

Módulo de análisis Somatic Variant (Variante somática) de Local Run Manager

Guía de flujo de trabajo para MiSeqDx

PARA USO DIAGNÓSTICO IN VITRO

Descripción general	3
Introducir la información sobre el experimento	3
Métodos de análisis	5
Ver un experimento y los datos sobre las muestras	7
Informe del análisis	7
Archivos de resultados de análisis	9
Asistencia técnica	17



Este documento y su contenido son propiedad 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 este 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.

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

© 2017 Illumina, Inc. Todos los derechos reservados.

Illumina, MiSeqDx y el diseño de las bases de streaming son marcas comerciales registradas o pendientes de Illumina, Inc. o sus afiliados en EE. UU. o en otros países. Todos los demás nombres, logotipos y marcas comerciales son propiedad de sus respectivos dueños.

Descripción general

El módulo de análisis Somatic Variant (variante somática) de Local Run Manager se debe utilizar con el ensayo del kit de amplicones personalizados TruSeq Dx de Illumina. Cuando se utiliza con el módulo Somatic Variant (Variante somática), este ensayo está diseñado para la preparación de bibliotecas que van a servir para la secuenciación de ADN de tejidos fijados en formol y embebidos en parafina (FFEP). El ensayo detecta mutaciones somáticas en frecuencias de variante bajas.

Este módulo de análisis evalúa regiones cortas de ADN amplificado, o amplicones, para detectar variantes. La secuenciación de amplicones enfocada permite una alta cobertura de regiones particulares a lo largo de un gran número de muestras. El módulo de análisis lleva a cabo un análisis secundario y genera informes teniendo en cuenta los experimentos de secuenciación desde un enfoque de cadena doble, que implica lecturas directas e inversas de los conjuntos de oligonucleótidos. Consulte el prospecto del *Kit de amplicones personalizados TruSeq Dx* (documento con n.º de referencia 1000000029772).

Para utilizar el módulo de análisis Somatic Variant (Variante somática) deberán estar disponibles los consumibles de secuenciación del Kit de reactivos MiSeqDx v. 3. Consulte el prospecto del *Kit de reactivos MiSeqDx v. 3* (documento con n.º de referencia 1000000030849).

Acerca de esta guía

Esta guía facilita las instrucciones necesarias para configurar los parámetros de experimentos de secuenciación y de análisis del módulo de análisis Somatic Variant (Variante somática). Para obtener más información sobre el panel de Local Run Manager y la configuración del sistema, consulte la *Guía de referencia del software Local Run Manager* (documento con n.º de referencia 1000000011880).

Introducir la información sobre el experimento

Configuración de parámetros

- 1 Inicie sesión en Local Run Manager.
- 2 Haga clic en **Create Run** (Crear experimento) y elija **Somatic Variant** (Variante somática).
- 3 Introduzca un nombre para el experimento de forma que se pueda identificar desde la secuenciación hasta el análisis.
Utilice caracteres alfanuméricos, espacios o guiones tanto normales como bajos.
- 4 (Opcional) Escriba una descripción que ayude a identificar el experimento.
Utilice caracteres alfanuméricos, espacios o guiones tanto normales como bajos.
- 5 Seleccione el número de muestras y la configuración del índice de la lista desplegable.

Importar los archivos de manifiesto para el experimento

- 1 Asegúrese de que los manifiestos que necesita importar están a su disposición a través de una ubicación accesible de la red o de un dispositivo USB.
- 2 Haga clic en **Import Manifest** (Importar manifiesto).
- 3 Vaya al archivo de manifiesto y elija los manifiestos que quiere añadir.



NOTA

Para que los archivos de manifiesto estén disponibles para cualquier experimento en el que se utilice el módulo de análisis Somatic Variant (Variante somática), añada los manifiestos por medio de la función Module Settings (Ajustes del módulo). Para acceder a esta función es necesario obtener permisos de administrador. Si desea obtener más información, consulte la *Guía de referencia del software Local Run Manager* (n.º de documento 1000000011880).


Definir las muestras para el experimento

Defina las muestras que se van a utilizar en el experimento por medio de una de las siguientes opciones.

- ▶ **Introducción manual de las muestras:** utilice la tabla en blanco que aparece en la pantalla Create Run (Crear experimento).
- ▶ **Importar las muestras:** vaya a un archivo externo que tenga un formato de valores separados por comas (*.csv). En la pantalla Create Run (Crear experimento) se ofrece una plantilla que se puede descargar.

Una vez se ha rellenado la tabla de muestras se puede exportar su información a un archivo externo para poder utilizarlo después como referencia a la hora de preparar las bibliotecas o se puede importar el archivo para llevar a cabo otro experimento.

