APPLICATION NOTE

NextSeq™ 550 exome sequencing solution

DNA-to-results workflow for robust, high-coverage exome analysis and variant calling

- Exome enrichment enables comprehensive coverage and uniformity across challenging regions
- Tunable sequencing platform with mid- and high-output modes supports faster turnaround times
- Rapid analysis pipeline offers industry-leading accuracy for calling common mutations and rare somatic variants

illumina®

Introduction

The NextSeq 550 exome sequencing solution offers a tried-and-true, cost-effective workflow for investigating the protein-coding regions of the genome. It leverages industry-leading Illumina next-generation sequencing (NGS) technology[1] for comprehensive exome coverage. The NextSeq 550 exome sequencing solution includes integrated library preparation and exome enrichment, push-button sequencing on the proven NextSeq 550 System (Figure 1), and streamlined data analysis with the DRAGEN™ Bio-IT Platform and BaseSpace™ Sequence Hub (Figure 2). This simple workflow delivers highly accurate data to identify exome variants for clinical research applications, such as precision medicine, genetic disease, and inherited cancers. With minimal hands-on time and tunable output, the NextSeq 550 exome sequencing solution is a robust and powerful workflow for efficient exome interrogation.

Figure 1: NextSeq 550 System—Proven platform offering the accuracy of Illumina NGS technology as part of a simplified DNA-to-results workflow.

Simple, efficient exome workflow

The solution begins with library preparation and exome enrichment using the Illumina DNA Prep with Enrichment Kit, Tagmentation and Illumina Exome Panel.* Prepared libraries are loaded onto a reagent cartridge and then onto the NextSeq 550 System for sequencing. The NextSeq 550 System features dual output modes (mid and high) that enable labs to scale exome studies according to their needs. Data analysis, including alignment and variant calling, is easily performed locally or in the cloud with the DRAGEN Enrichment Pipeline or other BaseSpace Sequence Hub apps.

* The open-platform NextSeq 550 System can accommodate library preparation and enrichment kits developed by third-party providers.

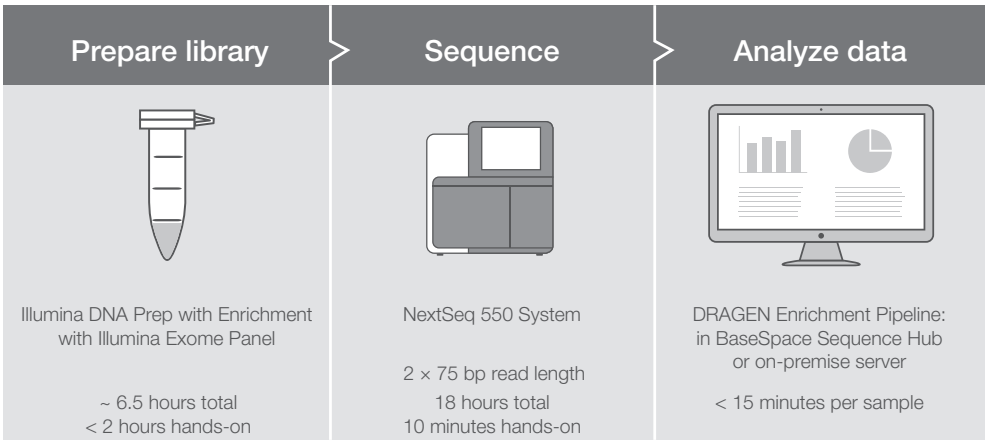

Figure 2: NextSeq 550 exome sequencing workflow—A simple workflow for exome analysis with minimal hands-on time and tunable output. The workflow includes integrated library preparation and exome enrichment, push-button sequencing, and streamlined data analysis.

Comprehensive exome coverage

Illumina DNA Prep with Enrichment combines rapid library preparation using Illumina bead-linked transposome chemistry and exome enrichment, enabling researchers to identify exome variants rapidly. On-bead tagmentation chemistry eliminates the need for mechanical shearing and shortens the workflow to a total time of about 6.5 hours with fewer than two hours hands-on time.

Providing comprehensive exome coverage from 50 ng input, Illumina DNA Prep with Enrichment allows labs to analyze precious DNA samples, while still producing high coverage uniformity and enrichment rates. Users can choose panel content from various vendors, including Illumina, Agilent, Twist, and IDT (Table 1). The highly sensitive detection of low-frequency variants enables labs to identify germline and rare somatic mutations accurately.

