

Análisis secundario DRAGEN™

Llamada de variantes precisa,
exhaustiva y eficiente con datos
de secuenciación de nueva
generación

illumina®

Introducción

Resulta fundamental liberar el potencial del genoma mediante la secuenciación de nueva generación (NGS, Next-Generation Sequencing) para permitir avances en la investigación y la medicina. A fin de obtener el máximo de conocimientos genéticos a través de la NGS, los investigadores precisan herramientas de análisis de datos que permitan una traducción precisa y eficiente de los datos sin procesar de secuenciación en resultados reveladores. Además, para aprovechar los beneficios de la NGS, las organizaciones necesitan soluciones fáciles de usar, que se adapten a distintos usuarios, y con el menor número posible de barreras tanto técnicas como económicas para su adopción.

El análisis secundario DRAGEN (Dynamic Read Analysis for GENomics) de Illumina se desarrolló para abordar importantes retos asociados al análisis de datos de NGS para una serie de aplicaciones, incluidos los estudios de genoma, exoma, transcriptoma y metiloma. La plataforma DRAGEN es un paquete de software de análisis secundario que procesa datos de NGS y permite el análisis terciario para obtener información. Las herramientas disponibles conforman una solución muy precisa, exhaustiva y eficiente que facilita a los laboratorios de cualquier tamaño y disciplina sacar más provecho de sus datos genómicos.

Resultados precisos

El análisis secundario DRAGEN genera resultados excepcionalmente precisos. En el Precision FDA Truth Challenge V2 (PrecisionFDA V2) organizado por la FDA en 2020, DRAGEN v3.7 fue galardonado con el premio a la mayor precisión en todas las regiones de referencia y en las regiones difíciles de asignar para los datos de secuenciación de Illumina.^{1,2} Las innovaciones en Graph Genomes e Illumina Machine Learning (ML) con el software DRAGEN 4.0 demuestran una precisión excepcional de los datos en todas las tecnologías de secuenciación, con una puntuación F1 del 99,83 % (medida combinada de precisión y recuperación) en todas las regiones de referencia (Figura 1).^{1,2} DRAGEN 4.0 + Graph (ML activado por defecto) también tiene la puntuación F1 más alta para las llamadas más precisas en comparación con todos los resultados presentados en PrecisionFDA V2 en las regiones del complejo principal de histocompatibilidad (MHC, Major Histocompatibility Complex).

Análisis exhaustivo

El análisis secundario DRAGEN satisface las necesidades de los laboratorios que trabajan en una amplia gama de aplicaciones de NGS, proporcionando una cobertura completa para un amplio juego de tipos de experimentos en una única plataforma.

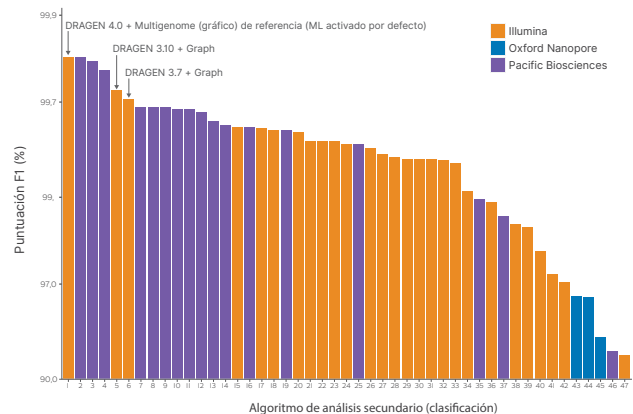


Figura 1: precisión de DRAGEN 4.0 + Graph (ML activado por defecto) en comparación con todos los resultados presentados en PrecisionFDA Truth Challenge v2 en el conjunto de datos de todas las regiones de referencia. DRAGEN 4.0 + Graph (ML activado por defecto) demuestra una precisión excepcional, igualada con Google DeepVariant en datos de secuenciación de Pacific Biosciences. DRAGEN 3.10 + Graph muestra una mejora respecto a DRAGEN 3.7 + Graph gracias a las mejoras en el gráfico y en la manipulación de referencias/cóntigo ALT. La puntuación F1 (%), en el eje de ordenadas, es un cálculo de resultados positivos verdaderos y negativos verdaderos, representado como una proporción de los resultados totales.^{3,4}

