

DRAGEN for Illumina DNA Prep with Enrichment Dx

Documentação do produto para a NovaSeq 6000Dx

Histórico de revisões

Documento	Data	Descrição da alteração
200014776 v02	Setembro de 2022	Formato corrigido de ficheiro manifest de texto (*.txt) para BED (*.bed), nas instruções para criar um ensaio. Corrigido consenso de ficheiros VCF para ficheiros VCF, na secção de saída de análises.
200014776 v01	Agosto de 2022	Adicionado: Predefinições da secção. Secção de filtragem do ruído sistemático. Ensaio atualizado, com instruções de criação para incluir um bom detalhe. Erros topográficos e gramaticais. Especificado que as instruções se destinam à aplicação, quando utilizada com a NovaSeq 6000Dx Instrument. Informação atualizada sobre os conteúdos do ficheiro VCF de saída.
200014776 v00	Março de 2022	Edição inicial.

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).

© de 2022 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.

Índice

Histórico de revisões	ii
Descrição geral	1
Métodos de análise	1
Criar um ensaio	4
Predefinições	5
Saída de análises	7
Ficheiros FASTQ	8
Ficheiros BAM	8
Ficheiros VCF	9
Ver resultados da análise	15
Assistência Técnica	16

Descrição geral

A aplicação DRAGEN™ for Illumina® DNA Prep with Enrichment Dx executa a demultiplexação, criação de FASTQ, interpretação de mapeamento e alinhamento a um genoma de referência e variante de identificação, dependendo do fluxo de trabalho selecionado.

Métodos de análise

DRAGEN for Illumina DNA Prep with Enrichment Dx executa a demultiplexação, criação de FASTQ, interpretação de mapeamento e alinhamento a um genoma de referência, dependendo do fluxo de trabalho selecionado:

- Criação de FASTQ
- Criação de Germline FASTQ e VCF
- Criação de FASTQ somático e VCF

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. 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

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, 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 bônus na pontuação e uma não-correspondência ou indel impõe uma penalidade. A pontuação global mais alta na matriz é o alinhamento escolhido.

Os valores específicos escolhidos por pontuação neste algoritmo indicam a forma como o equilíbrio, em relação a um alinhamento com várias interpretações possíveis, a possibilidade de um indel contrariamente a um ou mais SNP ou a preferência por um alinhamento sem clipagem. Os valores de pontuação DRAGEN por defeito são razoáveis para necessidades de alinhamento de comprimento moderado de alinhamento, para um genoma completo humano de referência para aplicações de variantes de identificação. Qualquer conjunto de parâmetros Smith-Waterman apresenta um modelo impreciso de pontuação de mutação genômica e erros de sequenciação. Valores diferentes afinados de pontuação podem ser mais adequados para algumas aplicações.

Identificação de variante de linha germinal DRAGEN

O DRAGEN Germline Small Variant Caller recolhe as leituras de ADN mapeadas e alinhadas, como entrada e identificações SNP e indels através de uma combinação de coluna inteligente de detecção e de conjunto locais *de novo* de haplótipos.

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 detecção 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.

No modo gVCF para identificação escalável de variantes multiamostras, o DRAGEN Germline Small Variant Caller pode ser executado por amostra, para criar um ficheiro intermédio de identificação de variantes genómicas (gVCF). O gVCF pode então ser utilizado por múltiplas amostras conjuntas eficientes de várias amostras, que permite um processamento rápido progressivo de amostras e recorrendo a uma coorte de grandes dimensões.

Porque o DRAGEN Germline Small Variant Caller tem algoritmos que possibilitam a distinção com precisão de erros correlacionados com verdadeiras variantes, as regras de filtragem são muito simples.

DRAGEN Somatic Variant Calling

O DRAGEN Somatic Small Variant Caller 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.

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 à volta de 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 pontuação TLOD, o DRAGEN Somatic Small Variant Caller analisa primeiramente por posição de referência, para cada evento somático candidato, assim como o evento referência na 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.

Criar um ensaio

Utilize as seguintes etapas para configurar um ensaio no Illumina Run Manager, seja no NovaSeq 6000Dx ou utilizando um navegador num computador ligado à rede. Podem ser introduzidos os dados das amostras manualmente ou importando uma folha de teste.

Predefinições de aplicação e ensaio

1. No ecrã Runs (ensaios), selecione **Create Runs** (Criar ensaios).
2. Selecione a aplicação DRAGEN for Illumina DNA Prep with Enrichment Dx e, em seguida, selecione **Next** (próximo).
3. No ecrã Predefinições de Ensaios, introduzir um nome do ensaio. O nome do ensaio é o nome que o identifica desde a sequenciação à análise.
4. **[Opcional]** Introduza uma descrição do ensaio, para ajudar a identificá-lo.
5. Certifique-se que o Kit de Preparação ao Banco selecionado é um Illumina DNA Prep With Enrichment Dx kit para esses efeitos.
6. Selecione o kit adaptadores de índice desejado.
7. Introduza o comprimento de leitura.
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.
8. **[Opcional]** Introduza o ID do tubo para o banco.
9. Selecione **Next** (próximo).

