illumina®

DATA SHEET

Illumina TruPath™ Genome

Simplest whole-genome sequencing workflow

Experience the simplest sample-to-data workflow with 10 min hands-on prep time

Resolve difficult-to-map regions of the genome with short reads and proximity data using the NovaSeq™ X Series

Uncover long-distance insights with improved detection of structural variants and ultra-long phasing

Paradigm-changing genome accessibility

Short-read next-generation sequencing (NGS) has been the primary method used by researchers to assay the majority of the genome. Yet, some regions and variant types remain challenging to characterize, such as those with extreme repetitiveness, high sequence homology, or structural complexity. Recent studies have shown that these regions harbor variants that have potential roles in human genetic disease,[1,2] making deciphering them of the utmost importance when searching for underlying causes of genetic diseases.

Alternative methods, such as long-read sequencing, offer improved resolution, furthering insights into many of these difficult-to-map regions and variants. However, long-read methods tend to have complex workflows with well-established accuracy challenges that lead to variable results.

Illumina TruPath Genome with proximity mapped read technology changes the whole-genome sequencing (WGS) workflow paradigm. This innovative assay captures long DNA fragments with a simple 10-minute hands-on assay preparation step. Sequencing occurs on the existing NovaSeq X Series with cutting-edge DRAGEN™ secondary analysis. The result is an accurate, comprehensive genome enabling researchers to generate insights that span greater distances than the read lengths typically generated using long-read methods.

With TruPath Genome, it is now possible to harness the ease and accuracy of short-read sequencing to resolve difficult-to-map regions of the genome, improve structural variant (SV) detection, and generate phased reads and variant calls.

Unprecedented workflow simplicity

TruPath Genome vastly simplifies WGS implementation and reduces hands-on time to 10 minutes (Figure 1). Trusted Illumina transposase chemistry eliminates traditional library prep steps. Advanced bioinformatic algorithms leverage nanowell proximity information with standard paired-end, short-read data* to resolve relationships between variants separated by up to millions of bases.

* Standard paired-end, short-read data refers to data generated using whole-genome sequencing performed with manual library preparation and standard sequencing by synthesis (SBS) methods.

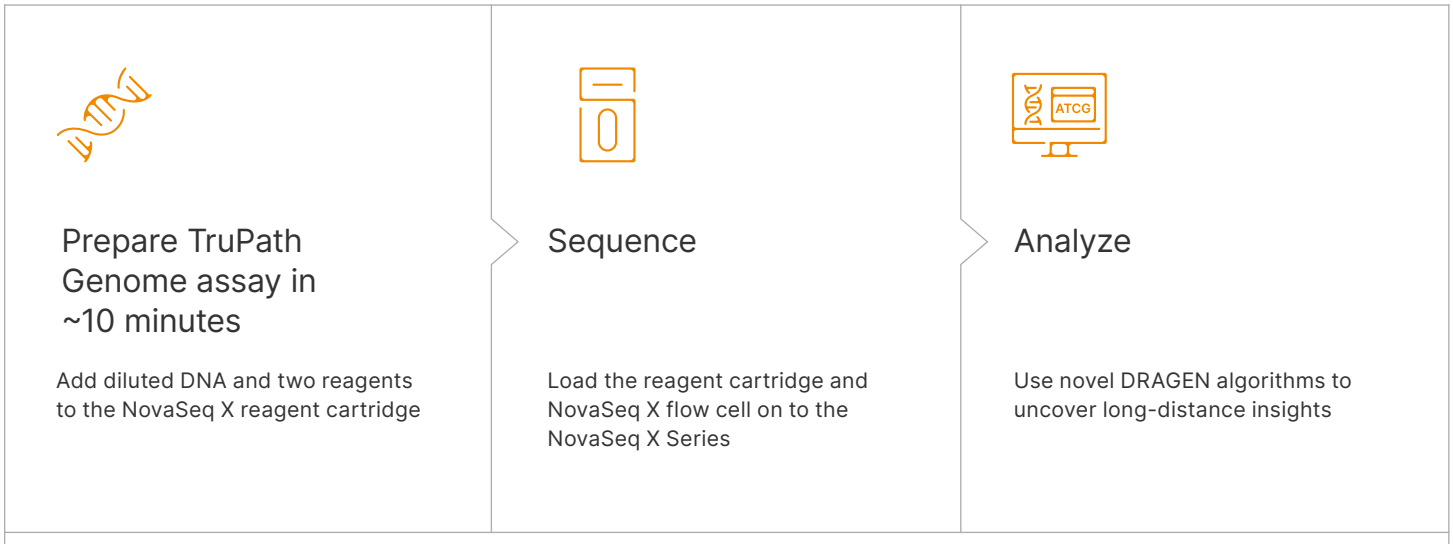


Figure 1: TruPath Genome workflow
DNA and one TruPath Genome reagent are added to the library tube strip, which is then inserted into the NovaSeq X reagent cartridge. Additional TruPath Genome reagents are then added to the cartridge. The cartridge and flow cell are loaded into the NovaSeq X Series and the run is started.

Samples

Purified double-stranded DNA (dsDNA) is extracted with standard or high molecular weight (HMW) commercial kits. The recommended input of 350 ng DNA is diluted with TruPath tagmentation buffer and added to the library tube strip. Input DNA can be isolated from cell lines or whole blood. DNA isolated from formalin-fixed paraffin-embedded (FFPE) and cell-free DNA (cfDNA) samples are not supported. For optimal performance, it is recommended that DNA is extracted using a method that preserves large fragments with 40% having lengths > 60 kb.[3]

For guidance on using additional sample types, varying sample quality, and optimal input amounts, read the TruPath Genome performance with various sample types, quality, and input technical note.

Sequencing

The library tube strip with diluted DNA and two TruPath Genome reagents are added to the NovaSeq X reagent cartridge and loaded into the NovaSeq X Series along with a NovaSeq X C2 or C8 flow cell, for sequencing two or eight human genomes, respectively.