Los procesos de DRAGEN admiten varios tipos de experimentos, incluida la secuenciación del genoma completo (WGS, Whole-Genome Sequencing), paneles de enriquecimiento, secuenciación de ARN de células únicas, secuenciación de ATAC de células únicas, multiómica de células únicas, secuenciación masiva de ARN y análisis de metilación (Tabla 1). Se necesitarían más de 30 herramientas de código abierto para reproducir parcialmente la amplitud de funciones del software DRAGEN.^{3,4}

Con el paquete incluido de llamadores de variantes (expansiones repetidas, variaciones estructurales [SV, Structural Variations], variantes en el número de copias [CNV, Copy-Number Variants], ExpansionHunter y llamadores selectivos, como *SMN*, *GBA*, *CYP2B6*, *CYP2D6* y *HLA*), el software DRAGEN ofrece una amplia cobertura genómica. Además, la referencia DRAGEN Multigenome (Graph) alarga de manera eficaz las lecturas de Illumina y alcanza regiones de baja complejidad, para resolver áreas del genoma difíciles de evaluar debido a secuencias repetidas. Esto mejora la cobertura de genes con potencial médicamente relevante y permite la llamada de variantes de nucleótido único, de número de copias y estructurales en regiones difíciles de asignar.

Tabla 1: el análisis secundario DRAGEN es compatible con multitud de aplicaciones de análisis secundario.

Aplicación	DRAGEN On-premise Server	Onboard NovaSeq X Series	Onboard NextSeq 1000, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics	
					Preconfigurado	Personalizado
Conversión de BCL	✓	✓	✓	✓	✓	✓
Compresión de DRAGEN ORA	✓	✓	✓			✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
Genoma completo	Germinal + somático	Solo germinal	Solo germinal	Germinal + somático	Germinal + somático	Germinal + somático
Enriquecimiento (incluido el exoma)	Germinal + somático	Germinal + somático	Germinal + somático	Germinal + somático	Germinal + somático	Germinal + somático
Amplicones de ADN	✓		✓	✓	✓	✓
ARN	✓	✓	✓	✓	✓	✓
ARN de células únicas	✓		✓	✓	✓	✓
Expresión diferencial		✓	✓	✓		
NanoString GeoMx NGS				✓		
Amplicones de ARN	✓			✓	Próximamente	Próximamente
Metilación	✓			✓	✓	✓
Metagenómica				✓		
Detección de patógenos de ARN				✓		
Estirpe de COVID	✓			✓	Próximamente	
TruSight Oncology 500	✓				✓	
ScATAC-Seq	✓			✓	✓	✓
Atribución	✓			✓	✓	✓
PGx Star Allele Caller	✓			✓	✓	✓
Illumina Complete Long Reads				✓		

Análisis eficiente

El software DRAGEN se ha diseñado específicamente para proporcionar a los laboratorios la velocidad de análisis de datos y las opciones de archivo que necesitan para obtener el mayor beneficio de los conjuntos de datos de NGS. El análisis secundario DRAGEN está optimizado con aceleración por hardware y utiliza una arquitectura de array de puertas programables por campo (FPGA, Field-Programmable Gate Array) para lograr tiempos de procesamiento rápidos. La eficiencia de los algoritmos de análisis DRAGEN dio lugar a dos récords mundiales de velocidad de análisis de datos genómicos.^{5,6} En aplicaciones prácticas, el análisis secundario DRAGEN en las instalaciones puede procesar datos de NGS de un genoma completo equivalente a una cobertura de 34x en aproximadamente 30 minutos en las instalaciones, en comparación con las más de 15 horas necesarias con un sistema tradicional basado en una CPU.⁷