Introducir las muestras de forma manual

- 1 Introduzca un nombre de muestra único en el campo Sample Name (Nombre de la muestra). Puede utilizar caracteres alfanuméricos y guiones tanto normales como bajos. El nombre de la muestra aparece automáticamente en el pocillo correspondiente en el otro conjunto.
- 2 (Opcional) Para las muestras de control tanto negativo como positivo, haga clic en el botón derecho del ratón y seleccione el tipo de control. El control que se asigna al pocillo de la muestra aparece automáticamente en el pocillo correspondiente en el otro conjunto y asigna ese mismo control.
- 3 (Opcional) Introduzca la descripción de la muestra en el campo Sample Description (Descripción de la muestra). Puede utilizar caracteres alfanuméricos y guiones tanto normales como bajos. La descripción de la muestra aparece automáticamente en el pocillo correspondiente en el otro conjunto. Las descripciones de la muestra se asocian con su ID. Las descripciones se sobrescribirán si se utiliza de nuevo el mismo ID de muestra en un experimento posterior.
- 4 Seleccione un adaptador del Index 1 (Índice 1) del menú desplegable que aparece en el Índice 1 (i7).
- 5 Seleccione un adaptador del Index 2 (Índice 2) del menú desplegable que aparece en el Índice 2 (i5).
- 6 Elija un archivo de manifiesto de la lista que aparece en el desplegable del campo Manifest (Manifiesto). Las muestras que se encuentran en el conjunto A precisan un manifiesto diferente al de las muestras del conjunto B.
- 7 Seleccione la opción correspondiente para visualizar, imprimir o guardar la disposición de la placa si desea utilizarla como referencia a la hora de preparar bibliotecas:
 - ▶ Haga clic en el icono  **Print** (Imprimir) para ver la disposición de la placa. Seleccione **Print** (Imprimir) para imprimir la disposición de la placa.
 - ▶ Haga clic en **Export** (Exportar) para exportar la información sobre las muestras a un archivo externo. Asegúrese de que el manifiesto y la información sobre la muestra son correctos. Si la información no es la

adecuada, los resultados se pueden ver afectados.

- 8 Haga clic en **Save Run** (Guardar Experimento).

Importar las muestras

- 1 Haga clic en **Import Samples** (Importar muestras) y busque la ubicación del archivo que contiene la información sobre la muestra. Se pueden importar dos tipos de archivo.
 - ▶ Para crear una nueva disposición de la placa, haga clic en **Template** (Plantilla) en la pantalla Create Run (Crear un experimento). El archivo que se usa como plantilla incluye los encabezados de columna adecuados para la importación. Introduzca en cada columna la información relativa a las muestras del experimento. Elimine la información que aparece como ejemplo en las celdas que no vaya a utilizar y guarde el archivo.
 - ▶ Emplee uno de los archivos de información sobre las muestras que se hayan exportado del módulo Somatic Variant (Variante somática) por medio de la función Export (Exportación).
- 2 Haga clic en el icono  **Print** (Imprimir) para ver la disposición de la placa.
- 3 Seleccione **Print** (Imprimir) para poder imprimir la disposición de la placa y utilizarla como referencia a la hora de preparar las bibliotecas.
- 4 (Opcional) Haga clic en **Export** (Exportar) para exportar la información sobre las muestras a un archivo externo.

Asegúrese de que el manifiesto y la información sobre la muestra son correctos. Si la información no es la adecuada, los resultados se pueden ver afectados.
- 5 Haga clic en **Save Run** (Guardar Experimento).

Editar un experimento

Para obtener instrucciones sobre cómo editar la información del experimento antes de comenzar la secuenciación, consulte la *Guía de referencia del software Local Run Manager (documento con n.º de referencia 1000000011880)*.

Métodos de análisis

El módulo de análisis Somatic Variant (Variante somática) lleva a cabo los siguientes pasos correspondientes al análisis y después genera los archivos de resultados que ubicará en la carpeta Alignment (Alineación).

- ▶ Lecturas de índice de demultiplexado
- ▶ Generación de archivos FASTQ
- ▶ Alineación de acuerdo con una referencia
- ▶ Identificación de variantes

Demultiplexado

El demultiplexado compara cada secuencia de lectura de índice con la secuencia de índice que se ha fijado para el experimento. En este paso, no se considera ningún valor de calidad.

Las lecturas de 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 de índice.

Generación de archivos FASTQ

Tras el demultiplexado, el software genera archivos de análisis intermedios en el formato FASTQ, que es un formato de texto empleado para representar secuencias. Los archivos FASTQ incluyen las lecturas de cada muestra con sus puntuaciones de calidad correspondientes. Los grupos que no han podido pasar los filtros quedan excluidos.

Cada archivo FASTQ contiene las lecturas de una sola muestra, cuyo nombre está incluido en el nombre del archivo FASTQ. Los archivos FASTQ son la entrada principal para la alineación. Se generan cuatro archivos FASTQ por cada muestra, dos con relación al conjunto A y dos al conjunto B.

