APPLICATION NOTE

High-accuracy next-generation sequencing with the NovaSeq™ X Series

XLEAP-SBS™ chemistry delivers exceptional data quality for high-throughput genomics applications

- Whole-genome sequencing
- Whole-exome sequencing
- Whole-transcriptome sequencing
- Whole-genome methylation sequencing
- Single-cell multiome sequencing

illumina®

Introduction

The technology advances built into the NovaSeq X and NovaSeq X Plus Systems provide throughput and productivity gains that transform the economics of production-scale sequencing. Advanced chemistry, ultrahigh-resolution optics, integrated secondary analysis, and operational simplicity combine to make the NovaSeq X Series our most powerful and cost-effective sequencing systems yet.

The NovaSeq X Series is powered by XLEAP-SBS chemistry—a faster, higher fidelity, and more robust advancement to proven Illumina sequencing by synthesis (SBS) chemistry. XLEAP-SBS reagents are optimized for performance and speed to maximize throughput without sacrificing the high-quality data users expect.

This application note demonstrates that the NovaSeq X Series delivers data quality that meets or exceeds that of the NovaSeq 6000 System for key methods, including whole-genome sequencing, whole-exome sequencing, whole-transcriptome sequencing, methylation sequencing, and single-cell multiomics.

Methods

Whole-genome sequencing

Whole-genome libraries were prepared from NA12878 genomic DNA (gDNA) (Coriell Institute for Medical Research) using the TruSeq™ PCR-Free Prep kit (Illumina, Catalog no. 20015963).

Sequencing was performed on the NovaSeq X Plus System with the NovaSeq X Series 10B Reagent Kit (300 cycles) (Illumina, Catalog no. 20085594) using the 2 × 151 bp run configuration (42 samples over multiple runs). For comparison, the same libraries were also sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, Catalog no. 20028312) using a 2 × 151 bp run configuration (24 samples in one run).

Secondary data analysis was performed using DRAGEN™ Germline pipeline v4.1 cloud-based workflow. Sequencing data were downsampled to 30× coverage to compare variant calling performance between the NovaSeq X Plus and NovaSeq 6000 Systems.

Exome sequencing

Exome libraries were prepared from NA12878 genomic DNA (gDNA) (Coriell Institute for Medical Research) using Illumina DNA Prep with Enrichment, (S) Tagmentation (Illumina, Catalog no. 2002554) to capture genomic regions targeted by the Twist Comprehensive Exome Panel (Twist Bioscience, Catalog no. 102033).

Sequencing was performed on the NovaSeq X Plus System with the NovaSeq X Series 10B Reagent Kit (300 cycles) using the 2 × 101 bp run configuration (753 samples over multiple runs). For comparison, the same libraries were also sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) using a 2 × 101 bp run configuration (48 samples in a single lane).

Secondary data analysis was performed using the DRAGEN Enrichment pipeline v4.0.3 cloud-based workflow. Variant calling accuracy was assessed against the Genome In A Bottle (GIAB) v3.3.2 truth set and hg19-alt-aware reference genome.[1,2] Sequencing data were downsampled to 30M read pairs per sample to compare variant calling performance between the NovaSeq X Plus and NovaSeq 6000 Systems.

Whole-transcriptome sequencing

Total RNA and messenger RNA (mRNA) libraries were prepared from leukemia cell line RNA: HL-60 (Thermo Fisher Scientific, Catalog no. AM7836) and K562 (BioChain, Catalog no. R1255820-50) and breast cancer cell line RNA: MCF7 (BioChain, Catalog no. R1255830-50) using Illumina Stranded Total RNA Prep with Ribo-Zero Plus (Illumina, Catalog no. 20040529) and Illumina Stranded mRNA Prep (Illumina, Catalog no. 20040534).

Sequencing was performed on the NovaSeq X Plus System with NovaSeq X Series 10B Reagent Kit (300 cycles) using a 2 × 75 bp run configuration and a custom dark-cycle recipe to avoid first-base T overhang from library preparation.[3] Sequencing included 573 samples for total RNA-Seq and 2304 samples for mRNA-Seq, over multiple runs. For comparison, the same libraries were also sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S2 Reagent Kit v1.5 (200 cycles) (Illumina, Catalog no. 20028315) using a 2 × 76 bp run configuration. For both total RNA-Seq and mRNA-Seq, 96 samples each were run in a single lane.

