ULTIMA GENOMICS

Application Note

Germline Whole-Genome Sequencing with Ultima Genomics

Highlights

- Whole-genome sequencing results show high accuracy for germline variant calling, with SNV F[1] of 99.78-99.84% and Indel F1 of 96.28-98.75% using the NIST HCR v4.2.1 using GIAB HG001-HG007 reference samples.1 The average Indel F1 score rises to 99.06% when excluding long homopolymers (length ≥ 12).
- Solaris 2.0 chemistry and an updated amplification process provide even coverage across the genome, reducing coverage bias in high-GC regions, improving coverage in clinically relevant genes as well as short tandem repeat (STR) regions.

Introduction

Whole-genome sequencing (WGS) is a method for determining the complete DNA sequence of an organism's genome in a single experiment. Whole genome sequencing of humans is commonly used as a research tool in the context of population studies to identify the functional effects of genomic variation. In clinical diagnostics, WGS plays a crucial role in identifying mutations in rare genetic disorders, predicting disease susceptibility, and drug response. As WGS becomes more affordable, it is redirecting diagnostic practices from traditional, cost-effective methods, such as SNP arrays, to the more encompassing and informative whole genome sequencing approach.

Here we demonstrate the utility of the Ultima sequencing platform to accurately determine germline variants by WGS profiling of genomic DNA, both in the context of well-characterized standard reference samples and sample-specific pathogenic variants. In addition, we show how Ultima Genomics' new Solaris 2.0 chemistry and amplification technology provides improvements in coverage across the genome, reducing coverage bias in high-GC regions, across several clinically relevant genes, as well as in short tandem repeat (STR) regions. The Solaris 2.0 workflows provide a simplified amplification workflow with hands-on time of only 10 minutes with one touch point, and a reduced overall system footprint.

Technology

Ultima Genomics utilizes a sequencing architecture that combines an open flow cell design on a circular wafer with large surface area, integrating rotational reagent delivery, optical end-point detection, and flow chemistry without reversible terminators.[2] The sequencing platform, running at current maximum capacity, enables sequencing billions of reads from genomic DNA at fast sequencing run times (6-11hrs).

Results

Short variant calling performance on GIAB samples

To evaluate short variant calling performance, PCR-free genomic DNA libraries were prepared from the seven standard Genome in a Bottle (GIAB) reference samples HG001-HG007[3,4] and sequenced using Ultima's new Solaris 2.0 library prep and amplification workflow. Sequencing data was downsampled to 120Gb, with key variant calling metrics summarized in Table 1. Short variants were called using a modified version of DeepVariant that was optimized for analysis of Ultima data, accounting for quality values that are unique to Ultima's flow chemistry base-calling. Variant calling performance was assessed in comparison to reference GIAB truth sets, using the full NIST HCR v4.2.1 for each reference sample. Table 1 shows overall average concordance of single nucleotide polymorphism (SNP) variant calls over the dataset was F1 = 99.79%, and for Indel variant calls it was F1= 96.99%. Variant calling accuracy for homopolymer indels was above 99% for all lengths up to 10 and remained high for all lengths up to 12 (F1 > 95% for n=12, Figure 1). The average whole genome Indel F1 score, excluding homopolymer regions longer than 12, jumps to F1 = 98.71%, indicating that the vast majority of Indel errors occur within long homopolymer regions. This is confirmed when examining the SNP and Indel F1 scores in coding regions of the genome, where the average SNP F1 = 99.60% and the Indel F1 = 98.84% (Table 1).

For Research Use Only. Not for use in diagnostic procedures.

© 2026 Ultima Genomics, Inc. All rights reserved.
D1001426 Rev 01

	WGS			Coding Regions	
	SNP	Indel	Indel (excluding hmer ≥ 12)	SNP	Indel
HG001	99.84%	96.75%	98.95%	99.70%	99.23%
HG002	99.80%	96.31%	98.90%	99.57%	98.70%
HG003	99.79%	96.28%	99.00%	99.63%	99.39%
HG004	99.79%	96.19%	98.98%	99.65%	99.26%
HG005	99.80%	98.75%	99.24%	99.57%	98.27%
HG006	99.79%	97.28%	99.20%	99.64%	98.73%
HG007	99.78%	97.10%	99.13%	99.55%	98.66%
Average	**99.79%**	**96.99%**	**99.06%**	**99.60%**	**98.84%**

Table 1. Genome in a Bottle (GIAB) reference whole-genome sequencing accuracy statistics across samples HG001-HG007. Whole genome variant calling performance was assessed in comparison to reference truth sets defined by the NIST v4.2.1 HCR unless otherwise noted.