Alineación

A lo largo del proceso de alineación, el algoritmo Smith-Waterman alinea las agrupaciones de cada muestra con las secuencias del amplicón que determina el archivo de manifiesto.

El algoritmo Smith-Waterman lleva a cabo alineaciones de secuencias semiglobales para establecer las regiones similares que hay en dos secuencias. En lugar de comparar la secuencia total, el algoritmo Smith-Waterman compara los segmentos de todas las longitudes posibles.

En el caso de la lectura "paired-end", se evalúa teniendo en cuenta su alineación respecto a las secuencias de sonda significativas de dicha lectura.

- ▶ La Lectura 1 se evalúa de acuerdo con el complemento inverso de los oligonucleótidos descendentes de locus específico (DLSO).
- ▶ La Lectura 2 se evalúa teniendo en cuenta oligonucleótidos ascendentes de locus específico (ULSO).
- ▶ Si el comienzo de una lectura observa una coincidencia con la secuencia de sonda con menos de tres diferencias (discrepancias o cambios producidos por las inserciones y deleciones), la lectura completa se alineará con el amplicón objetivo de esa secuencia.
- ▶ Las inserciones y deleciones dentro de los DLSO y los ULSO no se observan debido a la química del ensayo.

Las alineaciones se filtran teniendo en cuenta los resultados de alineación, que se basan en los índices de discrepancia que se encuentran en la región estudiada o en el amplicón completo, dependiendo de la longitud de este último. Las alineaciones filtradas se copian en los archivos de alineación como no alineadas y por tanto no se utilizan para la llamada de variantes.

Llamadas de variantes

Desarrollado por Illumina, Pisces Variant Caller (Llamador de variantes Pisces) identifica las variantes que están presentes con una frecuencia baja en una muestra de ADN.

El llamador de variantes Pisces identifica los SNP en tres pasos:

- ▶ Tiene en cuenta cada posición en el genoma de referencia de forma independiente
- ▶ Cuenta las bases en la posición que se ha determinado para las lecturas alineadas que se solapan en la posición.
- ▶ Calcula una puntuación de variante que mide la calidad de la llamada a través del modelo Poisson. Se excluyen todas las variantes cuya puntuación de calidad sea menor al Q30.

Al principio las variantes se llaman en cada conjunto por separado. Después, las variantes de cada conjunto se comparan y se combinan en un único archivo de resultados. Si una variante cumple los siguientes criterios, se señala como PASS (apta) en el archivo (VCF) de llamadas de variantes:

- ▶ La variante está presente en ambos conjuntos
- ▶ La profundidad acumulada es de 900x (al menos 450x por conjunto)
- ▶ La frecuencia de la variante que consta en el archivo VCF combinado es igual o mayor al 2,6 %

Pisces (Llamador de variantes somáticas)

Pisces lleva a cabo llamadas de variantes somáticas con el objetivo de identificar variantes con una frecuencia baja en las muestras de ADN. La aplicación funciona con todo tipo de muestras y genera archivos VCF y gVCF.

Si desea obtener más información, consulte la página github.com/Illumina/Pisces/wiki.

Ver un experimento y los datos sobre las muestras

- 1 Una vez en el panel principal de Local Run Manager, haga clic en el nombre del experimento.
- 2 En la ficha Run Overview (Resumen del experimento), revise los criterios de medición del experimento de secuenciación.
- 3 (Opcional) Haga clic en el icono **Copy to Clipboard**  (Copiar en el portapapeles) para copiar la ruta de la carpeta en la que se encuentran los resultados del experimento.
- 4 Haga clic en la ficha Sequencing Information (Información sobre la secuenciación) para verificar los parámetros del experimento y la información sobre los consumibles.
- 5 Haga clic en la ficha Samples and Results (Muestras y resultados) para ver dónde se encuentran ubicados los informes del análisis.
 - ▶ En caso de repetir el análisis, abra el menú desplegable Select Analysis (Seleccionar análisis) para seleccionar un análisis y escoja el que proceda.
- 6 Haga clic en el icono **Copy to Clipboard**  (Copiar en el portapapeles) para copiar la ruta de la carpeta Analysis (Análisis).

Para obtener más información sobre las fichas Run Overview (Resumen del experimento) y Sequencing Information (Información sobre la secuenciación), así como sobre cómo volver a poner un análisis en cola, consulte la *Guía de referencia del software Local Run Manager (documento con n.º de referencia 1000000011880)*.

Informe del análisis

Los resultados del análisis se encuentran resumidos en la ficha Samples and Results (Muestras y resultados) y en la carpeta Alignment (Alineación) en forma de informe global. Existe la posibilidad de encontrar un informe individual de cada muestra en archivos con formato PDF.