Nanowells on the flow cell surface are primed with Illumina transposomes.† The DNA is introduced to the flow cell where it is captured by the transposomes across multiple wells, creating a constellation-like pattern of seeded nanowells (Figure 2A). Transposomes cleave the captured DNA and attach the fragments to the flow cell surface in a process called tagmentation‡ (Figure 2B). A wash step removes unbound DNA and transposomes.

Inside the nanowells, tagmented DNA is used to create clusters§ for sequencing. DNA from the same original long DNA fragment will form clusters in nearby nanowells. Sequencing occurs on the NovaSeq X Series using proven Illumina XLEAP sequencing by synthesis (XLEAP-SBS™) chemistry and a 2 × 150 bp run.

Analysis

After sequencing is complete, the DRAGEN™ Germline pipeline uses a novel algorithm that combines sequencing data with nanowell proximity information to align the DNA with a reference genome. Leveraging proximity information enables proper assignment of reads that previously didn't map to the reference genome,

† Transposomes are DNA transposase complexes that exist as a dimer.
‡ Tagmentation is the process of cutting a fragment of DNA and adding an adapter sequence (tagging) using a transposome.
§ A cluster is an amplified spot of DNA on a flow cell that will be sequenced.

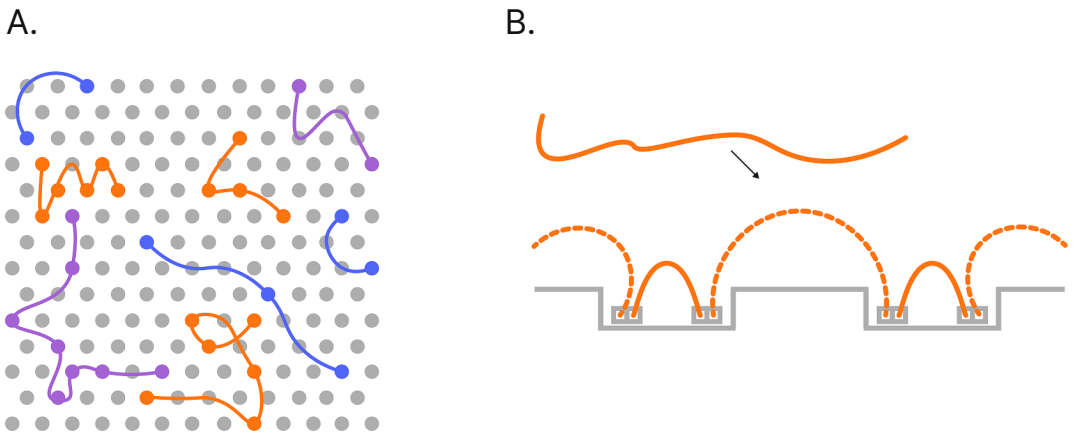


Figure 2: How proximity mapped read technology works
(A) DNA flows across multiple adjacent nanowells, enabling the use of proximity information to generate long-distance insights. (B) DNA is captured by transposomes in the nanowells (gray boxes) where it undergoes tagmentation and cluster generation.

or that could only be aligned with low confidence. In this way, standard short reads can be probabilistically associated with each other, enabling comprehensive variant detection for small variants (single-nucleotide variants (SNVs) and insertions/deletions (indels)), large variants (> 50 bp), known homologous regions, and phasing small variant calls that can be separated by millions of base pairs.

Unlock a comprehensive genome with high-accuracy short reads

Standard sequencing by synthesis (SBS) has demonstrated high-accuracy detection for small variants (eg, SNVs, indels, and copy number variations (CNVs)), but has struggled with specific regions and variant types such as structural variants. TruPath Genome introduces new capabilities to short-read sequencing that enable a more comprehensive genome. Short reads can now be used to address previously challenging regions of the genome and provide enhanced accuracy with additional variant types.

Improved coverage in difficult-to-map regions

Complex, highly polymorphic, or duplicated genomic areas can be difficult to assemble or align with a reference genome due to low mappability, either due to low coverage or poor mapping quality (MapQ). Previously, coverage of these “dark” regions[4] has led to low MapQ scores, gaps in the genome sequence data, and overall limited resolution for variant calling. To fill in these gaps, TruPath Genome uses proximity information from adjacent clusters to assign the correct genomic location. This enables high-confidence mapping of these ambiguously mapped reads (Figure 3 and Figure 4).

Higher accuracy variant calling

TruPath Genome was benchmarked with the HG002 genome and Telomere-to-Telomere" (T2T) Consortium T2T-Q100 v1.1 v0.019 truth set to demonstrate performance in small variant calling, SV detection, and phased sequencing. The increased mapping resolution enabled by proximity mapped read technology improved small variant calling performance, particularly in historically challenging-to-map regions, like homologous or repetitive regions, of the genome. Compared to standard short-read sequencing and variant calling,

