TECHNICAL NOTE

Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human

Exceptional performance in challenging, difficult-to-map regions

illumina®

Introduction

Next-generation sequencing (NGS) enables scientists to decipher the genome for a deeper understanding of biology. Proven Illumina sequencing by synthesis (SBS) chemistry combined with award-winning DRAGEN™ secondary analysis delivers whole-genome sequencing (WGS) data with outstanding accuracy.[1] Still, a small fraction of the genome can benefit from longer read lengths to enable more accurate resolution and mapping of these challenging regions. However, many long-read sequencing solutions have complex workflows, high DNA input requirements, and highly variable results.[2-5]

Illumina Complete Long Reads streamlines long-read sequencing and makes it accessible for genomic scientists. Illumina Complete Long Read Prep, Human is the first product based on this novel long-read chemistry. This high-performance WGS assay uses a standard NGS workflow to generate contiguous long-read sequences on the NovaSeq™ 6000 System and NovaSeq X Series (Figure 1). The efficient, single-day library preparation protocol is easy to scale for high-throughput studies and requires as little as 10 ng DNA input with no specialized extractions, shearing, or size selection.

This technical note demonstrates the exceptional performance of Illumina Complete Long Read Prep, Human with a wide range of DNA input, DNA quality, and sample types for highly accurate, comprehensive WGS for germline variant analysis.

Methods

The reference sample HG002/NA24385 was obtained from the Genome in a Bottle consortium (GIAB) as purified genomic DNA (gDNA). Separately, blood and saliva samples were obtained and gDNA was extracted using commercially available kits with either regular or high molecular weight extraction, following the manufacturer's instructions.

Library preparation

"Land-marked" libraries were prepared from varying amounts of input gDNA using Illumina Complete Long Read Prep, Human (Illumina, Catalog no. 20089108). Briefly, long, single-molecule DNA fragments are enzymatically marked with unique patterns (land-marks). Unmarked libraries, used during analysis to produce long contiguous reads representing the original single-molecule fragment, were prepared using Illumina DNA PCR-Free Prep, Tagmentation (Illumina, Catalog no. 20041794).

Sequencing

Land-marked and unmarked libraries were sequenced during separate runs on the NovaSeq 6000 System at 2 × 150 bp read length.

Data analysis

Assembly and rendering of land-marked reads was carried out using the Illumina Complete Long Read WGS app on BaseSpace™ Sequence Hub. Interactive Genomics Viewer (IGV) was used for data analysis and visualization.

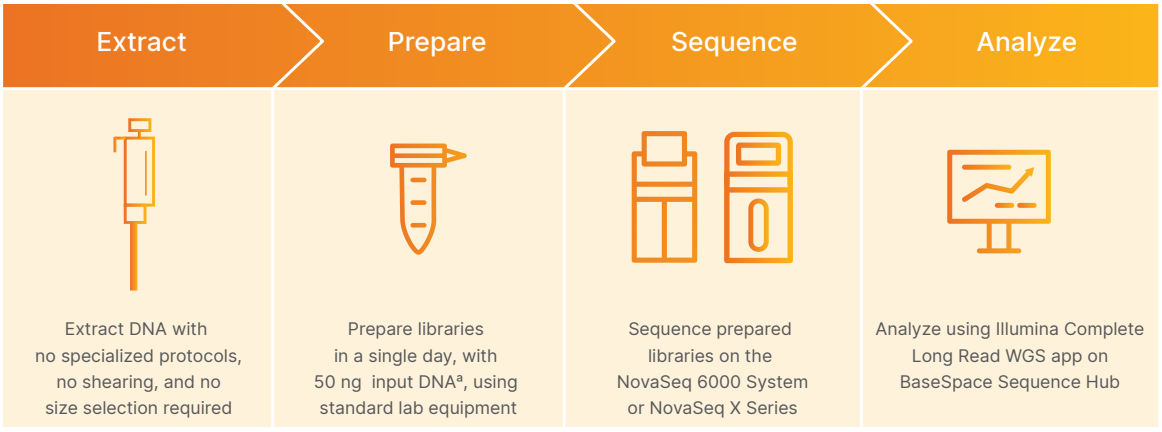

Figure 1: Illumina Complete Long Read Prep, Human workflow—Illumina Complete Long Read Prep, Human offers a streamlined workflow consisting of four steps: DNA extraction, library preparation, sequencing, and data analysis.

