illumina®

Data concordance between the NextSeq™ 1000, NextSeq 2000, and NextSeq 550 Sequencing Systems

The latest NGS systems from Illumina offer innovative design while maintaining the same data quality customers require for common sequencing applications.

Introduction

Over the last 20 years, Illumina has led the charge to improve next-generation sequencing (NGS) capabilities, making it easier and more affordable to explore the genome, transcriptome, and epigenome. The latest platforms from Illumina, the NextSeq 1000 and NextSeq 2000 Sequencing Systems, offer breakthrough system design, chemistry innovations, compatibility with an expansive list of library preparation options, and onboard integrated informatics.

The NextSeq 1000 and NextSeq 2000 Sequencing Systems use the same Illumina sequencing by synthesis (SBS) chemistry that powers all Illumina platforms, including the NextSeq 550 System. Illumina SBS chemistry is responsible for generating more than 90% of the world's sequencing data[1] and delivers the highest yield of error-free reads[2], enabling robust base calling across the genome.[3]

The NextSeq 1000 and NextSeq 2000 systems are optimized to increase cluster brightness, reduce channel cross talk, and improve signal-to-noise ratio. Combined with the latest advances in optics and instrument design, the NextSeq 1000 and NextSeq 2000 Sequencing Systems support increases in output and reduce the cost per run, all while maintaining the same high-quality data users have come to expect from the NextSeq 550 System.

This application note demonstrates that the NextSeq 1000 and NextSeq 2000 Sequencing Systems deliver equivalent data quality to the NextSeq 550 System for key applications, including exome, bulk, and single-cell RNA gene expression.

Methods

Exome sequencing

Exome libraries were prepared from NA12878 genomic DNA (gDNA) (Coriell Institute for Medical Research) using the Nextera™ DNA Flex Pre-Enrichment Library Prep and Enrichment Reagents Kit (Illumina, Catalog no. 20025524) to capture genomic regions targeted by the Illumina Exome Panel (Illumina, Catalog no. 20020183).

Sequencing was performed on the NextSeq 2000* System with the NextSeq 1000/2000 P2 Reagent Kit (200 cycles) (Illumina Catalog no. 20040557), using the 2 × 100 bp run configuration. For comparison, the same libraries were also sequenced on the NextSeq 550 System with the NextSeq 500/550 High-Output Kit v2.5 (300 cycles) (Catalog no. 20024908), using a 2 × 100 bp run configuration. Eight samples were multiplexed on a single run for each instrument.

Secondary data analysis was performed using the DRAGEN™ Enrichment Pipeline v3.4, available onboard the NextSeq 2000 System or through BaseSpace™ Sequence Hub. Variant calling accuracy was assessed against the Platinum Genomes 2016 v1.0 data set.[4]

Bulk mRNA sequencing

Messenger RNA (mRNA) libraries were prepared from Universal Human Reference RNA (Coriell Institute for Medical Research) using the TruSeq™ Stranded mRNA Library Prep Kit (Illumina, Catalog no. 20020594).

Sequencing was performed on the NextSeq 2000 System with NextSeq 1000/2000 P2 Reagent Kit (200 cycles) using a 2 × 76 bp run configuration. For comparison, the same libraries were also sequenced on the NextSeq 550 System with the NextSeq 500/550 High-Output Kit v2.5 (300 cycles) using a 2 × 76 bp run configuration. Twenty-four samples were multiplexed in a single run for each instrument.

Secondary data analysis was performed using the DRAGEN RNA Pipeline v3.5.112, available onboard the NextSeq 2000 System or in BaseSpace Sequence Hub. Data were aligned against the Genome Reference Consortium Human GRCh38 (h38 assembly).

Single-cell RNA sequencing

Samples for single-cell RNA sequencing (scRNA-Seq) were prepared from 4-plex peripheral blood mononuclear cells (PBMCs) isolated from whole blood from a single donor. Libraries were prepared from ~1000 PBMCs using the Chromium Single Cell Gene Expression v3 Solution (10x Genomics, Catalog no. 1000092) on the Chromium Controller (10x Genomics, Catalog no. 120223).

Sequencing was performed on the NextSeq 2000 System with NextSeq 1000/2000 P2 Reagent Kit (200 cycles). For comparison, the same libraries were also sequenced on the NextSeq 550 System with the NextSeq 500/550 High-Output Kit v2.5 (150 cycles) (Illumina, Catalog no. 20024907). Run configurations were set up according to parameters provided by 10x Genomics: 28-cycle read 1, 8-cycle index read, and 91-cycle read 2.

