

Análisis de CNV germinal con el flujo de trabajo de Illumina DNA Prep with Exome 2.5 Enrichment

Detección eficaz de variantes
de número de copias mediante
un panel de normales



Introducción

Las variantes en el número de copias (CNV, copy-number variants) son una fuente importante de diversidad genética en humanos.¹ Esta clase de variación genómica se produce cuando el número de copias de un gen o región genómica varía entre personas, e incluye grandes duplicaciones o deleciones genómicas. El análisis de CNV puede ayudar a los investigadores a identificar variantes asociadas a rasgos complejos o susceptibilidad a enfermedades, como el cáncer, enfermedades autoinmunitarias, trastornos genéticos hereditarios y más.¹⁻⁴

En comparación con el uso de microarrays para la detección de CNV, los métodos de secuenciación de nueva generación (NGS, next-generation sequencing) revelan más detalles sobre las variaciones estructurales. Además, la NGS puede detectar algunas variantes de menos de 50 kb, que los arrays no pueden detectar. Illumina DNA Prep with Exome 2.5 Enrichment forma parte de una solución de secuenciación del exoma completo (WES, whole-exome sequencing) humano de alto rendimiento, rápida y fiable que genera librerías compatibles con el análisis de CNV mediante NGS. En combinación con los sistemas de secuenciación y los procesos de análisis de Illumina, este kit de preparación de librerías se puede utilizar para detectar variantes de nucleótido único (SNV, single-nucleotide variants), pequeñas inserciones y deleciones (indel) y grandes CNV estructurales.

Esta nota de aplicación demuestra un flujo de trabajo de secuenciación del exoma optimizado para el análisis de CNV germinales, desde la preparación de librerías hasta la obtención de la información (figura 1). La solución de WES integra Illumina DNA Prep with Exome 2.5 Enrichment, la NGS de Illumina de eficacia probada y el análisis secundario de DRAGEN™ de alta precisión para la identificación de CNV.

Los usuarios también pueden emplear la plataforma de investigación de interpretación de variantes de Emedgene™, impulsada por inteligencia artificial explicable (XAI, eXplainable Artificial Intelligence). El software Emedgene permite el análisis y la obtención de información de variantes, la selección y la automatización de procedimientos operativos estándar (SOP, standard operating procedure) e incluye opciones optimizadas para la generación de informes de investigación.

Métodos

Muestras

Para facilitar la detección de CNV germinales en todo el exoma, los usuarios deben desarrollar un archivo de referencia de partida basado en un panel de muestras supuestamente normales para la comparación durante el análisis. Este panel de referencia de muestras normales se utiliza para normalizar los recuentos objetivo y permite la detección precisa de CNV durante la evaluación de muestras de casos. Las muestras para el panel de normales deben prepararse y secuenciarse en las mismas condiciones que las muestras del caso.

Para crear el panel de normales, se seleccionaron 54 muestras (Coriell Institute for Medical Research) sin CNV patógenas conocidas (tabla 1) para representar la diversidad de la variación natural entre superpoblaciones humanas (basadas en el 1000 Genomes Project, tabla 2).⁵ Para demostrar el rendimiento del flujo de trabajo de WES para la detección de CNV, se utilizaron 17 muestras de casos que contenían una CNV detectada previamente dentro de un gen o región de interés en un análisis con el panel de normales.



Figura 1: Flujo de trabajo completo de WES para llamadas de CNV. Combine Illumina DNA Prep with Exome 2.5 Enrichment, las plataformas de NGS de Illumina, el análisis de datos rápido con el análisis secundario de DRAGEN y la interpretación de variantes con el software Emedgene para lograr un flujo de trabajo fácil de usar y optimizado para una secuenciación precisa del exoma para interrogar las CNV.

