

# Diseño de paneles personalizados para Illumina Complete Long Read Prep with Enrichment, Human

Enriquecimiento de lectura larga selectivo y altamente flexible para genomas humanos



## Introducción

Al realizar la secuenciación del genoma completo (WGS, Whole-Genome Sequencing) humano, la asignación de diversas regiones puede ser difícil solo con lecturas cortas. La secuenciación de lectura larga puede complementar los datos estándar de lectura corta de la WGS para ayudar a abordar estas regiones de mayor complejidad. 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 con un solo proceso de análisis (figura 1).<sup>1-3</sup> Illumina Complete Long Read Prep with Enrichment, Human añade un enfoque selectivo para una secuenciación de lectura larga más rentable.\* La química de enriquecimiento de Illumina Complete Long Read proporciona una alta flexibilidad con objetivos y diseño de sonda para ayudar a resolver regiones difíciles de asignar o proporcionar información adicional de la secuenciación por fase de hebra retrasada.

## Diseño de paneles de sondas de enriquecimiento para lecturas largas

Illumina Complete Long Read Prep with Enrichment, Human utiliza una estrategia de diseño de sonda diferente para capturar fragmentos más largos (aproximadamente de 7 a 10 kb) que con la estrategia utilizada normalmente para capturar fragmentos cortos (aproximadamente de 200

a 500 pb). El software DesignStudio™ de Illumina es una herramienta gratuita y fácil de usar para diseñar paneles de sondas de enriquecimiento. El algoritmo de DesignStudio tiene en cuenta el contenido de GC, la especificidad del objetivo y la distancia entre sondas, es decir, cuántas sondas habrá en la región objetivo. La distancia estándar para los paneles de enriquecimiento de lectura corta de 120-mero es una ventana de sonda de 250 a 350 pb. Para el diseño de paneles de enriquecimiento de lectura larga, se analizó la distancia entre sondas en múltiples longitudes y se determinó que una ventana de una kilobase era óptima para una captura rentable y de alta eficiencia.

La eficacia del enriquecimiento de la hibridación depende en gran medida de la especificidad de la sonda. El porcentaje de enriquecimiento selectivo afecta directamente a la cantidad de secuenciones necesaria para lograr la profundidad de cobertura objetivo. Es más difícil lograr una alta especificidad en regiones repetitivas. Sin embargo, el uso de una ventana de sonda más amplia permite una mayor flexibilidad para excluir sondas de bajo rendimiento, evitar regiones repetidas (hasta un tamaño de ventana de 1 kb) y mantener la eficiencia de enriquecimiento con menos sondas (figura 2). El algoritmo de DesignStudio puede utilizar estas consideraciones para recomendar la colocación de la sonda. Los paneles de terceros deben utilizar directrices similares para un mejor rendimiento y una mejor rentabilidad. La distancia estándar de la sonda de enriquecimiento también es totalmente compatible.

## Flexibilidad en el diseño de sondas y la estrategia de objetivos

Illumina Complete Long Read Prep with Enrichment, Human ofrece una gran flexibilidad para elegir y diseñar paneles de sonda personalizados que se ajusten a los objetivos del estudio. Las regiones objetivo individuales pueden abarcar desde bases individuales hasta cientos de kilobases. El tamaño total del panel puede variar desde paneles personalizados de tan solo 2,5 Mb hasta más de 95 Mb.

\* Requiere datos de WGS de lectura corta estándar con una cobertura de al menos 30x de la misma muestra para el análisis. Se pueden utilizar archivos FASTQ de una muestra procesada previamente.



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 al menos 30x de la misma muestra para el análisis. Se pueden utilizar archivos FASTQ de una muestra procesada previamente.

Los investigadores pueden utilizar lecturas largas selectivas para mejorar la cobertura en regiones específicas que se sabe que son difíciles de asignar con datos de lectura corta. Como alternativa, las lecturas largas pueden estar dirigidas a cubrir genes completos o incluso regiones multigénicas largas para permitir la fase de hebra retrasada de variantes y la llamada de haplotipos.

