

# Custom panel design for Illumina Complete Long Read Prep with Enrichment, Human

Highly flexible, targeted  
long-read enrichment  
for human genomes



## Introduction

When performing human whole-genome sequencing (WGS), several regions may be difficult to map with short reads alone. Long-read sequencing can complement standard short-read WGS data to help address these challenging regions. Illumina Complete Long Reads technology uses a standard next-generation sequencing (NGS) workflow to generate contiguous long-read sequences on Illumina sequencing systems with a single analysis pipeline (Figure 1).<sup>1-3</sup> Illumina Complete Long Read Prep with Enrichment, Human adds a targeted approach for more cost-effective long-read sequencing.\* Illumina Complete Long Read enrichment chemistry provides high flexibility with targets and probe design to help resolve difficult-to-map regions or provide added insight from phased sequencing.

## Designing enrichment probe panels for long reads

Illumina Complete Long Read Prep with Enrichment, Human uses a different probe design strategy to capture longer fragments (~7–10 kb) than the approach typically used to capture short fragments (~200–500 bp).

\* Requires ≥ 30× standard short-read WGS data from the same sample for analysis. FASTQ files from a previously run sample can be used.

Illumina DesignStudio™ software is a free, user-friendly tool for designing enrichment probe panels. The DesignStudio algorithm considers GC content, target specificity, and probe spacing, ie, how many probes will be in the target region. Standard spacing for 120mer short-read enrichment panels is a 250–350-bp probe window. For long-read enrichment panel design, probe spacing was tested at multiple lengths and a one-kilobase window was found to be optimal for cost-effective, high-efficiency capture.

Hybridization enrichment effectiveness is highly dependent on probe specificity. The percent of on-target enrichment directly affects the amount of sequencing needed to achieve target coverage depth. High specificity is harder to achieve for repetitive regions. However, using a larger probe window allows greater flexibility to exclude poor-performing probes, avoid repeat regions (up to the window size of 1 kb), and maintain enrichment efficiency with fewer probes (Figure 2). The DesignStudio algorithm can use these considerations to recommend probe placement. Third party panels should use similar guidelines for best performance and cost efficiency. Standard enrichment probe spacing is also fully compatible.

### Flexibility in probe design and target strategy

Illumina Complete Long Read Prep with Enrichment, Human offers high flexibility to choose and design custom probe panels to match study goals. Individual targeted regions can span single bases up to hundreds of kilobases. The total panel size can range from custom panels as small

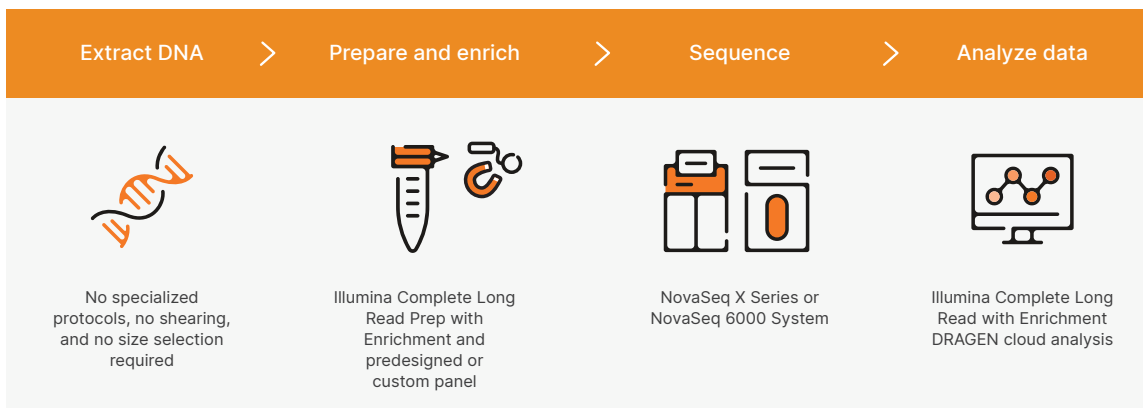


Figure 1: Part of an integrated workflow—Access cost-effective, targeted long-read WGS data using a scalable, optimized library prep with enrichment protocol, proven Illumina sequencing chemistry, and DRAGEN secondary analysis. Requires ≥ 30× standard short-read WGS data from the same sample for analysis. FASTQ files from a previously run sample can be used.

as 2.5 Mb up to > 95 Mb. Researchers can use targeted long reads to enhance coverage across specific regions known to have low mappability with short-read data. Alternatively, long reads can be targeted to cover entire genes up to long multigene regions to enable phasing of variants and calling haplotypes.

