illumina®

TECHNICAL NOTE

High-accuracy WGS with updates to the NovaSeq™ X Series

Comparable or improved data quality with software and hardware updates

Software update

NovaSeq X Series control software can be updated from v1.3 to v1.4

Laser system update

NovaSeq X Series control software v1.4 enables a new, more robust laser

Equivalent results

Comparable or better WGS data quality, as measured by primary and secondary metrics

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 the most powerful and cost-effective Illumina sequencing systems yet.

Continued support of the NovaSeq X Series includes an update of NovaSeq X Series control software from v1.3.1 (v1.3 hereafter) to v1.4. NovaSeq X Series control software v1.4 enables conversion of the Line Generation Module (LGM) optical component to the Alternate Laser Engine (ALE) component. The ALE component is more robust than the LGM component, leading to improved uptime and less unexpected downtime for the NovaSeq X Series.

This technical note demonstrates that these software and hardware updates deliver data quality that meets or exceeds the previous versions on the NovaSeq X Series.

Methods

Sample preparation

NA24385 genomic DNA (gDNA) was obtained from the Coriell Institute for Medical Research for evaluation of whole-genome sequencing (WGS).

Library preparation

Whole-genome libraries were prepared from NA24385 gDNA using the TruSeq™ PCR-Free Prep kit (Illumina, Catalog no. 20015963) following the standard protocol.

Sequencing

Sequencing was performed on the NovaSeq X Plus System with the NovaSeq X Series 1.5B Reagent Kit (300 cycles, Illumina, Catalog no. 20104705), the NovaSeq X Series 5B Reagent Kit (300 cycles, Illumina, Catalog no. 20145967), the NovaSeq X Series 10B Reagent Kit (300 cycles, Illumina, Catalog no. 20085594), and the NovaSeq X Series 25B Reagent Kit (300 cycles, Illumina, Catalog no. 20104706) using the 2 × 151 bp run configuration.

Data analysis

Secondary data analysis was performed using the DRAGEN™ Germline pipeline v4.3 cloud-based workflow for all v1.3 LGM runs and all v1.4 (LGM or ALE) were analyzed using DRAGEN Germline pipeline v4.4.

Results

Primary sequencing run metrics

Sequencing run metrics were evaluated, including percent passing filter, nonindexed yield, percent bases above Q30, and error rate. Updating the NovaSeq X Series control software from v1.3 to v1.4 resulted in comparable primary sequencing metrics (Table 1). Similarly, sequencing metrics were comparable between the LGM and ALE systems (Table 1).

Secondary metrics and variant calling

WGS analysis metrics were evaluated, including duplicate reads, insert size, GC coverage, and variant calling for both single nucleotide variants (SNVs) and insertions–deletions (indels). Updating the NovaSeq X Series control software from v1.3 to v1.4 resulted in comparable percentage of duplicate reads (Figure 1), insert size (Figure 2A), and GC coverage (Figure 2B). Similarly, the LGM and ALE systems resulted in comparable percentage of duplicate reads (Figure 1), insert size (Figure 2A), and GC coverage (Figure 2B).

Updating the NovaSeq X Series control software from v1.3 to v1.4 resulted in improved precision and recall for SNVs (Table 2, Figure 3) and indels (Table 2, Figure 4). Improvements in variant calling were also seen with reduced false positives and false negatives with the v1.4 control software compared to v1.3 (Figure 5). The ALE system delivered comparable or slightly better variant calling compared to the LGM system, as seen by SNV precision and recall (Table 2, Figure 3), indel precision and recall (Table 2, Figure 4), and total false positives and false negatives in variant calling (Figure 5). These data demonstrate that the software and hardware updates deliver WGS data quality that meets or exceeds the previous versions on the NovaSeq X Series.