Secondary data analysis was performed using the DRAGEN RNA Pipeline v4.1 cloud-based workflow. Sequencing data were downsampled to 10M reads for all samples to compare gene expression data. Data were aligned against the Genome Reference Consortium Human GRCh38 (h38 assembly).[2]

Whole-genome methylation sequencing

Methylation libraries were prepared from replicates of a matched human brain and spleen sample set (Zymo Research, Catalog no. D5018) (eight replicates each for NovaSeq X Plus System and five replicates each for NovaSeq 6000 System) using the Zymo-Seq WGBS Library Kit (Zymo Research, Catalog no. D5465), in combination with Illumina DNA Prep library prep kit (Illumina, Catalog no. 20060059).[4] Unmethylated *E. coli* control was spiked in at 0.25% to access cytosine conversion efficiency, which was estimated to be higher than 99.5%.

Sequencing was performed on the NovaSeq X Plus System with the NovaSeq X Series 10B Reagent Kit (300 cycles) using the 2 × 151 bp run configuration (16 samples in one run). For comparison, libraries were sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) using a 2 × 151 bp run configuration (10 samples in one run).

Methylation analysis was performed using the DRAGEN Methylation pipeline cloud-based workflow. Sequencing data were downsampled to 500M reads per sample for downstream analysis.

Single-cell multiomics

Chromium Single Cell Multiome ATAC + Gene Expression provides joint readouts of gene expression and epigenetic signatures at single-cell resolution. Samples for single-cell RNA sequencing (scRNA-Seq) and single-cell assay for transposase accessible chromatin (scATAC-Seq) were jointly prepared from cryopreserved human peripheral blood mononuclear cells (PBMCs) of a healthy male donor (aged < 35) obtained from AllCells. Nuclei were isolated as described in 10x Genomics Demonstrated Protocol-Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing (CG000365 Rev A). Paired scRNA-Seq and scATAC-Seq libraries were generated from the isolated nuclei as described in the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338 Rev B).

Sequencing was performed on the NovaSeq X Plus System with NovaSeq X Series 10B Reagent Kit (300 cycles) (80 replicates in one run). For comparison, the same libraries were also sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (10 replicates in one run). Run configurations were set up according to parameters provided by 10x Genomics: 28-cycle read 1, 10-cycle i7 and i5 index reads, and 90-cycle read 2 for multiome scRNA-Seq libraries; 50-cycle read 1, 8-cycle i7 index read, 24-cycle i5 index read, and 49-cycle read 2 for multiome scATAC-Seq libraries.

Data analysis was performed using the Cell Ranger ARC analysis pipeline v2.0 (10x Genomics) to count transcripts and chromatin accessibility peaks in single cells.

Results

The NovaSeq X Plus System enables significant gains in throughput compared to the NovaSeq 6000 System. The NovaSeq X 10B flow cell and the NovaSeq 6000 S4 flow cell can both output up to 3 Tb of 2 × 150 bp sequence data per flow cell. However, the run time of the NovaSeq X Series is almost half the run time of the NovaSeq 6000 System run time (Table 1).

Table 1: Comparable sequencing output in significantly shorter run times

Metric	NovaSeq 6000 S4 flow cell	NovaSeq X Plus 10B flow cell
2 × 100 bp output per run	1.6–4 Tb	~2–4 Tb
2 × 100 bp run time	~36 hr	~22 hr
2 × 150 bp output per run	2.4–6 Tb	~3–6 Tb
2 × 150 bp run time	~44 hr	~24 hr

Whole-genome sequencing

Whole-genome sequencing (WGS) analysis metrics were evaluated, including precision and recall for both single nucleotide variants (SNVs) and insertions–deletions (indels). Both the NovaSeq X Plus and NovaSeq 6000 Systems delivered high-quality data and highly accurate variant calling (Table 2, Table 3). These data demonstrate that WGS results on the NovaSeq X Series meet or exceed NovaSeq 6000 System performance.

