DATA SHEET

DRAGEN™ Bio-IT Platform onboard the NextSeq™ 1000 and NextSeq 2000 Systems

Accurate and efficient onboard solution for NGS data analysis

- Provides a streamlined workflow for highly accurate secondary analysis with an easy-to-use graphical interface
- Generates high-quality variant calls directly onboard NextSeq 1000 and NextSeq 2000 systems
- Reduces the need for additional bioinformatics infrastructure and requires fewer touchpoints than other DRAGEN software offerings

illumina®

For Research Use Only. Not for use in diagnostic procedures.

Introduction

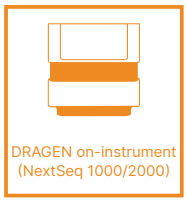
Unlocking the power of the genome is a goal shared by many labs around the world. Recent technology advances are improving access to the next-generation sequencing (NGS) platforms and methods that can help unlock the genome's secrets. NextSeq 1000 and NextSeq 2000 Systems support small- and mid-throughput applications, including, including sequencing exomes, transcriptomes, and small genomes; target enrichment; single-cell sequencing; and shotgun metagenomics.

A major advantage of the NextSeq 1000 and NextSeq 2000 Systems is the inclusion of an onboard implementation of the DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform. Onboard DRAGEN software provides labs with exceptionally fast, accurate, and comprehensive secondary NGS analysis, reducing the need for additional computing infrastructure and bioinformatics resources. With onboard DRAGEN software, secondary analysis can be configured during the sequencing run setup, enabling a streamlined workflow with fewer user touchpoints and a faster turnaround time. For labs looking for a high degree of configurability, Illumina offers additional DRAGEN software implementation options (Figure 1).

The onboard DRAGEN software is the perfect complement to the versatile and scalable NextSeq 1000 and NextSeq 2000 Systems. By including several best-in-class pipeline algorithms the onboard DRAGEN software allows investigators to overcome bottlenecks in data analysis and quickly generate accurate results for a wide range of applications (Table 1).

Efficiency

DRAGEN software onboard the NextSeq 1000 and NextSeq 2000 Systems allows labs to seamlessly integrate sequencing and analysis functions in their run setup. The onboard DRAGEN software enables labs to get FASTQ files and variant call format (VCF) files for downstream applications directly from the instrument as the run completes, saving file transfer and analysis time. Onboard DRAGEN pipeline algorithms help novice and expert users complete common analysis functions and reduce reliance on external informatics experts. Large data files are easily managed using DRAGEN Original Read Archive (ORA) file storage, which achieves an average 4× lossless compression of FASTQ files and uses 80% less energy for a more sustainable workplace.



DRAGEN on-instrument (NextSeq 1000/2000)

- High-quality variant calls directly off the sequencing system, reducing the need for additional bioinformatics infrastructure
- Intuitive graphical user interface
- Streamlined workflow with fewer touchpoints
- Fast turnaround time

DRAGEN on-premise server

- Scalable, high-throughput analysis on local enterprise-grade server
- Versatile command-line interface
- Fully configurable
- Frequent pipeline updates

DRAGEN on BaseSpace Sequence Hub

- Easy-to-use applications for secondary analysis
- Intuitive graphical user interface
- Easy integration with other downstream analysis tools
- Trusted and secure cloud-based platform, reducing the need for additional infrastructure

DRAGEN on Illumina Connected Analytics

- Highly scalable and configurable for higher-throughput customers
- Versatile command-line interface
- Easy integration with other downstream analysis tools

DRAGEN Multi-Cloud bring your own license (Amazon Web Services, Azure)

- Illumina license and software run on cloud provider of choice
- Versatile command-line interface

Figure 1: DRAGEN Bio-IT pipeline implementation options to fit the NGS analysis needs of every lab.

For Research Use Only. Not for use in diagnostic procedures.