Table 1: Sequencing metrics for 300-cycle WGS runs of TruSeq DNA PCR-Free libraries

Metric	NovaSeq X Series control software v1.3 with LGM				NovaSeq X Series control software v1.4 with LGM				NovaSeq X Series control software v1.4 with ALE[a]			
Flow cell	1.5B	5B[b]	10B	25B	1.5B	5B	10B	25B	1.5B	5B	10B	25B
Percent passing filter	76%	—	79%	74%	78%	75%	78%	74%	81%	77%	79%	76%
Yield per flow cell (Gb)	634	—	3680	9064	654	1760	3632	9072	674	1808	3704	9360
Percent bases ≥ Q30	92.7%	—	93.5%	92.7%	93.4%	93.4%	93.0%	93.7%	93.9%	93.6%	94.4%	94.8%
Read 1 error rate	0.2	—	0.2	0.3	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2
Read 2 error rate	0.2	—	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
No. runs averaged	51	—	58	80	9	6	7	25	9	8	13	21

a. The ALE system is only compatible with NovaSeq X Series control software v1.4 and higher.
b. 5B flow cells are enabled to run on NovaSeq X Series control software v1.4 and higher.
Metrics from single flow cell runs averaged across multiple flow cells with varying number of lanes. ALE, Alternate Laser Engine; LGM, Line Generation Module; WGS, whole-genome sequencing.

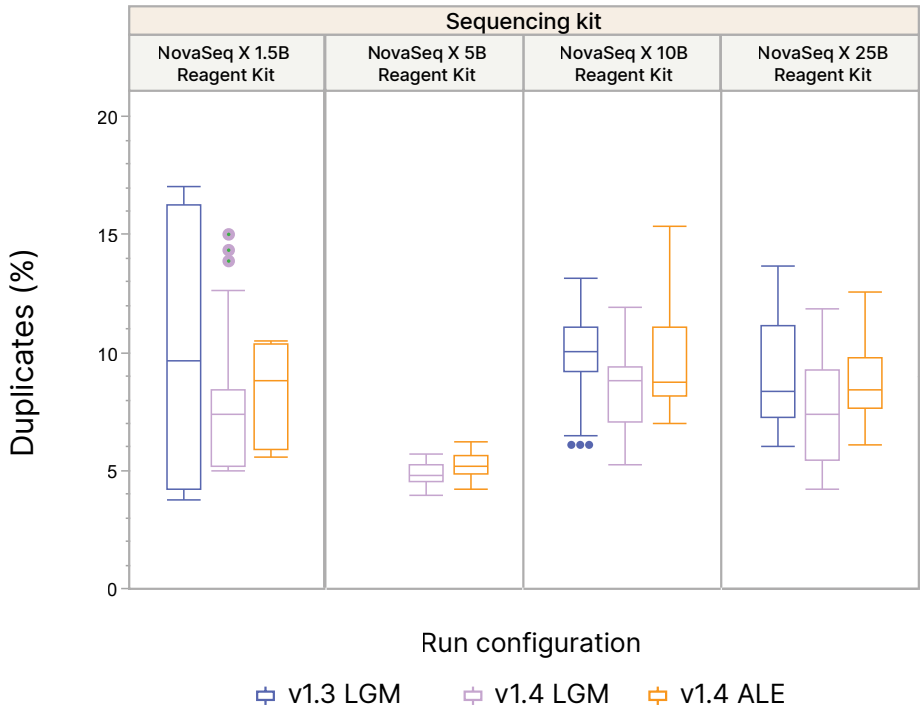

Figure 1: Comparable levels of duplicate reads between NovaSeq X Series configurations
TruSeq DNA PCR-Free libraries were sequenced on the NovaSeq X Series with various reagent kits. Updating the NovaSeq X Series control software from v1.3 (blue) to v1.4 (purple, orange) resulted in comparable levels of duplicate reads. Similarly, the ALE (orange) system resulted in comparable duplicate read levels compared to the LGM (purple). Small differences may be caused by inherent variation among library pools. ALE, Alternate Laser Engine; LGM, Line Generation Module.