Dados de amostra

Utilize a tabela no ecrã Sample Data (amostra de dados), para introduzir manualmente a informação da amostra. Em alternativa, selecione **Import Samples** (importar amostras) para carregar informação da amostra. Para informação sobre como importar uma amostra, consulte [Importar amostras na página 5](#).

Introduzir amostras manualmente

1. Introduza um ID único de amostra no campo Sample ID (ID da amostra).
2. Utilize a **Placa – Posição correta** para selecionar a posição correta.
Os campos dos índices i7, 1, i5 e 2 são preenchidos automaticamente.
3. **[Opcional]** Introduza um nomee do banco.
4. Adicione linhas e repita os passos 1–3, conforme necessário, até que todas as amostras tenham sido adicionadas.
5. Selecione **Next** (próximo).

Importar amostras

Está disponível um modelo (*.csv) para descarregar no ecrã Sample Data (dados da amostra), ao planificar um ensaio em Illumina Run Manager utilizando um navegador num computador com rede.

1. Selecione **Download Template** (Descarregar modelo) para descarregar um ficheiro CSV em branco.
2. No ficheiro CSV, introduza a informação da amostra e guarde o ficheiro.
A folha da amostra do ficheiro CVS inclui as seguintes colunas de dados: ID da amostra, placa – posição correta, **opcional** nome do banco.
3. Clique em **Import Samples** (Importar amostras) e aceda à localização do ficheiro de informações CSV.
4. Selecione **Next** (próximo).

Predefinições das análises

1. Selecione o fluxo de trabalho desejado das análises:
 - Criação de FASTQ
 - Germline FASTQ e produção VCF, para um fluxo de trabalho de linha germinal
 - Criação de FASTQ somático e VCF, para um fluxo de trabalho somático
2. **[Opcional]** Se desejado, selecione a caixa de seleção **Generate ORA compressed FASTQ** (criar ORA de FASTQ comprimidos), para permitir a compressão ORA dos FASTQ.
3. **[Fluxos de trabalho de criação de VCF]** Utilize o menu pendente **Manifest File Selection** (seleção de ficheiro manifest) para selecionar um ficheiro manifest.
Um ficheiro manifest é obrigatório para a entrada do DRAGEN for Illumina DNA Prep with Enrichment Dx. O manifesto é um ficheiro de texto delimitado por tabulações BED (*.bed), que especifica os nomes e localizações de regiões referência visadas.
4. **[Fluxo de trabalho de criação de FASTQ e VCF somáticos]** Utilize o menu pendente **Noise File Selection** (seleção de ficheiro noise) para selecionar um ficheiro noise.
Pode ser especificado um ficheiro BED com um determinado nível de noise, para filtrar noise sistemático. Consulte [Filtros do noise na página 6](#) para obter mais informações.
5. Selecione **Next** (próximo).

Run Rever

1. No ecrã Review (analisar), reveja a informação introduzida nos ecrãs Run Settings (preparação de ensaio), Sample Data (amostra de dados) e Analysis Settings (predefinições de análises).
2. Selecione **Save** (Guardar).
O ensaio é guardado no separador Planned (planeado) no ecrã Runs (ensaio).

Predefinições

Selecione a aplicação no ecrã Applications para visualizar as predefinições atuais e altere-as.

Configuração

O ecrã de configuração mostra as seguintes predefinições 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) — Predefinições 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.

Filtros do noise

Está disponível a filtragem sistemática do noise, quando utiliza o fluxo de trabalho somático. 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 noise sistemático BED deve ser criado de amostras normais. Recomenda-se a criação de ficheiros sistemáticos noise que são de preparação para bancos, sistema de sequenciação e painel específico. Para a criação de um ficheiro noise, recomenda-se a utilização aproximada de 50 amostras normais.

Saída de análises

DRAGEN for Illumina DNA Prep with Enrichment Dx guarda a seguinte informação, na pasta das análises. Apenas os fluxos de trabalho da linha germinal e somáticos produzem um PDF.

- 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 aplicação gera os ficheiros de saída que se seguem. Os ficheiros exatos gerados dependem do fluxo de trabalho usado. Os ficheiros de saída encontram-se na pasta de análises.

