illumina®

Human Whole-Genome Sequencing with the HiSeq X™ Sequencing System

Designed for population-level sequencing, the HiSeq X System delivers massive daily throughput with exceptional data quality.

HiSeq X Highlights

- **Unprecedented Speed and Throughput**
  Generate 1.6–1.8 Tb in < 3 days in dual flow cell mode to perform whole-genome sequencing on an unrivaled scale
- **Exceptional Data Quality**
  Highly accurate Illumina sequencing by synthesis (SBS) chemistry delivers proven industry-leading data quality
- **Population-Scale Genome Sequencing**
  Highest daily throughput delivers unmatched cost-effectiveness for population-scale projects

Table 1: HiSeq X Primary Run Metrics

	Per Lane Average (Flow Cell A)	*Per Lane* Average (Flow Cell B)	*Per Run* Total or Average %
Percent reads passing filter	70.9%	72.3%	71.6%
Clusters passing filter (million)	440	449	7,120
% Bases > Q30 (both reads)	89.7%	88.7%	89.2%
Yield (Gb)	132.2	134.7	2,134.4

Introduction

The revolutionary HiSeq X Ten is a set of 10 HiSeq X Sequencing Systems engineered for population-level, human whole-genome sequencing. Based on the same core architecture as the groundbreaking HiSeq 2500 system, but with advances such as patterned flow cells and faster chemistry, the HiSeq X Ten can generate an unprecedented level of daily throughput. In dual flow cell mode, a single HiSeq X instrument can deliver up to 1.8 terabases (Tb) of data per run in less than three days, or approximately 600 gigabases (Gb) of data per day.†

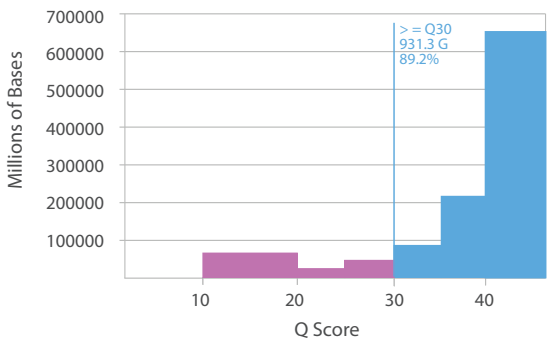
As with the HiSeq 2500 System, the HiSeq X System leverages proven Illumina sequencing by synthesis (SBS) chemistry—the most widely adopted next-generation sequencing (NGS) technology[1]. With SBS chemistry, the HiSeq X sequencer delivers exceptional data accuracy, the highest yield of error-free reads, and the highest percentage of base calls above Q30 in the industry[2,3].

This application note describes and compares human whole-genome sequencing from a single run on a development HiSeq X instrument to previous human whole-genome HiSeq 2500 runs. Here we show that the data quality meets or exceeds data quality generated from previous runs using standard TruSeq® v3 reagents and HiSeq v4 reagents on the HiSeq 2500 system.

Methods and Results

Libraries were prepared from 100 ng of Coriell sample NA12878 genomic DNA using the TruSeq Nano DNA library preparation protocol with a 350 bp insert size. Following PCR quantitation and dilution to 3 nM, cluster generation was performed on a cBot using HiSeq X HD clustering reagents (Illumina, HiSeq X HD Reagent Kit, FC-501-1001) and patterned flow cells with one sample loaded per lane. Once flowcells were loaded onto the HiSeq X instrument, all other sequencing steps, including Read 1, paired-end turn, and Read 2, proceeded automatically without user intervention. The run yielded 132–136 Gb per lane or 32×–38× genome coverage per lane after alignment and duplicate flagging. Data quality values were 89% of bases above Q30 (Figure 1). Additional primary sequencing metrics are shown in Table 1. BCL files generated by the HiSeq X were aligned against the human reference hg19[4], and BAM and variant call files were generated using Isaac[5]. Variant calls, build depth, and additional secondary analysis metrics are shown in Table 2. Raw data and variant call files from this run are available upon request. (Please contact your Illumina sales representative.)

Figure 1: HiSeq X Quality Score Distribution



Quality scores for a human genome 2 × 151 base pair run on a HiSeq X System. This example shows more than 89% of bases sequenced above Q30.

† The HiSeq X is only available as part of the The HiSeq X Ten set of sequencers.

Table 2: Secondary Build Metrics

Performance Parameter	HiSeq 2500	HiSeq 2500	HiSeq X
Run Configuration	2 × 101	2 × 126	2 × 151
Sequencing Chemistry	TruSeq SBS v3 Kit	HiSeq SBS v4 Kit	HiSeq X HD Reagent Kit
Lanes Used	4	2	1
Build depth	51.0	42.5	37.3
Build depth, downsampled	30×	30×	30×
Total SNPs	3,385,124	3,475,234	3,509,296
Heterozygous:Homozygous ratio	1.64	1.64	1.65
Transition:Transversion ratio	2.09	2.08	2.09
Matching position in dbSNP132	98.33	98.11	97.98
Reference genome coverage†	98.1%	98.7%	98.6%

† Percent of reference bases ≥10× mean coverage.

Conclusions

In addition to impressive speed and throughput, the HiSeq X delivers exceptional data quality comparable to the HiSeq 2500 System. While all HiSeq Systems employ our proven, highly accurate SBS chemistry, the HiSeq 2500 and HiSeq X sequencers have distinct features and benefits tailored to different research needs. The HiSeq 2500 System is a powerful, production-scale platform designed for maximum flexibility. With rapid or high-output modes and single or dual flow cell options, the HiSeq 2500 can be tuned to meet output needs ranging from 10 Gb to 1 Tb[6]. This flexibility supports a broad range of research applications including ChIP-Seq, RNA-Seq, exome sequencing, whole-epigenome sequencing, *de novo* sequencing, and more.

Designed specifically for, and supporting only human whole-genome sequencing, the HiSeq X Ten is revolutionizing the speed and throughput of human whole-genome sequencing. An individual HiSeq X System can generate up to 1.8 terabases (Tb) of data in less than three days, or approximately 600 gigabases (Gb) per day. When operating in parallel, the 10 instruments of a HiSeq X Ten System can generate tens of thousands of genomes per year for the ground-breaking price of $1000 per genome[7]. Aside from differences in throughput and supported applications, the HiSeq 2500 and HiSeq X Ten Systems both deliver industry-leading data accuracy and proven performance.

Learn More

For more information about the HiSeq X Ten, visit
www.illumina.com/systems/hiseq-x-sequencing-system.ilmn

To compare Illumina system specifications,
www.illumina.com/systems/sequencing.ilmn

References

1. Nakazato T, Ohta T, Bono H. (2013) Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. PLoS One 22;8(10): e77910.
2. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. (2013) Characterizing and measuring bias in sequence data. Gen Biol 14: R51.
3. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012: 251364.
4. UCSC Genome Bioinformatics Page (hgdownload.cse.ucsc.edu/downloads.html).
5. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, et al. (2013) Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 29(16): 2041–2043.
6. Illumina (2014) HiSeq 2500 system specification sheet. (www.illumina.com/documents/products/datasheets/datasheet_hiseq2500.pdf)
7. Illumina (2014) HiSeqX Ten preliminary system specification sheet. (www.illumina.com/documents/products/datasheets/datasheet-hiseq-x-ten.pdf)

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

FOR RESEARCH USE ONLY

Pub. No. 770-2014-008 Current as of 15 July 2014

illumina®