

DRAGEN for Illumina DNA Prep with Enrichment Dx

Documentación del producto para NovaSeq 6000Dx

Historial de revisiones

Documento	Fecha	Descripción del cambio
200014776 v02	Septiembre de 2022	<p>Se ha corregido el formato para el archivo del manifiesto, de texto (*.txt) a BED (*.bed), en las instrucciones sobre cómo crear una ejecución.</p> <p>Se ha corregido “archivos VCF de consenso” a “archivos VCF” en la sección de resultados de análisis.</p>
200014776 v01	Agosto de 2022	<p>Añadido:</p> <p>Sección de configuración.</p> <p>Sección sobre filtrado de ruido sistemático.</p> <p>Actualización de las instrucciones de creación de experimentos para incluir más detalles.</p> <p>Corrección de errores tipográficos y gramaticales.</p> <p>Se especifica que las instrucciones están destinadas a la aplicación cuando se utiliza con el instrumento NovaSeq 6000Dx.</p> <p>Actualización de la información sobre el contenido de los archivos de resultados VCF.</p>
200014776 v00	Marzo de 2022	Publicación inicial.

Este documento y su contenido son propiedad exclusiva de Illumina, Inc. y sus afiliados ("Illumina") y están previstos solamente para el uso contractual de sus clientes en conexión con el uso de los productos descritos en él y no para ningún otro fin. Este documento y su contenido no se utilizarán ni distribuirán con ningún otro fin ni tampoco se comunicarán, divulgarán ni reproducirán en ninguna otra forma sin el consentimiento previo por escrito de Illumina. Illumina no transfiere mediante este documento ninguna licencia bajo sus derechos de patente, marca comercial, copyright ni derechos de autor o similares derechos de terceros.

Para asegurar el uso correcto y seguro de los productos descritos en este documento, el personal cualificado y adecuadamente capacitado debe seguir las instrucciones incluidas en el mismo de manera rigurosa y expresa. Se debe leer y entender completamente todo el contenido de este documento antes de usar estos productos.

SI NO SE LEE COMPLETAMENTE EL DOCUMENTO Y NO SE SIGUEN EXPRESAMENTE TODAS LAS INSTRUCCIONES DESCRITAS EN ESTE, PODRÍAN PRODUCIRSE DAÑOS EN EL PRODUCTO, LESIONES PERSONALES, INCLUIDOS LOS USUARIOS U OTRAS PERSONAS Y DAÑOS EN OTROS BIENES Y QUEDARÁ ANULADA TODA GARANTÍA APLICABLE AL PRODUCTO.

ILLUMINA NO ASUME RESPONSABILIDAD ALGUNA DERIVADA DEL USO INCORRECTO DE LOS PRODUCTOS AQUÍ DESCRITOS (INCLUIDAS LAS PIEZAS O EL SOFTWARE).

© 2022 Illumina, Inc. Todos los derechos reservados.

Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea obtener información específica sobre las marcas comerciales, consulte www.illumina.com/company/legal.html.

Índice

Historial de revisiones	ii
Descripción general	1
Métodos de análisis	1
Creación de un experimento	5
Configuración	7
Resultados del análisis	8
Archivos FASTQ	9
Archivos BAM	9
Archivos VCF	10
Visualización de los resultados del análisis	16
Asistencia técnica	17

Descripción general

La aplicación DRAGEN™ for Illumina® DNA Prep with Enrichment Dx realiza demultiplexado, generación de FASTQ, cartografía de lecturas, alineación con un genoma de referencia y llamada de variantes dependiendo del flujo de trabajo de análisis seleccionado.

Métodos de análisis

DRAGEN for Illumina DNA Prep with Enrichment Dx realiza demultiplexado, generación de FASTQ, asignación de lecturas y alineación con un genoma de referencia dependiendo del flujo de trabajo seleccionado:

- Generación de FASTQ
- Generación de FASTQ y VCF Germline (germinal)
- Generación de FASTQ y VCF Somatic (somática)

Generación de FASTQ

Las secuencias ensambladas se escriben en archivos FASTQ por muestra. Los archivos FASTQ son archivos de texto que contienen datos de secuenciación y puntuaciones de calidad para una sola muestra. Para cada muestra, se generan archivos FASTQ independientes por carril de la celda de flujo y por lectura de secuenciación. El nombre de la muestra especificado durante la configuración del experimento se incluye en el nombre del archivo FASTQ. Los archivos FASTQ incluyen los datos principales para la alineación. El primer paso para la generación de FASTQ es el demultiplexado. El demultiplexado asigna los grupos que pasan el filtro a una muestra comparando cada secuencia de lectura del índice con las secuencias de índices especificadas en el experimento. En este paso, no se considera ningún valor de calidad. Las lecturas del índice se pueden identificar por medio de los siguientes pasos:

- Las muestras están numeradas comenzando por el número uno de acuerdo con el lugar que ocupan en el experimento.
- La muestra número cero se reserva para los grupos que no se han asignado al experimento.
- Los grupos se asignan a una muestra siempre que la secuencia de índice coincida de forma exacta o cuando exista solamente una discrepancia por lectura del índice.

