APPLICATION NOTE

RNA-Seq on the NovaSeq™ X Series

Streamlined, integrated workflow for gene expression profiling and transcriptome analysis

illumina®

Introduction

Multiomics provides an integrated approach to power discovery across multiple levels of biology. By combining data from genomics, transcriptomics, epigenetics, and proteomics, researchers can achieve a more comprehensive understanding of the molecular changes that contribute to normal development, cellular response, and disease.

Integrate -omes for multiomics insights

Multiomics goes beyond a singular -ome to unlock deeper biological insights:

- **Transcriptomics + genomics**—Annotate and prioritize variants uncovered in genome-wide association studies (GWAS) to understand mechanisms of disease.
- **Transcriptomics + epigenetics**—Measure the ties between gene regulation and gene expression to understand the mechanisms controlling interesting phenotypes.
- **Transcriptomics + proteomics**—Quantify RNA and protein expression simultaneously for a comprehensive view of cellular function and activity.

Using next-generation sequencing (NGS) technology, RNA sequencing (RNA-Seq) powers transcriptomics studies to capture changes in gene expression and characterize multiple forms of RNA to unravel its structure, variation, and activity.[1,2]

Advantages of RNA-Seq

RNA-Seq provides a detailed snapshot of the transcriptome at a given point in time and offers numerous advantages over quantitative PCR, including:

- Hypothesis-free experimental design, requiring no previous knowledge of the transcriptome
- Discovery power to detect known and novel transcripts with full sequence and variant information
- High-throughput capability to quantify hundreds to thousands of regions in each assay
- Broad dynamic range, providing more accurate measurement of gene expression

Figure 1: NovaSeq X and NovaSeq X Plus Sequencing Systems—Built to deliver the throughput and accuracy needed to perform data-intensive RNA-Seq applications.

RNA-Seq on the NovaSeq X Series

Illumina has a proven track record of developing RNA-Seq solutions that empower researchers to investigate the transcriptome. Illumina RNA library preparation kits deliver high-quality data with a streamlined workflow that can be completed within one standard working shift.

The NovaSeq X Series delivers extraordinary throughput and accuracy for data-intensive applications, such as RNA-Seq or multiomics studies at production-scale (Figure 1). Three flow cell types support multiple run configurations for added flexibility and scalability.

This application note showcases the use of the NovaSeq X Series to generate accurate RNA-Seq data at a quality that meets or exceeds that of data produced on the NovaSeq 6000 System.

XLEAP-SBS™ chemistry enables enhanced performance

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, included in the NovaSeq X Series Reagent kits, are optimized for performance and speed to maximize throughput without sacrificing data quality. The NovaSeq X Series enables significant gains in throughput with shorter run times, compared to the NovaSeq 6000 System (Table 1).

Methods

The NovaSeq X Series supports an RNA-Seq workflow that integrates RNA library preparation, sequencing, and streamlined, comprehensive data analysis (Figure 2).

Library preparation

Total RNA libraries were prepared from leukemia cell lines 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 the Illumina Stranded Total RNA Prep with Ribo-Zero™ Plus (Illumina, Catalog no. 20040529).

Table 1: Comparable sequencing output in significantly shorter run times

Metric	NovaSeq X Series 1.5B flow cell	NovaSeq X Series 10B flow cell	NovaSeq X Series 25B flow cell	NovaSeq 6000 S4 flow cell
2 × 100 bp output per run[a]	~330–660 Gb	~2–4 Tb	N/A	2–4 Tb
2 × 100 bp run time	~20 hr	~22 hr	N/A	~36 hr
2 × 150 bp output per run	~500–1000 Gb	~3–6 Tb	~8–16 Tb	3–6 Tb
2 × 150 bp run time	~23 hr	~25 hr	~48 hr	~44 hr
Single reads passing filter per flow cell	1.6B	10B	26B	8–10B
Total RNA-Seq libraries	~30	~200	~520	~400

a. Output range is for one or two flow cells.

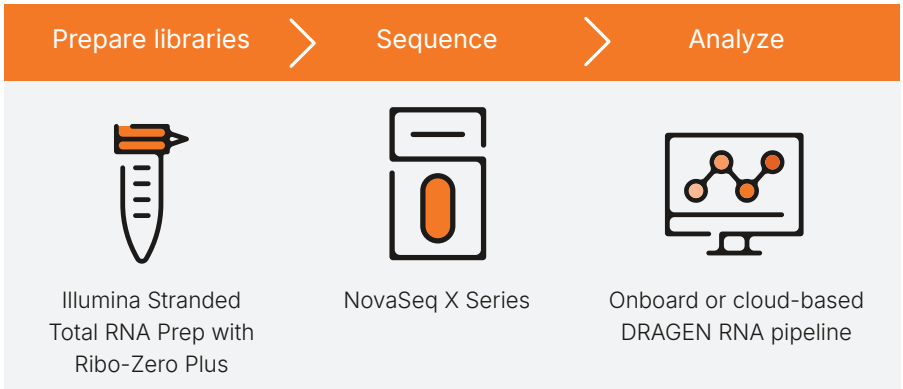

Figure 2: Simplified, integrated RNA-Seq workflow.