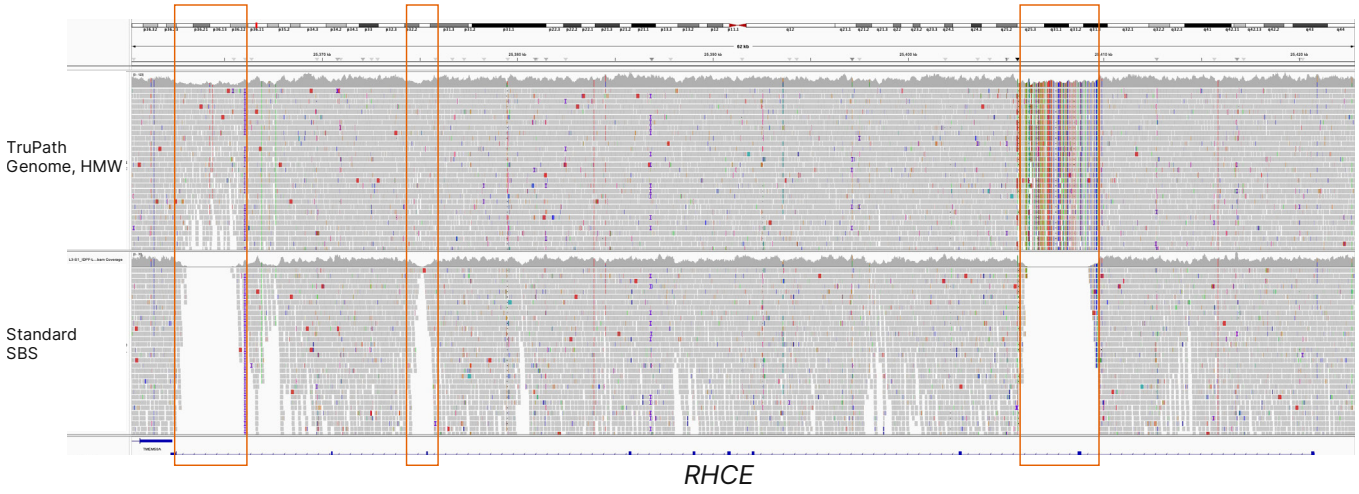


Figure 3: Effectively resolve challenging-to-map, medically relevant genes with TruPath Genome
Recovery of coverage in the *RHCE* gene, which lies in the Rh blood group locus adjacent to *RHD* and shares very high sequence homology with frequent gene conversion and hybrid rearrangements, making short-read mapping ambiguous in places. Accurate *RHCE* genotyping is clinically important for determining Rh antigens (C/c, E/e) as mismatches can increase the risk of hemolytic transfusion reactions and hemolytic disease of the fetus/newborn (HDFN). Standard SBS was performed using Illumina DNA PCR-Free Prep.

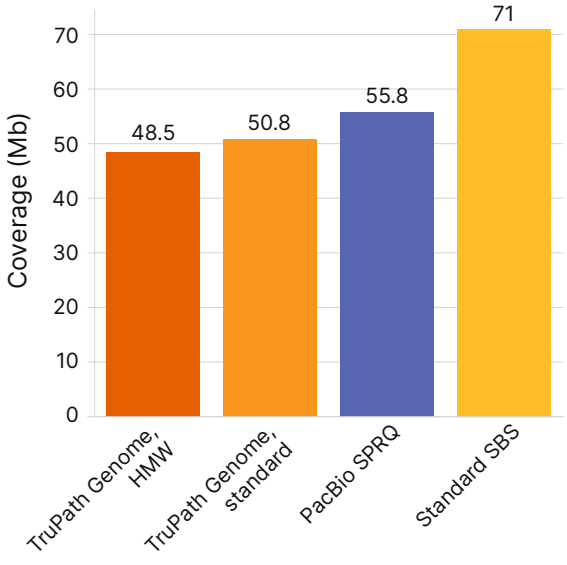


Figure 4: TruPath Genome enhances coverage of "dark regions" of the genome

TruPath Genome enhances coverage of "dark regions" defined as the "dark-by-MAPQ" regions in Ebbert et al,[1] in which 90% of the reads covering the region have a mapping quality (MAPQ) less than 10. TruPath Genome reduced the size of the genome that is dark due to improved mapping of reads in challenging regions of the genome. Standard SBS was performed using Illumina DNA PCR-Free Prep.

TruPath Genome shows enhanced accuracy via a substantial reduction in false negative (FN) and false positive (FP) variant calls. (Figure 5).

TruPath Genome shows a dramatic increase in SV (> 50 bp) calling performance compared to standard short-read methods. SV recall increased from 86% with standard SBS to 94% with proximity mapped read technology (Figure 6).

Ultralong phasing

Phased sequencing provides haplotype information that can distinguish between alleles on maternal and paternal chromosomes.[8] This information can help researchers resolve compound heterozygous variants, which is important for interpreting autosomal recessive conditions.

Proximity mapped read technology is particularly well suited for human genome phasing. DNA captured on the flow cell produces phased blocks that extend from hundreds of kilobases to several megabases (Figure 7 and Figure 8), contributing to greater insights into haplotypes and compound heterozygotes.