El software incluye compresión ORA para comprimir los archivos FASTQ. Cuando se utiliza el formato ORA (*.ora), la suma de comprobación md5 del contenido FASTQ se preserva tras un ciclo de compresión y descompresión para asegurar una compresión sin pérdida.

Cartografía y alineación de ADN

La primera etapa de la cartografía es la generación de semillas a partir de la lectura, y a continuación la búsqueda de coincidencias exactas en el genoma de referencia. Estos resultados se refinan posteriormente ejecutando alineaciones de Smith-Waterman completas en las ubicaciones con la mayor densidad de coincidencias con las semillas. Este algoritmo bien documentado compara cada posición de la lectura con todas las posiciones candidatas de la referencia. Estas comparaciones corresponden a una matriz de posibles alineaciones entre lectura y referencia. Para cada una de estas posiciones de alineación candidatas, Smith-Waterman genera puntuaciones que se utilizan para evaluar si la mejor alineación que pasa por esa celda de la matriz la alcanza por una coincidencia o discrepancia de nucleótidos (movimiento diagonal), una deleción (movimiento horizontal) o una inserción (movimiento vertical). Una coincidencia entre lectura y referencia proporciona una bonificación en la puntuación, y una discrepancia o indel impone una penalización. La ruta global a través de la matriz con una mayor puntuación es la alineación elegida.

Los valores específicos elegidos para las puntuaciones en este algoritmo indican cómo equilibrar, para una alineación con múltiples interpretaciones posibles, la posibilidad de un indel frente a uno o más SNP, o la preferencia por una alineación sin recorte. Los valores de puntuación predeterminados en DRAGEN son razonables para alinear lecturas de longitud moderada a un genoma de referencia humano completo para aplicaciones de llamada de variantes. Cualquier conjunto de parámetros de puntuación Smith-Waterman representa un modelo impreciso de mutación genómica y errores de secuenciación. Unos valores de puntuación de la alineación ajustados de forma diferente pueden ser más adecuados para algunas aplicaciones.

Llamada de variantes germinales de DRAGEN

El llamador de variantes pequeñas germinales de DRAGEN acepta como entradas lecturas de ADN cartografiado y alineado y llama SNP e indels mediante una combinación de detección por columnas y de ensamblado local *de novo* de haplotipos.

Las regiones de referencia llamables se identifican por vez primera con una suficiente cobertura de alineación. En estas regiones de referencia, una exploración rápida de las lecturas ordenadas identifica las regiones activas, centradas en torno a columnas de apilamiento con evidencias de una variante. Las regiones activas se rellenan con suficiente contexto para cubrir el contenido cercano significativo, no de referencia. Si hay evidencia de indels, las regiones activas reciben un relleno adicional.

Las lecturas alineadas se recortan dentro de cada región activa y se ensamblan en un gráfico de De Bruijn. Los bordes de las lecturas recortadas se ponderan mediante recuentos de observación, con la secuencia de referencia como eje. Tras una cierta limpieza y simplificación del gráfico, todas las rutas de origen a destino se extraen como haplotipos candidatos. Cada haplotipo se alinea al genoma de referencia con Smith-Waterman para identificar las variantes que representa. Este conjunto de eventos puede ser aumentado por una detección basada en la posición. Para cada par lectura-haplotipo, la probabilidad $P(r|H)$ de observar la lectura, asumiendo que el haplotipo es la verdadera muestra de inicio, se estima utilizando un modelo de Markov oculto (Hidden Markov Model, HMM) de pares.

Mediante una exploración por posiciones de referencia sobre la región activa, se forman genotipos candidatos a partir de combinaciones diploides de eventos de variantes (SNP o indels). Para cada evento (incluida la referencia), la probabilidad condicional $P(r|e)$ de observar cada lectura solapada se estima como la $P(r|H)$ máxima para los haplotipos que justifican el evento. Estos se combinan en la probabilidad condicional $P(r|e_1e_2)$ para un genotipo (par de eventos) y se multiplican para obtener la probabilidad condicional $P(R|e_1e_2)$ de observar el apilamiento de lecturas completo. Utilizando la fórmula de Bayes, se calcula la probabilidad posterior $P(e_1e_2|R)$ de cada genotipo diploide, y se obtiene el ganador.

En el modo gVCF utilizado para la llamada de variantes escalable de varias muestras, el llamador de variantes pequeñas germinales de DRAGEN puede ejecutarse por muestra para generar un archivo intermedio de llamada de variantes genómicas (gVCF). El gVCF puede utilizarse a continuación para un genotipado conjunto y eficiente de varias muestras, lo que permite el procesamiento incremental rápido de las muestras y el escalado a tamaños grandes de cohortes.