a. While inputs as low as 10 ng are possible, Illumina recommends 50 ng of DNA

High-quality data in challenging regions

Illumina Complete Long Read Prep, Human provides coverage in regions where short reads can't be mapped unambiguously. This enables contiguous sequencing of disease-associated genes and pseudogenes with known mapping issues with short reads only, including *STRC*, (Figure 2A), *SEMG1* (Figure 2B), and *SULTIA1* (Figure 2C).

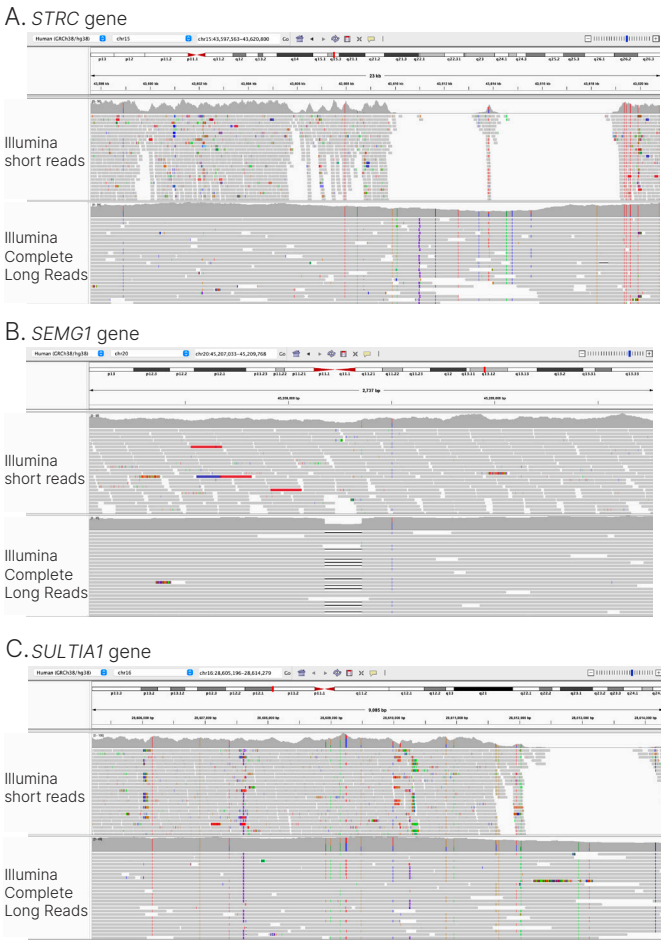

Figure 2: Comprehensive coverage enables accurate detection of multiple variant types in challenging regions—Illumina Complete Long Read Prep, Human enables accurate sequencing of disease-associated genes in challenging genomic regions, including (A) *STRC*, (B) *SEMG1*, and (C) *SULTIA1*.

Furthermore, Illumina Complete Long Read Prep, Human provides phase blocks significantly larger than typical gene size. This enables uniform coverage throughout the human leukocyte antigen (HLA) locus, a densely polymorphic region with high levels of sequence homology, pseudogenes, variability, and numerous alleles. While difficult to sequence accurately, Illumina Complete Long Read Prep, Human delivers uniform coverage throughout the HLA region and enables genome phasing to clearly visualize and distinguish alleles between maternal and paternal chromosomes (Figure 3).

Figure 3: Uniform coverage enables clear haplotype phasing—Illumina Complete Long Read Prep, Human provides comprehensive coverage in the HLA region and enables clear and accurate phasing of alleles within the HLA region.