Tabla 1: ID de muestras del Coriell Institute incluidas en el panel de normales

HG00096	HG01393	HG01950	HG03006	NA10830	NA12877	NA18502	NA18970	NA21112
HG00190	HG01441	HG01985	HG03870	NA10831	NA12878	NA18508	NA19681	NA24143
HG00262	HG01551	HG01990	HG03882	NA10835	NA12889	NA18622	NA19720	NA24149
HG00626	HG01599	HG02013	HG03898	NA10838	NA12890	NA18637	NA20509	NA24631
HG00628	HG01896	HG02348	HG04090	NA10839	NA12891	NA18942	NA20875	NA24694
HG01392	HG01914	HG02521	HG04214	NA12249	NA12892	NA18957	NA21098	NA24695

Preparación de librerías

Las librerías de secuenciación del exoma se prepararon por duplicado siguiendo el protocolo de enriquecimiento de Illumina DNA Prep with Exome 2.5 Enrichment (n.º de documento 1000000048041 v07) utilizando 50 ng de ADNg de muestras del Coriell Institute, a menos que se indique lo contrario. Se utilizaron IDT for Illumina DNA/RNA UD Indexes Sets A-D, Tagmentation (Illumina, n.º de catálogo 20027213, 20027214, 20042666 y 20042667) para permitir una indexación única en todas las librerías preparadas. Las librerías de preenriquecimiento se agruparon por masa (objetivo de 250 ng por muestra) a 12 unidades de plexado por grupo y se hibridaron durante la noche con Twist Bioscience for Illumina Exome 2.5 Panel, disponible como parte de los kits Illumina DNA Prep with Exome 2.5 Enrichment (Illumina, n.º de catálogo 20077595 y 20077596). Las librerías se prepararon con y sin adición de Twist Bioscience for Illumina Mitochondrial Panel (Illumina, n.º de catálogo 20093180). La adición del panel mitocondrial se diluyó en proporción 1:100 siguiendo la descripción de la Guía de referencia de Illumina DNA Prep with Exome 2.5 Enrichment (n.º de documento 1000000157112).

Secuenciación

Las librerías enriquecidas se secuenciaron en los sistemas NovaSeq™ 6000 y NextSeq™ 2000 utilizando reactivos de secuenciación por síntesis (SBS, sequencing by synthesis) y experimentos «paired-end» de 150 pb. Los resultados de datos más altos disponibles se utilizaron para lograr una alta profundidad de cobertura por muestra y permitir la reducción de muestras como parte del análisis. El panel de muestras normales y las muestras de casos se secuenciaron con NovaSeq 6000 S4 Reagent Kit v1.5 (300 ciclos) (Illumina, n.º de catálogo 20028312) y NextSeq 2000 P3 Reagents (300 ciclos) (Illumina, n.º de catálogo 20040561). Los kits de reactivos para NovaSeq 6000 System están diseñados para facilitar su uso con tres cartuchos listos para usar precargados con todos los reactivos necesarios.

Tabla 2: Resumen de atributos del panel de normales

Sexo	N.º de muestras
Masculino	27
Femenino	27
Superpoblación⁵	
Europea (EUR)	18
Americana mixta (AMR)	8
Africana (AFR)	7
Asiática del sur (SAS)	9
Asiática oriental (EAS)	12

Tabla 3: Rendimiento y cobertura de la secuenciación en todas las plataformas para panel de normales y muestras de CNV

Plataforma	Muestras por experimento	Cobertura media por muestra
NovaSeq 6000 System	72	~540×
NextSeq 2000 System	12	~280×

NextSeq 2000 System proporciona flexibilidad con múltiples configuraciones de celdas de flujo y ofrece análisis DRAGEN integrado. El número de muestras por experimento y la cobertura media alcanzada por muestra en todos los experimentos se muestran en la [tabla 3](#).

Análisis de datos

La aplicación DRAGEN Baseline Builder v4.2, accesible en BaseSpace™ Sequence Hub, se utilizó para crear el panel de archivos de recuentos de objetivos individuales de normales con el genoma de referencia no gráfico hg38. Los datos de secuenciación de las muestras de casos de CNV se redujeron a 80 M de lecturas (profundidad promedio de cobertura en el objetivo de aprox. 120-140×) y 50 M de lecturas (profundidad promedio de cobertura en el objetivo de aprox. 70-90×) y se analizaron con la aplicación DRAGEN Enrichment v4.2, también accesible en BaseSpace Sequence Hub, con la detección de CNV habilitada. Para cada plataforma, se utilizó el panel de normales correspondiente para la normalización de las muestras. La aplicación DRAGEN Enrichment realiza un análisis secundario preciso y eficiente para una llamada de variantes completa, incluidas, entre otras aplicaciones, SNV, indels, CNV y variantes estructurales (SV, structural variants). La llamada de variantes DRAGEN para Illumina DNA Prep with Exome 2.5 Enrichment también se puede realizar en el sistema NextSeq 2000 o totalmente integrada en el flujo de trabajo de investigación de interpretación de variantes del software Emedgene.