Several predesigned panels are available in the DesignStudio tool (Table 1). These panels target challenging medically relevant genes (CMRG),<sup>4</sup> genes commonly targeted by pharmacogenetic (PGx) testing assays,<sup>5-7</sup> genes on the American College of Medical Genetics and Genomics (ACMG) secondary findings list (ACMG SF v3.1),<sup>8</sup> or the full major histocompatibility complex (MHC) region.<sup>9</sup> The Illumina Human Comprehensive Panel, which primarily targets discrete low-coverage regions within protein-coding genes, is also available as a predesign or ready-to-ship premanufactured panel (Illumina, Catalog no. 20113836).<sup>10,11</sup> DesignStudio software supports design of custom panels from BED files<sup>†</sup> or modification of existing predesigns.

### Recommended sequencing depth for custom probe panels

Illumina Complete Long Read Prep with Enrichment, Human provides highly consistent and robust performance. For the predesigned panels tested, optimal performance was achieved with ~1.5 Gb of sequence data (~5M paired-end reads) per 1 Mb of target panel size (Figure 3). For newly designed panels with unknown performance, 3 Gb of sequence data (~10M paired-end reads) per 1 Mb of target panel size is a recommended starting point, subject to reduction with further optimization.

### High-accuracy coverage and phasing of difficult regions

Long-read enrichment probe panels focused on enhancing specific low-coverage regions, like the Illumina Human Comprehensive Panel and the CMRG panel, enhance variant calling accuracy in targeted challenging regions (Figure 4). Long-read enrichment with the CMRG panel also helps improve completeness of coverage and detection of variants across protein-coding regions (Figure 5, Figure 6).

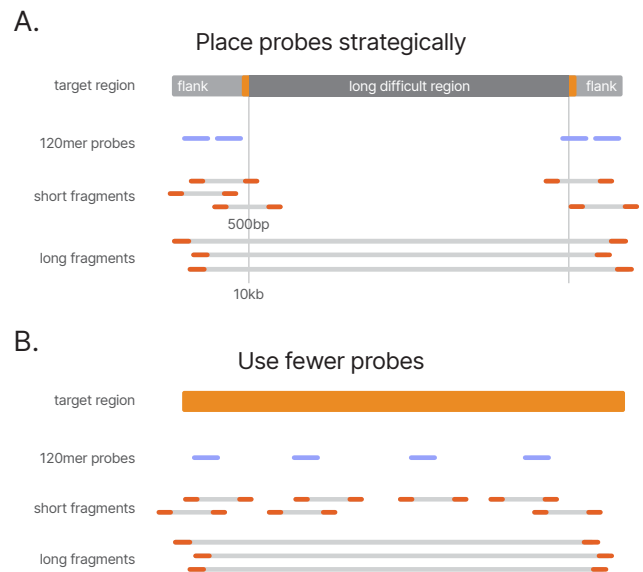


Figure 2: Long-fragment hybridization increases enrichment efficiency—Long-fragment hybridization offers advantages over short-fragment capture, including (A) strategic placement of probes outside of difficult regions for probe design, such as extreme GC content, low complexity, or repeats, and (B) fewer probes are needed to capture each target region. The DesignStudio algorithm searches across target regions in 1-kb sections for optimal GC content and highest specificity regions to place probes.

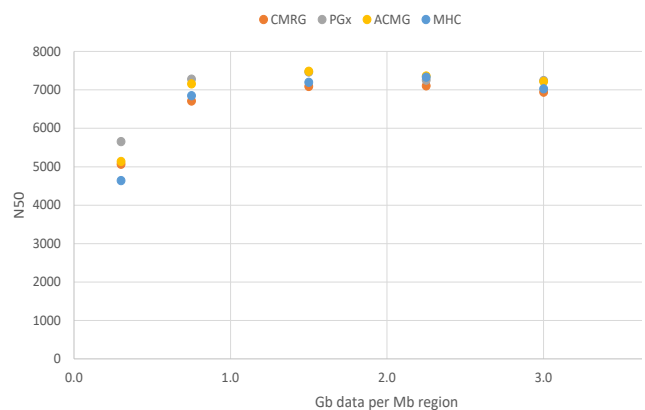


Figure 3: Sequencing requirements for custom probe panels— Titration of sequencing data required for maximum N50 shows that 1.5 Gb (~5M paired-end reads) per Mb target region efficiently resolves targeted regions for Illumina Complete Long Read data generation.

