NextSeq™ 2000 Exome Sequencing Solution

A cost-effective, high-coverage exome sequencing solution that delivers highly accurate variant calling.

Highlights

- **Integrated exome sequencing solution**
  Streamlined workflow delivers fast library preparation, push-button sequencing, and simple-to-use data analysis options
- **Comprehensive exome coverage**
  Exonic panel content from Illumina and third-party providers allows for highly uniform coverage of coding regions
- **Accurate variant calling**
  Industry-leading SBS read quality enables true coding variant calling to identify common mutations and rare somatic events
- **Expert Illumina technical support**
  Illumina scientists and engineers provide installation, training, and support for customers throughout the entire workflow

Introduction

The NextSeq 2000 exome sequencing solution delivers a clear, complete view of the exome, enabling researchers to investigate the protein-coding (exonic) regions of the genome. It harnesses industry-leading Illumina sequencing by synthesis (SBS) chemistry, responsible for > 90% of the world's next-generation sequencing (NGS) data.[1] By delivering exceptional data quality and high accuracy, the solution enables identification of true coding variants for a broad range of applications, including population genetics, genetic disease research, and cancer studies. The NextSeq 2000 exome sequencing solution begins with streamlined library preparation and exome enrichment, followed by push-button sequencing and rapid, accurate data analysis (Figure 1). With minimal hands-on time, the NextSeq 2000 exome sequencing solution is a highly flexible, comprehensive method for interrogating the exome efficiently and cost-effectively.

Figure 2: NextSeq 2000 Sequencing System—The NextSeq 2000 Sequencing System harnesses the latest advances in SBS chemistry and streamlines sequencing workflows.

Simple, efficient workflow

The 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 (formally known as Nextera™ Flex for Enrichment). Prepared libraries are loaded on to a flow cell and then onto the NextSeq 2000 System for sequencing (Figure 2). The NextSeq 2000 Sequencing System features dual sequencing output modes 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.

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 researchers to identify true coding variants rapidly. Providing comprehensive exome coverage from only 10 ng

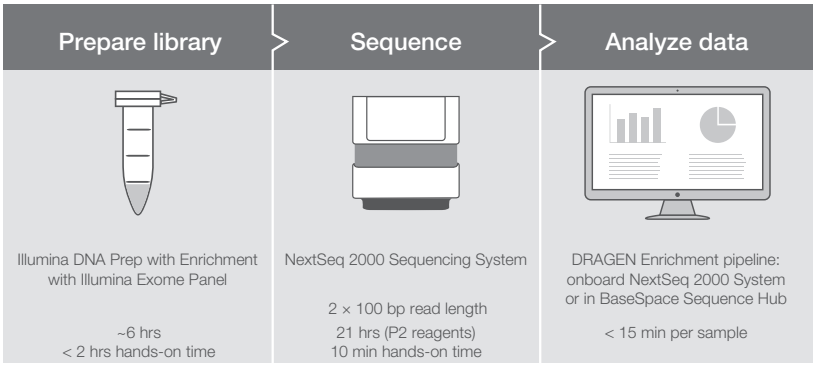

Figure 1: NextSeq 2000 Exome sequencing workflow—The NextSeq 2000 Sequencing System is part of a simple, integrated NGS workflow that delivers highly accurate exome sequencing data. Times vary by experiment and assay type.

of input, it 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 accurately.

Using Illumina DNA Prep with Enrichment, researchers can choose panel content from a variety of vendors, including Illumina, Twist, and IDT (Table 1). This means that researchers can keep 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 and streamlines the workflow to a total time of ~6 hours with < 2 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				
RefSeq[2]	99.83%	99.88%	99.08%	99.45%
GENCODE[3]	98.02%	97.29%	96.01%	96.82%
CCDS[4]	99.99%	99.91%	99.76%	99.67%
UCSC Known Genes[5]	99.89%	98.72%	97.63%	98.13%
ClinVar[6]	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).

The NextSeq 2000 Sequencing System

The NextSeq 2000 Sequencing System provides power and versatility to streamline and simplify the exome sequencing workflow. It takes less than 10 minutes to load and initiate the system. Sequencing is completed in ≤ 33 hours for up to 40 samples using NextSeq 2000 P3 Reagents and paired-end 100 bp read lengths.

The NextSeq 2000 Sequencing System is compatible with a wide range of library preparation kits from Illumina and third parties, and offers cross-application flexibility. Researchers can transition easily between sequencing projects, such as exome, bulk and single-cell RNA sequencing (RNA-Seq), and other methods (Figure 3, 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 both high-quality library preparation and enrichment and sequencing accuracy.

Figure 3: NextSeq 2000 sequencing applications—The NextSeq 2000 Sequencing System enables researchers to transition easily between applications to advance research.

Industry-leading SBS read quality

By harnessing industry-leading Illumina NGS technology[7] and the latest advances in SBS chemistry, the NextSeq 2000 Sequencing System delivers sequencing accuracy of ≥ 80% of sequenced bases over Q30* at 2 × 100 bp. It can accurately sequence even highly difficult regions (eg, GC-rich regions or homopolymers), yielding a high percentage of true coding variants. Its low false positive and false negative rates drastically reduce the time and cost of downstream validation. By offering exceptional data quality, the NextSeq 2000 Sequencing System offers the ideal option for comprehensive study of the exome.

By using the same SBS chemistry that powers all Illumina sequencing systems, the NextSeq 2000 Sequencing System enables researchers to compare and integrate data generated across systems. For example, 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).