Resultados

Detección de CNV en un intervalo de variaciones

El análisis con el proceso de DRAGEN CNV en la aplicación DRAGEN Enrichment reveló una sólida detección de CNV para las variantes evaluadas en este estudio con tan solo 50 M de lecturas, o una profundidad promedio de cobertura de aproximadamente 70-90×, en todas las plataformas. La detección de CNV se determina mediante la fórmula: (número de regiones llamadas)/(regiones que se solapan con el panel del exoma y la verdad). Las CNV se consideraron detectadas si este criterio de medición era >98 % (tabla 4).

Varias de las variantes analizadas se encuentran en el gen *DMD*, que es una diana importante en la investigación en enfermedades genéticas (tabla 4).⁶ Mediante Integrated Genomics Viewer, se pueden visualizar los eventos de CNV en *DMD* esperados detectados para las siete muestras afectadas analizadas (figura 2). Esto demuestra la capacidad de detectar CNV en este gen cuando se utiliza el flujo de trabajo de Illumina Exome 2.5. La pista del archivo BED del panel Exome 2.5 muestra las regiones cubiertas por el panel (figura 2, parte superior). Los resultados dependerán de varios factores y se recomienda secuenciar con una cobertura media superior a 150× para las pruebas iniciales de un flujo de trabajo de laboratorio para garantizar una cobertura suficiente para la mayoría de las variantes.

Análisis adicional posible con el software Emedgene

El software Emedgene optimiza y aumenta la escala de la interpretación de variantes, lo que supone un ahorro de tiempo de entre un 50 % y un 75 % por caso en investigación.⁷ Combina múltiples funciones para potenciar la interpretación definida por el usuario, incluida la XAI para clasificaciones automáticas transparentes y respaldadas por datos de variantes potencialmente causantes en las muestras, una anotación y una gráfica de conocimientos siempre actualizadas, visualización de variantes, selección de variantes, automatización definida por el usuario y muchas más opciones para facilitar una interpretación de las variantes fundamentada. El software Emedgene se ha diseñado para ofrecer una experiencia de usuario eficiente e intuitiva.

Para optimizar la revisión de casos, la priorización de variantes de IA o la funcionalidad «preselección», compila variantes, incluidas las CNV, que tienen más probabilidades de resolver un caso. Esta funcionalidad incluye pruebas de respaldo y proporciona un ahorro de tiempo significativo para el analista. En un estudio independiente⁸ de 51 singulones previamente resueltos por una variante de CNV, la variante de resolución se identificó en una preselección de 10 variantes en el 92 % de los casos. En el 6 % (n = 3), la variante de resolución estaba presente en la lista de candidatos. Durante la revisión del caso, el software Emedgene permite desetiquetar las variantes seleccionadas en la preselección o el etiquetado manual de las variantes no seleccionadas en dicha preselección. La clasificación automatizada del ACMG* también cubre las CNV.

Panel de acceso de archivos de normales

El panel de archivos de normales, generado con muestras del Coriell Institute, proporciona a los investigadores un conjunto de datos conocidos para la comparación con muestras de casos sin necesidad de dedicar tiempo y recursos a adquirir, secuenciar y analizar este conjunto de muestras. Estos archivos predeterminados se utilizan para las pruebas iniciales del flujo de trabajo de Illumina DNA Prep with Exome 2.5 Enrichment para el análisis de CNV y se recomienda a los laboratorios desarrollar un panel de normales específicos para sus protocolos de laboratorio. Si se utilizan estos archivos para las pruebas iniciales, las librerías para las muestras de casos deben prepararse siguiendo el mismo protocolo utilizado para preparar el panel de normales. Las desviaciones con respecto al protocolo pueden provocar diferencias en los perfiles de cobertura, lo que atenúa la eficacia del panel predeterminado de normales. Los archivos y la información adicional se pueden encontrar en el sitio de asistencia de Illumina.

*ACMG, American College of Medical Genetics and Genomics.