For a more comprehensive evaluation of variant calling accuracy, Illumina Long Read Prep, Human was compared against a third-party long-read platform and Illumina short reads in the PrecisionFDA Truth Challenge V2. Sponsored by PrecisionFDA, the GIAB, and the National Institute of Standards and Technology (NIST), this challenge was launched to assess small variant calling pipeline performance on a common frame of reference, with a focus on "difficult-to-map" regions. Using this data set, Illumina Complete Long Read Prep, Human demonstrates exceptional accuracy and outperfroms the alternative long-read platform with an F1 score of 99.87% (Figure 4).

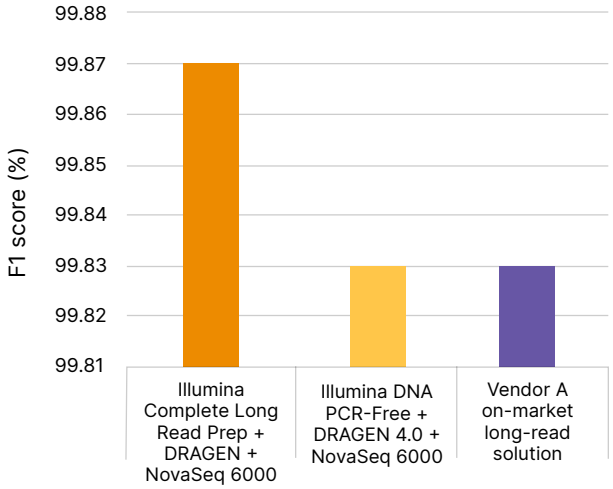

Figure 4: A new standard for accuracy—Illumina Complete Long Read Prep, Human delivers unprecendented accuracy for variant calling, as measured by F1 score (%), reflecting precision and recall for WGS.[1]

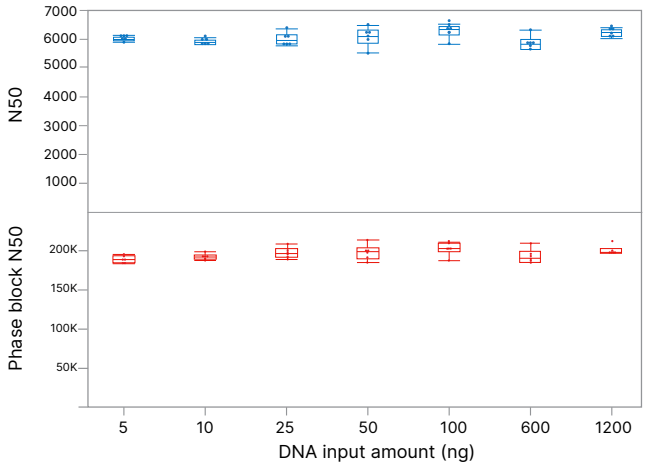

Figure 5: High-quality performance across a wide range of DNA input—Illumina Complete Long Read Prep, Human produces high-quality libraries prepared from DNA inputs from 5 ng to 1200 ng (in triplicate) generate similar data quality for N50 and phase block N50. N50 is defined as the sequence length of the shortest contig at 50% of the total assembly length; it can also be used as a measure of phase blocks.

Robust performance across DNA input amounts

The performance of Illumina Complete Long Read Prep, Human was evaluated across a range of DNA input amounts. Libraries were prepared from 10, 25, 100, 600, and 1200 ng of DNA in triplicate. Sequencing results demonstrated that libraries prepared from as little as 5 ng of input displayed high-quality metrics, including N50 and phasing block N50 (Figure 5). While Illumina recommends an input of 50 ng DNA, lower inputs can be used.

Support for varying DNA quality

To simulate damaged DNA, high-quality gDNA was subjected to shearing using a Covaris ultrasonicator for one or three seconds. DNA quality was evaluated using a Fragment Analyzer (Figure 6). Libraries were prepared using sheared gDNA input and Illumina Complete Long Read Prep, Human. Sequencing results showed that DNA sheared for one second generated high-quality data and enabled accurate variant calling (Figure 7). Additionally, input gDNA that was subjected to increasing cycles of freezing and thawing produced high-quality data (Figure 8).