Sequencing

A total of 682 RNA-Seq libraries were sequenced on the NovaSeq X Plus System using a 2 × 75 bp run configuration, as follows:

- 61 libraries were run on the 1.5B flow cell using the NovaSeq X Series 1.5B Reagent Kit (300 cycles) (Illumina, Catalog no. 20104705)
- 141 libraries were run on the 10B flow cell using the NovaSeq X Series 10B Reagent Kit (300 cycles) (Illumina, Catalog no. 20085594)
- 480 libraries were run on the 25B flow cell using the NovaSeq X Series 25B Reagent Kit (300 cycles) (Illumina, Catalog no. 20104706)

For comparison, 60 total RNA samples were also sequenced on the NovaSeq 6000 System with the NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycles) (Illumina, Catalog no. 20028313) using a 2 × 76 bp run configuration.

A custom dark cycle recipe was used for sequencing on the NovaSeq X Plus System to mitigate sequencing first-base T overhangs, a by-product from RNA library preparation that occurs during ligation of sequencing adapters. The dark sequencing cycle performs chemistry only and omits the imaging step to improve sequencing metrics.

Data analysis on board on in the cloud

Sequencing data were downsampled to 10M reads for all samples to compare gene expression data using the DRAGEN™ RNA pipeline v4.0.4 cloud-based app. Data were aligned against the Genome Reference Consortium Human GRCh38 (h38 assembly). The DRAGEN RNA pipeline v4.0.4 was also used to generate fusion calls. Heatmaps were generated using Morpheus, a tool from the Broad Institute.

While analysis for this application note was performed using the DRAGEN RNA pipeline, RNA-Seq analysis can also be performed using Partek Flow. The software offers an easy-to-use interface, robust statistical algorithms, and information-rich visualizations.

Results

Total RNA-Seq data generated on the NovaSeq X Plus System exceeded published specifications for data quality, as well as results produced using the NovaSeq 6000 System (Table 2). Sequencing alignment metrics, including percent duplicates, percent mapped reads, and percent unique reads, showed excellent concordance

Table 2: Sequencing run metrics for total RNA-Seq

Metric	NovaSeq X Series			NovaSeq 6000 System
	1.5B flow cell	10B flow cell	25B flow cell	S4 flow cell
Run configuration	2 × 75 bp	2 × 75 bp	2 × 75 bp	2 × 76 bp
Read 1 bases ≥ Q30	94.00%	93.75%	94.98%	95.02%
Read 2 bases ≥ Q30	93.62%	93.64%	95.02%	92.77%
Read 1 error rate	0.09%	0.10%	0.14%	0.58%
Read 2 error rate	0.13%	0.15%	0.18%	1.07%
No. samples averaged	64	144	360	60
Multiplex per flow cell	32	192	480	240
No. of lanes per run	4 lanes: 2 runs, 2 lanes each	6 lanes: 3 runs, 2 lanes each	6 lanes: 3 runs, 2 lanes each	1 run, 1 lane

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

between the two platforms (R^2 > 0.99) for total RNA-Seq across all NovaSeq X Series flow cells (Figure 3 and Figure 4). Similarly, fusion calling data calling performed well on both platforms (Figure 5). These data demonstrate that whole-transcriptome sequencing on the NovaSeq X Series produces data quality that meets or exceeds NovaSeq 6000 System performance.

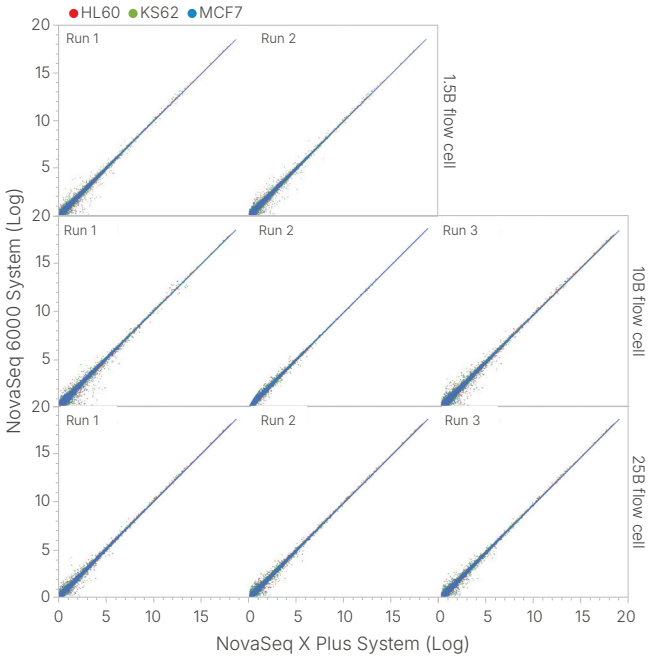

Figure 3: RNA-Seq data concordance between the NovaSeq X Plus and NovaSeq 6000 Systems—Sequencing data showed excellent concordance between the two platforms (R^2 > 0.99) for total RNA-Seq across all NovaSeq X Series flow cells and samples evaluated (HL60, KS62, and MCF7) across multiple runs.