Table 1: DRAGEN software applications onboard NextSeq 1000 and NextSeq 2000 Systems

Application[a]	Description
BCL conversion	Converts BCL files produced by Illumina sequencing systems to FASTQ files
DRAGEN ORA compression	Produces lossless, reference-based compression of FASTQ files
DRAGEN FASTQ + MultiQC	Performs hardware-accelerated FastQC metrics with no additional run time
Whole genome	Performs human genome mapping, alignment, and small variant calling. Germline only
Enrichment (including exome)	Performs small variant calling in germline samples, or low-frequency, variant calling in somatic samples
DNA amplicon	Analyzes genetic variation in specific genomic regions. Uses the DRAGEN DNA pipeline with additional step to soft-clip primers and rewrite alignments, ensuring that primer sequences do not contribute to variant calls
RNA	Offers an RNA-Seq (splice-aware) aligner with optional rRNA filtering during alignment, reducing run time and file size
Single-cell RNA	Processes a wide range of single-cell RNA-Seq data sets from reads to cell-by-gene expression matrices
Differential expression	Runs the DESeq2 algorithm on RNA quantification data produced by DRAGEN RNA pipeline. Outputs genes and transcripts that are differentially expressed between two sample groups
NanoString GeoMx NGS	Streamlines analysis for customers using both GeoMx NGS and Illumina instruments for spatial genomics workflows
Methylation	Handles bisulfite and tet-assisted pyridine sequencing (TAPS) methylation data

a. Additional application pipelines are available on BaseSpace Sequence Hub, Illumina Connected Analytics, DRAGEN Multi-Cloud, and DRAGEN on-premise server

Accuracy

DRAGEN analysis pipelines generate exceptionally accurate results. In the 2020 Precision FDA Truth Challenge V2 (PrecisionFDA V2), DRAGEN v3.7 won most accurate results for All Benchmark Regions and Difficult to Map regions for Illumina sequencing data. Further innovations in Graph Genomes and Illumina Machine Learning with DRAGEN v3.10 set a new standard for data accuracy across all sequencing technologies in All Benchmark Regions data set, achieving a 99.83% F1 score, (a calculation of true positive and true negative results as a proportion of total results) (Figure 2). DRAGEN 3.10, Graph, and Illumina Machine Learning also received the top F1 score for most accurate calling compared to all PrecisionFDA V2 submissions in the major histocompatibility complex (MHC) region.[1,2]

The exceptionally efficient DRAGEN Bio-IT Platform achieves rapid analysis times through use of field-programmable gate array (FPGA) technology. The integrated FPGA card, unique to NextSeq 1000 and NextSeq 2000 Systems, allows for hardware-accelerated secondary analysis informatics pipelines for a range of applications. These DRAGEN pipelines are continually improved and additional pipelines are released to deliver the best possible functionality, accuracy, and speed. The platform is designed around the analysis needs of investigators, enabling them to spend less time and effort running production-level pipelines so they can focus more on results.

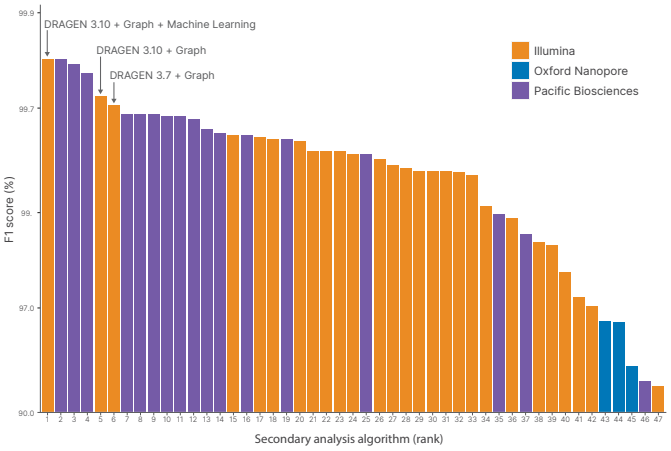


Figure 2: Accuracy of small variant calling achieved with DRAGEN 3.10, Graph, and Machine Learning compared to the PrecisionFDA Truth Challenge v2 submissions in the All Benchmark Regions data set—The DRAGEN 3.10 + Graph + Machine Learning analysis method comes in first place tied with the top performing Pacific Biosciences read submission. DRAGEN 3.10 + Graph shows improvement over the DRAGEN 3.7 + Graph due to graph and reference/alt-contig handling improvements.

