APPLICATION NOTE

NextSeq™ 1000 & NextSeq 2000 exome sequencing solution

Integrated workflow for cost-effective exome analysis and accurate variant calling

- Streamlined library preparation and exome enrichment for highly uniform coverage of coding regions
- Flexible, scalable benchtop sequencing systems for industry-leading data quality
- Onboard data analysis pipeline with award-winning performance for calling common mutations and rare somatic variants

illumina®

Introduction

The NextSeq 1000 and NextSeq 2000 exome sequencing solution offers a cost-effective, DNA-to-results workflow to investigate the protein-coding regions of the genome. The solution leverages industry-leading Illumina next-generation sequencing (NGS) technology and optimized sequencing by synthesis (SBS) chemistry to deliver exceptional data quality. This high-accuracy exome coverage enables identification of true coding variants for a broad range of applications, including population genetics, genetic disease research, and cancer studies. The integrated workflow provides streamlined library preparation and exome enrichment, push-button sequencing, and rapid, accurate data analysis (Figure 1). With minimal hands-on time, the NextSeq 1000 and NextSeq 2000 exome sequencing solution is a highly flexible, efficient method for interrogating the exome.

Figure 2: NextSeq 1000 and NextSeq 2000 Sequencing Systems—The NextSeq 1000 and NextSeq 2000 systems harness optimized SBS chemistry to streamline sequencing workflows.

Simple, efficient workflow

The NextSeq 1000 and NextSeq 2000 exome sequencing solution simplifies exome sequencing, enabling researchers to maximize their productivity. It begins with library preparation and exome enrichment using Illumina DNA Prep with Enrichment.* Prepared libraries are loaded on to a flow cell and then on to the NextSeq 1000 or NextSeq 2000 system for sequencing (Figure 2). The NextSeq 1000 and NextSeq 2000 Sequencing Systems feature multiple sequencing flow cell configurations that enable researchers to scale their exome studies according to their needs. Data analysis, including alignment and variant calling, is easily performed with the DRAGEN™ Enrichment pipeline onboard the instrument or in BaseSpace™ Sequence Hub, the cloud-based Illumina genomics computing environment.

* Illumina DNA Prep with Enrichment was previously known as Nextera™ Flex for Enrichment.

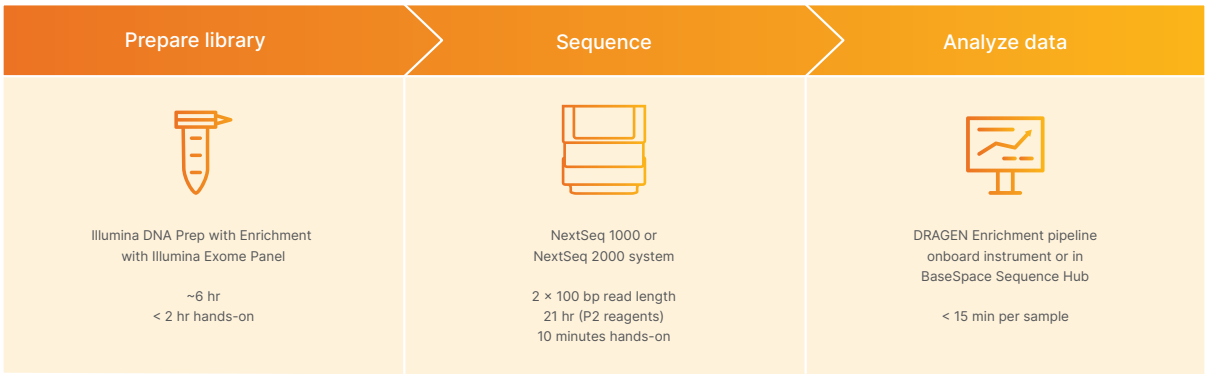

Figure 1: NextSeq 1000 and NextSeq 2000 exome sequencing workflow—The NextSeq 1000 and NextSeq 2000 Sequencing Systems are part of a simple, integrated NGS workflow that delivers highly accurate exome sequencing data. Times vary by experiment and assay type.

