

Illumina Human Comprehensive Panel

Cobertura mejorada de
regiones complejas en genes
codificantes de proteínas con
Illumina Complete Long Read
Prep with Enrichment, Human



Introducción

Al realizar la secuenciación del genoma completo (WGS, Whole-Genome Sequencing) humano, puede haber una pequeña cantidad de regiones génicas difíciles de asignar solo con lecturas cortas. Estas regiones complejas incluyen regiones altamente homólogas o repetitivas, variantes estructurales complejas, pseudogenes y grandes inserciones/delecciones (indel). En estos casos, la secuenciación de lectura larga puede complementar los datos estándar de la WGS de lectura corta para ayudar a ofrecer una cobertura mejorada a la hora de resolver regiones complejas, y facilitar la fase de hebra retrasada de variantes y la llamada de haplotipos.

La tecnología Illumina Complete Long Reads utiliza un flujo de trabajo de secuenciación de nueva generación (NGS, next-generation sequencing) estándar para generar secuencias de lectura larga contiguas en los sistemas de secuenciación de Illumina (figura 1). Illumina Complete Long Read Prep with Enrichment y el Human Comprehensive Panel ofrecen un enfoque rentable y flexible que ayuda a los investigadores a resolver regiones genómicas complejas conocidas. Las lecturas largas selectivas se analizan con datos estándar de WGS de alta precisión para permitir la asignación de un mayor número de regiones.

Panel de sondas completo y optimizado

Illumina Human Comprehensive Panel es un panel de sonda de captura híbrida optimizado diseñado para abordar regiones génicas que pueden beneficiarse de una mayor asignación de lecturas más largas. El panel se dirige a regiones de baja asignación en más de 6500 genes codificantes de proteínas (tabla 1)¹ para permitir lecturas largas complementarias que mejoren la resolución de áreas que son difíciles de asignar utilizando métodos estándar de WGS de lectura corta.

Tabla 1: Parámetros del Human Comprehensive Panel^a

Sistema	NovaSeq X Series NovaSeq 6000 System
Tamaño de la región objetivo	>95 Mb
Rendimiento de secuenciación por muestra ^b	90-120 Gb
N.º de sondas	~40 000
N.º de genes objetivo	>6500
Tipo de muestra	ADN genómico
Entrada de ADN recomendada	50 ng
Tiempo total de preparación de librerías	~2 días
Tiempo de participación activa	~6 h
Multiplexado	Hasta 64 muestras
N50	6,3 kb
Bloque de fase de hebra retrasada N50	15,6 kb
Uniformidad	95 %
Enriquecimiento de lectura completada ^c	83 %
% de SNV heterocigóticas en fase de hebra retrasada ^d	98 %

- Datos generados con 50 ng de ADN genómico de HG002 (Coriell, n.º de catálogo NA24385). El rendimiento puede variar en función de la entrada de ADN y la calidad de la muestra.
- Requiere un experimento de secuenciación de 2×150 pb, lo que genera una cobertura final de aproximadamente $30\times$ de Illumina Complete Long Reads.
- Uniformidad de la cobertura calculada como $\% >0,2 * \text{media}$.
- Enriquecimiento de lectura completada calculado como $100 * (\text{lecturas completadas alineadas con el objetivo} / \text{lecturas alineadas totales})$.
- SNV, variante de nucleótido único



Figura 1: Parte de un flujo de trabajo integrado. Acceda a datos de WGS de lectura larga selectivos y rentables mediante un protocolo de preparación de librerías con enriquecimiento flexible y optimizado, la química de secuenciación de Illumina probada y el análisis secundario DRAGEN. Requiere datos de WGS de lectura corta estándar con una cobertura de $30\times$ de la misma muestra para el análisis. Se pueden utilizar archivos FASTQ de una muestra procesada previamente.

En el desarrollo de este panel, Illumina consideró el conjunto completo de más de 20 000 genes codificantes de proteínas, incluidos los intrones y las regiones no traducidas (UTR, untranslated regions), y se dirigió de manera selectiva al 70 % de todos los pares de bases difíciles de asignar en estas regiones. Se excluyeron los genes que están cubiertos exhaustivamente solo por lecturas cortas. Este panel se ha probado y optimizado en laboratorio de investigación para maximizar la eficiencia y el rendimiento del diseño y la secuenciación.