Para lidiar con las demandas de almacenamiento de grandes archivos de datos de NGS, la tecnología DRAGEN Original Read Archive (ORA) ofrece una compresión 5x sin pérdidas para archivos FASTQ. La compresión sin pérdidas de DRAGEN ORA mantiene los detalles de los archivos FASTQ y es considerablemente más rápida; se requieren aproximadamente 8 minutos para comprimir archivos FASTQ de 50 a 70 GB. El análisis secundario DRAGEN cuenta con un juego versátil de procesos que también pueden aceptar archivos de datos de entrada y generar archivos de resultados en diferentes etapas de los procesos (Figura 2).

FPGA y optimización con aceleración por hardware

La FPGA de gran capacidad de configuración permite implementaciones ultraeficientes optimizadas con aceleración por hardware de algoritmos de análisis genómico, como la conversión de archivos de llamada de bases (BCL, Base Call), asignaciones, alineaciones, clasificaciones, marcado de duplicados y llamada de variantes de haplotipo. La naturaleza flexible de las FPGA permite a Illumina desarrollar un amplio paquete de procesos de DRAGEN para diversas aplicaciones, con actualizaciones y adiciones frecuentes para ofrecer la mayor precisión, exhaustividad y eficiencia posibles.

Referencias personalizadas

DRAGEN Reference Builder, también conocido como tabla fragmentada, permite a los usuarios generar una referencia humana, no humana o no estándar. Las referencias creadas se pueden utilizar como entrada en todas las aplicaciones de DRAGEN compatibles con los archivos de referencia del cliente. La aplicación DRAGEN Reference Builder en BaseSpace™ Sequence Hub requiere un archivo FASTA. La mayoría de procesos de DRAGEN incluyen compatibilidad incorporada para hg19, hg38 (con o sin HLA*), GRCh37 y Hs37d5. DRAGEN Graph Toolkit permite a los usuarios ampliar las capacidades de referencia de gráficos también a referencias de gráficos humanos más diversas.