Dado que el llamador de variantes pequeñas germinales de DRAGEN tiene algoritmos que le permiten distinguir de forma eficiente los errores de correlación de las variantes verdaderas, las reglas de filtrado son muy simples.

Llamada de variantes somáticas de DRAGEN

El llamador de variantes pequeñas somáticas de DRAGEN acepta como entradas lecturas de ADN cartografiado y alineado y llama SNV e indels mediante una combinación local *de novo* de haplotipos en una región activa.

Las regiones de referencia llamables se identifican por vez primera con una suficiente cobertura de alineación. En estas regiones de referencia, una exploración de las lecturas ordenadas identifica las regiones activas, centradas en torno a columnas de apilamiento con evidencias de una variante en las lecturas de tumores. Las regiones activas se rellenan con suficiente contexto para cubrir el contenido cercano significativo, no de referencia. Si hay evidencia de indels, las regiones activas reciben un relleno adicional.

Las lecturas alineadas se recortan dentro de cada región activa y se ensamblan en un gráfico de De Bruijn. Los bordes de las lecturas recortadas se ponderan mediante recuentos de observación, con la secuencia de referencia como eje. Tras una cierta limpieza y simplificación del gráfico, todas las rutas de origen a destino se extraen como haplotipos candidatos. Cada haplotipo se alinea al genoma de referencia con Smith-Waterman para identificar las variantes que representa. Para cada par lectura-haplotipo, la probabilidad $P(r|H)$ de observar la lectura se estima utilizando un modelo de Markov oculto (HMM) de pares, asumiendo que el haplotipo es la verdadera muestra de inicio.

Para determinar la puntuación TLOD, el llamador de variantes pequeñas somáticas de DRAGEN explora primero en base a las posiciones de referencia cada evento somático candidato, así como el evento de referencia, sobre la región activa. La probabilidad condicional $P(r|e)$ de observar cada lectura solapada se estima como la $P(r|H)$ máxima para los haplotipos que justifican el evento. Estas se combinan en la probabilidad condicional $P(r|E)$ para una hipótesis de evento, E , que implica una mezcla de los alelos somáticos candidatos y de referencia en un intervalo de posibles frecuencias alélicas, y se multiplican

para obtener la probabilidad condicional $P(R|E)$ de observar el apilamiento de lecturas completo. A partir de lo anterior, se calcula una puntuación TLOD como evidencia de que un alelo ALT está presente en la muestra tumoral en un locus dado.

Creación de un experimento

Siga estos pasos para configurar un experimento en Illumina Run Manager, ya sea en el NovaSeq 6000Dx o utilizando un navegador en un ordenador conectado a la red. Los datos de la muestra pueden introducirse manualmente o importando una hoja de muestras.

Ajustes de configuración del experimento y de la aplicación

1. En la pantalla Runs (Experimentos), seleccione **Create Run** (Crear experimento).
2. Seleccione la aplicación DRAGEN for Illumina DNA Prep with Enrichment Dx y, a continuación, seleccione **Next** (Siguiente).
3. En la pantalla Run Settings (Ajustes de configuración del experimento), introduzca un nombre para el experimento. El nombre del experimento lo identifica desde la secuenciación hasta el análisis.
4. **[Opcional]** Escriba una descripción que ayude a identificar mejor el experimento.
5. Asegúrese de que el kit de preparación de bibliotecas seleccionado sea un kit de preparación de bibliotecas Illumina DNA Prep with Enrichment Dx.
6. Seleccione el kit de adaptadores de índices que desee.
7. Introduzca la longitud de la lectura.
La Lectura 1 y la Lectura 2 tienen un valor predeterminado de 151 ciclos.
El Índice 1 y el Índice 2 tienen un valor fijo de 10 ciclos.
8. **[Opcional]** Indique un ID de tubo de bibliotecas.
9. Seleccione **Next** (Siguiente).

Datos de la muestra

Utilice la tabla que aparece en la pantalla Sample Data (Datos de la muestra) para introducir manualmente la información de la muestra. Como alternativa, seleccione **Import Samples** (Importar muestras) para cargar la información de la muestra. Para obtener más información sobre la importación de muestras, consulte [Importar muestras en la página 6](#).

Introducción de las muestras de forma manual

1. Introduzca un ID de muestra único en el campo Sample ID (ID de muestra).
2. Utilice **Plate - Well Position** (Placa - Posición del pocillo) para seleccionar la posición del pocillo. Los campos i7 Index, Index 1, i5 Index e Index 2 se rellenan automáticamente.
3. **[Opcional]** Introduzca un nombre de biblioteca.
4. Añada filas y repita los pasos 1–3 según sea necesario hasta que se hayan añadido todas las muestras a la tabla.
5. Seleccione **Next** (Siguiente).

Importar muestras

Puede descargarse una plantilla (*.csv) en la pantalla Sample Data (Datos de la muestra) cuando se planea un experimento en Illumina Run Manager utilizando un navegador en un ordenador conectado a la red.