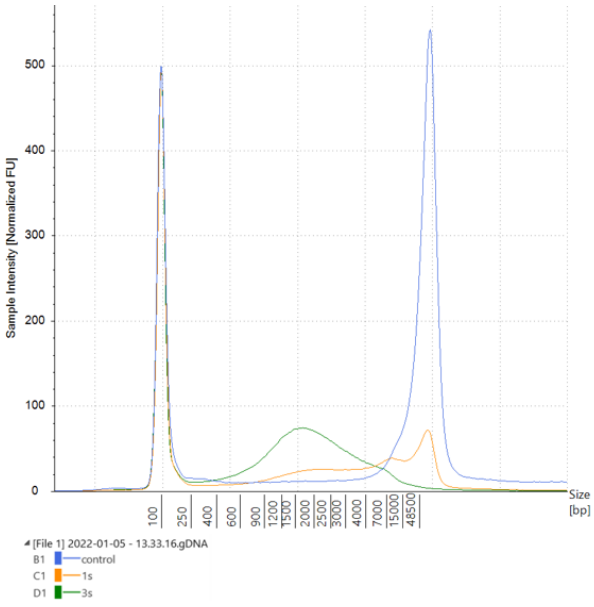

Figure 6: Sheared DNA quality—Analysis of DNA on a Fragment Analyzer shows decreased quality with increased shearing time. DNA samples were subjected to Covaris shearing for 1 or 3 seconds and quality was measured using a Fragment Analyzer.

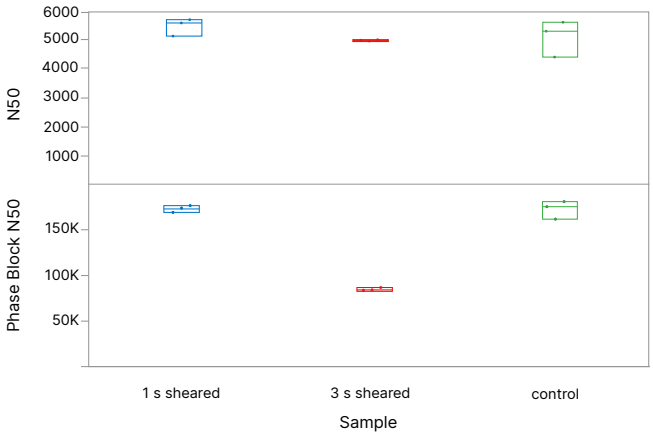

Figure 7: Sheared DNA sequencing metrics —DNA sheared for 1 and 3 seconds was used as input for Illumina Complete Long Read Prep, Human. Resulting libraries generated high-quality long-read sequencing data, as measured by N50 and phase block N50.

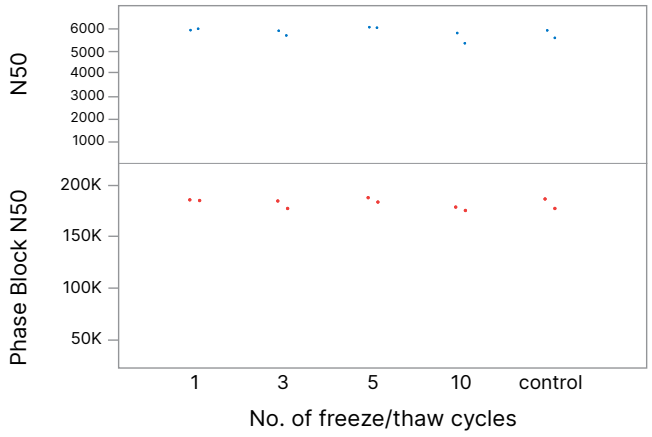

Figure 8: Freeze-thawed DNA sequencing metrics —DNA subjected to 1, 3, 5, and 10 freeze-thaw cycles was used as input for Illumina Complete Long Read Prep, Human. Resulting libraries generated high-quality long-read sequencing data, as measured by N50 and phase block N50.