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[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).
b. Percentages refer to how much of the databases each exome panel covers.

Exceptional variant call accuracy

Obtaining true coding variant calls is as much a function of sequencing accuracy as it is of high-quality library preparation and enrichment.† At the core of the NextSeq 550 System is proven Illumina sequencing by synthesis (SBS) chemistry, which is used to generate > 90% of the world's sequencing data.[1] NextSeq v2.5 reagent kits are optimized to provide greater stability and robustness, high signal-to-noise ratio, and consistent run-to-run performance.

The NextSeq 550 System delivers industry-leading sequencing accuracy of ≥ 80% of sequenced bases over Q30‡ at 2 × 75 bp. It can successfully sequence even highly difficult GC-rich or homopolymer regions, 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.

Tunable, robust sequencing performance

The NextSeq 550 System streamlines the exome sequencing workflow. It takes less than 10 minutes to load and initiate the system. Sequencing is completed in as few as 18 hours for up to 16 samples using the high-output mode and paired-end 75-bp read lengths. Two flow cell configurations (high- and mid-output) let labs shift easily with each sequencing run to handle different sample numbers and achieve faster turnaround times (Table 2).

> "Recent stat exomes are done in three to four days and that's a significant improvement… It has allowed us flexibility to push certain samples to the front of the line."
>
> —Dr. Mike Friez, Diagnostic Laboratories, Greenwood Genetic Center

† 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).

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

Table 2: NextSeq 550 System sequencing performance[a]

Flow cell configuration	Read length	Output	Run time	Data quality	Throughput[b]	Required input
High-output flow cell Up to 400M single reads Up to 800M paired-end reads	2 × 150 bp	100-120 Gb	29 hours	> 80% bases above Q30 at 2 × 75 bp	16 exomes per run at 50× coverage	Illumina DNA Prep with Enrichment: 10-1000 ng high-quality genomic DNA 50-100 ng FFPE DNA
	2 × 75 bp	50-60 Gb	18 hours			
	1 × 75 bp	25-30 Gb	11 hours			
Mid-output flow cell Up to 130M single reads Up to 260M paired-end reads	2 × 150 bp	32.5-39 Gb	26 hours	> 75% bases above Q30 at 2 × 150 bp	5 exomes per run at 50× coverage	
	2 × 75 bp	16.25-19.5 Gb	15 hours			

a. Run time includes cluster generation, sequencing, and base calling on the NextSeq 550 System. Install specifications are based on Illumina PhiX control library at supported cluster densities (between 129 and 165 K/mm² clusters passing filter). Actual performance parameters may vary based on sample type, sample quality, and clusters passing filter. The percentage of bases > Q30 is averaged over the entire run.

b. Exome calculations performed with 2 × 75 bp read lengths. Based on 50× mean targeted coverage; 90% target coverage at 20×.

Fast, accurate exome analysis

The DRAGEN Bio-IT Platform

Labs can perform exome sequencing data analysis using tools from the DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform on BaseSpace Sequence Hub, the Illumina genomics computing environment. The DRAGEN-Bio-IT Platform uses optimized, hardware-accelerated algorithms for ultrarapid secondary analysis. This PrecisionFDA award-winning informatics solution[§] uses best-in-class pipeline algorithms to help users overcome bottlenecks in data analysis and reduce reliance on external informatics experts.

The DRAGEN Enrichment Pipeline (Figure 3) analyzes output from the NextSeq 550 System to perform alignment and variant calling with speed and high accuracy. Users can launch the DRAGEN Enrichment Pipeline in BaseSpace Sequence Hub or on-premise using a DRAGEN Server. Whether in the cloud or onsite, the pipeline provides industry-leading accuracy in calling multiple variant types, including small variants, indels, and copy number variants (CNVs, germline only).[7-8]

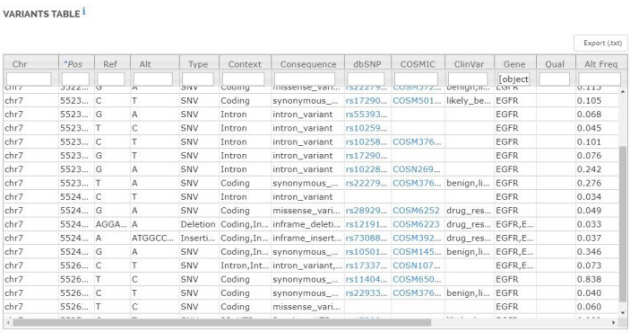

Figure 3: DRAGEN Enrichment Pipeline— Accessible on BaseSpace Sequence Hub or an on-premise server, the DRAGEN Enrichment Pipeline provides industry-leading accuracy for variant calling with advanced visualization in a user-friendly interface.