1. Seleccione **Download Template** (Descargar plantilla) para descargar un archivo CSV en blanco.
2. Desde el archivo CSV, introduzca la información de la muestra y guarde el archivo.
El archivo CSV de hoja de muestras incluye las siguientes columnas de datos: ID de muestra, Placa - Posición del pocillo, **Opcional** Nombre de biblioteca opcional.
3. Seleccione **Import Samples** (Importar muestras) y busque la ubicación del archivo CSV.
4. Seleccione **Next** (Siguiente).

Configuración de los ajustes del análisis

1. Seleccione el flujo de trabajo de análisis que desee:
 - Generación de FASTQ
 - Generación de FASTQ y VCF Germline para un flujo de trabajo germinal
 - Generación de FASTQ y VCF Somatic para un flujo de trabajo somático
2. **[Opcional]** Si lo desea, seleccione la casilla **Generate ORA compressed FASTQs** (Generar FASTQ comprimidos con ORA) para habilitar la compresión ORA de los FASTQ.
3. **[Flujos de trabajo de generación de VCF]** Use el menú desplegable **Manifest File Selection** (Selección de archivo de manifiesto) para seleccionar un archivo de manifiesto.
Un archivo de manifiesto es un archivo de entrada requerido por el DRAGEN for Illumina DNA Prep with Enrichment Dx. El manifiesto es un archivo BED (*.bed) delimitado por tabuladores que especifica los nombres y ubicaciones de las regiones de referencia de interés.
4. **[Flujo de trabajo de generación de FASTQ y VCF Somatic (somática)]** Use el menú desplegable **Noise File Selection** (Selección de archivo de ruido) para seleccionar un archivo de ruido.
Puede especificarse un archivo BED con un nivel de ruido específico para el centro para filtrar el ruido sistemático. Para obtener más información, consulte [Filtrado del ruido en la página 7](#).
5. Seleccione **Next** (Siguiente).

Experimento Revisión

1. En la pantalla Review (Revisión), revise la información introducida en las pantallas Run Settings (Ajustes de configuración del experimento), Sample Data (Datos de la muestra) y Analysis Settings (Configuración de los ajustes del análisis).
2. Seleccione **Save** (Guardar).
El experimento se guarda en la pestaña Planned (Planeados) de la pantalla Runs (Experimentos).

Configuración

Seleccione la aplicación en la pantalla Applications (Aplicaciones) para ver la configuración actual y realizar cambios.

Configuración

La pantalla de configuración muestra los siguientes ajustes de la aplicación:

- **Library Prep Kits** (Kits de preparación de bibliotecas): muestra el kit de preparación de bibliotecas predeterminado para la aplicación. Este ajuste no se puede cambiar.
- **Index Adapter Kits** (Kits de adaptadores de índices): muestra el kit de adaptadores de índices predeterminado para la aplicación. Este ajuste no se puede cambiar.
- **Read lengths** (Longitudes de la lectura): las longitudes de la lectura se han establecido en 151 para la aplicación de forma predeterminada, pero pueden cambiarse durante la creación de un experimento.
- **Manifest and Noise Files** (Archivos de manifiesto y ruido): cargar y cambiar la configuración para los archivos de manifiesto y ruido.
 - Seleccione **Upload File** (Cargar archivo) para cargar archivos para utilizarlos en el análisis.
 - Seleccione el botón de opción **Default** (Predeterminado) para fijar el archivo como archivo de manifiesto o de ruido seleccionado por defecto durante la creación del experimento cuando se selecciona la aplicación.
 - Seleccione la casilla **Enabled** (Habilitado) para que el archivo se muestre en el menú desplegable durante la creación del experimento.

Permisos

Utilice las casillas de la pantalla Permissions (Permisos) para gestionar el acceso de los usuarios a la aplicación.

Filtrado del ruido

Está disponible el filtrado sistemático del ruido cuando se utiliza el flujo de trabajo somático. El filtro puede utilizarse en modo Tumor-Normal, pero es especialmente útil en experimentos Tumor-Only (Solo tumor) en los que no está disponible una muestra normal emparejada.

El BED de ruido sistemático debe generarse a partir de muestras normales. Se recomienda crear archivos de ruido sistemático específicos para la preparación de bibliotecas, el sistema de secuenciación y el panel. Para generar un archivo de ruido, se recomienda utilizar aproximadamente 50 muestras normales.

Resultados del análisis

DRAGEN for Illumina DNA Prep with Enrichment Dx guarda la información siguiente en la carpeta de análisis. Solo los flujos de trabajo germinales y somáticos producen un PDF.