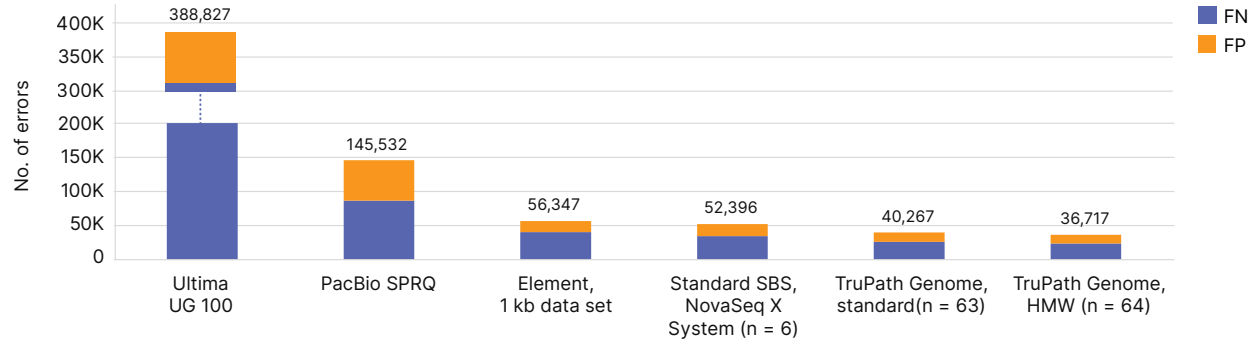


Figure 5: Greater accuracy for small variant calling with TruPath Genome

Small variant calling performance across various NGS systems and assays was benchmarked against the Genome in a Bottle NIST T2T-Q100 HG002 SV v1.1 truth set. Data for Ultima UG 100, PacBio with SPRQ chemistry, and Element AVITI with a 1 kb data set were sourced from published materials.[5–7] Standard SBS data was generated with Illumina DNA PCR-Free Prep libraries sequenced on a NovaSeq X System with v1.4 software and the 10B flow cell and analyzed with DRAGEN Germline v4.5.2 (six technical replicates). TruPath Genome, standard data was generated with DNA extracted using standard methods and sequenced on a NovaSeq X System with v1.4 software and the C8 flow cell and analyzed with DRAGEN Germline v4.5.2 (63 technical replicates). TruPath Genome, HMW was generated with DNA extracted using HMW methods and sequenced on a NovaSeq X System with v1.4 software and the C8 flow cell and analyzed with DRAGEN Germline v4.5.2 (64 technical replicates). FN, false negative variant calls; FP, false positive variant calls.

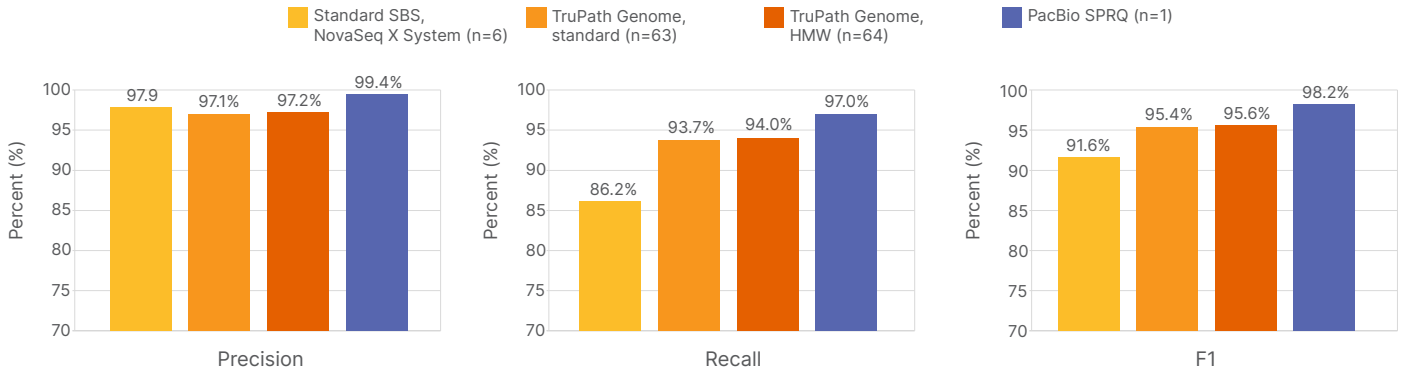


Figure 6: Improved structural variant calling with TruPath Genome

TruPath Genome structural variant (SV) performance as compared to standard SBS using DRAGEN Germline v4.5.2. Analysis uses the Genome in a Bottle NIST T2T-Q100 HG002 SV v1.1 truth set with the SV confident BED file. Benchmarking was performed in accordance with Genome in a Bottle guidance for structural variant benchmarking using "bench" and "refine" commands from Truvari v4.2.2. Standard SBS data was generated with Illumina DNA PCR-Free Prep libraries sequenced on a NovaSeq X System with v1.4 software and the 10B flow cell (six technical replicates). TruPath Genome, standard data was generated with DNA extracted using standard methods and sequenced on a NovaSeq X System with v1.4 software and the C8 flow cell (63 technical replicates). TruPath Genome, HMW was generated with DNA extracted using HMW methods and sequenced on a NovaSeq X System with v1.4 software and the C8 flow cell (64 technical replicates). PacBio with SPRQ chemistry data sourced from published material.[6] Precision measures the proportion of true positives. Recall represents the percent of true negatives compared to the total number of variants in a sample. F1 score (%) is a calculation of true positive and true negative results as a proportion of total results. Higher scores indicate improved accuracy based on reference data.

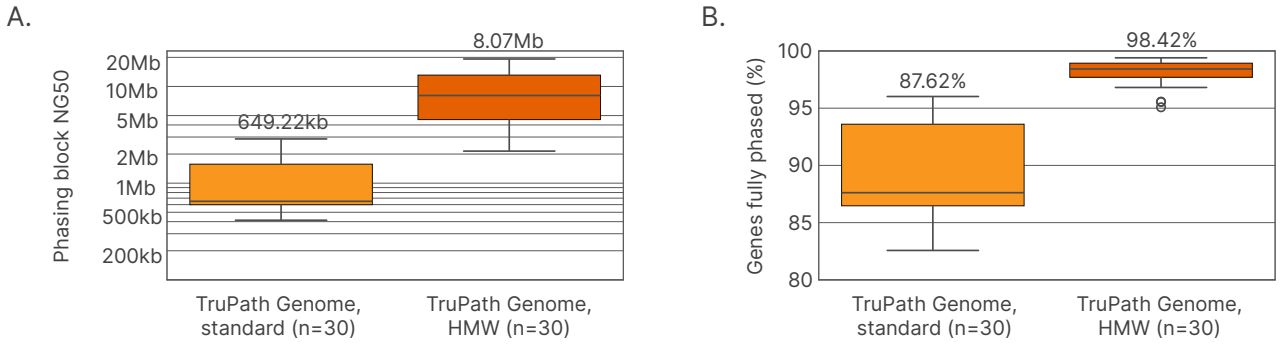


Figure 7: Ultralong phasing with TruPath Genome on 30 matched HMW and standard extraction cell lines

Phase block is measured over chromosomes 1–22 using DRAGEN phasing in DRAGEN Germline v4.5.2. TruPath Genome data was obtained using 30 different cell line Coriell samples with both high molecular weight (HMW) and standard extractions. Phase block NG50 is the length of the phase block once 50% of the target region (genome or other) has been phased. A technology that is unable to phase 50% of a given target region will have an NG50 of zero base pairs. Percent genes fully phased is the percentage of genic regions from a specified gene list (Gencode v44 genes.gtf) that are completely contained within a single phasing block.