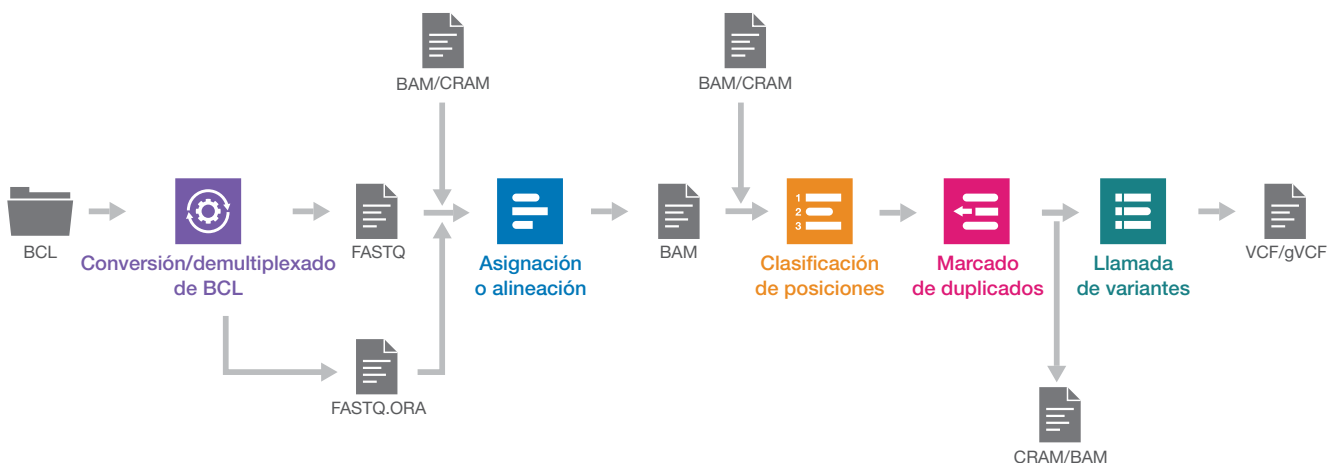


Figura 2: flexibilidad de los procesos de DRAGEN. Cada proceso de DRAGEN contiene un juego específico de etapas para respaldar un análisis preciso y eficiente. El proceso de DRAGEN proporciona flexibilidad para aceptar diversos archivos de entrada y producir una amplia variedad de tipos de resultados, lo que permite a los usuarios personalizar su experiencia y generar el formato de archivo deseado.

Flexibilidad

El análisis secundario DRAGEN permite flexibilizar las operaciones del laboratorio según las necesidades, al tiempo que mantienen bajos los costes y los tiempos de procesamiento. El software DRAGEN puede facilitar la expansión de las funciones de investigación de varias formas:

- 1. Seguir el ritmo de NovaSeq™ X Series.** El software DRAGEN integrado en el instrumento puede realizar hasta cuatro aplicaciones simultáneas por celda de flujo en un experimento único.
- 2. Capacidad de picos de trabajo.** Durante las ocasiones de aumento de la carga de trabajo con volúmenes de muestra elevados, los laboratorios se pueden beneficiar de una capacidad adicional a través de las opciones disponibles de acceso paralelo al software DRAGEN (Figura 3).
- 3. Ampliación de las operaciones.** Se puede utilizar una única instancia DRAGEN para ejecutar todos los procesos de DRAGEN y los tipos de muestras compatibles. La precisión, exhaustividad y eficiencia del software DRAGEN permiten a los usuarios ampliar las operaciones, sin comprometer los tiempos de procesamiento ni la calidad de los resultados.
- 4. De exomas a genomas.** El paso de la secuenciación del exoma completo (WES, Whole-Exome Sequencing) a la WGS implica un gran aumento de los datos generados. El software DRAGEN permite a los clientes pasar fácilmente de los análisis de exomas a los análisis de genomas sin grandes inversiones en infraestructura de hardware adicional ni en soluciones basadas en la nube.

- 5. Conjuntos de datos de gran tamaño.** El análisis secundario DRAGEN ofrece un flujo de trabajo simplificado para análisis de cohortes a gran escala, con múltiples procesos que se utilizan conjuntamente para la llamada de variantes pequeñas y grandes, con una alta precisión a partir del muestreo de una cohorte. El software DRAGEN permite la agregación y el genotipado de miles a millones de archivos de formato de llamada de variantes genómicas (gVCF, Genomic Variant Call Format) y agrega nuevos lotes sin volver a procesar los existentes. DRAGEN Joint Genotyping Pipeline realiza llamadas de variantes de forma conjunta en múltiples genomas y se adapta a grandes cohortes con un análisis rápido y una precisión rigurosa.⁸ El análisis secundario DRAGEN de los datos del [1000 Genomes Project](#), por ejemplo, permitió realizar llamadas de variantes precisas y a gran escala en diversas muestras e identificar regiones en las que los datos de cobertura no son uniformes o se desvían de los supuestos.

Accesibilidad a través de múltiples plataformas

El paquete de procesos de DRAGEN es accesible a través de soluciones en las instalaciones, integradas en el instrumento o basadas en la nube, para que cada laboratorio elija la solución que mejor se adapte a sus necesidades (Figura 3).