- Archivo de manifiesto utilizado
- Versión del software
- ID de la muestra
- Total de lecturas alineadas
- Porcentaje de lecturas alineadas por muestra
- Número de SNV llamadas por muestra
- Número de indels llamadas por muestra
- Estadísticas de cobertura

Archivos de resultados de análisis

Los archivos de resultados siguientes son generados por la aplicación. Los archivos exactos generados dependen del flujo de trabajo de análisis que se utilice. Los archivos de resultados se encuentran en la carpeta de análisis.

Archivo de resultados	Descripción
FASTQ (*.fastq.gz o *.fastq.ora)	Son archivos intermedios que contienen las puntuaciones de calidad de las llamadas a las bases. Los archivos FASTQ son la entrada principal para el paso de alineación. Si se selecciona la compresión ORA, el nombre de archivo lo refleja.
Archivos de alineación BAM (*.bam)	Contiene lecturas alineadas para una muestra dada.
Archivos de genoma VCF (*.gvcf.gz)	Contiene el genotipo de cada posición, ya sea llamada como variante o como referencia.
Archivos VCF (*.vcf.gz)	Contiene variantes llamadas en cada posición.

Archivo de resultados	Descripción
Informe de mediciones del experimento (*.csv)	Contiene mediciones de calidad sobre el experimento, incluyendo el rendimiento total y la puntuación Q30.

Archivos FASTQ

El formato FASTQ (*.fastq.gz, *.fastq.ora) es un formato de archivo de texto que contiene las llamadas de bases y los valores de calidad por lectura. Cada archivo contiene la siguiente información:

- El identificador de la muestra
- La secuencia
- Un signo de suma (+)
- Una puntuación de calidad según la escala Phred en formato encriptado ASCII + 33

El identificador de la muestra presenta el siguiente formato.

```
@Instrumento:IDExperimento:IDCeldaFlujo:Carril:Placa:X:Y
NúmeroLectura:IndicadorFiltro:0:NúmeroMuestra
Ejemplo:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAAA9#:<#<;<<<????#=#
```

Archivos BAM

Un archivo BAM (*.bam) es la versión binaria comprimida de un archivo SAM (mapa de alineación de secuencias) que se utiliza para representar secuencias alineadas de hasta 128 Mb. Los archivos BAM utilizan el formato de nomenclatura de archivos `SampleName_S#.bam`. # es el número de muestra determinado por el orden en que aparecen las muestras para el experimento. En modo multinodo, S# se fija en S1, con independencia del orden de la muestra.

Los archivos BAM contienen una sección de encabezado y una sección de alineación:

- Encabezado: contiene información sobre todo el archivo, como el nombre de la muestra, la longitud de la muestra y el método de alineación. Las alineaciones en la sección de alineaciones están asociadas con información específica en la sección de encabezado.

- **Alineaciones:** contiene el nombre de la lectura, la secuencia de lectura, la calidad de lectura, la información de alineación y marcadores personalizados. El nombre de lectura incluye el cromosoma, la coordenada de inicio, la calidad de alineación y la cadena del descriptor de coincidencias.

La sección de alineaciones incluye la siguiente información para cada lectura o par de lectura:

- **AS:** Calidad de la alineación "Paired-end".
- **RG:** Grupo de lectura, que indica el número de lecturas para una muestra específica.
- **BC:** Marcador de código de barras, que indica el ID de muestra demultiplexado asociado a la lectura.
- **SM:** Calidad de la alineación "Single-end".
- **XC:** Cadena del descriptor de coincidencias.
- **XN:** El marcador de nombre del amplicón, que registra el ID del amplicón asociado con los archivos de índices BAM (*.bam.bai) de la lectura, proporciona un índice del archivo BAM correspondiente.

Archivos VCF

Los archivos de formato de llamada de variantes (*.vcf) contienen información sobre las variantes que se encuentran en posiciones específicas de un genoma de referencia.

El encabezado del archivo VCF incluye la versión de formato del archivo VCF y la versión del llamador de variantes y detalla las anotaciones utilizadas en el resto del archivo. El encabezado del VCF incluye también el archivo del genoma de referencia y el archivo BAM. La última línea del encabezado contiene los encabezados de las columnas para las líneas de datos. Cada una de las líneas de datos del archivo VCF contiene información sobre una sola variante.

Tabla 1 Encabezados del archivo VCF

Encabezado	Descripción
CHROM	El cromosoma del genoma de referencia. Los cromosomas aparecen en el mismo orden que en el archivo FASTA de referencia.
POS	La posición de base individual de la variante en el cromosoma de referencia. Para las variantes de nucleótido único (SNV), esta posición es la base de referencia con la variante. Para las indels, esta posición es la base de referencia que precede inmediatamente a la variante.
ID	El número rs (SNP de referencia) para el SNP obtenido de <code>dbSNP.txt</code> , si es aplicable. Si existen varios números rs en esta ubicación, la lista se delimita con punto y coma. Si no existe una entrada dbSNP en esta posición, se utiliza un marcador de valor que falta ('.').
REF	El genotipo de referencia. Por ejemplo, una delección de una sola T se representa como TT de referencia y T alternativa. Una variante de nucleótido único de A a T se representa como A de referencia y T alternativa.