Data analysis was performed using Cell Ranger scRNA-Seq analysis pipelines (10x Genomics, Inc.) to align reads, perform clustering, and analyze gene expression.

* While the NextSeq 2000 System has higher throughput than the NextSeq 1000 System, performance and data quality are equivalent. Therefore, only the NextSeq 2000 System was used for data comparison studies in this application note.

Results

Exome sequencing

Primary and secondary analysis metrics, including precision and recall for both single nucleotide variants (SNVs) and insertions/deletions (indels), run output (yield), error rate, % aligned reads, mean target coverage depth, and uniformity of coverage, were evaluated. Both the NextSeq 550 and NextSeq 2000 Systems exceeded published specifications for sequencing output and data quality, with excellent performance on both sequencing platforms (Table 1). These data demonstrate that exome sequencing results on the NextSeq 2000 System are equivalent to the NextSeq 550 System, with both platforms producing high-quality data and highly accurate variant calling.

Table 1: Equivalent performance for exome sequencing

	NextSeq 550 System	NextSeq 2000 System Run 1	NextSeq 2000 System Run 2
Primary run metrics			
Clusters PF	514.2M	573.0M	565.1M
Yield	103.9 Gb	114.6 Gb	113.1 Gb
% Q30	92.20%	94.14%	94.15%
Error rate	0.34%	0.16%	0.16%
Secondary analysis metrics[a]			
Autosome callability	95.61%	96.90%	96.90%
% aligned reads	99.15%	99.53%	99.50%
Mean target coverage depth	78.05	81.87	82.42
Uniformity of coverage	93.20%	96.92%	96.94%
SNV precision	99.3%	99.56%	99.59%
SNV recall	94.1%	94.88%	94.94%
Indel precision	88.0%	95.34%	95.71%
Indel recall	83.3%	84.90%	84.97%

a. Averaged over 12 samples.

Bulk mRNA sequencing

Both the NextSeq 550 and NextSeq 2000 Systems exceeded published specifications for sequencing output and data quality (Table 2). Quantification of specific RNA targets by bulk mRNA-Seq showed excellent concordance between the two platforms ($R^2 > 0.99$) across four replicates (Figure 1). These data reinforce the findings that the NextSeq 2000 System produces data quality equivalent to the NextSeq 550 System for quantifying mRNA expression levels.

Table 2: Equivalent performance for bulk mRNA-Seq

Metric	NextSeq 550 System	NextSeq 2000 System Run 1	NextSeq 2000 System Run 2
Clusters PF	422.4M	537.8M	531.4M
Yield	63.6 Gb	80.2 Gb	79.75 Gb
% Q30	92.85%	92.93%	92.28%
Error rate	0.27%	0.14%	0.16%
Mean no. of genes detected (n = 24)	24,355	26,507	24,739
Mean no. of transcripts detected (n = 24)	74,387	88,125	81,948

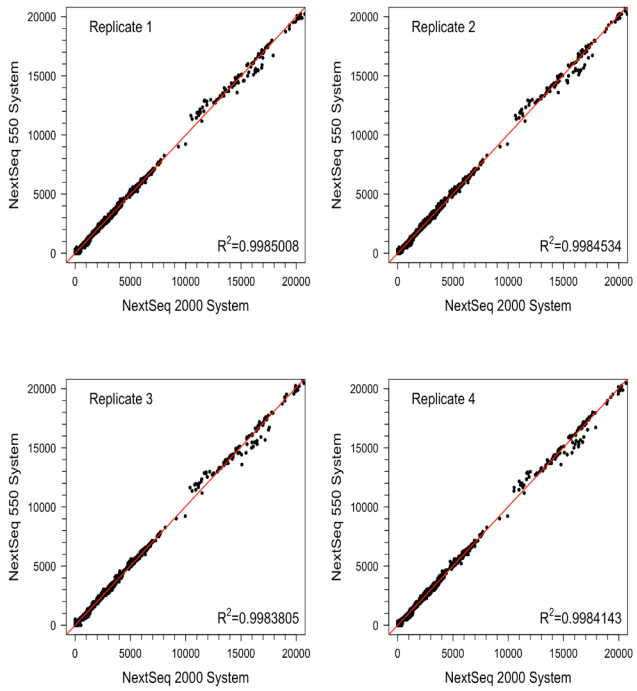

Figure 1: Highly concordant results for mRNA-Seq—Quantification of specific RNA targets by bulk mRNA-Seq using the NextSeq 550 System are plotted on the y-axis. Results using the NextSeq 2000 System are plotted on the x-axis. Concordance of data is observed along the y = x trendline ($R^2 > 0.99$ across four replicates).