† BED, browser extensible data format.

Table 1: Predesigned enrichment probe panels for Illumina Complete Long Read Prep with Enrichment, Human

Panels <sup>a</sup>	CMRG panel	PGx panel	ACMG panel	MHC panel
Genes targeted	391 medically relevant genes known to be challenging to resolve with short reads <sup>5</sup>	98 genes commonly targeted by pharmacogenetic testing assays <sup>6-8</sup>	78 unique genes from the ACMG secondary findings list (ACMG SF v3.1) <sup>9</sup>	> 140 genes across the full MHC region in the GRCh38.p14 assembly <sup>10</sup>
Target region size <sup>b</sup>	22.5 Mb	8.1 Mb	7 Mb	4.9 Mb
Sequencing output per sample <sup>c</sup>	~67.5 Gb	~24.3 Gb	~21 Gb	~14.7 Gb
No. of probes	~22.5K	~8.2K	~6.9K	~5.0K
N50 <sup>d</sup>	6.1 kb	7.3 kb	7.3 kb	7.3 kb
Phase block N50 <sup>d,e</sup>	82.8 kb	94.4 kb	94.4 kb	357 kb
Mean target region size <sup>e</sup>	58 kb	83 kb	88 kb	5000 kb
Uniformity <sup>d,f</sup>	97.9%	99.0%	99.5%	97.8%
Padded read enrichment (PRE) <sup>d,f</sup>	80.1%	79.3%	66.3%	67.5%
% Heterozygous SNV phased <sup>d</sup>	98.9%	98.9%	99.6%	98.6%

a. CMRG, challenging medically relevant genes; PGx, pharmacogenomics; ACMG, American College of Medical Genetics and Genomics; MHC, major histocompatibility complex.  
 b. Target region size is the sum of padded probe location lengths, merged where they overlap.  
 c. Requires 2 × 150 bp sequencing run and 5M–10M paired-end reads (~1.5–3 Gb data) per Mb target region, generating approximately 30× final coverage of Illumina Complete Long Reads. Custom panel data requirements per sample are only a recommended starting point. Users may optimize allotted data based on panel performance.  
 d. Data generated using 50 ng HG002 genomic DNA (Coriell, Catalog no. NA24385). Performance may vary with DNA input and sample quality.  
 e. Phase block sizes are limited to the sizes of individual contiguous targets regions.  
 f. Uniformity of coverage calculated as % > 0.2 \* mean. PRE calculated as 100 \* (padded target aligned reads / total aligned reads).