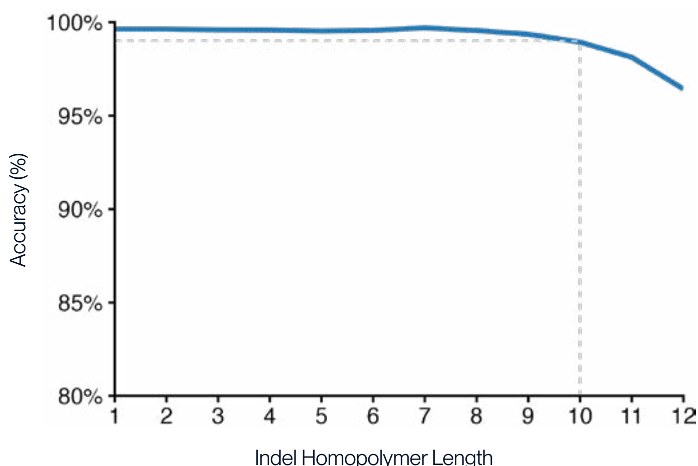

Figure 1. Average Indel accuracy (F1) per homopolymer length for HG001-HG007. Horizontal dashed line corresponds to F1 = 99%.

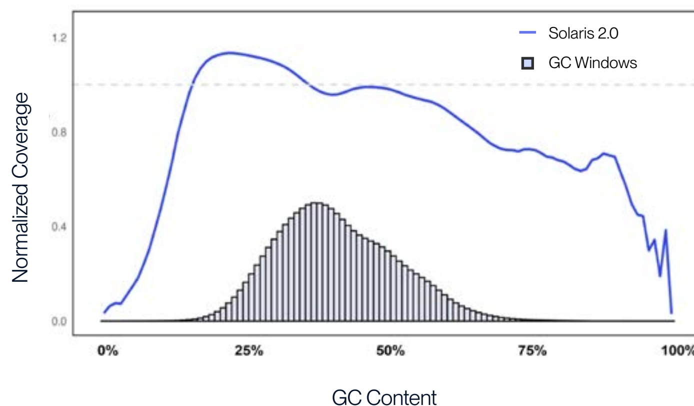

Figure 2. Normalized coverage (blue line) in HG002 amplified using the Solaris 2.0 workflow and plotted as a function of GC content across the genome. The underlying histogram depicts the GC-content distribution across the genome.

Improved GC coverage with Solaris 2.0

Ultima introduces new amplification technology and Solaris 2.0 chemistry which provide improvements in coverage across the genome, reducing coverage bias in high-GC regions, including across several clinically relevant genes, as well as in short tandem repeat (STR) regions. The Solaris 2.0 workflows provide a simplified amplification workflow with hands-on time of only 10 minutes with one touch point, and a reduced overall system footprint. Figure 2 shows normalized coverage by GC content measured in HG002 in a sample amplified using the Solaris 2.0 workflow.

The improvement in high-GC coverage translates into more consistent depth across GC-rich, clinically relevant loci that are traditionally challenging for short-read sequencing workflows. By reducing dropout and uneven amplification in these regions, Solaris 2.0 enables more reliable variant detection, improved confidence in calling across promoter regions and GC-dense exons, and stronger performance in genes associated with inherited disease and oncology applications. Enhanced uniformity at high GC content also supports improved characterization of STR regions, where amplification bias can otherwise complicate repeat length determination and downstream analysis.

Improvement of coverage in high-GC, clinically relevant genes is highlighted in Table 2. The Solaris 2.0 chemistry and amplification workflow demonstrates substantially improved normalized coverage (normalized to the chromosome median coverage) across nearly all evaluated loci, including challenging genes such as FMR1, HOXA13, HTT, and MUC16, with normalized coverage increasing, on average, from 40% to 88%.

Short tandem repeats (STRs) are among the most challenging regions of the genome to interrogate using short-read sequencing due to their repetitive structure, GC richness, and propensity for amplification

For Research Use Only. Not for use in diagnostic procedures.

© 2026 Ultima Genomics, Inc. All rights reserved.
D1001426 Rev 01

bias. For many technologies, these features can lead to uneven coverage, reduced mapping confidence, and allele dropout—for example, in clinically relevant loci such as the CAG repeat expansion in the HTT gene associated with Huntington's disease. Ultima's new amplification chemistry reduces sequence-dependent bias, enhancing coverage uniformity across STR regions, increasing confidence in repeat detection and sizing. In the following example, we demonstrate how improved amplification technology from Ultima delivers more consistent coverage across the HTT locus, enabling clearer characterization of this clinically important STR.