Table 2: Sequencing run metrics for WGS

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration	2 × 151 bp	2 × 151 bp
Read 1 bases ≥ Q30	92.17%	95.89%
Read 2 bases ≥ Q30	89.60%	94.30%
Read 1 error rate	0.25%	0.15%
Read 2 error rate	0.24%	0.23%

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

Table 3: Secondary analysis metrics for WGS

Metric	NovaSeq 6000	NovaSeq X Plus
Build depth	30×	30×
Autosome coverage	31×	31×
Total SNPs	3,041,268	3,041,454
Het:Hom ratio	1.59	1.60
Ti:Tv ratio	1.99	1.98
SNP precision	99.95%	99.95%
SNP recall	99.95%	99.96%
Indel precision	99.64%	99.60%
Indel recall	99.61%	99.57%
Read 1 mismatched bases	0.48%	0.36%
Read 2 mismatched bases	0.61%	0.43%
No. samples averaged	24	42

Whole-exome sequencing

Whole-exome sequencing (WES) primary and secondary analysis metrics were evaluated, including quality scores, error rate, autosome callability, percent aligned reads, coverage uniformity, and precision and recall for both SNVs and indels. Both the NovaSeq X Plus and NovaSeq 6000 Systems delivered high-quality data and highly accurate variant calling (Table 4, Table 5). These data demonstrate that WES results on the NovaSeq X Series are equivalent to NovaSeq 6000 System performance.

Table 4: Sequencing run metrics for WES[a]

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration	2 × 101 bp	2 × 101 bp
Read 1 bases ≥ Q30	92.06%	96.63%
Read 2 bases ≥ Q30	90.96%	96.24%
Read 1 error rate	N/A[b]	0.10%
Read 2 error rate	N/A[b]	0.21%

a. Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.
b. PhiX control was not used for these NovaSeq 6000 runs, so no error rates shown.

Table 5: Secondary analysis metrics for WES

Metric	NovaSeq 6000	NovaSeq X Plus
Autosome callability	97.49%	97.53%
Aligned reads	99.28%	99.11%
Coverage uniformity	97.17%	97.29%
SNP precision	99.77%	99.77%
SNP recall	98.20%	98.30%
Indel precision	97.57%	97.36%
Indel recall	88.53%	89.05%
No. samples averaged	48	753

Whole-transcriptome sequencing

Both the NovaSeq X Plus and NovaSeq 6000 Systems exceeded published specifications for data quality for total RNA-Seq (Table 6) and mRNA-Seq (Table 7). Quantification of transcripts showed excellent concordance between the two platforms ($R^2 > 0.99$) for total RNA-Seq (Figure 1) and for mRNA-Seq (Figure 2). These data demonstrate that whole-transcriptome sequencing on the NovaSeq X Series produces data quality that meets or exceeds NovaSeq 6000 System performance.

Table 6: Sequencing run metrics for total RNA-Seq

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration	2 × 76 bp	2 × 75 bp
Read 1 bases ≥ Q30	91.83%	96.82%
Read 2 bases ≥ Q30	90.52%	96.37%
Read 1 error rate	0.44%	0.07%
Read 2 error rate	1.17%	0.15%
No. samples averaged	96	573

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

Table 7: Sequencing run metrics for mRNA-Seq

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration	2 × 76 bp	2 × 75 bp
Read 1 bases ≥ Q30	91.47%	96.03%
Read 2 bases ≥ Q30	89.92%	95.65%
Read 1 error rate	0.74%	0.09%
Read 2 error rate	1.32%	0.16%
No. samples averaged	96	2304