Ficheiro de saída	Descrição
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. Se for selecionada a compressão ORA, isto será refletido no nome do ficheiro.
Ficheiros de alinhamento BAM (*.bam)	Contêm 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.
Relatório de métrica do ensaio (*.csv)	Contém métricas de qualidade sobre o ensaio, incluindo o rendimento total e 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
- Um sinal de adição (+)
- 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=AAAAAAAAAAA9#:<#<;<<<????#=#
```

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 n.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:

- 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.
- Alinhamentos — Contêm 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 desmultiplexagem, associado à leitura.

- SM: Extremidade única com qualidade de alinhamento.
- XC: Descritor de correspondências.
- XN: Etiqueta com o nome do amplicão, que regista o respetivo ID associado aos ficheiros de índice de leitura BAM (*.bam.bai), facultando 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, a versão identificadora da variante e lista as anotações no 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 genoma 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 de ficheiros VCF

Cabeçalho	Descrição
FILTRO	<p>Se todos os filtros passarem, PASS encontra-se escrito na coluna dos filtros. Eventuais entradas de FILTER no fluxo de trabalho da linha germinal 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

Cabeçalho	Descrição
FILTER (continuação)	<ul style="list-style-type: none"> • 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 <p>Eventuais entradas de FILTER no fluxo de trabalho somático incluem:</p> <ul style="list-style-type: none"> • 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 • 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

Cabeçalho	Descrição
FILTER (continuação)	<ul style="list-style-type: none"> • weak_evidence — Pontuação somática da variante não atinge o limite • systematic_noise — Filtrado no local com base em evidência de ruído sistemático nos normais
INFO	<p>Eventuais entradas INFO no fluxo de trabalho da linha germinal 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. <p>Eventuais entradas de INFO no fluxo de trabalho somático 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 ruído sistemático. • hotspot — Local somático conhecido, utilizado para aumentar a fiabilidade na identificação. • SOMATIC — Pelo menos uma variante nesta posição é somática.

Cabeçalho	Descrição
FORMAT	<p>O formato da coluna indica os campos separados por dois pontos. Por exemplo, GT:GQ.</p> <p>Os campos disponíveis no fluxo de trabalho da linha germinal 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. <p>Os campos disponíveis no fluxo de trabalho somático 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.

Cabeçalho	Descrição
FORMAT (continuação)	<ul style="list-style-type: none"> • 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.
SAMPLE (AMOSTRA)	A coluna amostra dá-lhe os valores especificados na coluna FORMAT.

Ficheiros genoma VCF

Os ficheiros VCF (* gvcf.gc) 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.

Ver resultados da análise

Os ensaios em curso atualmente são mostrados no separador Active. Os ensaios concluídos são mostrados no separador Completed. Para obter mais informações sobre a visualização dos resultados, consulte [NovaSeq 6000Dx, Documentação do produto \(documento n.º 200010105\)](#).

Assistência Técnica

Para obter assistência técnica, contacte o Illumina Suporte Técnico.

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

Números de telefone do Suporte Técnico da Illumina

Região	Número gratuito	Internacional
Austrália	+61 1800 775 688	
Áustria	+43 800 006249	+43 1 9286540
Bélgica	+32 800 77 160	+32 3 400 29 73
Canadá	+1 800 809 4566	
China		+86 400 066 5835
Dinamarca	+45 80 82 01 83	+45 89 87 11 56
Finlândia	+358 800 918 363	+358 9 7479 0110
França	+33 8 05 10 21 93	+33 1 70 77 04 46
Alemanha	+49 800 101 4940	+49 89 3803 5677
Hong Kong, China	+852 800 960 230	
Índia	+91 8006500375	
Indonésia		0078036510048
Irlanda	+353 1800 936608	+353 1 695 0506
Itália	+39 800 985513	+39 236003759
Japão	+81 0800 111 5011	
Malásia	+60 1800 80 6789	
Países Baixos	+31 800 022 2493	+31 20 713 2960
Nova Zelândia	+64 800 451 650	
Noruega	+47 800 16 836	+47 21 93 96 93
Filipinas	+63 180016510798	
Singapura	1 800 5792 745	
Coreia do Sul	+82 80 234 5300	
Espanha	+34 800 300 143	+34 911 899 417

Região	Número gratuito	Internacional
Suécia	+46 2 00883979	+46 8 50619671
Suíça	+41 800 200 442	+41 56 580 00 00
Taiwan, China	+886 8 06651752	
Tailândia	+66 1800 011 304	
Reino Unido	+44 800 012 6019	+44 20 7305 7197
Estados Unidos	+1 800 809 4566	+1 858 202 4566
Vietname	+84 1206 5263	

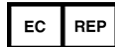
Fichas de dados de segurança (FDS) — Disponíveis no sítio Web da Illumina em support.illumina.com/sds.html.

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

CE



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

Promotor australiano

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

PARA UTILIZAÇÃO EM DIAGNÓSTICO IN VITRO

© de 2022 Illumina, Inc. Todos os direitos reservados.

illumina[®]