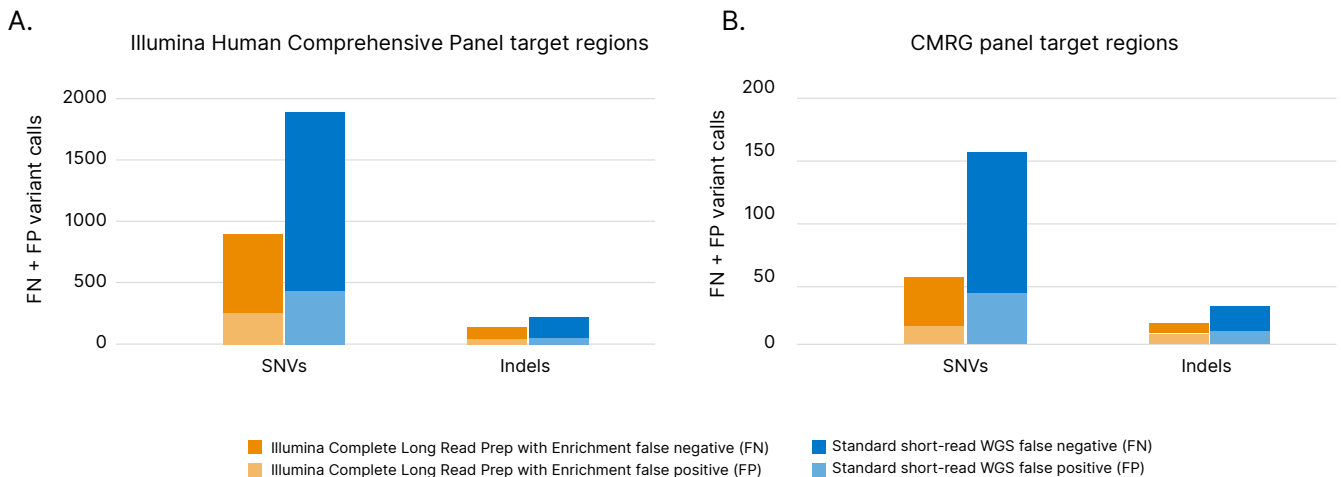


Figure 4: Targeted long reads to enhance variant calling accuracy in challenging regions—False negative (FN) plus false positive (FP) variant calls for single nucleotide variants (SNVs) and insertion–deletions (indels) in HG002 genic regions targeted by the (A) Human Comprehensive Panel or (B) CMRG panel, using Illumina Complete Long Read Prep with Enrichment (orange) compared to standard short-read WGS (blue).

PANEL DESIGN FOR ILLUMINA COMPLETE LONG READS ENRICHMENT



Figure 5: Targeted long reads enhance regions of low coverage—Integrative Genomics Viewer (IGV) plots from long-read sequencing of *HBG1* using Illumina Complete Long Read Prep, Human WGS (top) and Illumina Complete Long Read Prep with Enrichment, Human and the CMRG panel (middle) compared to standard short-read WGS (bottom).

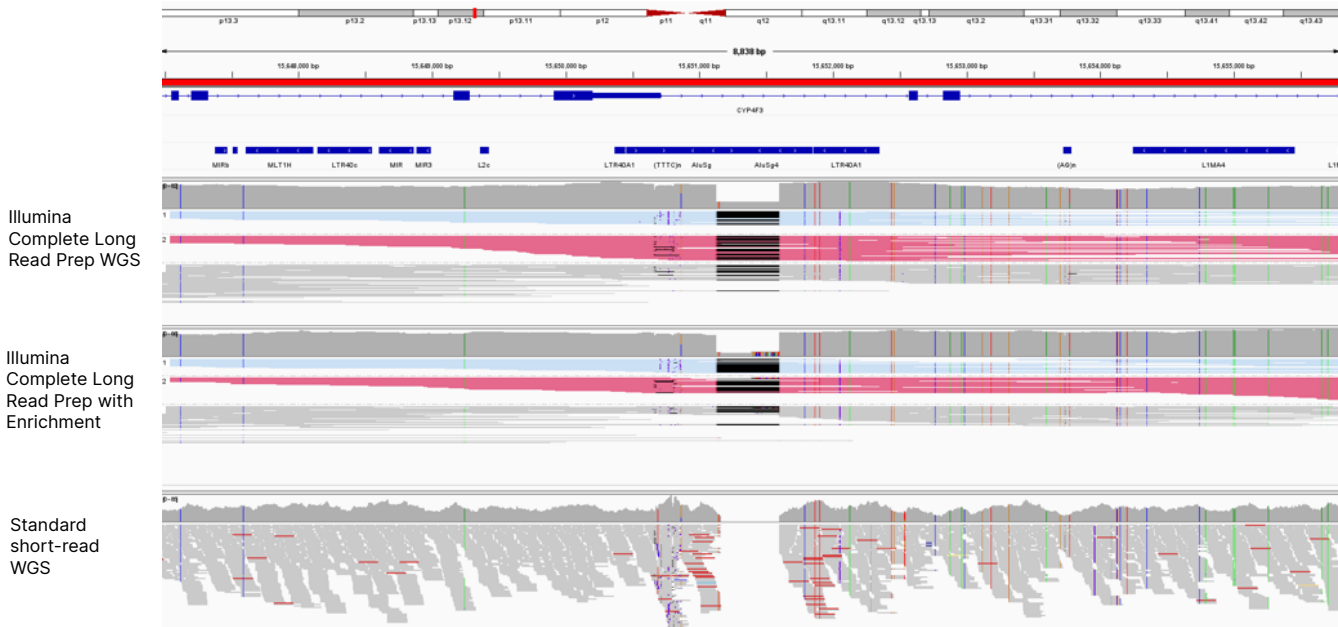


Figure 6: Clear resolution of deletion borders with targeted long reads—IGV plots from long-read sequencing and phasing of *CYP4F3* using Illumina Complete Long Read Prep, Human WGS (top) and Illumina Complete Long Read Prep with Enrichment, Human and the CMRG panel (middle) compared to standard short-read WGS (bottom). Allele 1 in blue, allele 2 in pink.