Flexible support for various sample types

Illumina Complete Long Read Prep, Human supports various sample types, including blood and saliva. After extraction using either a standard or high molecular weight (HMW) kits, DNA quality was evaluated using a Fragment Analyzer (Figure 9). Sequencing results showed that DNA from blood and saliva generated high-quality data (Figure 10). Of note, there was no difference in performance between DNA extracted with a standard kit and a HMW kit (Figure 10).

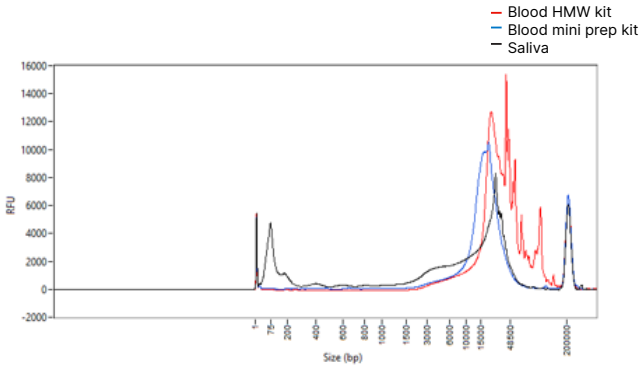

Figure 9: Blood and saliva DNA quality—DNA extracted from blood (HMW, mini) or saliva was analyzed on a Fragment Analyzer.

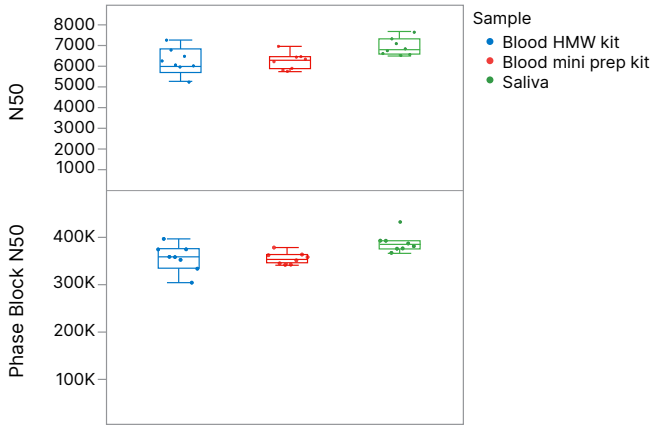

Figure 10: High-quality performance with different sample types—Illumina Complete Long Read Prep, Human libraries generated from blood (HMW, mini) and saliva deliver high-quality data, as measured by N50 and phase block N50.

Summary

Long-read sequencing chemistry provides additional information for regions of the genome that are challenging to resolve with short-read NGS. Illumina Complete Long Reads make long-read sequencing accessible and streamlined for genomic scientists. Illumina Complete Long Read Prep, Human, the first product built on this novel chemistry, offers a streamlined, familiar workflow and synergy with proven Illumina SBS chemistry and DRAGEN analysis. This highly innovative chemistry enables short- and long-read sequencing to be performed on a single instrument, making long-read NGS accessible for genomic labs. Illumina Complete Long Read Prep, Human demonstrates robust performance with DNA of varying quantity and quality from different sample sources. The result is a highly scalable and accurate human WGS solution.

Learn more

Illumina Complete Long Read Prep, Human

References

1. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at Precision-FDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html. Accessed January 12, 2023.
2. Pacific Biosciences. Preparing DNA for PacBio HiFi sequencing—Extraction and quality control. pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf. Published 2022. Accessed January 12, 2023.
3. Pacific Biosciences. Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0. pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf. Published 2022. Accessed January 12, 2023.
4. Oxford Nanopore Technologies. Ligation Sequencing Kit. store.nanoporetech.com/us/ligation-sequencing-kit110.html. Accessed January 12, 2023.
5. Pacific Biosciences. Low Yield Troubleshooting Guide. pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf. Published 2018. Accessed January 12, 2023.

illumina®

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

M-GL-01421 v1.0