En la herramienta DesignStudio hay diversos paneles prediseñados disponibles (tabla 1). Estos paneles si selectivos de genes complejos de relevancia médica (CMRG, challenging medically-relevant genes),<sup>4</sup> genes a los que suelen dirigirse los ensayos de pruebas farmacogenéticas (PGx),<sup>5-7</sup> genes en la lista de hallazgos secundarios del Colegio Americano de Genética Médica y Genómica (ACMG, American College of Medical Genetics and Genomics) (ACMG SF v3.1)<sup>8</sup> o a la región completa del complejo principal de histocompatibilidad (MHC, major histocompatibility complex).<sup>9</sup> El panel Illumina Human Comprehensive Panel, que se dirige principalmente a regiones aisladas de baja cobertura dentro de genes codificantes de proteínas, también está disponible como panel prediseñado o prefabricado listo para enviar (Illumina, N.º de catálogo 20113836).<sup>10,11</sup> El software DesignStudio admite el diseño de paneles personalizados a partir de archivos BED† o la modificación de prediseños existentes.

### Profundidad de secuenciación recomendada para paneles de sonda personalizados

Illumina Complete Long Read Prep with Enrichment, Human proporciona un rendimiento muy uniforme y sólido. Para los paneles prediseñados probados, se logró un rendimiento óptimo con aproximadamente 1,5 Gb de datos de secuencias (unos 5 millones de lecturas «paired-end») por panel objetivo de 1 Mb de tamaño (figura 3). En el caso de los paneles de nuevo diseño con una rendimiento desconocido, se recomiendan 3 Gb de datos de secuencias (unos 10 millones de lecturas «paired-end») por panel objetivo de 1 Mb de tamaño, sujeto a reducción con mayor optimización.

## Cobertura de alta precisión y fase de hebra retrasada de regiones difíciles

Los paneles de sondas de enriquecimiento de lectura larga centrados en mejorar regiones específicas de baja cobertura, como Illumina Human Comprehensive Panel y el panel de CMRG, mejoran la precisión de las llamadas de variantes en regiones objetivo complejas (figura 4). El enriquecimiento de lectura larga con el panel de CMRG también ayuda a mejorar la integridad de la cobertura y la detección de variantes en las regiones codificantes de proteínas (figura 5, figura 6).

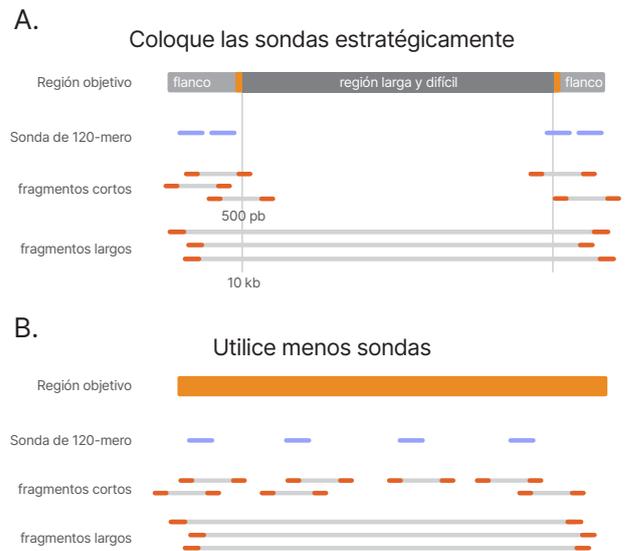


Figura 2: La hibridación de fragmentos largos aumenta la eficiencia del enriquecimiento. La hibridación de fragmentos largos ofrece ventajas respecto a la captura de fragmentos cortos, incluida (A) la colocación estratégica de sondas fuera de regiones difíciles para el diseño de sondas, como contenido extremo de GC, baja complejidad o repeticiones, y (B) una menor necesidad de sondas para capturar cada región objetivo. El algoritmo de DesignStudio realiza una búsqueda en regiones objetivo en secciones de 1 kb para obtener un contenido óptimo de GC y las regiones de mayor especificidad para la colocación de sondas.