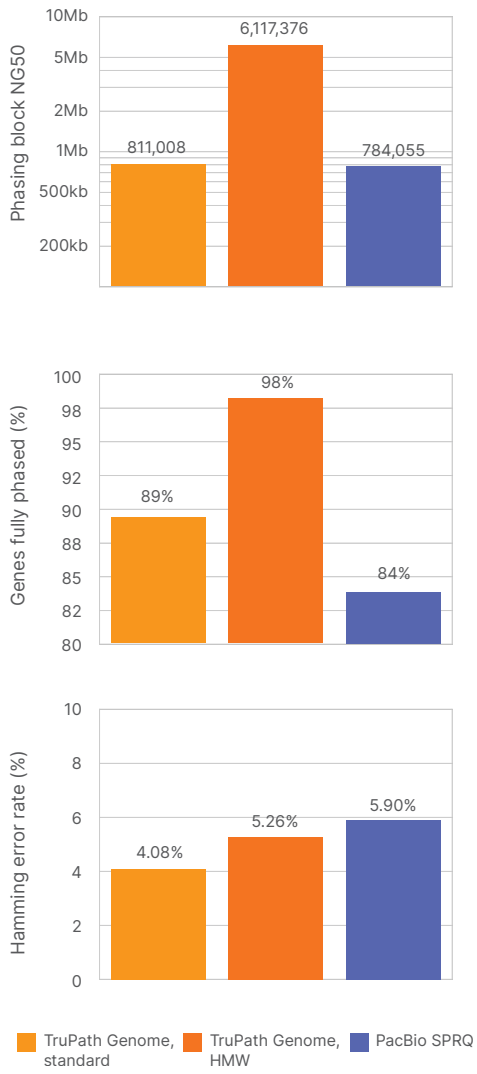


Figure 8: Phased sequencing achieved with TruPath Genome benchmarked with HG002 truth set data

Phase block is measured over chromosomes 1-22 using DRAGEN phasing in DRAGEN Germline v4.5.2. HiFi data (PacBio SPRQ) Phased VCF was obtained from https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. TruPath Genome data was obtained using sequencing replicates from HG002 cell line samples with HMW and standard extraction samples. Phase block NG50 is the length of the phase block once 50% of the target region (genome or other) has been phased. Percent genes fully phased is the percentage of genic regions from specified gene list (Gencode v44 genes.gtf) that are completely contained within a single phasing block. Phasing hamming error rate is benchmarked against the T2T Q100 truth set[9] downloaded from https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_HG002_DraftBenchmark_defrabbV0.020-20250117/GRCh38_HG2-T2TQ100-V1.1_smvar.vcf.gz.

TruPath Genome fully phases a median of ~91% of all genes with standard DNA extraction and ~97% of genes with high molecular weight (HMW) DNA extraction. Additionally, TruPath Genome phases ~98%¶ of all heterozygous SNVs in both standard and HMW DNA extractions.

TruPath Genome data with DRAGEN phasing for HMW DNA reduces hamming errors** (~4%) compared to publicly available tools like HapCut2 (~6%).[10]

Variant detection in high-homology regions

Accurate variant detection in paralogous regions using standard short reads is difficult due to high sequence homology, resulting in ambiguous read mapping and variant detection errors. Resolution often requires costly reflex assays like multiplex ligation-dependent probe amplification (MLPA) or longe-range polymerase chain reaction (LR-PCR). TruPath Genome combines flow cell proximity information with a novel DRAGEN Multi-Region Joint Detection (MRJD) algorithm to provide reliable, *de novo*, haplotype-resolved, and copy-number-aware small variant calling in paralogous genes (Figure 9 and Figure 10).

TruPath Genome targets 15 medically relevant and paralogous genes (Table 1) known to have tandem or nontandem copies distributed throughout the genome for high-accuracy mapping with MRJD and produces haplotype-resolved small variant calling (Figure 10).

Table 1: Genes targeted with TruPath Genome for MRJD analysis

CFHR1	*CFHR2*	*CFHR3*
CFHR4	*CYP2D6*	*CYP11B1/2*
NCF1	*PMS2*	*RCCX (CYP21A2 and TNXB)*
SMN1/2	*STRC*	*USP18*

¶Percent heterozygous variants phased is the percentage of phased heterozygous SNVs, calculated as the number of phased SNVs divided by the number of heterozygous SNVs.

**Hamming errors occur when variants are assigned to the incorrect maternal or paternal strand.

Improved resolution for short tandem repeats (STRs)

STRs are repetitive regions of the genome that can expand to lengths beyond the normal range and cause mutations associated with many genetic diseases, including Fragile X syndrome, amyotrophic lateral sclerosis, and Huntington's disease.[11] Accurate size estimation and read recovery of these regions has been difficult because the most relevant expanded alleles often exceed the read length of standard short-read sequencing data. Through a combination of proximity information and high-quality short-read data, TruPath Genome provides enhanced recovery of the reads needed for accurate STR sizing and improved resolution by genotyping the two copies using phased information (Figure 11).

Ultralong-distance insights for complex SVs

The ability of TruPath Genome to resolve large structural rearrangements includes novel capabilities that extend beyond traditional variant calling performance benchmarks. TruPath Genome captures information about reads from proximal clusters†† between any pair of regions of the genome, making it possible to build high-resolution visual representations of genome structure maps—termed “colocation plots” (Figure 12).

†† Proximal clusters are clusters that are physically near each other on the flow cell.