Información sobre la ficha Sample and Details (Muestra e información detallada)

1 Haga clic en una de las muestras del listado para ver el informe de muestras.

Tabla 1 Tabla Run Information (Información del experimento)

Encabezado de columna	Descripción
Estado del experimento	Determina si el experimento de secuenciación es apto o no.
Rendimiento total (GB)	Número de bases que se han llamado en el experimento de secuenciación. Muestra el umbral determinado para considerar el experimento como apto y su estado como apto o no apto.
% ≥ Q30 (Porcentaje ≥ Q30)	Porcentaje de lecturas realizadas en un experimento de secuenciación que han logrado una puntuación de calidad igual o mayor que 30 (Q30). Muestra el umbral determinado para considerar el experimento como apto y su estado como apto o no apto.
Sample Name (Nombre de la muestra)	Nombre de la muestra que se introdujo al crear el experimento.
Total PF Reads (Total de lecturas PF)	Cantidad total de lecturas que han pasado el filtro.
Read 1% ≥ Q30 (Porcentaje de lectura 1 ≥ Q30)	Porcentaje de lecturas llevadas a cabo en la lectura 1 que han logrado una puntuación de calidad para la muestra igual o mayor que 30 (Q30).
Read 2% ≥ Q30 (Porcentaje de lectura 2 ≥ Q30)	Porcentaje de lecturas llevadas a cabo en la lectura 2 que han logrado una puntuación de calidad para la muestra igual o mayor que 30 (Q30).
Autosome call rate (Índice de llamada de autosomas)	Número de posiciones genómicas en los autosomas (desde el cromosoma 1 hasta el 22) que cumplen un umbral de valor de confianza predeterminado dividido entre la cantidad total de posiciones genómicas autosómicas interrogadas. El índice de llamada se describe por muestra y se incluye en el informe en forma de porcentaje que se calcula como uno menos (número de posiciones autosómicas con llamadas incompletas dividido entre el número total de posiciones autosómicas secuenciadas).

Tabla 2 Tabla Sample Reports (Informes de muestra)

Encabezado de columna	Descripción
Muestra	Nombre de la muestra que se introdujo al crear el experimento.
Report Date (Fecha del informe)	Fecha en la que se generó el informe.
Información sobre la muestra	ID de muestra que se proporcionó en el momento de la creación del experimento, la cantidad total de lecturas de la muestra que han pasado el filtro, el porcentaje de lecturas que han obtenido una puntuación de calidad igual o mayor que 30 (Q30) y el índice de llamada autosómico.
Amplicon Summary (Resumen del amplicón)	Cantidad total de regiones de amplicón que se han secuenciado y longitud total en pares de bases de los amplicones que se han secuenciado en las regiones objetivo, tanto de las muestras del conjunto A como de las del conjunto B, así como el archivo de manifiesto que se ha utilizado para cada grupo. El archivo de manifiesto indica el genoma de referencia y las regiones de referencia objetivo que se han utilizado en el paso de alineación.

Encabezado de columna	Descripción
Read Level Statistics (Estadísticas de lectura)	Cantidad y porcentaje de lecturas llevadas a cabo en la muestra que cubren todas las posiciones en la referencia incluidos en la lectura 1 y la lectura 2 en el conjunto A y en el conjunto B.
Variants Summary (Resumen de variantes)	Indica la cantidad total de SNV, inserciones y deleciones que se han detectado en la muestra y que han superado los valores propuestos con el objetivo de definir si la calidad de los resultados se encuentra dentro del margen admisible.
Coverage Summary (Resumen de la cobertura)	Muestra la cantidad total de bases alineadas dividida por el tamaño de la región objetivo y el porcentaje de las regiones del amplicón con valores de cobertura mayores que los del umbral de baja cobertura ($0,2 * \text{cobertura media del amplicón}$) de la muestra en el conjunto A y en el conjunto B.
Coverage Plot (Diagrama de cobertura)	El diagrama de la cobertura por cada región del amplicón muestra la cobertura de todas las regiones del amplicón de la muestra. Las regiones cuyos valores de cobertura son menores que el valor del umbral de cobertura se destacan en rojo. La media de la totalidad de los valores se indica con una línea naranja. Se presenta un diagrama con la cobertura del conjunto A y del conjunto B.
Software Versions (Versiones del software)	Muestra las versiones del software en el momento que se ha secuenciado la muestra. Contiene la versión de MiSeq Operating Software, del software Local Run Manager, del software RTA y la versión del módulo Somatic Variant (Variante somática).

Archivos de resultados de análisis

Se generan los siguientes archivos de resultado de análisis para el módulo de análisis Somatic Variant (Variante somática), que proporcionan los resultados de los análisis de alineación y de llamada de variantes. Los archivos de resultados de análisis están ubicados en la carpeta Alignment (Alineación).