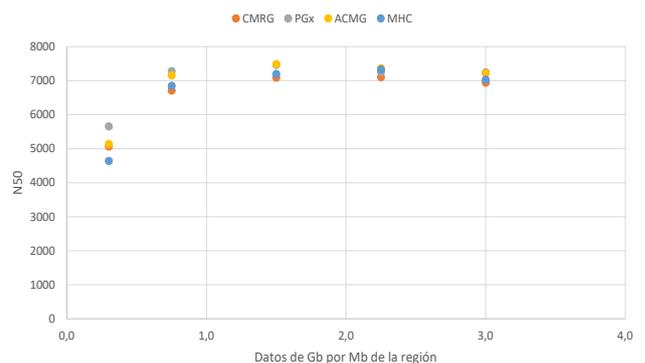


Figura 3: Requisitos de secuenciación para paneles de sondas personalizados. La valoración de los datos de secuenciación necesarios para un N50 máximo muestra que 1,5 Gb (unos 5 millones de lecturas «paired-end») por región objetivo de 1 Mb resuelve de forma eficiente las regiones objetivo para la generación de datos con Illumina Complete Long Read.

† BED, formato de datos extensibles del navegador (browser extensible data).

Tabla 1: Paneles de sondas de enriquecimiento prediseñados para Illumina Complete Long Read Prep with Enrichment, Human

Paneles <sup>a</sup>	Panel de CMRG	Panel de PGx	Panel de ACMG	Panel de MHC
Genes objetivo	391 genes médicamente relevantes que se sabe que son difíciles de resolver con lecturas cortas <sup>5</sup>	98 genes a los que suelen dirigirse los ensayos de pruebas farmacogenéticas <sup>6-8</sup>	78 genes únicos de la lista de hallazgos secundarios del ACMG (ACMG SF v3.1) <sup>9</sup>	>140 genes en la región completa del MHC en el conjunto GRCh38.p14 <sup>10</sup>
Tamaño de la región objetivo <sup>b</sup>	22,5 Mb	8,1 Mb	7 Mb	4,9 Mb
Rendimiento de secuenciación por muestra <sup>c</sup>	~67,5 Gb	~24,3 Gb	~21 Gb	~14,7 Gb
N.º de sondas	~22 500	~8200	~6900	~5000
N50 <sup>d</sup>	6,1 kb	7,3 kb	7,3 kb	7,3 kb
Bloque de fase de hebra retrasada N50 <sup>d,e</sup>	82,8 kb	94,4 kb	94,4 kb	357 kb
Tamaño medio de la región objetivo <sup>e</sup>	58 kb	83 kb	88 kb	5000 kb
Uniformidad <sup>d,f</sup>	97,9 %	99,0 %	99,5 %	97,8 %
Enriquecimiento de lectura completada <sup>d,f</sup>	80,1 %	79,3 %	66,3 %	67,5 %
% de SNV heterocigóticas en fase de hebra retrasada <sup>d</sup>	98,9 %	98,9 %	99,6 %	98,6 %

- a. CMRG, genes complejos de relevancia médica; PGx, farmacogenómica; ACMG, Colegio Americano de Genética Médica y Genómica; MHC, complejo principal de histocompatibilidad.
- b. El tamaño de la región objetivo es la suma de las longitudes de ubicación de la sonda completada, fusionadas donde se superponen.
- c. Requiere un experimento de secuenciación de 2 x 150 pb y de 5 A 10 M de lecturas «paired-end» (~1,5-3 Gb de datos) por Mb de región objetivo, lo que genera una cobertura final de aproximadamente 30x de las Illumina Complete Long Reads. Los requisitos de datos de paneles personalizados por muestra son solo un punto de partida recomendado. Los usuarios pueden optimizar los datos asignados en función del rendimiento del panel.
- d. 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.
- e. Los tamaños de los bloques de fase de hebra retrasada se limitan a los tamaños de las regiones objetivo contiguas individuales.
- f. Uniformidad de la cobertura calculada como % >0,2 \* media. Enriquecimiento de lectura completada calculado como 100 \* (lecturas completadas alineadas con el objetivo/ lecturas alineadas totales).