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.

BaseSpace Sequence Hub app ecosystem

Output from the DRAGEN Enrichment Pipeline can be input directly into a wide selection of downstream analysis tools available in BaseSpace Sequence Hub. Beyond the DRAGEN platform, BaseSpace Sequence Hub includes a growing community of Illumina and third-party bioinformatics tools for visualization, analysis, and sharing.

§ The DRAGEN Bio-IT Platform was awarded Best Performance for difficult-to-map regions and Best Performance for all benchmark regions in the 2020 PrecisionFDA Truth Challenge V2.[8]

World-class service and support

With a NextSeq 550 System in their laboratory, users join a global community of thousands of scientists using Illumina technology. 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 application 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 access online support 24/7, worldwide and in multiple languages.

With this unmatched service and support, Illumina helps users maximize the effectiveness of their NextSeq 550 system, train new employees, and learn the latest techniques and best practices.

Summary

The NextSeq 550 exome sequencing solution offers users a simple and reliable exome sequencing workflow for identifying variants in coding regions. Countless publications and adoption by leading genomics centers demonstrate that the NextSeq 550 System is a reliable addition to any fleet. The streamlined workflow, comprehensive exome coverage, and tunability enable users to efficiently and accurately interrogate more of the exome.

Learn more

To learn more about exome sequencing, visit 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. ncbi.nlm.nih.gov/refseq. Accessed January 11, 2021.
3. GENCODE - The GENCODE Project: Encyclopedia of genes and gene variants. gencodegenes.org. Accessed January 11, 2021.
4. CCDS - Consensus CDS (CCDS) Database. ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi. Accessed January 11, 2021.
5. UCSC Known Genes - UCSC Genome Browser. genome.ucsc.edu. Accessed January 11, 2021.
6. ClinVar Database. ncbi.nlm.nih.gov/clinvar. Accessed January 11, 2021.
7. Illumina. Accuracy improvements in germline small variant calling with the DRAGEN Platform. Accessed January 11, 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 January 11, 2021.

Ordering information

Product	Catalog no.
NextSeq 550 System	SY-415-1002
NextSeq 500/550 Mid-Output v2.5 Kit (150 cycles)	20024904
NextSeq 500/550 Mid-Output v2.5 Kit (300 cycles)	20024905
NextSeq 500/550 High-Output v2.5 Kit (75 cycles)	20024906
NextSeq 500/550 High-Output v2.5 Kit (150 cycles)	20024907
NextSeq 500/550 High-Output v2.5 Kit (300 cycles)	20024908
TG NextSeq 500/550 Mid-Output Kit v2.5 (150 cycles)[a]	20024909
TG NextSeq 500/550 Mid-Output Kit v2.5 (300 cycles)	20024910
TG NextSeq 500/550 High-Output Kit v2.5 (75 cycles)	20024911
TG NextSeq 500/550 High-Output Kit v2.5 (150 cycles)	20024912
TG NextSeq 500/550 High-Output Kit v2.5 (300 cycles)	20024913
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
IDT for Illumina DNA/RNA Unique Dual Indexes Set A, Tagmentation (96 indexes, 96 samples)	20027213
IDT for Illumina DNA/RNA Unique Dual Indexes Set B, Tagmentation (96 indexes, 96 samples)	20027214
IDT for Illumina DNA Unique Dual Indexes Set C, Tagmentation (96 indexes, 96 samples)	20027215
IDT for Illumina DNA Unique Dual Indexes Set D, Tagmentation (96 indexes, 96 samples)	20027216

a. TG-labeled consumables have features that help customers reduce the frequency of revalidation. These consumables are available only under supply agreement and require customers to provide a binding forecast. Contact your account manager for more.

illumina®

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

Pub. no. 770-2013-047-C. QB11178.