Nombre de archivo	Descripción
Demultiplexado (*.txt)	Son archivos intermedios que contienen un resumen de los resultados del demultiplexado.
FASTQ (*.fastq.gz)	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.
Archivos de alineación en formato BAM (*.bam)	Contienen las lecturas alineadas de una muestra específica.
Archivos de las llamadas de variantes por conjunto en formato VCF (*.vcf)	Contiene las variantes que se han llamado en cada posición en cualquiera de los dos conjuntos teniendo en cuenta la lectura directa o la lectura inversa.
Archivos de llamada de variantes en formato VCF de genoma (*.genome.vcf.gz)	Contiene el genotipo para cada posición, ya sea porque se ha llamado como variante o como referencia.
Archivos de consenso de llamadas de variantes en formato VCF (*.vcf.gz)	Incluye las variantes que se han llamado por cada posición en ambos conjuntos.
AmpliconCoverage_M1.tsv	Incluye la información sobre la cobertura por cada amplicón de cada muestra para cada manifiesto que se ha proporcionado. M# hace referencia al número de manifiesto.

Formato de archivo del demultiplexado

El proceso de demultiplexado lee la secuencia de índice asociada a cada grupo para determinar en qué muestra se ha originado. La correspondencia entre los grupos y los números de muestra se registra en un archivo de demultiplexado (*.demux) por cada placa de la celda de flujo.

El formato de nomenclatura del archivo de demultiplexado es s_1_X.demux, en el que X representa el número de placa.

Los archivos de demultiplexado comienzan con un encabezado:

- ▶ Versión (número entero de cuatro bytes), que actualmente es uno.
- ▶ Recuento de grupos (número entero de 4 bytes).

El resto del archivo contiene los números de muestra de cada grupo de la placa.

Cuando finaliza en proceso de demultiplexado, el software genera un archivo de demultiplexado que se denominará ResumendemultiplexadoF1L1.txt.

- ▶ En el nombre del archivo, F1 corresponde al número de la celda de flujo.
- ▶ En el nombre del archivo, L1 corresponde al número del carril.
- ▶ El demultiplexado da como resultado una tabla con una fila por cada placa y una columna por cada muestra, incluida la muestra cero.
- ▶ Las secuencias más habituales en las lecturas de índice.

Formato de archivo FASTQ

El formato FASTQ es un formato de archivo de texto que engloba las llamadas a las bases y los valores de calidad por cada lectura. Cada registro incluye cuatro líneas:

- ▶ Identificador
- ▶ Secuencia
- ▶ Signo de suma (+)
- ▶ Una puntuación de calidad según la escala Phred en formato encriptado ASCII + 33