### Long phase blocks to resolve haplotypes

The phase block N50<sup>‡</sup> of each panel is related to the contiguous length of the target regions (Figure 7, Table 1). The CMRG, PGx, and ACMG panels are designed to target full-length genes of interest and yielded a mean phase block N50 of ~80–95 kb to enable full phasing of heterozygous alleles (Figure 8). The MHC panel targets a single ~4.9 Mb contiguous region and yielded a mean phase block N50 of over 350 kb to enable resolution of the full-length genic region (Figure 9).

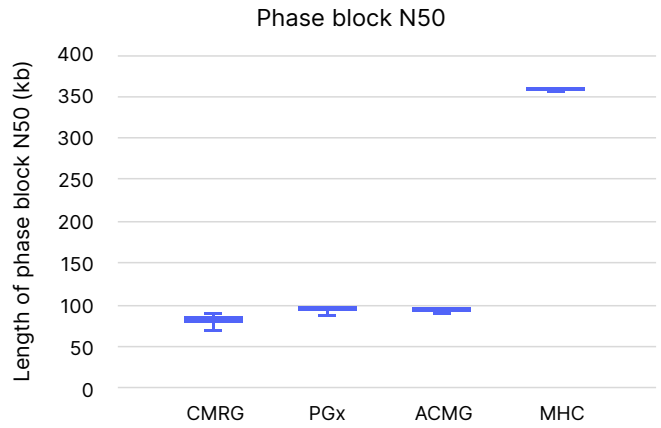


Figure 7: Phase block N50 depends on the length of contiguous target regions—The CMRG, PGx, and ACMG panels target full-length genes of interest and yielded mean phase block N50 of ~80–95 kb. The MHC panel targets the full major histocompatibility complex genic region and yielded a mean phase block N50 of over 350 kb. The mean target region size is 58 kb for the CMRG panel, 83 kb for the PGx panel, 88 kb for the ACMG panel, and 5000 kb for the MHC panel.

<sup>‡</sup> Phase block N50 reflects the length of the shortest block of contiguous sequence at 50% of the total assembly length of target regions.