Single-cell whole-transcriptome sequencing

The NextSeq 550 and NextSeq 2000 Systems exceeded published specifications for sequencing output and data quality (Table 3). Quantification of individual cells using unique molecular identifiers (UMIs) in scRNA-Seq showed excellent concordance between the two platforms ($R^2 > 0.99$) across four replicates (Figure 2). These data further demonstrate that the NextSeq 2000 System produces data quality equivalent to the NextSeq 550 System for scRNA-Seq studies.

Table 3: Equivalent performance for scRNA-Seq

Metric	NextSeq 550 System	NextSeq 2000 System Run 1	NextSeq 2000 System Run 2
Clusters PF	454.8M	498.7M	530.1M
Yield	53.2 Gb	58.3 Gb	62.0 Gb
% Q30 R1	96.28%	94.05%	93.91%
% Q30 R2	87.73%	91.35%	91.51%
Error rate	0.29%	0.14%	0.14%
No. of genes detected	18,529	18,541	18,438
Median UMI counts per cell	5220	5175	5132
Fraction of reads in cells	95.05%	94.95%	94.93%
Estimated no. of cells	1181	1185	1184

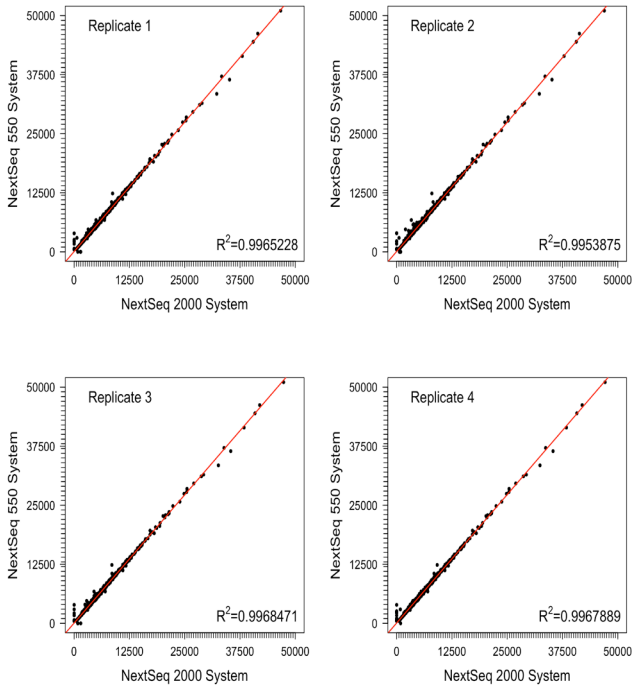

Figure 2: Highly concordant results for scRNA-Seq—Quantification of individual cells (UMIs) by scRNA-Seq using the NextSeq 550 System are plotted on the y-axis. Results using the NextSeq 2000 System are plotted on the x-axis. Concordance of data is observed along the y = x trendline ($R^2 > 0.99$ across four replicates).

Summary

The NextSeq 1000 and NextSeq 2000 Sequencing Systems feature breakthrough system design, chemistry innovations, and onboard informatics that revolutionize what can be accomplished with a benchtop sequencing system. These platforms deliver improvements in sequencing costs and run capacity while maintaining the same high-quality data users expect from Illumina. Data from key applications commonly run on the NextSeq 550 System, including exome, bulk mRNA, and single-cell RNA sequencing, were directly compared to data generated using the NextSeq 2000 System. Results show that performance on both the NextSeq 550 and NextSeq 2000 Systems is highly concordant and reflects the commitment of Illumina to deliver consistent, high-quality sequencing performance reliably.

Learn more

To learn more about the NextSeq 1000 and NextSeq 2000 Sequencing Systems, visit www.illumina.com/systems/sequencing-platforms/nextseq-1000-2000.html

To download the data sets referenced in this note, visit basespace.illumina.com/datacentral.

References

1. Data calculations on file. Illumina, Inc. 2017.
2. Based on a comparison of the top three industry-leading NGS platforms. Data calculations on file. Illumina, Inc. 2016.
3. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013;14:R51.
4. Eberle MA, Fritzilas E, Krusche P, et al. A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Res*. 2017;27:157–164.

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