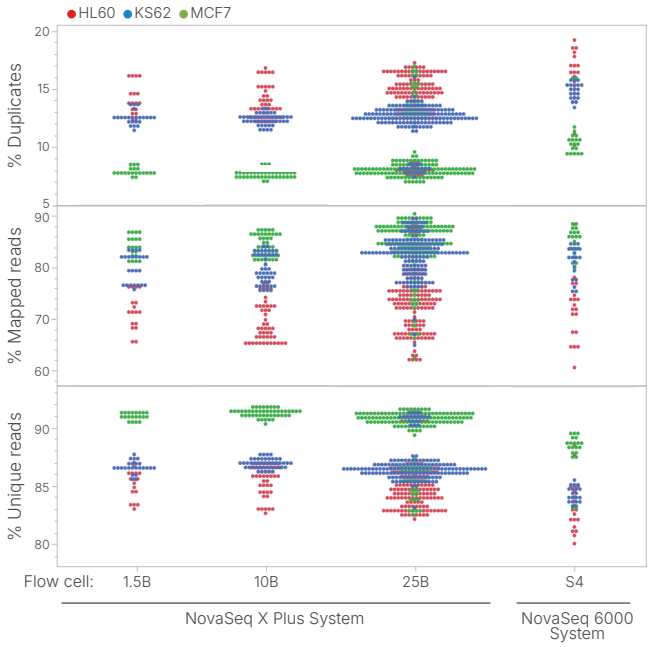

Figure 4: RNA-Seq alignment metrics on the NovaSeq X Plus and NovaSeq 6000 Systems—Sequencing alignment metrics, including percent duplicates, percent mapped reads, and percent unique reads showed excellent concordance between the two platforms (R^2 > 0.99) for total RNA-Seq across all NovaSeq X Series flow cells and samples evaluated (HL60, KS62, and MCF7).

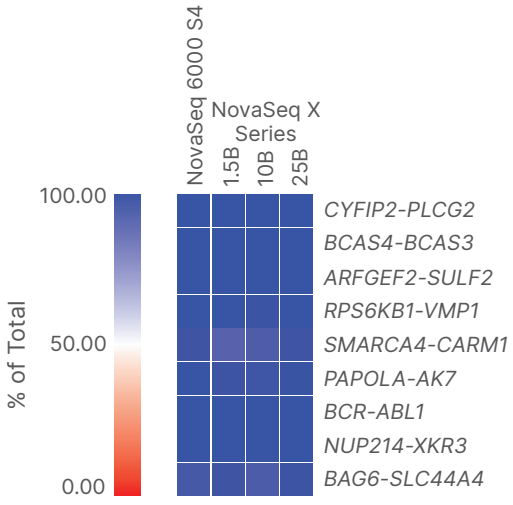

Figure 5: Fusion calling data concordance between the NovaSeq X Plus and NovaSeq 6000 Systems—Fusion calling data calling performed well on both platforms across flow cells, indicated in parentheses.

Summary

RNA-Seq on the NovaSeq X Series offers a streamlined RNA-to-results workflow that combines power, speed, and flexibility for high-throughput transcriptome analysis. Whole-transcriptome sequencing data run on the NovaSeq X Plus System were directly compared to data generated using the NovaSeq 6000 System. Results show that performance on the NovaSeq X Series meets or exceeds NovaSeq 6000 System performance across all three flow cell configurations on the NovaSeq X Plus System, offering scalability across RNA-Seq project types, from gene expression profiling to whole-transcriptome discovery. Transcriptomics profiling on the NovaSeq X Series can be integrated with other omes studies to drive a deeper understanding of biology.

Learn more

NovaSeq X Series

Data concordance on the NovaSeq X Series

Multiomics

References

1. Geraci F, Saha I, Bianchini M. Editorial: RNA-Seq Analysis: Methods, Applications and Challenges. *Front Genet*. 2020;11:220. Published 2020 Mar 17. doi:10.3389/fgene.2020.00220.
2. Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep*. 2020;10(1):19737. Published 2020 Nov 12. doi:10.1038/s41598-020-76881-x.

illumina®

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

M-GL-02636 v1.0