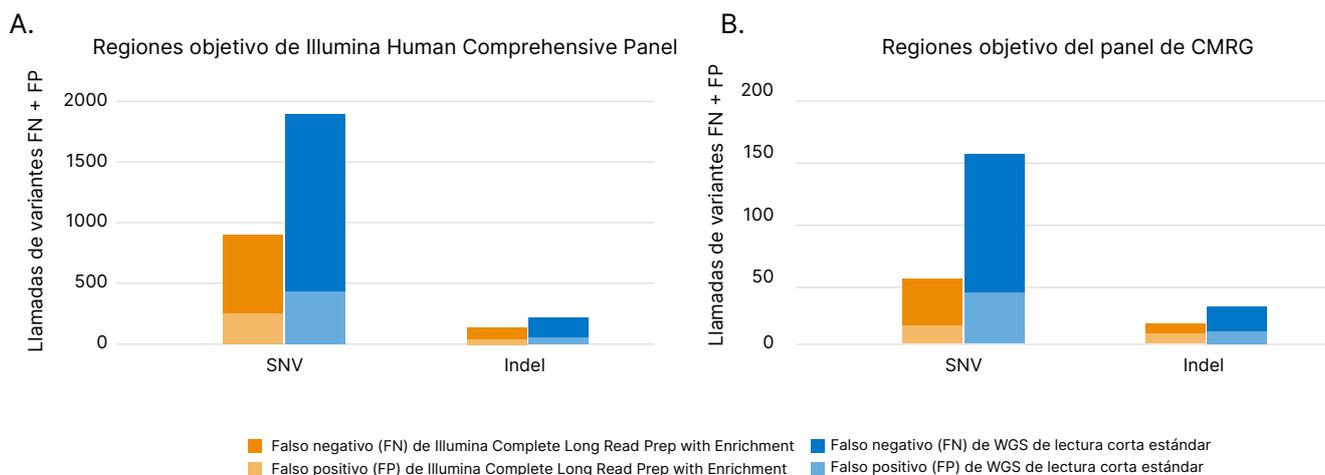


Figura 4: Lecturas largas selectivas para mejorar la precisión de la llamada de variantes en regiones complejas. Llamadas de variantes falsas negativas (FN) más falsas positivas (FP) para variantes de nucleótido único (SNV, single nucleotide variants) e inserciones/delecciones (indel) en regiones génicas HG002 a las que se dirigen (A) Illumina Human Comprehensive Panel o (B) el panel de CMRG, utilizando Illumina Complete Long Read Prep with Enrichment (naranja) en comparación con la WGS de lectura corta estándar (azul).



Figura 5: Las lecturas largas selectivas mejoran las regiones de baja cobertura. Gráficos de Integrative Genomics Viewer (IGV) a partir de la secuenciación de lectura larga de *HBG1* con Illumina Complete Long Read Prep, Human WGS (arriba) e Illumina Complete Long Read Prep with Enrichment, Human y el panel de CMRG (centro) en comparación con la WGS de lectura corta estándar (abajo).