Streamlined library preparation and exome enrichment

Illumina DNA Prep with Enrichment combines rapid library preparation using Illumina bead-bound transposome chemistry and exome enrichment, enabling efficient identification of true coding variants. Providing comprehensive exome coverage from only 10 ng of input, Illumina DNA Prep with Enrichment allows labs to analyze precious DNA samples, while still producing high coverage uniformity and enrichment rates. The highly sensitive detection of low-frequency variants enables labs to identify germline and rare somatic mutations with accuracy.

Using Illumina DNA Prep with Enrichment, researchers can choose panel content from various vendors, including Illumina, Twist, and IDT (Table 1). This means that researchers can take advantage of the workflow and data quality benefits of Illumina DNA Prep with Enrichment with multiple exome panels. On-bead tagmentation eliminates the need for mechanical shearing to fragment DNA. This streamlines the workflow to a total time of about six hours with less than two hours of hands-on time.

Table 1: Exome panel specifications

Panel features[a]	Illumina Exome Panel	Agilent	Twist	IDT
Panel size	45.2 Mb	36 Mb	33 Mb	39 Mb
Probe size	80 bp	N/A	120 bp	120 bp
Probe type	ssDNA	RNA	dsDNA	ssDNA
Enrichment (Hyb) time	1.5 hr	16 hr	1.5 hr	1.5–16 hr
Databases used for exome panel design[b]				
RefSeq[1]	99.83%	99.88%	99.08%	99.45%
GENCODE[2]	98.02%	97.29%	96.01%	96.82%
CCDS[3]	99.99%	99.91%	99.76%	99.67%
UCSC Known Genes[4]	99.89%	98.72%	97.63%	98.13%
ClinVar[5]	84.95%	73.41%	72.56%	72.90%

a. Panel size = the total length of sequence in the target regions; probe size = length of enrichment hybridization (Hyb) probe; probe type = probe oligonucleotides can be RNA, DNA, single stranded (ss), or double stranded (ds).
b. Percentages refer to how much of the databases each exome panel covers.

The NextSeq 1000 and NextSeq 2000 Sequencing Systems

The NextSeq 1000 and NextSeq 2000 Sequencing Systems provide power and versatility to streamline and simplify the exome sequencing workflow. With no washes required, it takes less than 10 minutes to load and initiate the system. For the NextSeq 2000 System, sequencing is completed in approximately 33 hours for up to 48 samples using NextSeq 2000 P3 Reagents and paired-end 100 bp read lengths.

The NextSeq 1000 and NextSeq 2000 Sequencing Systems are compatible with a wide range of library preparation kits from Illumina and third parties, and offer cross-application flexibility. Researchers can transition easily between sequencing projects, such as exome, bulk and single-cell RNA sequencing (RNA-Seq), and other methods (Table 2). For example, researchers can pair exome sequencing with transcriptome sequencing to assess whether identified variants alter transcript expression. A wide range of customizable Illumina targeted resequencing solutions are also available to validate variants discovered from any sequencing application.

Delivers “true coding variant” calls

A true coding variant is an accurate base call that differs from the consensus sequence within a coding region. It is not a false positive (where a variant is called but does not truly exist) or a false negative (where a variant that truly exists is not called). A system with a high false positive call rate requires extensive downstream validation, increasing costs and experimental time. A system with a high false negative call rate is failing to detect potentially important findings, often in regions that are highly repetitive or that contain homopolymer stretches. Obtaining true coding variant calls is a function of high-quality library preparation and enrichment, sequencing accuracy, and secondary analysis accuracy.