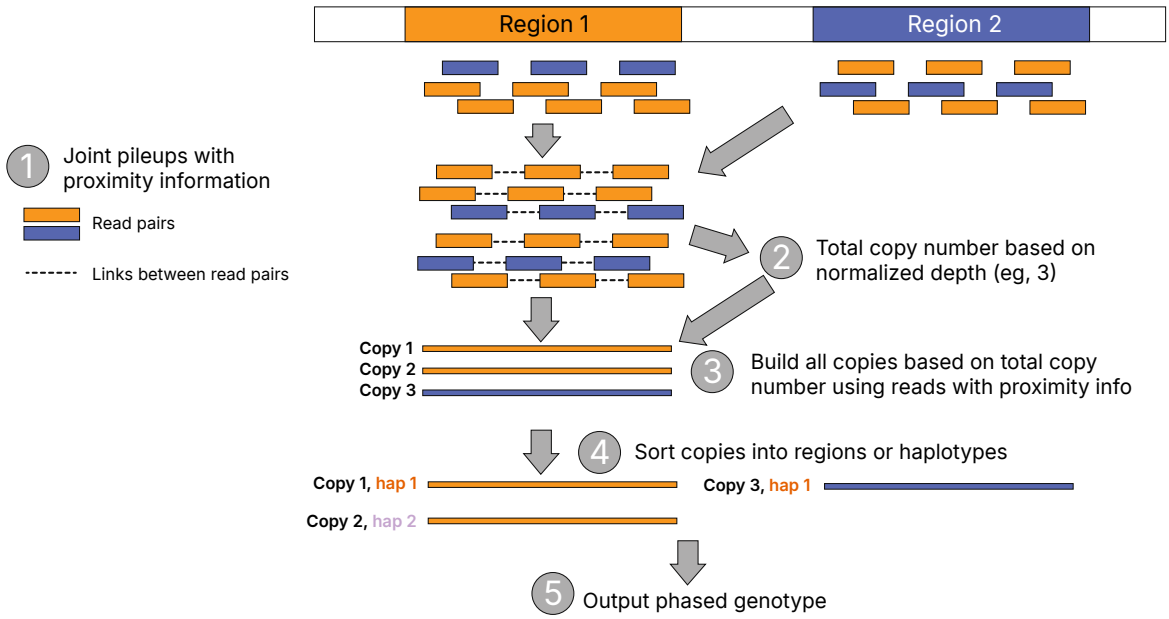
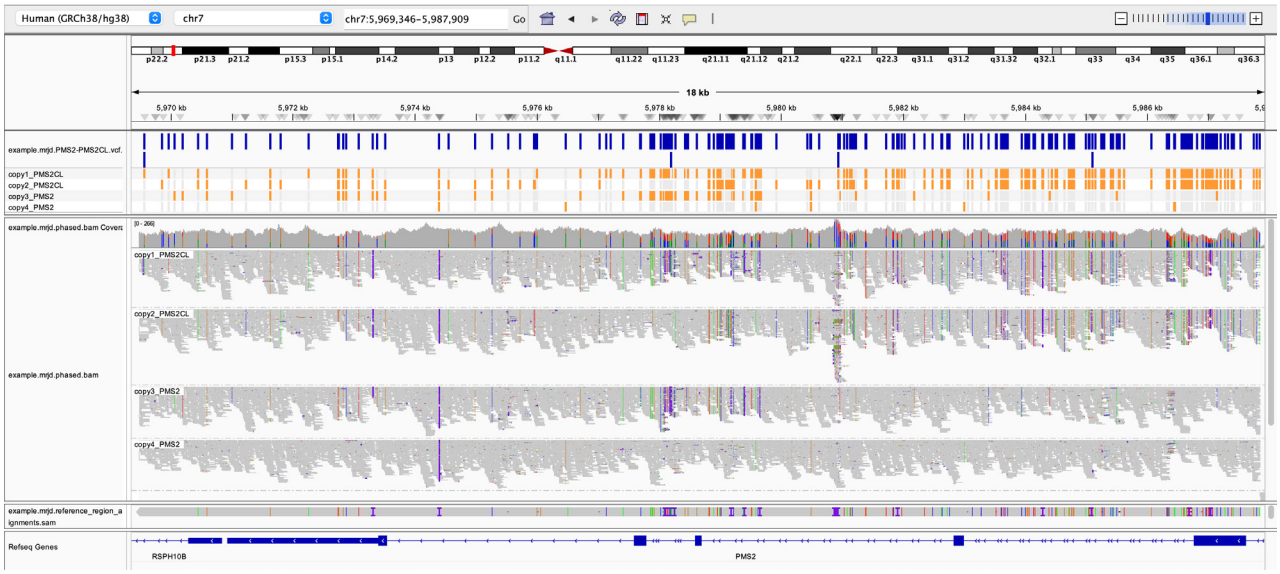


Figure 9: Schematic of the MRJD method used to detect variants in segmental duplication regions

1. For a given set of segmentally duplicated regions, MRJD extracts all reads mapped to any of the copies in that region set for joint analysis. Proximity information is leveraged to establish links between such reads and identify read sets originating from the same original molecule; 2. The count of all reads extracted in step 1 is then normalized, GC-corrected, and used to estimate the total number of copies of the segmentally duplicated sequence in the input sample; 3. A joint genotyping approach informed by the number of copies of the region estimated in step 2 and applied to the full set of proximity-linked reads is used to identify and phase all the variants in each copy of the segmentally duplicated sequence in the input sample. 4. Each fully phased copy is assigned to a specific reference genome location (when regions are in distal reference locations) or phased to the other copies into haplotypes (only possible when segmental duplication regions are in tandem/nearby in the reference). 5. Fully phased variants across all copies, the haplotype assignment of each copy, the assignment of copies to each reference genome location, and the reads assigned to each copy are output to the MRJD-specific VCF, JSON, and BAM files.