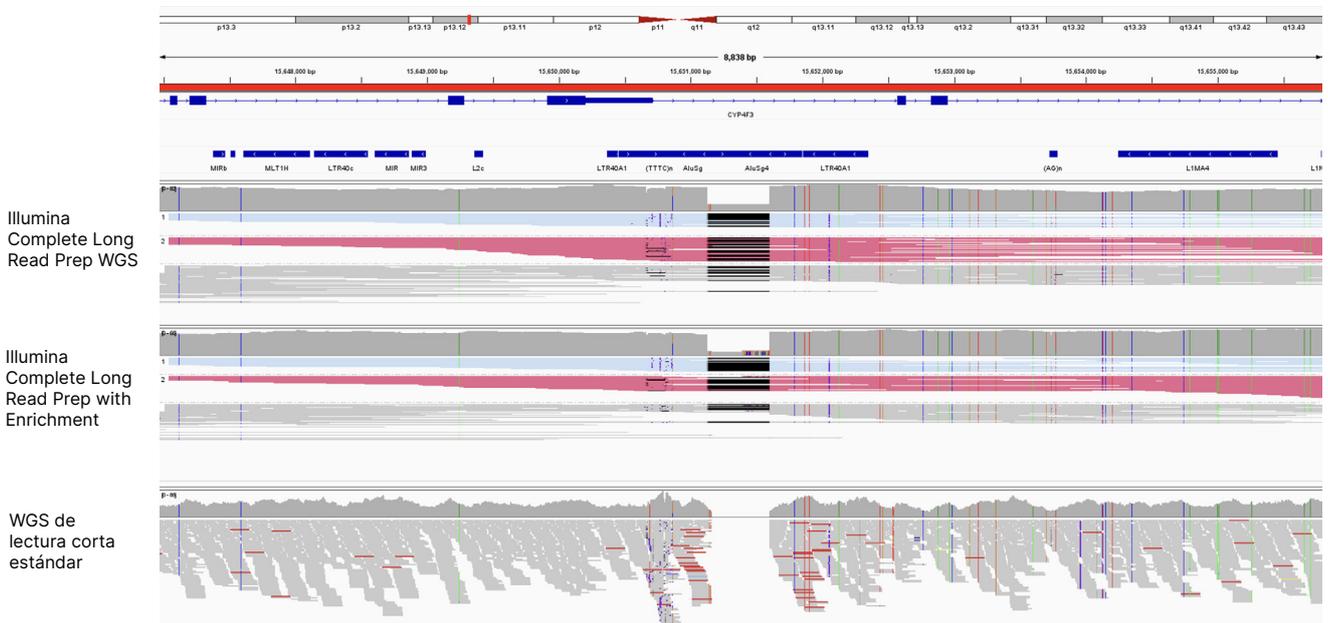


Figura 6: Clara resolución de los bordes de delección con lecturas largas selectivas. Gráficos de IGV a partir de la secuenciación de lectura larga y fase de hebra retrasada de *CYP4F3* con Illumina Complete Long Read Prep, Human WGS (arriba) e Illumina Complete Long Read Prep with Enrichment, Human y el panel de CMRG (centro) en comparación con la WGS de lectura corta estándar (abajo). Alelo 1 en azul, alelo 2 en rosa.

### Bloques de fase de hebra retrasada larga para resolver haplotipos

El bloque de fase de hebra retrasada N50<sup>†</sup> de cada panel está relacionado con la longitud contigua de las regiones objetivo (figura 7, tabla 1). Los paneles de CMRG, PGx y ACMG están diseñados para dirigirse a genes de interés de longitud completa y produjeron un bloque de fase de hebra retrasada N50 medio de entre unas 80 y 95 kb para permitir la fase de hebra retrasada completa de alelos heterocigóticos (figura 8). El panel de MHC se dirige a una única región contigua de aproximadamente 4,9 Mb y produjo un bloque de fase de hebra retrasada N50 medio de más de 350 kb para permitir la resolución de la región génica de longitud completa (figura 9).

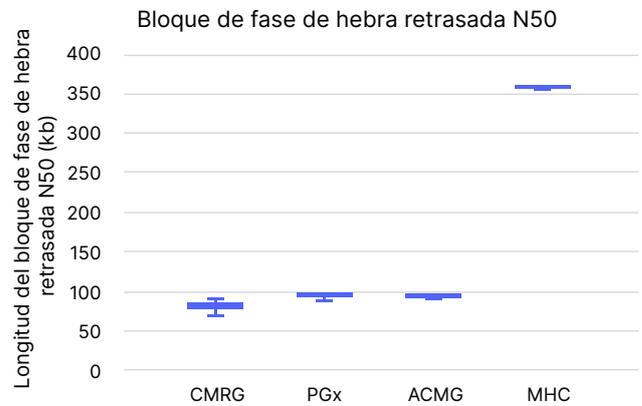


Figura 7: El bloque de fase de hebra retrasada N50 depende de la longitud de las regiones objetivo contiguas. Los paneles de CMRG, PGx y ACMG se dirigen a genes de interés de longitud completa y produjeron un bloque de fase de hebra retrasada N50 medio de unas 80 a 95 kb. El panel de MHC se dirige a la región génica completa del complejo principal de histocompatibilidad y produjo un bloque de fase de hebra retrasada N50 medio de más de 350 kb. El tamaño medio de la región objetivo es de 58 kb para el panel de CMRG, 83 kb para el panel de PGx, 88 kb para el panel de ACMG y 5000 kb para el panel de MHC.

† El bloque de fase de hebra retrasada N50 refleja la longitud del bloque más corto de secuencia contigua al 50 % de la longitud total de ensamblaje de las regiones objetivo.