Encabezado	Descripción
ALT	Los alelos que difieren de la lectura de referencia. Por ejemplo, una inserción de una sola T se representa como A de referencia y AT alternativa. Una variante de nucleótido único de A a T se representa como A de referencia y T alternativa.
QUAL	Una puntuación de calidad en la escala Phred asignada por el llamador de variantes. Puntuaciones más altas indican mayor confianza en la variante y una menor probabilidad de errores. Para una puntuación de calidad de Q, la probabilidad estimada de un error es $10^{-(Q/10)}$. Por ejemplo, el conjunto de llamadas Q30 tiene una tasa de errores del 0,1 %. Muchos llamadores de variantes asignan puntuaciones de calidad a partir de sus modelos estadísticos, que son altas con relación a la tasa de errores observada.

Tabla 2 Anotaciones del archivo VCF

Encabezado	Descripción
FILTER (Filtro)	<p>Si se superan todos los filtros, se escribe PASS (Pasa) en la columna Filtro. Las entradas posibles para FILTER (Filtro) en el flujo de trabajo Germline (germinal) incluyen:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL: se aplica si la puntuación QUAL de una variante SNP no alcanza el umbral • DRAGENIndelHardQUAL: se aplica si la puntuación QUAL de una variante indel no alcanza el umbral • LowDepth: sitio filtrado porque la profundidad de cobertura no alcanza el umbral • LowGQ: sitio filtrado porque la calidad del genotipo no alcanza el umbral • PloidyConflict: la llamada de genotipo del llamador de variantes no es coherente con la ploidía de cromosomas • base_quality: sitio filtrado porque la mediana de la calidad de las bases de las lecturas alternativas en este locus no alcanza el umbral • filtered_reads: Sitio filtrado porque una fracción de las lecturas demasiado grande se ha excluido. • fragment_length: sitio filtrado porque la diferencia absoluta entre la mediana de la longitud de los fragmentos de las lecturas alternativas y la mediana de la longitud de los fragmentos de las lecturas de referencia en este locus supera el umbral • low_depth: sitio filtrado porque la profundidad de lectura es demasiado baja • low_frac_info_reads: sitio filtrado porque la fracción de lecturas informativas está por debajo del umbral • low_normal_depth: sitio filtrado porque la profundidad de lectura de las muestras normales es demasiado baja • long_indel: sitio filtrado porque la longitud de indel es demasiado larga • mapping_quality: sitio filtrado porque la mediana de la calidad de la cartografía de las lecturas alternativas en este locus no alcanza el umbral • multiallelic: sitio filtrado porque más de dos alelos alternativos pasan el LOD tumoral • non_homref_normal: sitio filtrado porque el genotipo de la muestra normal no es una referencia homocigótica • no_reliable_supporting_read: sitio filtrado porque no existe una lectura somática de apoyo fiable • panel_of_normals: observado en al menos una muestra en el vcf del panel de normales • read_position: sitio filtrado porque la mediana de las distancias entre el inicio y el fin de la lectura en este locus está por debajo del umbral • RMxNRepeatRegion: sitio filtrado porque todo o parte del alelo de la variante es una repetición de la referencia • strand_artifact: sitio filtrado debido a cortes importantes en la cadena

Encabezado	Descripción
FILTER (continuación)	<ul style="list-style-type: none"> • str_contraction: sitio filtrado debido a la sospecha de un error de PCR en donde el alelo alternativo es una unidad de repetición menor que la referencia • too_few_supporting_reads: sitio filtrado porque hay demasiadas pocas lecturas de apoyo en la muestra tumoral • weak_evidence: la puntuación de la variante somática no alcanza el umbral <p>Las entradas posibles para FILTER (Filtro) en el flujo de trabajo Somatic (somática) incluyen:</p> <ul style="list-style-type: none"> • base_quality: sitio filtrado porque la mediana de la calidad de las bases de las lecturas alternativas en este locus no alcanza el umbral • filtered_reads: sitio filtrado porque una fracción de las lecturas demasiado grande ha sido filtrada • fragment_length: sitio filtrado porque la diferencia absoluta entre la mediana de la longitud de los fragmentos de las lecturas alternativas y la mediana de la longitud de los fragmentos de las lecturas de referencia en este locus supera el umbral • low_depth: sitio filtrado porque la profundidad de lectura es demasiado baja • low_frac_info_reads: sitio filtrado porque la fracción de lecturas informativas está por debajo del umbral • low_normal_depth: sitio filtrado porque la profundidad de lectura de las muestras normales es demasiado baja • long_indel: sitio filtrado porque la longitud de indel es demasiado larga • mapping_quality: sitio filtrado porque la mediana de la calidad de la cartografía de las lecturas alternativas en este locus no alcanza el umbral • multiallelic: sitio filtrado porque más de dos alelos alternativos pasan el LOD tumoral • non_homref_normal: sitio filtrado porque el genotipo de la muestra normal no es una referencia homocigótica • no_reliable_supporting_read: sitio filtrado porque no existe una lectura somática de apoyo fiable • panel_of_normals: observado en al menos una muestra en el vcf del panel de normales • read_position: sitio filtrado porque la mediana de las distancias entre el inicio y el fin de la lectura en este locus está por debajo del umbral • RMxNRepeatRegion: sitio filtrado porque todo o parte del alelo de la variante es una repetición de la referencia • strand_artifact: sitio filtrado debido a cortes importantes en la cadena • str_contraction: sitio filtrado debido a la sospecha de un error de PCR en donde el alelo alternativo es una unidad de repetición menor que la referencia • too_few_supporting_reads: sitio filtrado porque hay demasiadas pocas lecturas de apoyo en la muestra tumoral