66 genomic DNA samples obtained from the Coriell Institute for Medical Research were amplified using the Ultima Genomics Solaris 2.0 workflow and sequenced to assess coverage across STR regions. Samples were selected to represent a range of repeat lengths (normal, intermediate, and pathogenic repeat expansions), including four samples classified as clinically affected by Huntington's disease based on repeat expansions in the HTT gene[5-8]: NA03643, NA13514, NA13510, and NA13515. These samples contain expanded CAG repeats consistent with disease-associated alleles. For all 66 samples, PCR-free libraries were prepared and sequenced (to ~40x coverage) to evaluate coverage uniformity, read distribution across STR regions, and repeat count.

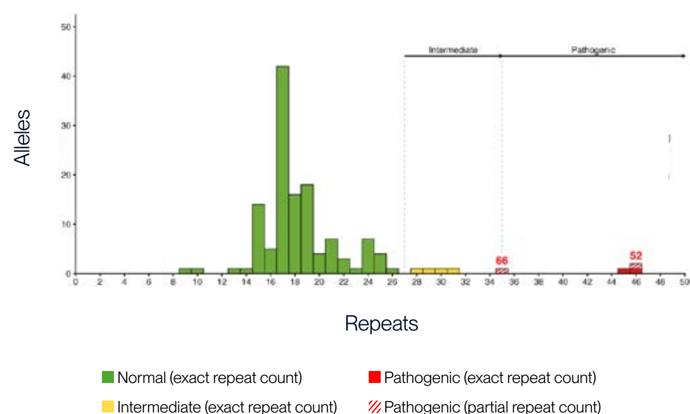

Figure 3. Distribution of observed (CAG) repeat counts at the HTT locus across alleles sampled from 66 Coriell genomic DNA samples. Green bars represent normal-range repeat counts, yellow bars indicate intermediate alleles, and red bars denote pathogenic repeat expansions. Hatched red bars represent partially-measured repeat counts for correctly characterized pathogenic samples with the expected repeat length shown above the bar.

Figure 3 shows the distribution of CAG repeat lengths at the HTT locus. Most observed alleles fall within the normal repeat range (green bars), clustering tightly around the typical non-pathogenic repeat counts. A smaller number of alleles extend into the intermediate range (yellow), while clearly expanded pathogenic alleles (red) are observed beyond the established clinical threshold. Hatched red bars represent partially-measured repeat counts for correctly characterized pathogenic samples with the expected repeat length shown above the bar in Figure 3. The separation between normal, intermediate, and pathogenic ranges highlights the ability of the sequencing workflow to accurately cover and resolve repeat length across the clinical spectrum.

	Solaris 1.0	Solaris 2.0
ATXN1	81%	65%
ATXN7	48%	71%
B3GALT6	39%	93%
CACNA1A	56%	91%
CBL	45%	65%
DIP2B	39%	93%
FMR1	17%	119%
GIPC1	42%	91%
GLS	56%	93%
HOXA13	17%	54%
HTT	25%	82%
LRP12	22%	88%
MARCHF6	62%	161%
NIPA1	22%	76%
PABPN1	62%	91%
PHOX2B	31%	99%
PRDM12	25%	79%
ZIC2_1	42%	85%
ZIC2_2	22%	82%
Average	40%	88%

Median normalized coverage calculated across a selection of GC-rich, clinically relevant genes comparing Solaris to Solaris 2.0 chemistry and amplification workflows.

Conclusions

This study demonstrates the capabilities and advantages of WGS sequencing with Ultima Genomics for germline variant detection as well as the improved coverage gained from Ultima's improved Solaris 2.0 chemistry and amplification workflow. Integration of advanced variant calling tools, such as DeepVariant tailored specifically for Ultima Genomics data, has shown significant improvements in both precision and recall rates.
Ultima Genomics' unique sequencing technology offers a significant reduction in cost without compromising the quality of data, a pivotal advancement in the field of genomics. This cost-effectiveness, coupled with high throughput and accuracy, makes Ultima an invaluable partner in both clinical and research settings. Ultima facilitates a more inclusive approach to WGS studies, allowing for broader population studies and more comprehensive genetic screenings in clinical diagnostics.

Methods

UG PCR-free Library Prep

Ultima Genomics (UG) compatible, PCR-free libraries were generated from approximately 500 ng of genomic DNA. The DNA was enzymatically fragmented, end-repaired, and A-tailed using the NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs). xGEN PCR-free Adapters for Ultima Genomics (Integrated DNA Technologies) were ligated; and resulting library molecules were size selected using a double-sided SPRI to capture average insert sizes of approximately 350-400 bp. Library size was determined by running 1 µL on a High Sensitivity D1000 TapeStation (Agilent Technologies), and library concentration was quantified using the NEBNext Library Quantification Kit (New England Biolabs).