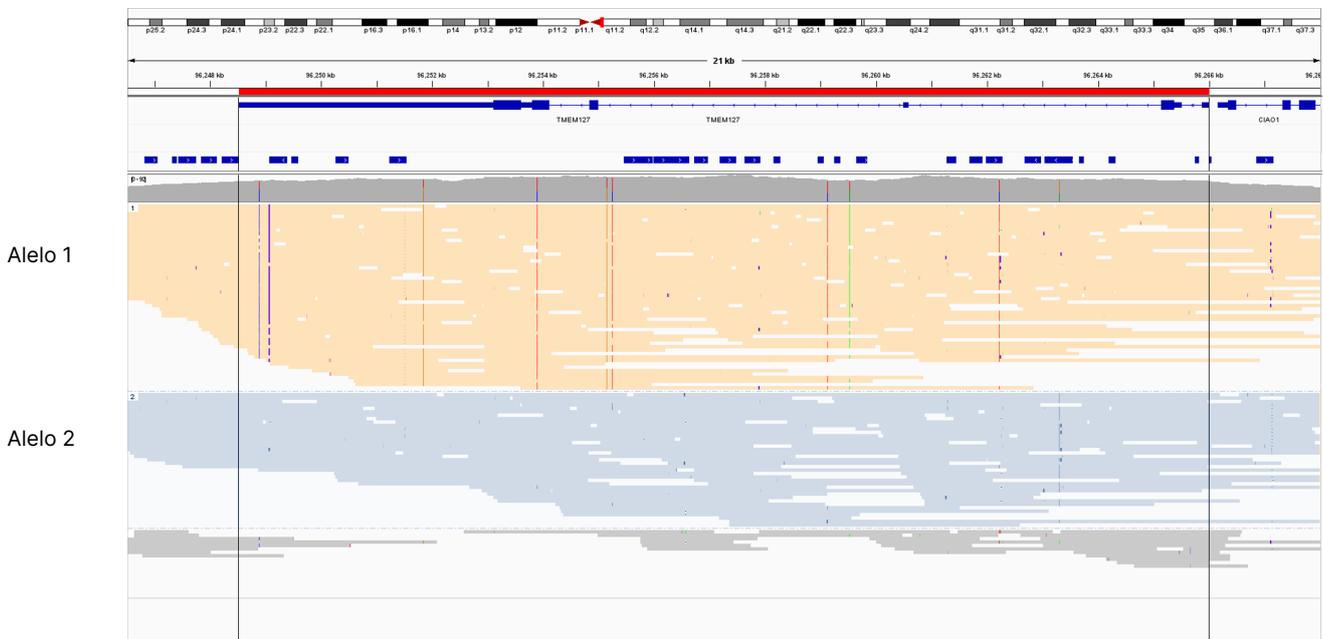


Figura 8: Las lecturas largas selectivas permiten la fase de hebra retrasada de regiones con SNV heterocigóticas. Los gráficos de IGV de la secuenciación de lectura larga muestran la fase de hebra retrasada completa en un bloque de fase de hebra retrasada para *TMEM127*, un gen de 21 kb, utilizando Illumina Complete Long Read Prep with Enrichment, Human y el panel de ACMG. Alelo 1 en amarillo. Alelo 2 en azul.