El identificador tiene 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=AAAAAAAAAA9#:<#<;<<<????#=#
```

Formato de archivo BAM

Un archivo BAM (*.bam) es la versión binaria comprimida de un archivo SAM que se utiliza para representar secuencias alineadas de hasta 128 Mb. Puede encontrar la descripción detallada de los formatos SAM y BAM en samtools.github.io/hts-specs/SAMv1.pdf.

Los archivos BAM emplean el formato de nomenclatura de archivos SampleName_S#.bam, donde # es el número de la muestra determinado por el orden en que aparecen las muestras en el experimento.

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

- ▶ **Encabezado:** contiene información sobre todo el archivo como, por ejemplo, el nombre de la muestra, su longitud y el método de alineación. Las alineaciones de la sección de alineaciones están asociadas a la información específica de la sección del encabezado.
- ▶ **Alineaciones:** contiene el nombre, la secuencia y la calidad de la lectura e información sobre la alineación, así como etiquetas personalizadas. El nombre de la lectura incluye el cromosoma, la coordenada de inicio, la calidad de la alineación y la secuencia del descriptor de la coincidencia.

La sección de alineaciones contiene la siguiente información por cada lectura o par de lectura:

- ▶ **AS:** calidad de alineación "paired-end".
- ▶ **BC:** etiqueta de código de barras, que indica el ID de la muestra demultiplexada que está vinculado a la lectura.
- ▶ **SM:** calidad de alineación "single-end".
- ▶ **XC:** secuencia del descriptor de la coincidencia.
- ▶ **XN:** etiqueta con el nombre del amplicón, que registra el ID del amplicón vinculado a la lectura.

Los archivos de índice BAM (*.bam.bai) facilitan un índice del archivo BAM correspondiente.

Formato de archivo VCF

Variant Call Format (Formato de llamada de variantes) (VCF) es un formato de archivo que ha desarrollado la comunidad científica que trabaja con la genómica. Contiene información sobre las variantes que se encuentran en posiciones específicas de un genoma de referencia. Los archivos VCF terminan con el sufijo ".vcf".

El encabezado del archivo VCF incluye la versión del formato de archivo VCF y la versión del llamador de variantes, además de mostrar un listado de las anotaciones que se utilizan en el resto del archivo. El encabezado del VCF también contiene el archivo del genoma de referencia y un archivo BAM. La última línea del encabezado incluye los encabezados de columna para las líneas de datos. Cada una de las líneas de datos del archivo VCF contiene información sobre una variante.

Encabezados de archivo VCF

Encabezado	Descripción
CHROM	El cromosoma del genoma de referencia. Los cromosomas aparecen en el mismo orden que el archivo FASTQ de referencia.
POS	La posición de base única de la variante en el cromosoma de referencia. En el caso de los SNP, esta posición es la base de referencia con la variante. En el caso de las inserciones y deleciones o las deleciones, esta posición es la base de referencia justo antes de la variante.
ID	Corresponde al número rs para la variante obtenida a partir del dbSNP.txt, si procede. Si existen varios números rs en esta posición, la lista queda delimitada por punto y coma. Si no existen entradas de dbSNP en esta posición, se utiliza un marcador de valor ausente ('.').
REF	El genotipo de referencia. Por ejemplo, una deleción de una única 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 única 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	Puntuación de calidad según la escala Phred asignada por el llamador de variantes. Las puntuaciones más altas indican una mayor seguridad 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 de Q30 tiene un índice de error del 0,1 %. Muchos llamadores de variantes asignan puntuaciones de calidad según sus modelos estadísticos, que son altas en relación con el índice de error observado.

Anotaciones de archivos VCF

Encabezado	Descripción
FILTER (FILTRO)	Si se superan todos los filtros, aparece PASS (apto) en la columna de filtrado. <ul style="list-style-type: none"> • LowDP (DP baja): se aplica a los lugares con una profundidad de cobertura por debajo de 450x en cualquiera de los dos conjuntos. Esta medida es equivalente a una cobertura de 900x en una sola lectura cuando nos referimos a las posiciones del amplicón que están cubiertas por las lecturas que se realizan tanto hacia adelante como hacia atrás. • LowGQ (CG baja): la calidad del genotipado (CG) se encuentra por debajo del corte. • q30: la puntuación de calidad es inferior a 30. • LowVariantFreq (Frecuencia de variante baja): la frecuencia de la variante es inferior al umbral asignado. • PB: sesgo en el conjunto de sonda. No se ha encontrado la variante o bien se ha encontrado con una frecuencia muy baja en uno o ambos conjuntos de sonda. • R3x6: la cantidad de repeticiones adyacentes (de una longitud de entre 1 y 3 pb) en las llamadas de variantes es igual o mayor que seis. • SB: el sesgo en la cadena es mayor que el umbral que se ha determinado.