Table 2: NextSeq 1000 and NextSeq 2000 sequencing applications

Application	NextSeq 1000/2000 P1 Reagents		NextSeq 1000/2000 P2 Reagents		NextSeq 2000 P3 Reagents[a]	
	No. samples	Time	No. samples	Time	No. samples	Time
Small whole-genome sequencing (300 cycles) 130 Mb genome; > 30× coverage	7	~19 hr	30	~29 hr	90	~48 hr
Whole-exome sequencing (200 cycles) 50× mean targeted coverage; 90% targeted coverage at 20×	4 (300 cycles)	~19 hr	16	~21 hr	48	~33 hr
Single-cell RNA-Seq (100 cycles)[b] 5K cells, 20K reads/cell	1[c]	~10 hr–	4	~13 hr	11	~19 hr

a. P3 Reagents are available for the NextSeq 2000 system only.
b. Recommended sequencing depth will largely depend on sample type and experimental objective and will need to be optimized for each study.
c. P1 reagents are a good option for single-cell quality control experiments.

Industry-leading SBS read quality

By harnessing industry-leading Illumina NGS technology and the latest advances in SBS chemistry, the NextSeq 1000 and NextSeq 2000 Sequencing Systems deliver sequencing accuracy of at least 85% of sequenced bases higher than Q30† at 2 × 100 bp.[6] The systems can accurately sequence even highly difficult regions (eg, GC-rich regions or homopolymers), yielding a high percentage of true coding variants. The low false positive and false negative rates drastically reduce the time and cost of downstream validation. By offering exceptional data quality, the NextSeq 1000 and NextSeq 2000 Sequencing Systems offer the ideal option for comprehensive study of the exome.

Using the same SBS chemistry that powers all Illumina sequencing systems, the NextSeq 1000 and NextSeq 2000 Sequencing Systems enable researchers to compare and integrate data generated across systems. For example, NextSeq 1000 and NextSeq 2000 exome sequencing data can be integrated with data from follow-up studies performed with targeted panels or large-scale exome sequencing studies run on a NovaSeq™ 6000 System (Table 3).

Table 3: Illumina exome sequencing throughput by system

Sequencing system	Sequencing reagents	No. of exomes per run
NextSeq 1000 and NextSeq 2000 Systems	P1 300 cycles	4
	P2 200 cycles	16
	P3[a] 200 cycles	48
NovaSeq 6000 System	SP 200 cycles	40
	S1 200 cycles	80
	S2 200 cycles	200
	S4 200 cycles	500[b]

a. P3 Reagents are available for the NextSeq 2000 system only.
b. A maximum of 384 unique dual indexes is available.

† Q30 = 1 error in 1000 base calls or an accuracy of 99.9%

Simplified analysis with the DRAGEN Bio-IT Platform

Labs can perform exome sequencing data analysis using tools from the Illumina DRAGEN Bio-IT Platform, a suite of fast and accurate data analysis pipelines available onboard the NextSeq 1000 and NextSeq 2000 Sequencing Systems. This PrecisionFDA award-winning informatics solution‡ uses optimized, hardware-accelerated algorithms to help users overcome bottlenecks in data analysis and reduce reliance on external informatics experts.

The DRAGEN Enrichment pipeline analyzes output from the NextSeq 1000 and NextSeq 2000 Sequencing Systems and performs accurate variant calling in less than two hours after a sequencing run is complete. The pipeline provides industry-leading accuracy in mapping and small variant calling and is available in Germline and Somatic modes.[7-9] With the onboard DRAGEN Enrichment App, analysis can be set up during run planning, streamlining the workflow from sample to answer. The DRAGEN Enrichment App on BaseSpace Sequence Hub features advanced results visualization and table sorting capabilities packaged in an intuitive interface suitable for both new-to-NGS and experienced users.

Output from the DRAGEN Enrichment pipeline can be directly input into a broad range of available downstream analysis tools in BaseSpace Sequence Hub. Beyond the DRAGEN platform, BaseSpace Sequence Hub includes a growing community of software tools for visualization, analysis, and sharing.

Comprehensive Illumina technical support

Illumina provides a world-class support team comprised of experienced scientists who are experts in library preparation, sequencing, and analysis. This dedicated team includes highly qualified field service engineers (FSE), technical applications scientists (TAS), field applications scientists (FAS), system support engineers, bioinformaticians, and IT network experts, all deeply familiar with NGS and the applications that Illumina customers perform around the globe. Technical support is available via phone five days a week or via online support 24/7, worldwide and in multiple languages.