Tabla 4: Detección de CNV ordenada por tamaño aproximado del evento en función de las coordenadas de eventos esperados

ID de Coriell	Cromosoma	Localización génica (exones afectados) o cromosómica	Evento esperado	Tamaño aproximado del evento (kb) ^a	¿Detectado con 50 M de lecturas?
NA04315	X	<i>DMD</i> (44)	Pérdida	0,15	✓
NA05115	X	<i>DMD</i> (45)	Pérdida	0,18	✓
NA05117	X	<i>DMD</i> (45)	Pérdida	0,18	✓
NA23599	X	<i>MECP2</i> (3–4)	Pérdida	2,25	✓
NA18949	17	<i>BRCA1</i> (15–16)	Pérdida	3,59	✓
NA21939	15	<i>FBN1</i> (42–43)	Pérdida	3,70	✓
HG03857	16	<i>PALB2</i> (5–7)	Pérdida	4,23	✓
HG00343	22	<i>CHEK2</i> (9–10)	Pérdida	4,25	✓
NA23127	X	<i>DMD</i> (27–28)	Ganancia	7,47	✓
NA04520	16	<i>TSC2</i> (1–15)	Pérdida	16,28	✓
NA04327	X	<i>DMD</i> (5–7)	Ganancia	22,70	✓
NA10283	X	<i>DMD</i> (72–79)	Pérdida	54,38	✓
HG00500	2	<i>SPAST</i> (4–17)	Ganancia	58,84	✓
NA23675	X	<i>MECP2</i> (todos)	Ganancia	76,14	✓
NA23087	X	<i>DMD</i> (2–30)	Ganancia	608,45	✓
NA06870	18	18p11.32-18p11.1	Ganancia	15 390,21	✓
NA20125	10	10q23.1-10q26.3	Ganancia	52 877,99	✓

a. En función de las coordenadas de eventos esperados de la ubicación cromosómica o del exón afectado.

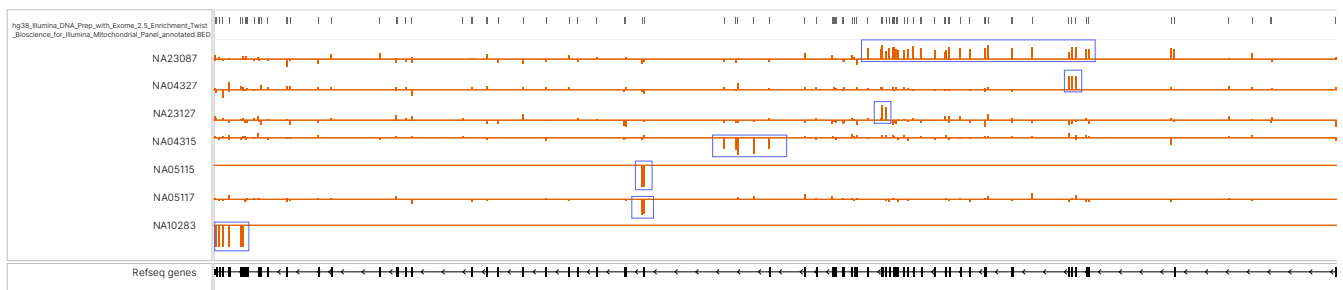


Figura 2: Visualización de las CNV detectadas en siete muestras que contienen eventos esperados en el gen *DMD*. Los archivos que proporcionan representación BigWig de la señal normalizada de la tangente (*.tn.bw) se generan como parte del análisis de la aplicación DRAGEN Enrichment y muestran las ganancias y pérdidas en función de las regiones objetivo en Twist Bioscience for Illumina Exome 2.5 Enrichment Panel. Las pistas se muestran a escala automática como un gráfico de barras.

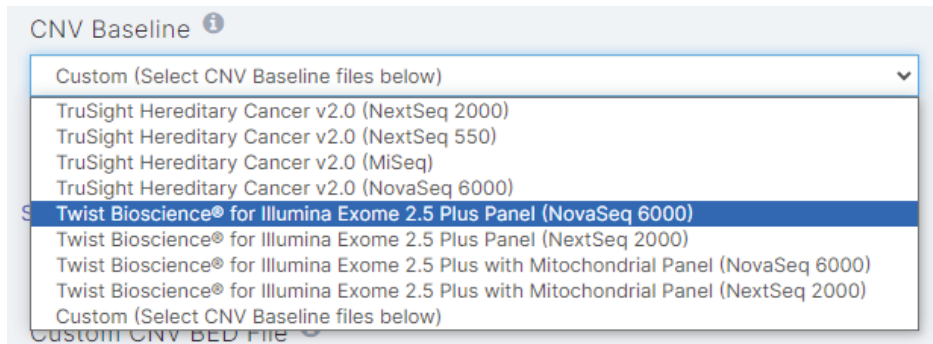


Figura 3: Menú desplegable del panel de normales en la aplicación DRAGEN Enrichment v4.3 en BaseSpace Sequence Hub.