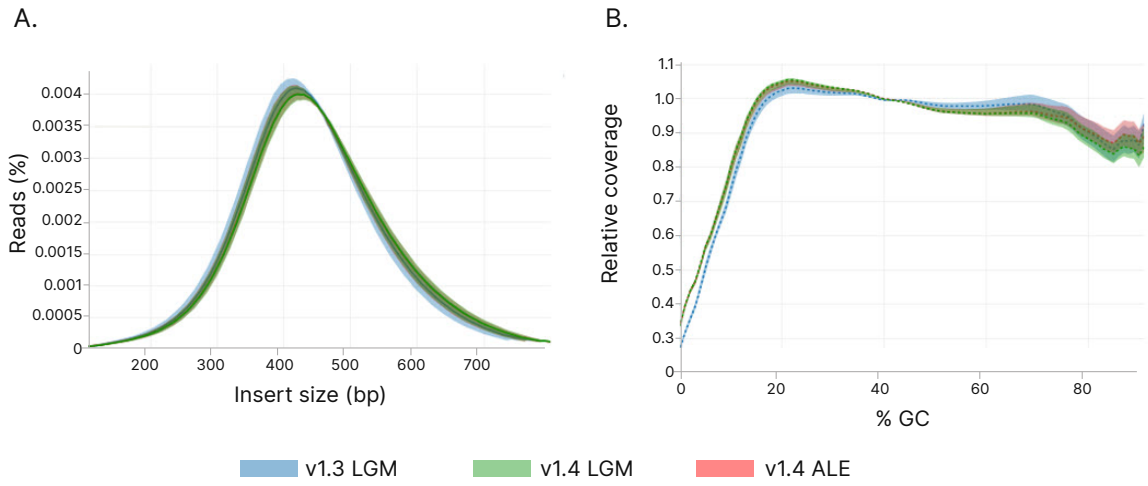

Figure 2: Comparable insert sizes and GC coverage between NovaSeq X Series configurations
Updating the NovaSeq X Series control software from v1.3 (blue) to v1.4 (red, green) resulted in comparable (A) insert sizes and (B) GC coverage. Similarly, the ALE system (red) resulted in equivalent (A) insert sizes and (B) GC coverage, compared to the LGM system (green). Insert size was determined via the DRAGEN Germline Pipeline. ALE, Alternate Laser Engine; LGM, Line Generation Module.

Table 2: Secondary analysis metrics for 300-cycle WGS runs of TruSeq DNA PCR-Free libraries

Metric	NovaSeq X Series control software v1.3 with LGM[a]				NovaSeq X Series control software v1.4 with LGM[b]				NovaSeq X Series control software v1.4 with ALE[b,c]			
Flow cell	1.5B	5B[d]	10B	25B	1.5B	5B	10B	25B	1.5B	5B	10B	25B
SNV precision	99.77%	—	99.77%	99.76%	99.79%	99.78%	99.78%	99.77%	99.80%	99.79%	99.80%	99.78%
SNV recall	99.41%	—	99.39%	99.39%	99.51%	99.48%	99.49%	99.47%	99.41%	99.49%	99.52%	99.50%
Indel precision	97.54%	—	97.31%	97.32%	98.34%	98.14%	98.12%	98.05%	98.48%	98.19%	98.30%	98.32%
Indel recall	95.72%	—	95.29%	95.42%	97.42%	97.07%	97.06%	96.93%	97.66%	97.17%	97.36%	97.37%
No. runs averaged	2	—	5	3	4	1	4	8	4	1	4	4

a. Sequencing data for the sample HG002 (NA24385) was downsampled to 30× effective coverage and analyzed using DRAGEN Germline pipeline v4.3.
b. Sequencing data for the sample HG002 (NA24385) was downsampled to 30× effective coverage and analyzed using DRAGEN Germline pipeline v4.4.
c. The ALE system is only compatible with NovaSeq X Series control software v1.4 and higher.
d. 5B flow cells are not enabled to run on NovaSeq X Series control software v1.3.
ALE, Alternate Laser Engine; Indel, insertion/deletion; LGM, Line Generation Module; SNV, single nucleotide variant; WGS, whole-genome sequencing.