With this unmatched service and support, Illumina helps users maximize the efficacy of their NextSeq 1000 or NextSeq 2000 Sequencing Systems, train new employees, and learn the latest techniques and best practices.

Summary

The NextSeq 1000 and NextSeq 2000 exome sequencing solution offers a cost-effective, scalable workflow for identifying variants in coding regions. The solution combines the power, speed, and flexibility of the NextSeq 1000 and NextSeq 2000 Sequencing Systems with high-quality library preparation and enrichment options and rapid, user-friendly analysis software. Exome sequencing on the NextSeq 1000 and NextSeq 2000 Sequencing Systems enables researchers to interrogate the exome with high accuracy and efficiency.

Learn more

Exome sequencing, illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/exome-sequencing.html

Illumina DNA Prep with Enrichment, illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-flex-enrichment.html

NextSeq 1000 and NextSeq 2000 Sequencing Systems, illumina.com/systems/sequencing-platforms/nextseq-1000-2000.html

DRAGEN Bio-IT Platform, illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html

‡ The DRAGEN Bio-IT Platform was awarded Best Performance for difficult-to-map regions and Best Performance for all benchmark regions on Illumina sequencing data in the 2020 PrecisionFDA Truth Challenge V2.[8,9]

References

1. RefSeq - NCBI Reference Sequence Database. ncbi.nlm.nih.gov/refseq. Accessed November 1, 2021.
2. GENCODE - The GENCODE Project: Encyclopedia of genes and gene variants. gencodegenes.org/. Accessed November 1, 2021.
3. CCDS - Consensus coding sequences (CCDS) Database. ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi. Accessed November 1, 2021.
4. UCSC Known Genes - University of California, Santa Cruz Genome Browser. genome.ucsc.edu/. Accessed November 1, 2021.
5. ClinVar Database. ncbi.nlm.nih.gov/clinvar. Accessed November 1, 2021.
6. Illumina. NextSeq 1000 and NextSeq 2000 Sequencing Systems Specification Sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-1000-2000-spec-sheet-m-na-00008/nextseq-1000-2000-spec-sheet-m-na-00008.pdf. Accessed November 29, 2021.
7. Illumina. Accuracy improvements in germline small variant calling with the DRAGEN Platform. Accessed October 29, 2021.
8. Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. PrecisionFDA website. precision.fda.gov/challenges/10. Accessed October 29, 2021.
9. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html. Accessed October 27, 2021.

Ordering information

Product	Catalog no.
NextSeq 2000 Sequencing System	20038897
NextSeq 1000 Sequencing System	20038898
NextSeq 1000 to NextSeq 2000 upgrade	20047256
NextSeq 1000/2000 P1 Reagents (100 cycles)[a]	20074933
NextSeq 1000/2000 P1 Reagents (300 cycles)	20050264
NextSeq 1000/2000 P2 Reagents (100 cycles)	20046811
NextSeq 1000/2000 P2 Reagents (200 cycles)	20046812
NextSeq 1000/2000 P2 Reagents (300 cycles)	20046813
NextSeq 2000 P3 Reagents (50 cycles)	20046810
NextSeq 2000 P3 Reagents (100 cycles)	20040559
NextSeq 2000 P3 Reagents (200 cycles)	20040560
NextSeq 1000/2000 Read and Index Primers	20046115
NextSeq 1000/2000 Index Primer Kit	20046116
NextSeq 1000/2000 Read Primer Kit	20046117
Illumina DNA Prep with Exome 2.0 Plus Enrichment, (S) Tagmentation Set B (96 samples, 12-plex)	20077595
Illumina DNA Prep with Exome 2.0 Plus Enrichment, (S) Tagmentation Set D (96 samples, 12-plex)	20077596
Illumina DNA Prep with Enrichment, (S) Tagmentation (16 samples)	20025523
Illumina DNA Prep with Enrichment, (S) Tagmentation (96 samples)	20025524
Illumina Exome Panel (8 enrichment reactions)	20020183

a. For QC analysis.

illumina®

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

M-GL-00479 v2.0