit para Amplicon personalizado TruSeq (documento n.º 1000000029772)*.

Durante a criação da execução no Local Run Manager você escolherá o número de amostras a serem testadas. As combinações de índices sugeridas que atendem aos requisitos de diversidade de índice são automaticamente preenchidas pelo software. Embora você não seja obrigado a usar as combinações de índices sugeridas, elas são recomendadas.

Histórico de revisões

Documento	Data	Descrição da alteração
Documento n.º 1000000030330 v04	Agosto de 2021	Atualizado o endereço do Representante autorizado da UE.
Documento n.º 1000000030330 v03	Abril de 2020	Atualizado o endereço do Representante autorizado da UE. Atualizado o endereço do Patrocinador australiano.
Documento n.º 1000000030330 v02	Janeiro de 2019	Adicionadas informações sobre os kits de reagentes v2.5.
Documento n.º 1000000030330 v01	Agosto de 2018	Atualizadas as marcas regulatórias.
Documento n.º 1000000030330 v00	Novembro de 2017	Versão inicial.

Assistência técnica

Para obter assistência técnica, entre em contato com o Suporte técnico da Illumina.

Site: www.illumina.com
E-mail: techsupport@illumina.com

Telefones do suporte ao cliente da Illumina

Região	Ligação gratuita	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	
Cingapura	+1.800.579.2745	
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
Suécia	+46 850619671	+46 200883979
Suíça	+41 565800000	+41 800200442
Taiwan	00806651752	
Outros países	+44.1799.534000	

Fichas de dados de segurança (SDSs) — Disponíveis no site da Illumina em support.illumina.com/sds.html.

Documentação do produto — Disponível para download em PDF no site da Illumina. Acesse support.illumina.com, selecione um produto e depois selecione **Documentation & Literature** (Documentação e literatura).



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

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 DE DIAGNÓSTICO IN VITRO

© 2021 Illumina, Inc. Todos os direitos reservados.

illumina®