A. *PMS2* gene



B. *SMN* gene

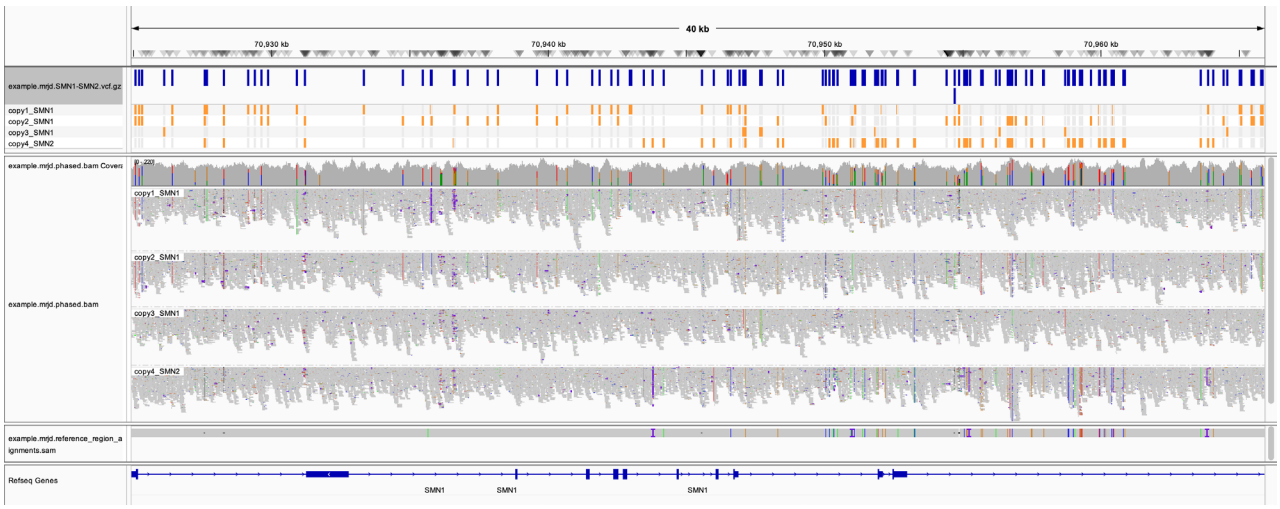


Figure 10: Example output of variant calls and read assignment in paralogous genes using TruPath Genome and MJRD

(A) Phased variant calls and associated supporting reads across all copies of *PMS2*/*PMS2CL* gene-pseudogene pair represented in the *PMS2* reference genome location. Two of the four copies are identified as *PMS2CL* (pseudogene) copies and two are identified as *PMS2* (active gene) copies based on proximity links to the *PMS2*/*PMS2CL* flanks. Copy3 (*PMS2*) contains what looks like a gene conversion event overlapping exons 13 and 14 indicated by the presence of variants matching the *PMS2CL* reference sequence. The alignment between the *PMS2CL* reference sequence to the *PMS2* reference sequence is represented in the third pane (top to bottom). (B) Phased variant calls and associated supporting reads across all copies of *SMN1*/*SMN2* gene-pseudogene pair represented in the *SMN1* reference genome location. Three of the four copies are identified as *SMN1* (active gene) and two are identified as *SMN2* (inactive gene) according to the base at the SMN1 c.840 position in each of the phased copies (*SMN1* copies have C base and *SMN2* copies have T base at that position).

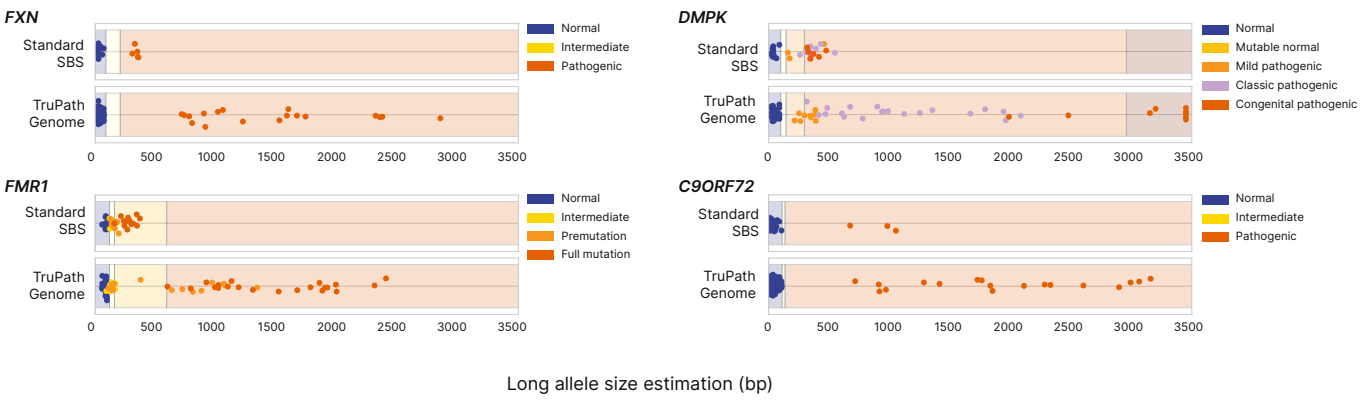


Figure 11: More accurate measurements of STR expansions can be achieved using TruPath Genome

True classifications of STRs are represented by the color of each dot. Colored bands represent the different classification ranges used to define the different classes of expansions in each locus. The x-axis indicates the STR size estimate from DRAGEN STR using either TruPath Genome or standard SBS (Illumina DNA PCR-Free Prep) as inputs. TruPath Genome classifications are significantly more consistent with the true classifications and span a broader range of STR lengths. Total replicates for standard SBS: 78 biosamples (78 unique individual). Total replicates forTruPath Genome: 135 biosamples (42 unique individual).