INFO (INFORMACIÓN)	Las entradas que se pueden encontrar en la columna INFO (Información) incluyen: <ul style="list-style-type: none"> • AC: recuento de alelos en genotipos de cada alelo ALT, en el mismo orden en que se enumeran. • AF: frecuencia de alelos de cada alelo ALT, en el mismo orden en que se enumeran. • AN: número total de alelos en genotipos llamados. • CD: señal indicativa de que el SNP se encuentra dentro de la región de codificación de, al menos, una entrada de RefGene. • DP: profundidad (número de llamadas de bases alineadas en una posición y que se utilizan en las llamadas de variantes). • Exon: lista de valores separados por comas que incluye las regiones de exones leídos en RefGene. • FC: consecuencia funcional. • GI: lista de valores separados por comas que incluye los ID de genes leídos en RefGene. • QD: seguridad de la variante y calidad por profundidad. • TI: lista de valores separados por comas que incluye los ID de transcripción leídos en RefGene.

Encabezado	Descripción
FORMAT (FORMATO)	<p>La columna de formato muestra en una lista los campos separados por dos puntos. Por ejemplo, GT:GQ. La lista de campos que se muestra depende del llamador de variantes que se haya utilizado. Los campos de los que se dispone incluyen:</p> <ul style="list-style-type: none"> • AD: entrada con formato X,Y, donde X es el número de llamadas de referencia e Y es el número de llamadas alternativas. • DP: profundidad de lectura aproximada; se filtran las lecturas con MQ = 255 o con emparejamientos erróneos. • GQ: calidad del genotipo. • GQX: calidad del genotipo. GQX es el mínimo del valor de GQ y la columna QUAL. En general, estos valores son similares. Si se tiene en cuenta el mínimo, GQX indica la medición más fiable relativa a la calidad del genotipo. • GT: genotipo. 0 corresponde a la base de referencia, 1 corresponde a la primera entrada en la columna ALT y, así, sucesivamente. La barra inclinada (/) indica que no hay información de hebras retrasadas disponible. • NC: proporción de las bases a las que no se ha llamado o una calidad de llamada de bases inferior al umbral mínimo. • NL: nivel de ruido. Estimación del ruido en la llamada a las bases en la posición correspondiente. • PB: sesgo en el conjunto de sonda. Los valores que se acercan a cero indican un sesgo mayor hacia uno de los conjuntos de sonda y menor grado de confianza respecto a la llamada de variantes. • SB: sesgo de la cadena en la posición correspondiente. Los valores negativos más grandes indican un sesgo menor, mientras que las posiciones cercanas a cero indican uno mayor. • VF: frecuencia de variante. Se trata del porcentaje de lecturas del alelo alternativo.
SAMPLE (MUESTRA)	La columna SAMPLE (MUESTRA) incluye los valores especificados en la columna FORMAT (FORMATO).

Archivos VCF de genoma

Los archivos VCF de genoma (gVCF) son archivos VCF en versión 4.1 que obedecen a una serie de convenciones para representar todas las posiciones que hay en un genoma en un formato bastante compacto. Los archivos gVCF (*.genome.vcf.gz) albergan todas las posiciones dentro de cada región estudiada en un solo archivo por cada muestra.

Los archivos gVCF no muestran las llamadas en las posiciones que no han pasado los filtros. Etiqueta de genotipo (GT) de ./.. Indica que no se ha hecho llamada.

Para obtener más información, consulte sites.google.com/site/gvcftools/home/about-gvcf.

Archivos VCF por conjunto y de consenso

El flujo de trabajo de Somatic Variant (Variante somática) genera dos grupos de archivos de llamadas de variantes.

- ▶ **Archivos VCF por conjunto**: incluye las variantes que se han llamado o en la lectura directa del conjunto o en la lectura inversa. Los archivos por conjunto se generan en la carpeta VariantCallingLogs (Registros de llamada de variantes).
- ▶ **Archivos VCF de consenso**: contienen las variantes de ambos conjuntos. Los archivos de consenso se generan en la carpeta Alignment (Alineación).

Los archivos VCF por conjunto o de consenso contienen archivos VCF (*.vcf) y gVCF (*.genome.vcf) y su denominación se basa en la siguiente convención de nomenclatura, en la que los caracteres S# representan el orden que ocupa la muestra en el experimento:

- ▶ **Informes de todas las posiciones**: SampleName_S#.genome.vcf
- ▶ **Informe únicamente de variantes**: SampleName_S#.vcf

El software compara los archivos VCF por conjunto y combina los datos de cada posición para generar un archivo VCF de consenso para la muestra.

Las llamadas de variantes de cada grupo se incorporan a los archivos VCF de consenso por medio de los siguientes criterios:

Criterios	Resultado
Una llamada de referencia en cada conjunto	Llamada de referencia
Una llamada de referencia en un conjunto y una llamada de variante en el otro conjunto	Llamada de variantes con filtro
Llamadas de variantes coincidentes con frecuencias parecidas en cada conjunto	Llamadas de variantes
Llamadas de variante coincidentes con frecuencias sustancialmente distintas en cada conjunto	Llamada de variantes con filtro
Llamadas de variantes en las que no hay coincidencias en cada conjunto	Llamada de variantes con filtro

Los criterios de medición de cada conjunto se combinan teniendo en cuenta los valores que se exponen a continuación.

Criterio de medición	Valor
Profundidad	Incorporación de profundidades de ambos conjuntos
Frecuencia de variantes	Cálculo de la cantidad total de variantes dividido por la profundidad de cobertura total
Puntuación Q	Valor mínimo de ambos conjuntos

Archivo de cobertura de amplicones

Se genera un archivo de cobertura de amplicones por cada archivo de manifiesto. Los caracteres M# del nombre del archivo representan el número de manifiesto tal y como se muestra en la tabla de muestras del experimento.

Cada archivo contiene una fila de encabezado que incluye los ID de las muestras asociadas con el manifiesto. Debajo del encabezado aparecen tres columnas que muestran la siguiente información:

- ▶ El ID objetivo tal y como aparece en el manifiesto.
- ▶ La profundidad de cobertura de las lecturas que han pasado el filtro.
- ▶ La profundidad de cobertura total.

Archivos de resultados complementarios

Los siguientes archivos de resultados proporcionan información complementaria. También pueden incluir resúmenes de errores de análisis y resultados de experimentos. Aunque estos archivos no resultan necesarios para evaluar los resultados de los análisis, se pueden utilizar para la solución de posibles problemas. Todos los archivos están ubicados en la carpeta Alignment (Alineación) a menos que se indique lo contrario.

Nombre de archivo	Descripción
AnalysisLog.txt	Registro de procesamiento que describe todos los pasos que se han seguido durante el análisis de la carpeta del experimento actual. Este archivo no contiene mensajes de error. Está ubicado en la carpeta Alignment (Alineación).
AnalysisError.txt	Registro de procesamiento que enumera todos los errores que se han producido durante el análisis. Este archivo aparecerá vacío si no se produjo ningún error. Está ubicado en la carpeta Alignment (Alineación).
DemultiplexSummaryF1L1#.txt	Informa de los resultados del demultiplexado en una tabla con una fila por cada placa y una columna por cada muestra. El símbolo # indica el carril de la celda de flujo: uno, dos, tres o cuatro. Está ubicado en la carpeta Alignment (Alineación).
AmpliconRunStatistics.xml	Contiene las estadísticas de resumen específicas del experimento. Está ubicado en la carpeta Alignment (Alineación).

Carpeta de análisis

La carpeta de análisis contiene los archivos que genera el software Local Run Manager.


La relación entre la carpeta de resultados y la carpeta de análisis se resume de la siguiente manera:


- ▶ Durante la secuenciación, el análisis en tiempo real (RTA) llena la carpeta de resultados con los archivos que se han generado en los procesos de análisis de imágenes, de llamada a las bases y de puntuación de calidad.
- ▶ El RTA hace una copia en tiempo real de los archivos en la carpeta de análisis. Una vez el RTA ha vinculado la puntuación de calidad a cada base de cada ciclo, el software rellena el archivo RTAComplete.txt que se encuentra en ambas carpetas.
- ▶ Una vez existe el archivo RTAComplete.txt, el análisis puede comenzar a ejecutarse.
- ▶ Durante el análisis, Local Run Manager genera los archivos con los resultados y los envía a la carpeta de análisis al tiempo que hace copias de ellos en la carpeta de resultados.

Carpetas Alignment (Alineación)

Cada vez que se vuelve a poner en cola un experimento, Local Run Manager genera una carpeta Alignment (Alineación) con el nombre **Alignment_N**, donde N es un número secuencial.

Estructura de la carpeta

 **Alignment** (Alineación): contiene los archivos *.bam, *.vcf y FASTQ, así como los archivos específicos del módulo de análisis.

 **Date and Time Stamp** (Marca de fecha y hora): marca la fecha y la hora del análisis en el formato AAAAMMDD_HHMMSS.

 AnalysisError.txt

 AnalysisLog.txt

 AmpliconRunStatistics.xml

 Sample1.genome.vcf.gz

 Sample1.coverage.csv

 Sample1.report.pdf

 Sample1.summary.csv

 Sample1.vcf.gz

 Sample1.bam

 **FASTQ**

 **Sample1**

 **Stats**

 DemuxSummaryF1L1.txt

 FastqSummaryF1L1.txt

 **Data** (Datos)


 **Intensities** (Intensidades)


 **BaseCalls**

 **L001**: incluye una subcarpeta por ciclo en la que se pueden encontrar los archivos *.bcl.

 **L001**: contiene los archivos *.locs, uno por cada placa.

 **RTA Logs** (Registros RTA): incluye los archivos de registro del software de análisis RTA.

 **InterOp**: contiene los archivos binarios que se utilizan para informar sobre los criterios de medición de los experimentos de secuenciación.

 **Logs** (Registros): son los archivos de registro que señalan los pasos que se han seguido durante la secuenciación.

 RTAComplete.txt

 RunInfo.xml

 runParameters.xml

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 al cliente de Illumina

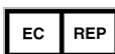
Zona	Teléfono gratuito	Regional
Norteamérica	+1.800.809.4566	
Alemania	+49 8001014940	+49 8938035677
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Bélgica	+32 80077160	+32 34002973
China	400.635.9898	
Dinamarca	+45 80820183	+45 89871156
España	+34 911899417	+34 800300143
Finlandia	+358 800918363	+358 974790110
Francia	+33 805102193	+33 170770446
Hong Kong	800960230	
Irlanda	+353 1800936608	+353 016950506
Italia	+39 800985513	+39 236003759
Japón	0800.111.5011	
Noruega	+47 800 16836	+47 21939693
Nueva Zelanda	0800451650	
Países Bajos	+31 8000222493	+31 207132960
Reino Unido	+44 8000126019	+44 2073057197
Singapur	+1.800.579.2745	
Suecia	+46 850619671	+46 200883979
Suiza	+41 565800000	+41 800200442
Taiwán	00806651752	
Otros países	+44.1799.534000	

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 en formato PDF en el sitio web de Illumina. Vaya a support.illumina.com, seleccione un producto y, a continuación, seleccione **Documentation & Literature** (Documentación y literatura).



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



Illumina Cambridge Limited
Chesterford Research Park, Little Chesterford
Saffron Walden, CB10 1XL
REINO UNIDO

Patrocinador australiano:
Illumina Australia
1 International Court
Scoresby, Victoria, 3179
Australia

PARA USO DIAGNÓSTICO IN VITRO

© 2017 Illumina, Inc. Todos los derechos reservados.

illumina®