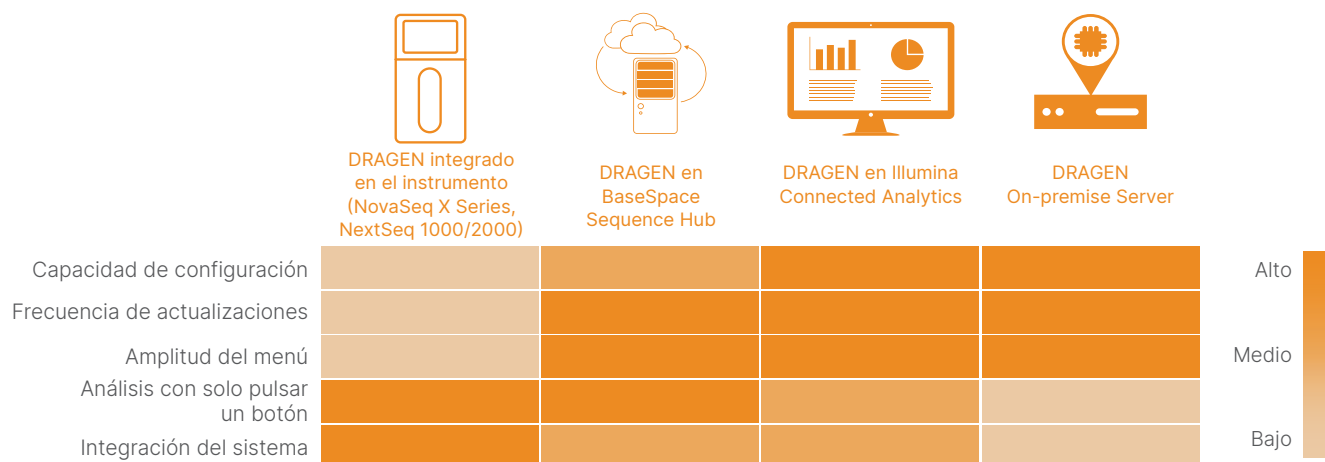


Figura 3: opciones de acceso a los procesos de DRAGEN con funciones diseñadas para adaptarse a las necesidades de análisis de NGS de cada laboratorio.

Servidor DRAGEN en las instalaciones

El servidor DRAGEN en las instalaciones depende de una solución de almacenamiento local para recopilar y almacenar datos de NGS. Una vez que los datos de secuenciación sin procesar se hayan transferido del instrumento de secuenciación al almacenamiento local mediante una conexión de red local, se transfieren al servidor DRAGEN para ejecutar el flujo de trabajo seleccionado. Después del análisis, el software guarda los archivos de resultados generados en la ubicación de almacenamiento local. El servidor DRAGEN en las instalaciones:

- Admite distintos niveles de interfaz de línea de comandos
- Sustituye hasta 30 instancias computacionales tradicionales
- Procesa datos de NGS para un genoma humano completo a una cobertura de 34x en aproximadamente 30 minutos

DRAGEN integrado en el instrumento en NovaSeq™ X Series

NovaSeq X Series incluye el software DRAGEN más potente hasta la fecha, que ofrece un análisis secundario preciso, automatizado y exhaustivo. El paquete de software DRAGEN integrado en el instrumento proporciona la llamada de variantes y compresión de ORA con aplicaciones de NGS (Tabla 1) que abarcan la conversión de BCL, la línea germinal, el ARN y el enriquecimiento. DRAGEN integrado en el instrumento:

- Ejecuta varios procesos de análisis secundario en paralelo
- Realiza hasta cuatro aplicaciones simultáneas por celda de flujo en un experimento único
- Aporta una compresión de datos de hasta 5x sin pérdidas y un análisis con aplicaciones compatibles
- Proporciona una reducción del coste de los análisis que, en cinco años, puede superar el precio de compra de NovaSeq X System

DRAGEN integrado en el instrumento en NextSeq™ 1000 System y NextSeq 2000 System

NextSeq 1000 System y NextSeq 2000 System incluyen el software DRAGEN integrado en el instrumento para un análisis secundario rápido y preciso. El paquete de software DRAGEN integrado en el instrumento ofrece un juego específico de procesos para cubrir diversas aplicaciones habituales de NGS (Tabla 1) con una interfaz de usuario sencilla que permite a usuarios expertos e inexpertos realizar los análisis necesarios y obtener resultados rápidamente.