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

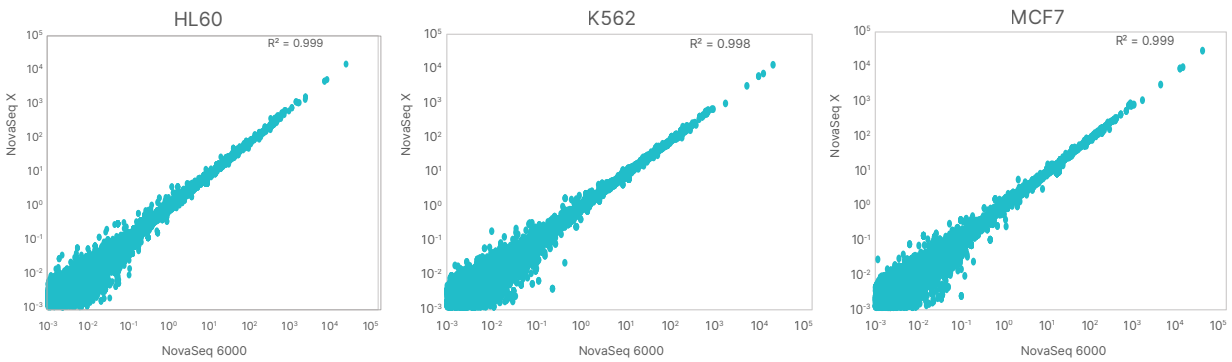

Figure 1: Whole transcriptome total RNA-Seq correlations—Transcripts per million (TPM) for cancer cell lines: HL-60, K562, and MCF7.

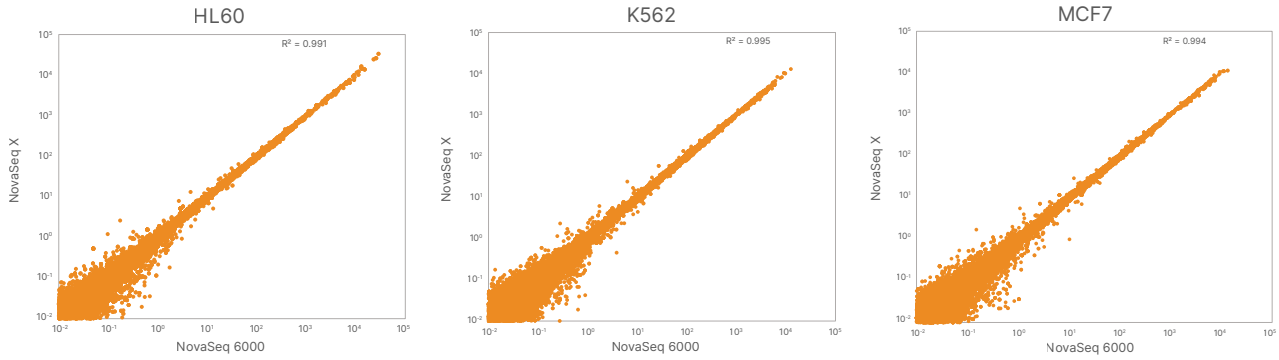

Figure 2: Whole transcriptome mRNA-Seq correlations—Transcripts per million (TPM) for cancer cell lines: HL-60, K562, and MCF7.

Methylation sequencing

Whole-genome methylation metrics were evaluated to compare performance of the NovaSeq X Series and the NovaSeq 6000 System. Both the NovaSeq X Plus System and the NovaSeq 6000 System demonstrated numbers for quantifying percent methylated cytosines in line with the expected based on product documentation (Figure 3A). Higher mapping efficiency was detected on the NovaSeq X Plus System for matched libraries (Figure 3B). Whole-genome coverage distribution plots show comparable results with an increase in high-coverage CpGs (> 50×) between the NovaSeq X Plus System and NovaSeq 6000 System (Figure 4).