Secuenciación y análisis altamente flexibles

El flujo de trabajo de Illumina Complete Long Read Prep with Enrichment es muy flexible y fácil de automatizar para respaldar procesos exhaustivos de WGS con un mayor número de muestras (tabla 2). En NovaSeq™ X Plus System, los usuarios pueden generar hasta 15 000 genomas de alta precisión al año.* Para reducir los requisitos de generación de lotes, secuencie una menor cantidad de muestras en consumibles de menos productividad, como la celda de flujo NovaSeq X 1.5B. Illumina Complete Long Read Prep with Enrichment y el Human Comprehensive Panel se pueden utilizar para aumentar los conjuntos de datos de WGS existentes como una herramienta alternativa para una detección de variantes más amplia.

* Productividad potencial cuando se usan Illumina Complete Long Read Prep with Enrichment y el Human Comprehensive Panel con NovaSeq X Plus System en experimentos de celda de flujo doble con celdas de flujo 25B.

El análisis de datos de Illumina Complete Long Read Prep with Enrichment, Human está disponible como aplicación BaseSpace™ Sequence Hub o a través de Illumina Connected Analytics. Los datos se transmiten a la nube y se analizan junto con datos de WGS de lectura corta estándar con una cobertura de al menos 30× de la misma muestra (generados previamente o en paralelo). El proceso DRAGEN™ combina los resultados en un único conjunto de archivos de resultados, incluidos los llamadores selectivos de DRAGEN.³

Cobertura rentable y mejorada

Illumina Complete Long Read Prep with Enrichment y el Human Comprehensive Panel mejoran la precisión de la llamada de variantes y la cobertura en las regiones objetivo, en comparación con la WGS de lectura corta estándar (figura 2, figura 3, tabla 3). Las lecturas largas selectivas con el Human Comprehensive Panel proporcionan una precisión exhaustiva del genoma completo con una puntuación F1 (SNV + indel) del 99,87 %,² lo que supone una mejora considerable con respecto a las lecturas cortas por sí solas. Esto proporciona un rendimiento comparable al de Illumina Complete Long Read Prep, Human y otros genomas completos de lectura larga en genes codificantes de proteínas, aunque con un menor coste y una mayor productividad (tabla 3).

Tabla 2: Productividad de muestras recomendada para generar una cobertura final de 30× para Human Comprehensive Panel e Illumina Complete Long Read Prep with Enrichment, Human^{a,b,c,d}

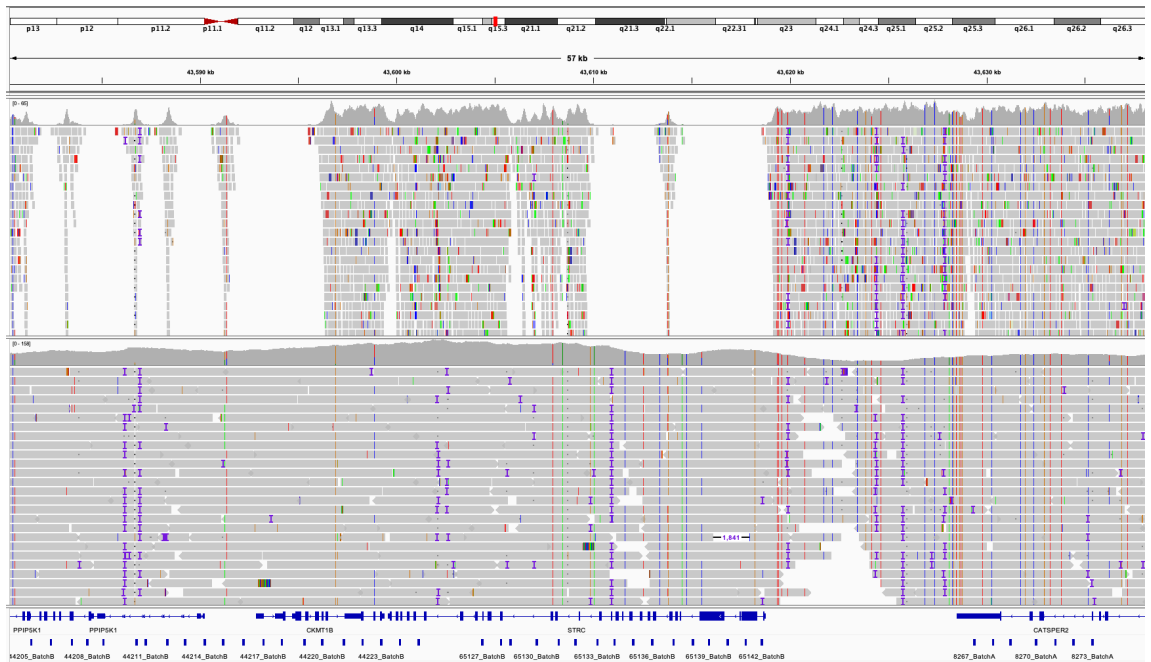
	NovaSeq 6000				NovaSeq X		
	SP	S1	S2	S4	1.5B	10B	25B
Kits de reactivos de 300 ciclos							
Muestras por celda de flujo	2	4	10	24	4	24	64
Rendimiento por celda de flujo	~250 Gb	~500 Gb	~1,25 Tb	~3 Tb	~500 Gb	~3 Tb	~8 Tb
Duración del experimento	~25 h	~25 h	~36 h	~44 h	~21 h	~25 h	~48 h