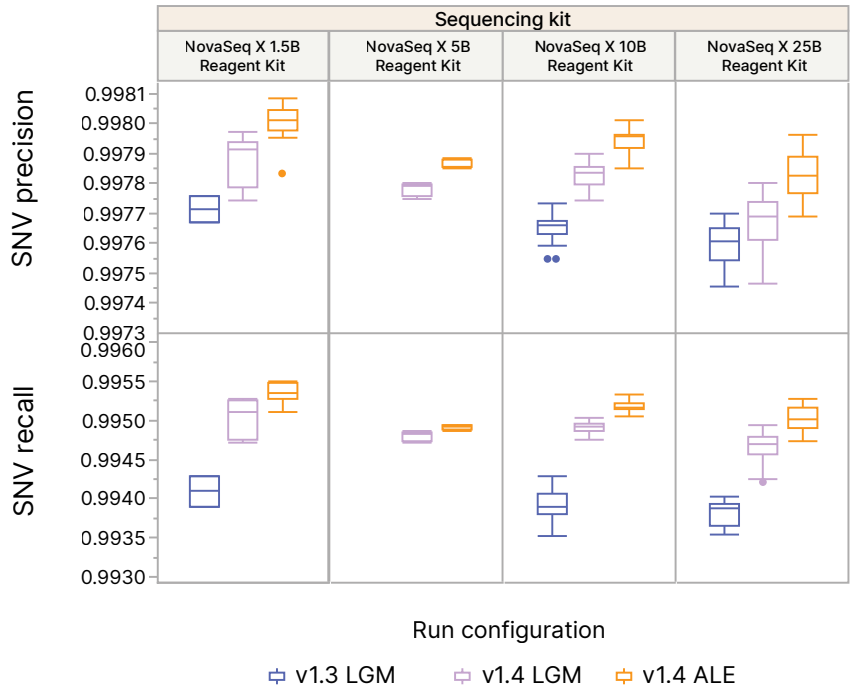

Figure 3: Comparable or improved SNV calling between NovaSeq X Series configurations

TruSeq DNA PCR-Free libraries were sequenced on the NovaSeq X Series with various reagent kits. Updating the NovaSeq X Series control software from v1.3 (blue) to v1.4 (purple, orange) resulted in improved SNV precision and recall. The ALE system (orange) resulted in equivalent or slightly improved SNV precision and recall, compared to the LGM system (purple). ALE, Alternate Laser Engine; LGM, Line Generation Module; SNV, single nucleotide variant.