DRAGEN integrado en el instrumento:

- Proporciona acceso a determinados procesos informáticos de DRAGEN
- Permite a los usuarios generar resultados en tan solo dos horas
- Utiliza algoritmos de procesos intuitivos para reducir la dependencia de expertos informáticos externos

BaseSpace Sequence Hub

El paquete DRAGEN basado en la nube disponible en BaseSpace Sequence Hub combina el análisis preciso y eficiente con un ecosistema seguro y una gran flexibilidad. El software DRAGEN en BaseSpace Sequence Hub permite realizar análisis secundarios con solo pulsar un botón en laboratorios de cualquier tamaño y disciplina. BaseSpace Sequence Hub es una extensión directa de sus instrumentos de Illumina. Los datos cifrados se transfieren del instrumento a BaseSpace Sequence Hub, para que pueda gestionar y analizar fácilmente sus datos con un juego seleccionado de aplicaciones. BaseSpace Sequence Hub, respaldado por Amazon Web Services (AWS):

- Ofrece una solución fácil de usar con solo pulsar un botón para análisis DRAGEN.
- Utiliza una interfaz de usuario gráfica intuitiva para un funcionamiento eficiente por parte de usuarios expertos e inexpertos.
- Proporciona acceso a potentes recursos informáticos sin necesidad de invertir en infraestructuras adicionales.

Illumina Connected Analytics

El análisis secundario DRAGEN en Illumina Connected Analytics es una exhaustiva plataforma bioinformática basada en la nube que permite a los investigadores gestionar, analizar e interpretar grandes volúmenes de datos multiómicos en un entorno seguro, flexible y adaptable. Illumina Connected Analytics:

- Proporciona acceso al software DRAGEN completo, disponible como procesos predeterminados o herramientas individuales para procesos personalizados
- Es compatible con flujos de trabajo muy automatizados y soluciones personalizadas para estudios de alta productividad optimizados
- Ofrece un entorno muy seguro con residencia de datos garantizada, acceso con inicio de sesión individual, registros de auditoría y control de acceso compatible con el cumplimiento de la Ley de portabilidad y responsabilidad del seguro médico (Health Insurance Portability and Accountability Act, HIPAA) de los Estados Unidos y los principios del Reglamento General de Protección de Datos (RGPD) de la Unión Europea

Resumen

El análisis secundario DRAGEN es un potente paquete de herramientas de software que proporciona un análisis preciso, exhaustivo y eficiente de los datos de NGS. Las múltiples opciones del software DRAGEN permiten a los laboratorios seleccionar la solución que mejor se adapte al tipo y a la escala de sus proyectos. A tenor de la continua evolución de la tecnología de NGS, las actualizaciones puntuales del análisis secundario DRAGEN garantizan el mejor rendimiento posible de los procesos actuales, sin cesar en la incorporación de nuevos procesos a medida que se disponen de nuevas aplicaciones.

Información adicional

[Análisis secundario DRAGEN](#)

[Página de asistencia del análisis secundario DRAGEN](#)

[Póngase en contacto con nosotros](#)

Bibliografía

1. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Fecha de consulta: 14 de marzo de 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Fecha de consulta: 14 de marzo de 2022.
3. Illumina. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Fecha de consulta: 14 de marzo de 2022.
4. Datos internos archivados. Illumina, Inc. 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Fecha de consulta: 14 de marzo de 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Fecha de publicación: 12 de febrero de 2018. Fecha de consulta: 14 de marzo de 2022.
7. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015; 7: 100. doi: 10.1186/s13073-015-0221-8.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Fecha de consulta: 14 de marzo de 2022.

illumina®

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

© 2023 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-00680-ESP v5.0