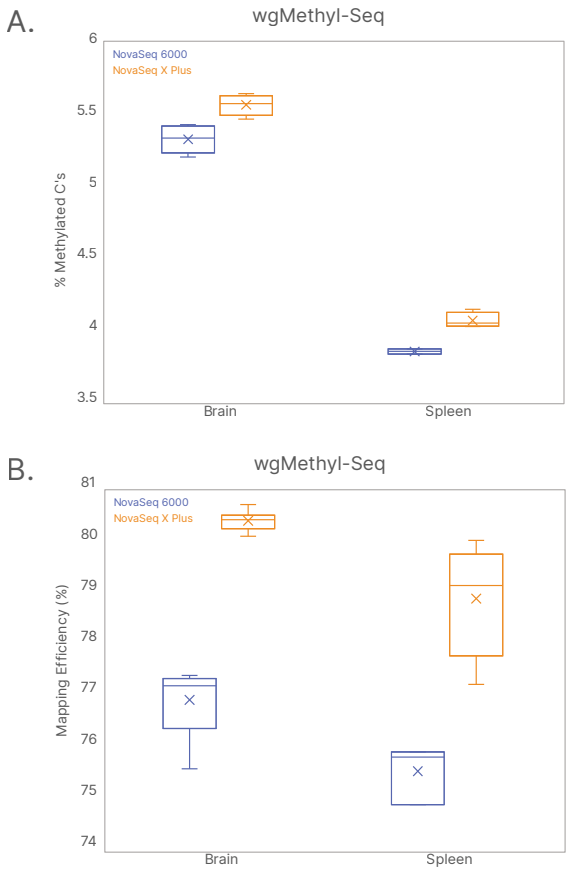

Figure 3: Whole-genome methylation sequencing—Zymo-Seq WGBS results for NovaSeq X Plus System and NovaSeq 6000 System showing (A) percent methylated cytosines and (B) mapping efficiency.

Bisulfite or enzymatic conversion changes unmethylated cytosines to uracil during library preparation. This results in unbalanced libraries that have been traditionally challenging to sequence and usually required a high percentage (> 5%) of PhiX to enhance base diversity. On the NovaSeq X Series a low percentage (1%) of PhiX was sufficient to achieve high-quality runs (Table 8).

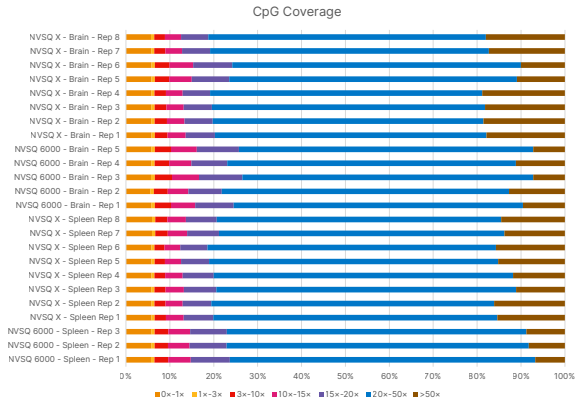

Figure 4: Genome coverage for whole-genome methylation sequencing—Zymo-Seq WGBS results for NovaSeq X Plus System and NovaSeq 6000 System showing CpG coverage distribution.

Table 8: Sequencing run metrics for methylation sequencing

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration	2 × 151 bp	2 × 151 bp
Read 1 bases ≥ Q30	89.01%	91.95%
Read 2 bases ≥ Q30	86.75%	93.09%
Read 1 error rate	0.30%	0.14%
Read 2 error rate	0.59%	0.25%
No. samples averaged	10	16

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

Single-cell multiomics

Performance metrics for the single-cell multiomics assay, including scRNA-Seq to measure gene expression and scATAC-Seq to measure chromatin accessibility, were evaluated. The NovaSeq X Plus and NovaSeq 6000 Systems exceeded published specifications for data quality (Table 9, Table 10). t-SNE plots for scRNA-Seq gene expression (Figure 5) and scATAC-Seq chromatin accessibility (Figure 6) show an excellent correlation between the NovaSeq X Plus System and the NovaSeq 6000 System.

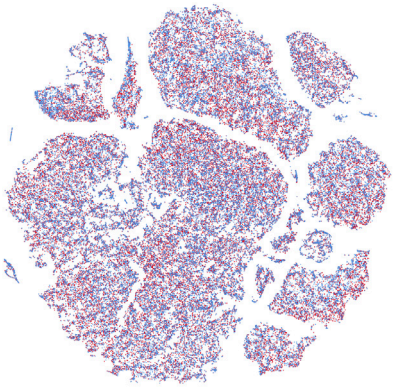

Figure 5: Single-cell multiome gene expression—t-SNE plots for scRNA-Seq libraries sequenced on the NovaSeq X Plus System (blue) and NovaSeq 6000 System (red).