Encabezado	Descripción
FILTER (continuación)	<ul style="list-style-type: none"> • weak_evidence: la puntuación de la variante somática no alcanza el umbral • systematic_noise: Sitio filtrado a partir de las pruebas de que hay ruido sistemático en las muestras normales.
INFO	<p>Las entradas posibles para INFO (Información) en el flujo de trabajo Germline (germinal) incluyen:</p> <ul style="list-style-type: none"> • AC: número de alelos en los genotipos para cada alelo alternativo, en el mismo orden en que aparecen. • AF: frecuencia alélica para cada alelo alternativo, en el mismo orden en que aparecen. • AN: número total de alelos en los genotipos llamados. • DB: miembro de dbSNP. • FS: valor de p en la escala Phred utilizando la prueba exacta de Fisher para detectar cortes en la cadena. • QD: confianza de la variante/calidad por profundidad. • R2_5P_bias: puntuación basada en el sesgo de emparejamiento (“mate bias”) y la distancia desde el extremo de cebado en 5. • SOR: cociente de posibilidades simétrico de una tabla de contingencia de 2x2 para detectar cortes en la cadena. • DP: profundidad aproximada de lectura (informativa y no informativa); algunas lecturas pueden haber sido filtradas según mapq, etc. • END: posición de parada del intervalo. • FractionInformativeReads: la fracción de lecturas informativas con respecto al número total de lecturas. • MQ: calidad de la cartografía RMS. • MQRankSum: puntuación Z de la prueba de suma de rangos de Wilcoxon de las cualidades de cartografía de las lecturas alternativas frente a las de referencia. • ReadPosRankSum: puntuación Z de la prueba de suma de rangos de Wilcoxon del sesgo de posición de las lecturas alternativas frente a las de referencia. • SOMATIC: al menos una variante en esta posición es somática. <p>Las entradas posibles para INFO (Información) en el flujo de trabajo Somatic (somática) incluyen:</p> <ul style="list-style-type: none"> • DP: profundidad aproximada de lectura (informativa y no informativa); algunas lecturas pueden haber sido filtradas según mapq, etc. • END: posición de parada del intervalo. • FractionInformativeReads: la fracción de lecturas informativas con respecto al número total de lecturas. • MQ: calidad de la cartografía RMS. • MQRankSum: puntuación Z de la prueba de suma de rangos de Wilcoxon de las cualidades de cartografía de las lecturas alternativas frente a las de referencia.

Encabezado	Descripción
INFO (continuación)	<ul style="list-style-type: none"> • ReadPosRankSum: puntuación Z de la prueba de suma de rangos de Wilcoxon del sesgo de posición de las lecturas alternativas frente a las de referencia. • AQ: puntuación de ruido sistemático. • hotspot: sitio somático conocido, utilizado para aumentar la confianza en la llamada. • SOMATIC: al menos una variante en esta posición es somática. <p>La columna Formato muestra campos separados por dos puntos. Por ejemplo, GT:GQ.</p>
FORMAT (Formato)	<p>Los campos disponibles para el flujo de trabajo Germline (germinal) incluyen:</p> <ul style="list-style-type: none"> • AD: profundidades alélicas (contando solo las lecturas informativas del total de lecturas) para los alelos de referencia y alternativos en el orden en que aparecen. • AF: fracciones de alelos para los alelos alternativos en el orden en que aparecen. • DP: profundidad aproximada de lectura (las lecturas con MQ=255 o con malas parejas se filtran). • F1R2: número de lecturas en orientación de pares F1R2 que apoyan cada alelo. • F2R1: número de lecturas en orientación de pares F2R1 que apoyan cada alelo. • GP: probabilidades posteriores en la escala Phred para los genotipos según se define en la especificación VCF. • GQ: calidad de genotipos. • GT: genotipo. 0 corresponde a la base de referencia, 1 corresponde a la primera entrada en la columna ALT, etc. La barra inclinada (/) indica que no hay disponible información sobre la fase de hebra retrasada. • MB: estadísticas de los componentes por muestra para detectar el sesgo de emparejamiento. • PL: verosimilitudes normalizadas en la escala Phred para los genotipos según se define en la especificación VCF. • PRI: probabilidades anteriores en la escala Phred para los genotipos. • PS: información de ID de fase de hebra retrasada física, en donde cada ID único en una muestra dada (pero no entre muestras) conecta los registros dentro de un grupo en fase de hebra retrasada. • SB: estadísticas de componentes por muestra que incluyen la prueba exacta de Fisher para detectar cortes en la cadena. • SQ: calidad somática. <p>Los campos disponibles en el flujo de trabajo somático incluyen:</p> <ul style="list-style-type: none"> • AD: profundidades alélicas (contando solo las lecturas informativas del total de lecturas) para los alelos de referencia y alternativos en el orden en que aparecen.