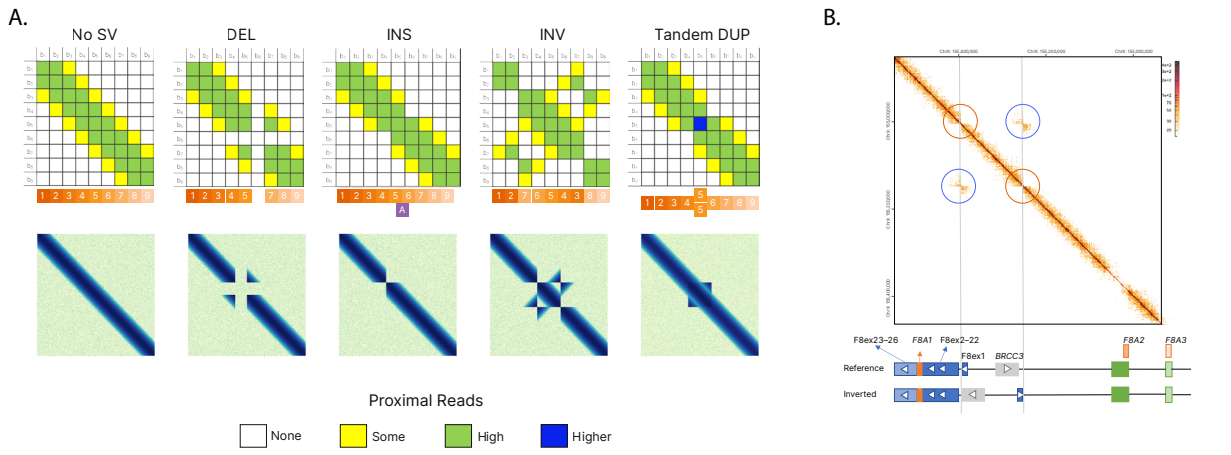


Figure 12: Visualization of structural variant events using TruPath Genome colocation plots

(A) Counts of proximal reads (flow cell space) mapping to all pairs of genomic bins are assessed to generate a 2-dimensional colocation map. X and Y axes represent reference genome bins. A high number of connections between two bins indicates that such bins are near each other in genomic space in the input sample. In a region with no structural variants, genomic bins that are nearby in the reference genome are also nearby in the input sample and so appear as a high number of connections near the diagonal of the colocation plot. When a structural variant is present, genomic bins that are nearby in the reference genome may no longer be nearby in the input sample's genome and genomic bins that are far in the reference may be proximal in the input sample. Such changes in connectivity are reflected on the colocation maps as off-diagonal signals of different shapes as well as lack of diagonal connectivity. (B) An example of a colocation map for an intron 1 gene inversion in the *F8* gene locus, associated with severe hemophilia A. Note the expected hourglass-shaped signal in the off-diagonal region (blue circles) and the lack of diagonal signal (orange circles) in the event boundaries. A schematic representation of the locus in the reference genome and the inverted gene region is shown below the plot.

The simplest whole-genome sequencing workflow

Illumina TruPath Genome, powered by foundational proximity mapped read technology, transforms human genome sequencing. The innovative workflow eliminates traditional library preparation for unprecedented simplicity while proximity information preserves the long-distance DNA strand information with highly accurate, standard short-read sequencing data. The result is an assay that combines the data quality strengths of short reads, high accuracy detection of small variants, with improved mapping coverage of the human genome and a powerful tool for understanding genetic disease.

Learn more →

Illumina TruPath Genome

Proximity mapped read technology

NovaSeq X Series

Ordering information	
Product	**Catalog no.**
Illumina TruPath Genome (NovaSeq X C8 flow cell)	20157405
Illumina TruPath Genome (NovaSeq X C2 flow cell)	20157406

References

1. Ebbert MTW, Jensen TD, Jansen-West K, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 2019;20(1):97. Published 2019 May 20. doi:10.1186/s13059-019-1707-2
2. Ryan NM, Corvin A. Investigating the dark-side of the genome: a barrier to human disease variant discovery?. *Biol Res*. 2023;56(1):42. Published 2023 Jul 20. doi:10.1186/s40659-023-00455-0
3. Illumina. TruPath Genome performance with samples of varying type and quality. illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932. Published February 2026. Accessed February XX, 2026.
4. Wadsworth ME, Page ML, Aguzzoli Heberle B, Miller JB, Steely C, Ebbert MTW. Sequencing the gaps: dark genomic regions persist in CHM13 despite long-read advances. Preprint. *bioRxiv*. 2025;2025.05.23.655776. Published 2025 May 28. doi:10.1101/2025.05.23.655776
5. Ultima Genomics. Genome-in-a-Bottle (GIAB) samples HG001-HG005. https://cdn.sanity.io/files/l7780ks7/production-2024/0a1b6a62a6da3e3fcafb81cad4c8ff2ffe85dd41.pdf Published February 2025. Accessed January 29, 2026.
6. PacBio. Index of /public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2. https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. Accessed January 29, 2026.
7. Asri M, Chang PC, Mier JC, et al. Pangenome-aware DeepVariant. Preprint. *bioRxiv*. 2025;2025.06.05.657102. Published 2025 Jun 6. doi:10.1101/2025.06.05.657102
8. Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 2011; 12(10):703-714. doi: 10.1038/nrg3054
9. Hansen NF, Dwarshuis N, Ji HJ, et al. A complete diploid human genome benchmark for personalized genomics. Preprint. *bioRxiv*. 2025;2025.09.21.677443. Published 2025 Sep 21. doi:10.1101/2025.09.21.677443
10. Data on file. Illumina, Inc., 2026.
11. Steely CJ, Watkins WS, Baird L, Jorde LB. The mutational dynamics of short tandem repeats in large, multigenerational families. *Genome Biol*. 2022;23(1):253. Published 2022 Dec 12. doi:10.1186/s13059-022-02818-4

illumina®

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

M-GL-03931 v1.0