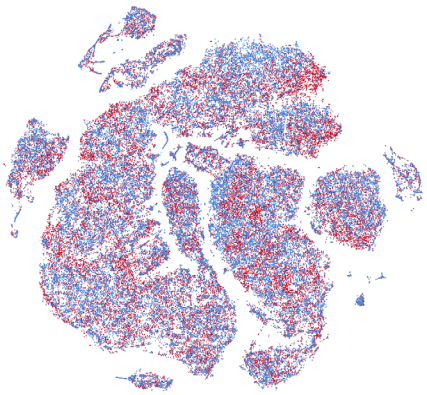

Figure 6: Single-cell multiome chromatin accessibility—t-SNE plots for scATAC-Seq libraries sequenced on the NovaSeq X Plus System (blue) and NovaSeq 6000 System (red).

Table 9: Sequencing run metrics for single-cell multiome scRNA-Seq

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration		
Read 1	28 bp	28 bp
Index 1	10 bp	10 bp
Index 2	10 bp	10 bp
Read 2	90 bp	90 bp
Read 1 bases ≥ Q30	97.11%	97.35%
Read 2 bases ≥ Q30	95.01%	95.93%
Read 1 error rate	0.04%	0.04%
Read 2 error rate	0.17%	0.15%
No. samples averaged	10	80
No. genes detected	25,975	25,743
Median UMI counts per cell	2790	2571
Estimated no. cells per sample	3880	3882

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

Table 10: Sequencing run metrics for single-cell multiome scATAC-Seq

Metric	NovaSeq 6000	NovaSeq X Plus
Run configuration		
Read 1	50 bp	50 bp
Index 1	8 bp	8 bp
Index 2	24 bp	24 bp
Read 2	49 bp	49 bp
Read 1 bases ≥ Q30	93.35%	95.58%
Read 2 bases ≥ Q30	92.21%	94.24%
Read 1 error rate	0.08%	0.09%
Read 2 error rate	0.28%	0.16%
No. samples averaged	10	80
Estimated no. cells per sample	3880	3882

Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. All runs met published specifications for yield. Yield per lane is not equivalent between NovaSeq 6000 S4 flow cells and NovaSeq X 10B flow cells.

Summary

The NovaSeq X and NovaSeq X Plus Sequencing Systems feature breakthrough chemistry, optics, informatics, and operational simplicity to transform the economics of high-throughput sequencing. The NovaSeq X Series delivers extraordinary throughput while delivering the high-quality data users expect from Illumina. XLEAP-SBS chemistry enables significant improvements in sequencing run times and output without sacrificing data quality. Data from key methods commonly run on the NovaSeq 6000 System, including whole-genome sequencing, whole-exome sequencing, whole-transcriptome sequencing, methylation sequencing, and single-cell multiomics, were directly compared to data generated using the NovaSeq X Plus System. Results show that performance on the NovaSeq X Series meets or exceeds NovaSeq 6000 System performance and will support more data-intensive applications.

Learn more

NovaSeq X and NovaSeq X Plus Sequencing Systems, illumina.com/systems/sequencing-platforms/novaseq-x-plus.html

Data sets referenced in this note, basespace.illumina.com/datacentral

References

1. National Institute of Standards and Technology. Genome in a Bottle. nist.gov/programs-projects/genome-bottle. Accessed July 27, 2023.
2. Genome Reference Consortium. Human Genome Overview. NCBI website. ncbi.nlm.nih.gov/grc/human. Accessed July 27, 2023.
3. Illumina. Best practices for read trimming for Illumina Stranded mRNA and Total RNA workflows. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-stranded-rna-t-overhang-tech-note-470-2020-010/illumina-stranded-rna-t-overhang-tech-note-470-2020-010.pdf. Published 2020. Accessed August 1, 2023.
4. Wehrkamp-Richter S. Illumina. BaseSpace™ Sequence Hub now supports whole genome bisulfite sequencing (WGBS) data with Zymo-Seq Library Kit running on NovaSeq X Series. developer.illumina.com/news-updates/whole-genome-bisulfite-sequencing-zymo-seq-data-now-available-on-novaseqtm-x-series. Published June 6, 2023. Accessed August 1, 2023.

illumina®

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

M-US-00201 v1.0