Encabezado	Descripción
FORMAT (continuación)	<ul style="list-style-type: none"> • AF: fracciones de alelos para los alelos alternativos en el orden en que aparecen. • DP: profundidad aproximada de lectura (las lecturas con MQ=255 o con malas parejas se filtran). • F1R2: número de lecturas en orientación de pares F1R2 que apoyan cada alelo. • F2R1: número de lecturas en orientación de pares F2R1 que apoyan cada alelo. • GT: genotipo. 0 corresponde a la base de referencia, 1 corresponde a la primera entrada en la columna ALT, etc. La barra inclinada (/) indica que no hay disponible información sobre la fase de hebra retrasada. • MB: estadísticas de los componentes por muestra para detectar el sesgo de emparejamiento. • PS: información de ID de fase de hebra retrasada física, en donde cada ID único en una muestra dada (pero no entre muestras) conecta los registros dentro de un grupo en fase de hebra retrasada. • SB: estadísticas de componentes por muestra que incluyen la prueba exacta de Fisher para detectar cortes en la cadena. • SQ: calidad somática.
SAMPLE (Muestra)	La columna Muestra da los valores especificados en la columna FORMAT (Formato).

Archivos VCF de genoma

Los archivos VCF (*.gvcf.gz) de genoma siguen una serie de convenciones para representar todos los sitios dentro del genoma en un formato razonablemente compacto. Los archivos gVCF incluyen todos los sitios de la región de interés en un solo archivo para cada muestra. El archivo gVCF muestra ausencia de llamadas en las posiciones que no superan todos los filtros. Un marcador ./ de genotipo (GT) indica una ausencia de llamadas.

Visualización de los resultados del análisis

Los experimentos que estén en curso se muestran en la pestaña Active (Activos). Los experimentos completados se muestran en la pestaña Completed (Finalizados). Consulte el [Documentación del producto NovaSeq 6000Dx \(n.º de documento 200010105\)](#) para obtener más información sobre cómo ver los resultados.

Asistencia técnica

Si necesita asistencia técnica, póngase en contacto con el servicio de asistencia técnica de Illumina.

Sitio web: www.illumina.com
Correo electrónico: techsupport@illumina.com

Números del servicio de asistencia técnica de Illumina

Región	Teléfono gratuito	Internacional
Australia	+61 1800 775 688	
Austria	+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
Finlandia	+358 800 918 363	+358 9 7479 0110
Francia	+33 8 05 10 21 93	+33 1 70 77 04 46
Alemania	+49 800 101 4940	+49 89 3803 5677
Hong Kong (China)	+852 800 960 230	
India	+91 8006500375	
Indonesia		0078036510048
Irlanda	+353 1800 936608	+353 1 695 0506
Italia	+39 800 985513	+39 236 00 37 59
Japón	+81 0800 111 5011	
Malasia	+60 1800 80 6789	
Países Bajos	+31 800 022 2493	+31 20 713 2960
Nueva Zelanda	+64 800 451 650	
Noruega	+47 800 16 836	+47 21 93 96 93
Filipinas	+63 180016510798	
Singapur	+1 800 579 27 45	
Corea del Sur	+82 80 234 53 00	

Región	Teléfono gratuito	Internacional
España	+34 800 300 143	+34 911 899 417
Suecia	+46 2 00883979	+46 8 50619671
Suiza	+41 800 200 442	+41 56 580 00 00
Taiwán, China	+886 8 06651752	
Tailandia	+66 1800 011 304	
Reino Unido	+44 800 012 6019	+44 20 7305 7197
Estados Unidos	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

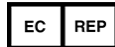
Hojas de datos de seguridad (SDS): Disponibles en el sitio web de Illumina, support.illumina.com/sds.html.

Documentación del producto: Disponible para su descarga de support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 (EE. UU.)
+1 800 809 ILMN (4566)
+1 858 202 4566 (fuera de Norteamérica)
techsupport@illumina.com
www.illumina.com

CE



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

Patrocinador australiano

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

PARA USO DIAGNÓSTICO IN VITRO

© 2022 Illumina, Inc. Todos los derechos reservados.

illumina[®]