Cuando se utiliza BaseSpace Sequence Hub, la aplicación DRAGEN Enrichment v4.3 proporciona opciones desplegables para realizar el análisis inicial de CNV con Twist Bioscience for Illumina Exome 2.5 Enrichment Panel para múltiples referencias (gráficas y no gráficas hg19, hg38 y hs37d5) en múltiples plataformas (NextSeq 2000 System y NovaSeq 6000 System) utilizando múltiples librerías (Twist Bioscience for Illumina Exome 2.5 Enrichment Panel con y sin Twist Bioscience for Illumina Mitochondrial Panel como panel de adición) (figura 3). Las muestras de casos utilizadas con este flujo de trabajo de análisis deben prepararse como se describe en la sección Métodos de esta nota de aplicación.

Resumen

En el flujo de trabajo de secuenciación del exoma mediante NGS para analizar CNV se utiliza Illumina DNA Prep with Exome 2.5 Enrichment para la preparación de librerías, la secuenciación en NovaSeq 6000 System o NextSeq 2000 System y aplicaciones DRAGEN para el análisis de datos. Los resultados demuestran la aplicación de un panel de normales como referencia para el análisis de CNV de muestras de casos utilizando el flujo de trabajo de WES descrito. La elevada recuperación de CNV por parte del software de análisis secundario DRAGEN guarda buena correlación con otros métodos. Este flujo de trabajo, incluido el uso del panel de normales, para el análisis de CNV también se puede adaptar a otras plataformas de secuenciación de Illumina. El software Emedgene ayuda a los laboratorios a realizar análisis adicionales, como la interpretación de variantes y la generación de informes de investigación.

Más información

[Panel de datos de referencia de normales](#)

[Análisis de variantes del número de copias](#)

[Illumina DNA Prep with Exome 2.5 Enrichment](#)

[Plataformas de secuenciación de Illumina](#)

[Análisis secundario de DRAGEN](#)

[Aplicación DRAGEN Germline](#)

[Software Emedgene](#)

Bibliografía

1. Zhang F, Gu W, Hurles ME, Lupski JR. [Copy number variation in human health, disease, and evolution](#). *Annu Rev Genomics Hum Genet.* 2009;10:451-481. doi:10.1146/annurev.genom.9.081307.164217
2. Mehawej C, Maalouf JE, Abdelkhalik M, Mahfouz P, Chouery E, Megarbane A. [CNV Analysis through Exome Sequencing Reveals a Large Duplication Involved in Sex Reversal, Neurodevelopmental Delay, Epilepsy and Optic Atrophy](#). *Genes (Basel).* 2024;15(7):901. doi:10.3390/genes15070901
3. Fang X, Ma M, Rong W, et al. [Exome sequencing confirms the clinical diagnosis of both joubert syndrome and klinefelter syndrome with keratoconus in a han Chinese family](#). *Front Genet.* 2024;15:1417584. doi:10.3389/fgene.2024.1417584
4. Zeng Y, Ding H, Wang X, et al. [High positive predictive value of CNVs detected by clinical exome sequencing in suspected genetic diseases](#). *J Transl Med.* 2024;22(1):644. doi:10.1186/s12967-024-05468-1
5. Harrison PW, Amode MR, Austine-Orimoloye O, et al. [Ensembl 2024](#). *Nucleic Acids Res.* 2024;52(D1):D891-D899. doi:10.1093/nar/gkad1049
6. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. [Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model](#). *BMC Med Genomics.* 2018;11(1):91. doi:10.1186/s12920-018-0404-4
7. Greenwood Genetic Center. [GGCC reduces turn around time on genomic analysis by 75% with Emedgene's AI platform](#). ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform. Fecha de publicación: 12 de septiembre de 2019. Fecha de consulta: 14 de agosto de 2024.
8. Datos en archivo. Illumina, Inc. 2024.



1 800 809 4566 (llamada gratuita, EE. UU.) | tel.: +1 858 202 4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, consulte www.illumina.com/company/legal.html.
M-GL-01453 ESP v1.0