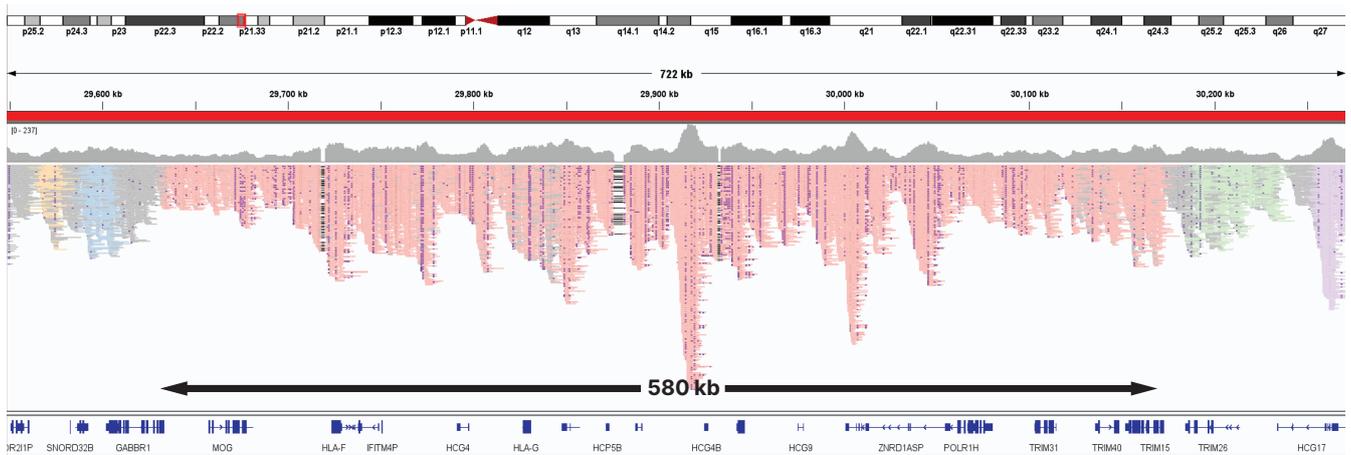


Figura 9: Las lecturas largas selectivas ayudan a resolver haplotipos en genes polimórficos. Gráficos de IGV de secuenciación de lectura larga utilizando Illumina Complete Long Read Prep with Enrichment, Human. Fase de hebra retrasada sobre una región de 722 kb en el locus del MHC. Hay una región de 580 kb (rosa) encapsulada en un bloque de fase de hebra retrasada.

## Resumen

Illumina Complete Long Read Prep with Enrichment, Human complementa la WGS de lectura corta probada de Illumina y se centra en la secuenciación de lectura larga, donde proporciona el mayor valor. Los investigadores tienen una gran flexibilidad a la hora de elegir paneles prediseñados o utilizar el algoritmo DesignStudio para diseñar paneles personalizados para un enriquecimiento selectivo de lectura larga. Los paneles de sonda de enriquecimiento se pueden dirigir de forma selectiva para mejorar la cobertura o aportar información mediante la fase de hebra retrasada de genes completos para una WGS rentable y de alta precisión con una solución de flujo de trabajo completa.

## Información adicional

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

[Herramienta de diseño de ensayos DesignStudio](#)

[Tecnología de secuenciación de lectura larga](#)

## Bibliografía

1. Illumina. Illumina Complete Long Read Prep, Human data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf). Año de publicación: 2022. Fecha de consulta: 22 de septiembre de 2023.
2. Illumina. Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human technical note. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf). Año de publicación: 2022. Fecha de consulta: 22 de septiembre de 2023.
3. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. <https://www.illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html>. Año de publicación: 2023. Fecha de consulta: 22 de septiembre de 2023.
4. Wagner J, Olson ND, Harris L, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.* 2022;40(5):672-680. doi:10.1038/s41587-021-01158-1
5. PharmGKB. VIPs: Very Important Pharmacogenes. [pharmgkb.org/vips](https://www.pharmgkb.org/vips). Fecha de consulta: 22 de septiembre de 2023.
6. National Library of Medicine. GTR: Genetic Testing Registry. Precision HealthPGx Panel (25 Genes). [ncbi.nlm.nih.gov/gtr/tests/593428/](https://www.ncbi.nlm.nih.gov/gtr/tests/593428/). Fecha de actualización: 29 de noviembre de 2022. Fecha de consulta: 22 de septiembre de 2023.

7. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn.* 2016;18(1):109-123. doi:10.1016/j.jmoldx.2015.08.005
8. Miller DT, Lee K, Abul-Husn NS, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2022;24(7):1407-1414. doi:10.1016/j.gim.2022.04.006
9. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var.* 2022;9(1):49. doi:10.1038/s41439-022-00226-5
10. Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. Año de publicación: 2021. Fecha de consulta: 30 de agosto de 2023.
11. Illumina. Illumina Human Comprehensive Panel data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf](https://illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf). Año de publicación: 2024. Fecha de consulta: 26 de enero de 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](http://www.illumina.com/company/legal.html).  
M-GL-02189 ESP v1.0