Variant calling

Reads were aligned to a human pangenome reference graph to better represent population-scale genetic diversity and reduce reference bias in polymorphic and structurally complex regions. Alignment was performed using Giraffe[9], a haplotype-aware graph mapper that leverages indexed haplotypes to prioritize biologically plausible paths while maintaining computational efficiency comparable to linear aligners.

Variant calling was performed using an adapted implementation of Google DeepVariant (DV)[10] optimized for Ultima Genomics flow-based sequencing data. DeepVariant uses a two-step process: high-recall candidate variant generation followed by convolutional neural network (CNN) classification of each candidate as homozygous reference, heterozygous, or homozygous variant. To ensure compatibility with Ultima data, DeepVariant was customized in three ways: candidate-generation parameters were tuned to reflect platform-specific error profiles; additional Ultima-specific signal channels were incorporated into the CNN input; and the model was retrained on UG100 sequencing datasets. These modifications improved accuracy while reducing runtime by approximately tenfold. Given the distinct prior probability of homopolymer-associated deletions, particularly in exome data, a calibrated post-calling quality threshold was applied to such events. For low-confidence predictions, model ensembling across shuffled and augmented image representations was used to improve robustness.

Variant calling performance evaluation: Variant calling performance metrics (recall, precision, and F1) for reference samples HG001-7 were calculated for the NIST HCR v4.2.1 using vcfeval.[11]

Variant annotation: Mutation details were extracted from the Coriell database[12], annotated using coordinates from ClinVar,[13] and verified with IGV visualization, to establish an accurate mutation dataset.

Calls evaluation: Short variants were validated with bcftools[14] using DeepVariant VCF files and the variants coordinates.

STR analysis: Short tandem repeat (STR) genotypes were determined using an alignment-based caller that performs local Smith–Waterman realignment of reads to locus-specific templates containing flanking sequence and variable repeat lengths. For each candidate allele size, per-read likelihoods were derived from alignment scores, and diploid genotypes were inferred by maximum-likelihood estimation across all overlapping reads.

References

1. National Institute of Standards and Technology (NIST). Genome in a Bottle Consortium: High-confidence small variant benchmark set v4.2.1. 2020. Available from: https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/
2. Almogy G, et al. Cost-efficient whole genome sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform. bioRxiv 2022;2022.05.29.493900. doi:10.1101/2022.05.29.493900.
3. Zook JM, McDaniel J, Olson ND, et al. An open resource for accurately benchmarking small variant and reference calls. *Nat Biotechnol*. 2019;37(5):561–566. doi:10.1038/s41587-019-0074-6.
4. Wagner J, Olson ND, Harris L, et al. Benchmarking challenging small variants with linked and long reads. *Cell Genom*. 2022;2(5):100128. doi:10.1016/j.xgen.2022.100128.
5. Coriell Institute for Medical Research. NA03643. Camden (NJ): Coriell Institute for Medical Research; Available from: https://www.coriell.org/Search?Grid=0&q=-NA03643
6. Coriell Institute for Medical Research. NA13514. Camden (NJ): Coriell Institute for Medical Research; Available from: https://www.coriell.org/Search?Grid=0&q=-NA13514
7. Coriell Institute for Medical Research. NA13510. Camden (NJ): Coriell Institute for Medical Research; Available from: https://www.coriell.org/Search?Grid=0&q=-NA13510
8. Coriell Institute for Medical Research. NA13515. Camden (NJ): Coriell Institute for Medical Research; Available from: https://www.coriell.org/Search?Grid=0&q=-NA13515
9. Sirén J, Garrison E, Novak AM, Paten B, Durbin R. Giraffe: a haplotype-aware pangenome short-read mapper. *bioRxiv*. 2025;2025.09.29.678807. doi:10.1101/2025.09.29.678807.
10. Asri M, Chang PC, Mier JC, Sirén J, Eskandar P, Kolesnikov A, et al. Pangenome-aware DeepVariant. *bioRxiv*. 2025;2025.06.05.657102. doi:10.1101/2025.06.05.657102.
11. Cleary JG, Braithwaite R, Gaastra K, et al. Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines. *bioRxiv* 023754; 2015. doi:10.1101/023754.
12. Coriell Institute for Medical Research. NIGMS Human Genetic Cell Repository. Available from: https://www.coriell.org/1/NIGMS
13. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062–D1067. doi:10.1093/nar/gkx1153.
14. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. doi:10.1093/gigascience/giab008.

For Research Use Only. Not for use in diagnostic procedures.

© 2026 Ultima Genomics, Inc. All rights reserved.
D1001426 Rev 01