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

Table 2: NextSeq 2000 sequencing applications

Application	NextSeq 1000/2000 P2 Reagents		NextSeq 2000 P3 Reagents	
	No. of samples	Time	No. of samples	Time
Small whole-genome sequencing (300 cycles) 130 Mb genome; > 30× coverage	30	~29 hours	75	~48 hours
Whole-exome sequencing (200 cycles) 50× mean targeted coverage; 90% targeted coverage at 20×	16	~21 hours	40	~33 hours
Single-cell RNA-Seq (100 cycles) 4K cells, 50K reads/cell	2	~13 hours	5	~19 hours

Table 3: Illumina exome sequencing throughput by system

Sequencing system	Sequencing reagents	No. of exomes per run
NextSeq 2000 System	P2 200 cycles	16
	P3 200 cycles	40
NovaSeq 6000 System	SP 200 cycles	40
	S1 200 cycles	80
	S2 200 cycles	200
	S4 200 cycles	500[a]

a. A maximum of 384 unique dual indexes is available.

Simplified analysis with the DRAGEN Bio-IT Platform

Exome sequencing data analysis can be performed using tools from the Illumina DRAGEN Bio-IT Platform, a suite of fast and accurate data analysis pipelines, and a broad ecosystem of commercial and open-source data analysis software tools. The Illumina DRAGEN Enrichment pipeline analyzes output from the NextSeq 2000 Sequencing System and performs accurate variant calling in less than two hours after a sequencing run is complete. The DRAGEN Enrichment pipeline uses the same alignment and variant calling modules developed in partnership with the Broad Institute for the DRAGEN-GATK pipeline, which is used for whole-genome sequencing analysis.

The DRAGEN Enrichment pipeline can be launched in BaseSpace Sequence Hub or as an on-instrument workflow using the NextSeq 2000 onboard DRAGEN hardware. The pipeline provides high-quality data packaged in an intuitive user interface. Simple-to-follow prompts guide users through the entire process, from selecting the files generated by the sequencer, to viewing analyzed data and results, including coverage statistics and annotated SNPs and indels, in easy-to-interpret reports. 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.

Complete Illumina technical support

With a NextSeq 2000 Sequencing System in their laboratory, researchers join a worldwide community of thousands of scientists using Illumina technology for their research studies. Illumina schedules community events throughout the year, bringing researchers together to share ideas. User group meetings, scientific symposiums, and blog forums provide venues to discuss new research methods and breakthrough studies.

An integral part of the Illumina community is our dedicated service and support team, consisting of hundreds of people worldwide, 75% of whom have advanced degrees. Illumina technical support begins when the NextSeq 2000 System is delivered, with Illumina scientists and engineers assisting with system installation and setup, and the training of laboratory personnel. They are there 24/7 globally to answer questions every step of the way, giving researchers the peace of mind to focus on their next research study.

As researchers' needs change, new systems are brought into the laboratory, or new methods are undertaken, the Illumina support and training teams are there to provide assistance. In addition to on-site support, training courses (via webinar or at an Illumina facility) are available to bring laboratory personnel quickly up to speed.

Summary

The NextSeq 2000 exome sequencing solution combines the power, speed, and flexibility of the NextSeq 2000 Sequencing System with high-quality library preparation and enrichment options and user-friendly analysis software to provide a clear, comprehensive view of the exome. Exome sequencing on the NextSeq 2000 Sequencing System enables researchers to call true coding variants and gain a deeper understanding of the exome.

Learn more

To learn more about exome sequencing, visit www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/exome-sequencing.html

References

1. Data calculations on file. Illumina, Inc., 2017.
2. RefSeq - NCBI Reference Sequence Database. www.ncbi.nlm.nih.gov/refseq. Accessed June 18, 2020.
3. GENCODE - The GENCODE Project: Encyclopedia of genes and gene variants. www.gencodegenes.org/. Accessed June 18, 2020.
4. CCDS - Consensus CDS (CCDS) Database. www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi. Accessed June 18, 2020.
5. UCSC Known Genes - UCSC Genome Browser. genome.ucsc.edu/. Accessed June 18, 2020.
6. ClinVar Database. www.ncbi.nlm.nih.gov/clinvar. Accessed June 18, 2020.
7. Based on a comparison of the top three industry-leading NGS platforms. Data calculations on file. Illumina, Inc., 2016.

Appendix

Recommended library preparation, enrichment, and sequencing reagents for exome sequencing

Sequencing system and reagents		Catalog no.
NextSeq 2000 System		20038897
NextSeq 1000/2000 P2 Reagents (200 cycles)		20040557
NextSeq 2000 P3 Reagents (200 cycles)		20040560
Library preparation	**Description**	**Catalog no.**
Illumina DNA Prep with Enrichment, (S) Tagmentation, 96 samples	96 samples, 8 12-plex enrichment reactions	20025524
Illumina DNA Prep with Enrichment, (S) Tagmentation, 16 samples	16 samples, 16 1-plex enrichment reactions	20025523
Illumina DNA Prep, (S) Tagmentation, 96 samples	96 samples	20025520
Illumina DNA Prep, (S) Tagmentation, 16 samples	16 samples	20025519
Illumina Exome Panel	8 enrichment reactions	20020183
Indexes	**Description**	**Catalog no.**
IDT for Illumina Nextera DNA Unique Dual Indexes Set A, Tagmentation (96 indexes, 96 samples)	96 indexes, 96 samples	20027213
IDT for Illumina Nextera DNA Unique Dual Indexes Set B, Tagmentation (96 indexes, 96 samples)	96 indexes, 96 samples	20027214
IDT for Illumina Nextera DNA Unique Dual Indexes Set C, Tagmentation (96 indexes, 96 samples)	96 indexes, 96 samples	20027215
IDT for Illumina Nextera DNA Unique Dual Indexes Set D, Tagmentation (96 indexes, 96 samples)	96 indexes, 96 samples	20027216

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