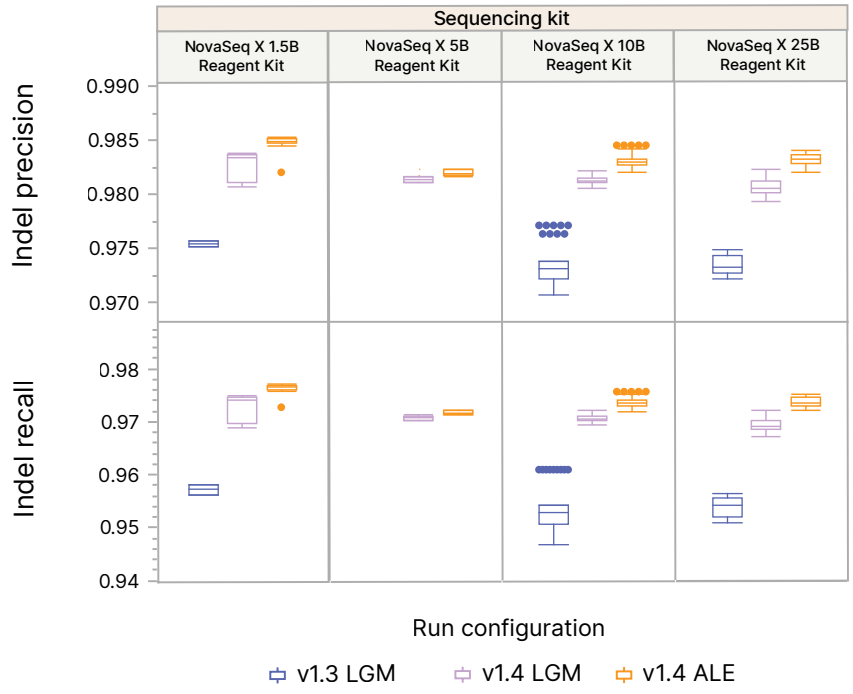

Figure 4: Comparable or improved indel calling between NovaSeq X Series configurations

TruSeq DNA PCR-Free libraries were sequenced on the NovaSeq X Series with various reagent kits. Updating the NovaSeq X Series control software from v1.3 (blue) to v1.4 (purple, orange) resulted in improved indel precision and recall. The ALE system (orange) resulted in equivalent or slightly improved indel precision and recall, compared to the LGM system (purple). ALE, Alternate Laser Engine; LGM, Line Generation Module; indel, insertion/deletion.

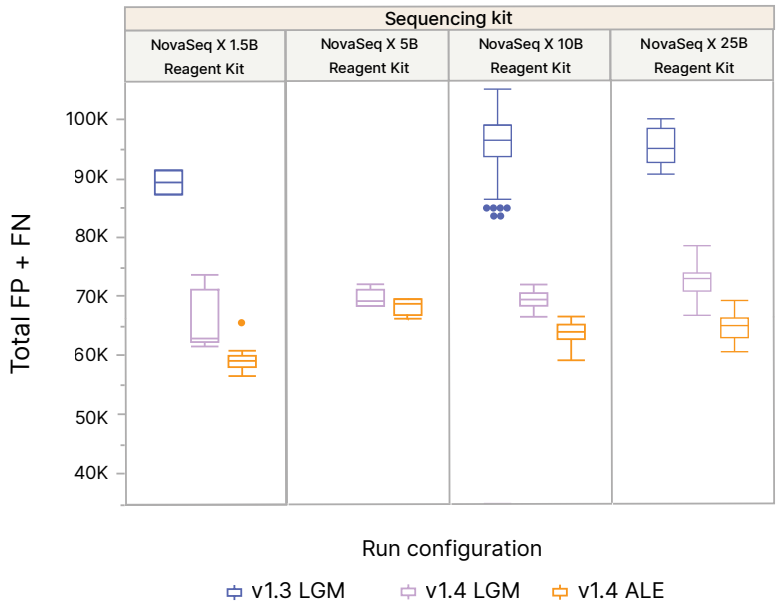

Figure 5: Reduced levels of false positives and false negatives between NovaSeq X Series configurations
TruSeq DNA PCR-Free libraries were sequenced on the NovaSeq X Series with various reagent kits. Updating the NovaSeq X Series control software from v1.3 (blue) to v1.4 (purple, orange) resulted in improved variant calling, as observed by reduced numbers of false positives and false negatives. The ALE system (orange) resulted in equivalent or slightly improved variant calling compared to the LGM system (purple), as observed by fewer false positive and false negative calls. ALE, Alternate Laser Engine; LGM, Line Generation Module; FP, false positive; FN, false negative.

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. Users can expect comparable or improved data quality when upgrading from v1.3 to v1.4 control software. Users should expect equivalent data quality when comparing data between an LGM and ALE system on v1.4 and may see slightly improved variant calling.

Learn more →

NovaSeq X Series

Demo data sets

illumina®

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

M-GL-03907 v1.0