- El tamaño de la región objetivo del Human Comprehensive Panel es de más de 95 Mb y requiere un rendimiento de secuenciación de 90 a 120 Gb por muestra.
- Requiere un experimento de secuenciación de 2 × 150 pb, lo que genera una cobertura final de aproximadamente 30× de Illumina Complete Long Reads.
- Requiere datos de genoma completo humano de lectura corta estándar con una cobertura de 30× de la misma muestra para el análisis. Se recomienda [Illumina DNA PCR-Free Prep](#). Los kits de WGS de terceros también son compatibles. No es necesario preparar o secuenciar en paralelo una librería no marcada; se pueden usar archivos FASTQ de una muestra analizada previamente.
- La secuenciación de librerías de Illumina Complete Long Read en plataformas NovaSeq puede hacer que la puntuación Q30 de un experimento notificada caiga por debajo de la especificación de NovaSeq. Esto no indica un problema de rendimiento ni con el experimento de secuenciación ni con la librería.

A.

WGS de lectura corta estándar

Illumina Complete Long Read Prep with Enrichment



B.

WGS de lectura corta estándar

Illumina Complete Long Read Prep with Enrichment

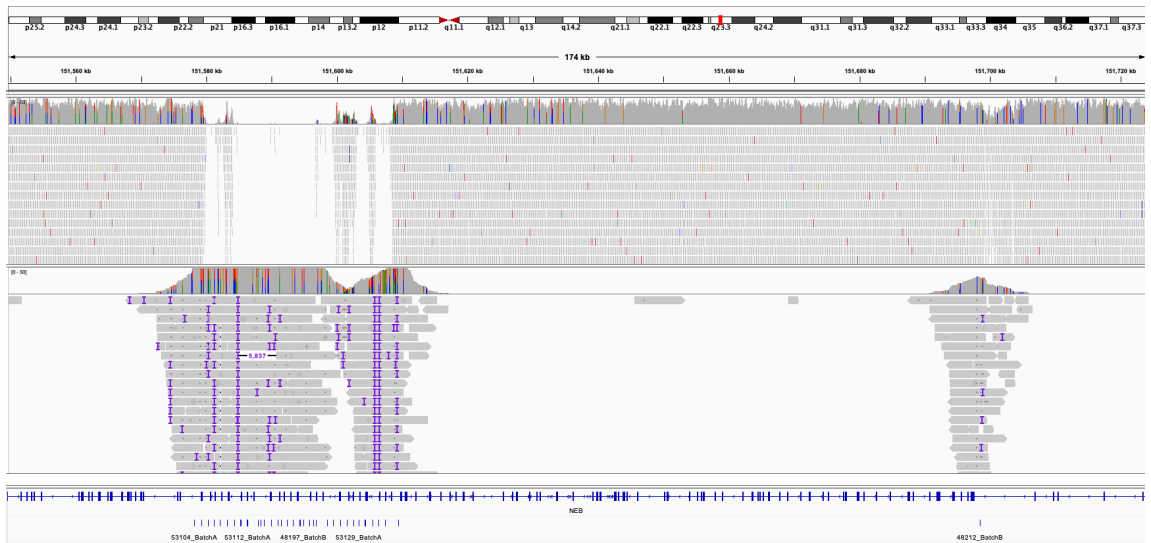


Figura 2: Resolución de regiones difíciles de asignar con lecturas largas selectivas con el Human Comprehensive Panel. Illumina Complete Long Read Prep with Enrichment, Human y el Human Comprehensive Panel ayudan a mejorar la cobertura en regiones génicas difíciles para complementar la WGS estándar de lectura corta en material de origen humano. Gráficos de Integrative Genomics Viewer (IGV) de (A) *STRC* y (B) *NEB* con uso de WGS de lectura corta estándar e Illumina Complete Long Reads with Enrichment.