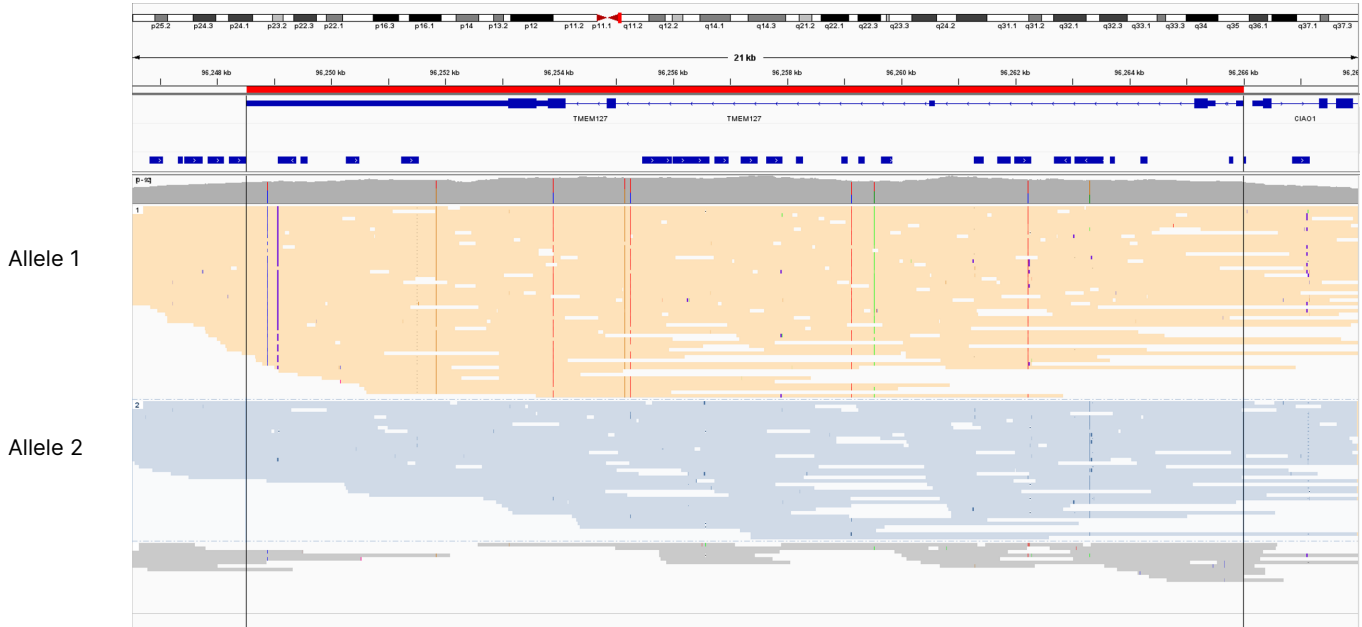


Figure 8: Targeted long reads enable phasing of regions with heterozygous SNVs—IGV plots from long-read sequencing shows full phasing in one phase block for *TMEM127*, a 21 kb gene, using Illumina Complete Long Read Prep with Enrichment, Human and the ACMG panel. Allele 1 in yellow. Allele 2 in blue.

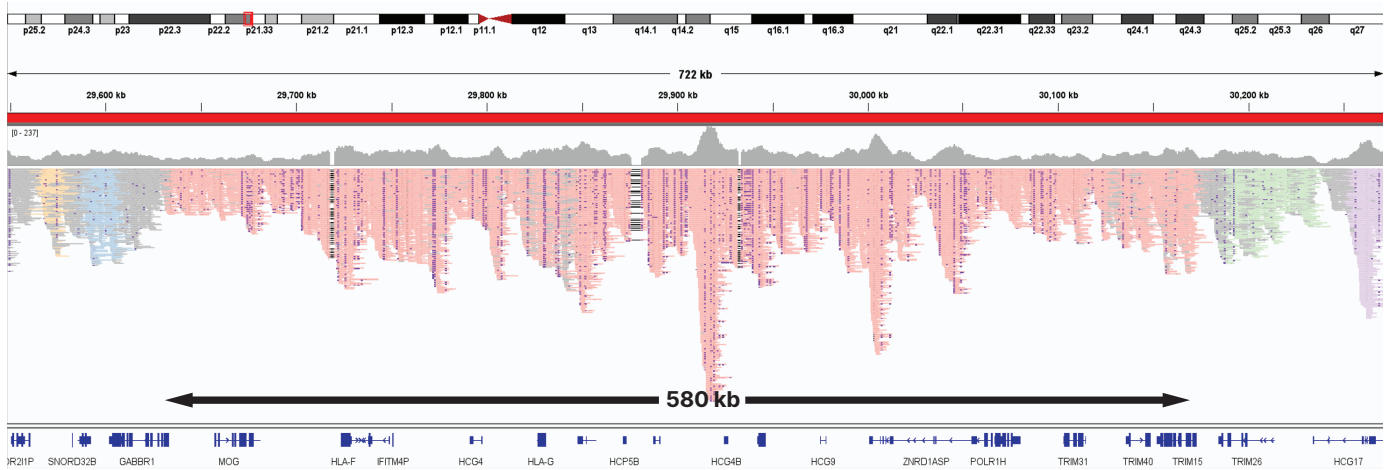


Figure 9: Targeted long reads help resolve haplotypes in polymorphic genes—IGV plots from long-read sequencing using Illumina Complete Long Read Prep with Enrichment, Human. Phasing over a 722 kb region in the MHC locus. A 580 kb region (pink) is encapsulated in one phase block.

## Summary

Illumina Complete Long Read Prep with Enrichment, Human complements proven Illumina short-read WGS and focuses long-read sequencing where it provides greatest value. Researchers have high flexibility to choose predesigned panels or use the DesignStudio algorithm to design custom panels for long-read targeted enrichment. Enrichment probe panels can be targeted to enhance coverage or add insight by phasing entire genes for cost-effective, highly accurate WGS with a full workflow solution.

## Learn more

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

[DesignStudio assay design tool](#)

[Long-read sequencing technology](#)

## References

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). Published 2022. Accessed September 22, 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). Published 2022. Accessed September 22, 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>. Published 2023. Accessed September 22, 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). Accessed September 22, 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/). Updated November 29, 2022. Accessed September 22, 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. Published 2021. Accessed August 30, 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://www.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). Published 2024. Accessed January 26, 2024.



1.800.809.4566 toll-free (US) | +1.858.202.4566 tel  
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners.  
For specific trademark information, see [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-02189 v1.0