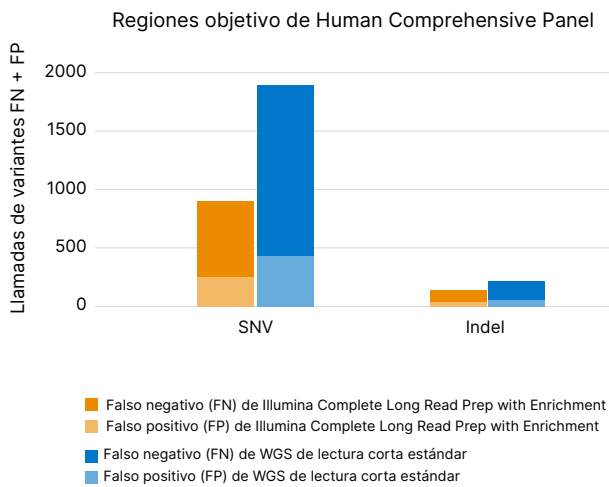


Figura 3: Lecturas largas selectivas para mejorar la precisión de la llamada de variantes. Llamadas de variantes falsas negativas (FN) más falsas positivas (FP) para SN e indel en regiones génicas a las que se dirige el Human Comprehensive Panel utilizando Illumina Complete Long Read Prep with Enrichment y el Human Comprehensive Panel (naranja) en comparación con la WGS de lectura corta estándar (azul).

Resumen

Illumina Human Comprehensive Panel ofrece una cobertura de lectura larga de alta precisión para esas pocas regiones génicas que son difíciles de asignar solo con lecturas cortas. Illumina Complete Long Read Prep with Enrichment y Human Comprehensive Panel ofrecen un análisis del genoma completo optimizado y rentable que complementa la WGS de Illumina y se centra en lecturas largas en las que proporcionan el mayor valor con una solución de flujo de trabajo completa.

Información adicional

[Illumina Complete Long Read Prep with Enrichment, Human](#)

[Human Comprehensive Panel](#)

[Tecnología de secuenciación de lectura larga](#)

Tabla 3: Rendimiento del Human Comprehensive Panel

Precisión para las regiones objetivo del Human Comprehensive Panel ^a		
	Illumina Complete Long Read Prep with Enrichment y Human Comprehensive Panel ^b	WGS de lectura corta estándar ^c
Puntuación F1 (SNV)	99,54 %	99,04 %
Puntuación F1 (Indel)	99,47 %	99,15 %
Puntuación F1 (SV)	80,43 %	60,18 %

Precisión para el genoma completo, todas las regiones de referencia ^d Puntuación F1 (SNV + indel)			
	Illumina Complete Long Read Prep with Enrichment y Human Comprehensive Panel ^b	Illumina Complete Long Read Prep, Human ^e	WGS de lectura corta estándar ^c
	99,87 %	99,90 %	99,84 %

- a. Solo autosomas. SNV: variantes de un solo nucleótido; SV: variantes estructurales.
- b. Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel + Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.
- c. Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.
- d. Según la medición frente a todas las soluciones participantes con datos de referencia de PrecisionFDA Truth Challenge v2,² datos de archivo internos para DRAGEN v4.
- e. Illumina Complete Long Read Prep, Human + Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.

Bibliografía

- Bekritsky MA, Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. illumina.com/science/genomics-research/articles/identifying-genomic-regions-with-high-quality-single-nucleotide-.html. Año de publicación: 2021. Fecha de consulta: 30 de agosto de 2023.
- PrecisionFDA. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Fecha de consulta: 2 de octubre de 2023.
- Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html. Año de publicación: 2023. Fecha de consulta: 22 de septiembre de 2023.

Datos para realizar pedidos

Producto	N.º de catálogo
Illumina Complete Long Read Prep with Enrichment, Human (24 samples)	20113832
Illumina Complete Long Read Prep with Enrichment, Human (96 samples)	20113833
Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel (24 samples)	20113834
Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel (96 samples)	20113835
Illumina Human Comprehensive Panel (24 samples)	20113836
Illumina Human Comprehensive Panel (96 samples)	20113837
Illumina Unique Dual Indexes, LT (48 indexes, 48 samples)	20098166
Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 indexes, 96 samples)	20091654
Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 indexes, 96 samples)	20091656
Illumina DNA/RNA UD Indexes Set C, Tagmentation (96 indexes, 96 samples)	20091658
Illumina DNA/RNA UD Indexes Set D, Tagmentation (96 Indexes, 96 Samples)	20091660
Illumina Analytics - 1 iCredit	20042038



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-02191 ESP v1.0