Broad application selection

DRAGEN software onboard NextSeq 1000 and NextSeq 2000 Systems features versatile analysis pipelines and can create various output files at different stages of the workflow (Figure 3). The included selection of analysis pipelines supports a wide variety of analysis types types, including but not limited to single-cell, exome, and RNA analysis (Table 1).

Onboard DRAGEN Single-Cell RNA pipelines

The DRAGEN Single-Cell RNA pipelines are fast and scalable solutions that can process a range of single-cell RNA-Seq data sets with easy-to-use output formats, including cell-type clustering plots based on RNA expression activities (Figure 4). Additionally, it features useful extensions for processing multiplexed data sets consisting of several samples (eg, using genotype demultiplexing or cell hashing) and counting the expression of cell-surface proteins. It also supports a range of input library prep types for compatibility with downstream analysis tools.

Onboard QC of single-cell expression libraries and the single-cell analysis pipeline are executed in less time compared to cloud-based solutions and without sacrificing accuracy. The single-cell pipeline goes from run setup to quantified expression per cell on the NextSeq 1000 and NextSeq 2000 Systems from a single touchpoint, reducing the need for additional computing resources.

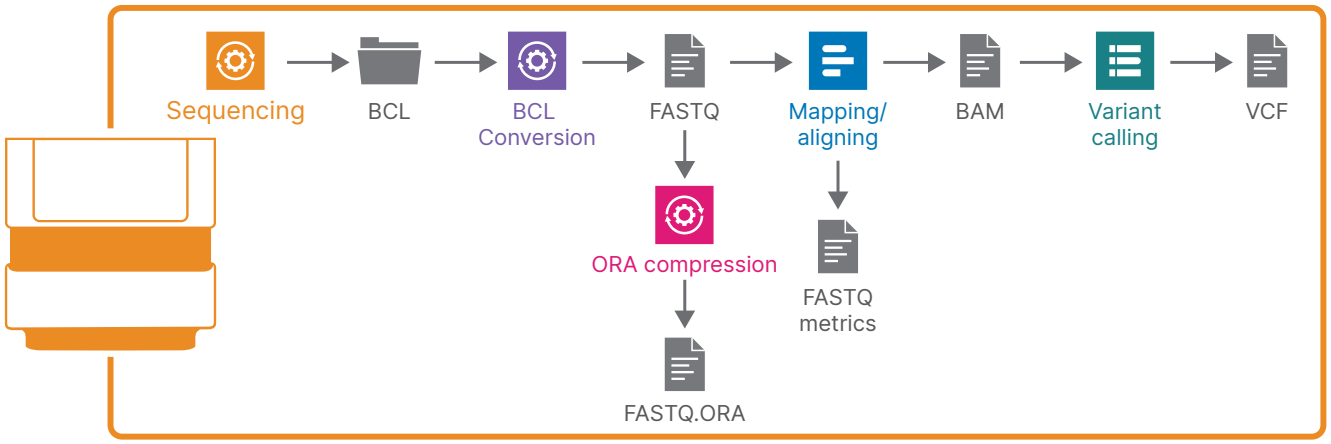


Figure 3: Onboard DRAGEN software on NextSeq 1000 and NextSeq 2000 Systems provides streamlined data processing workflow and secondary analysis—The integrated onboard DRAGEN software solution provides a highly efficient tool for producing the results in the formats labs require for downstream applications.

For Research Use Only. Not for use in diagnostic procedures.

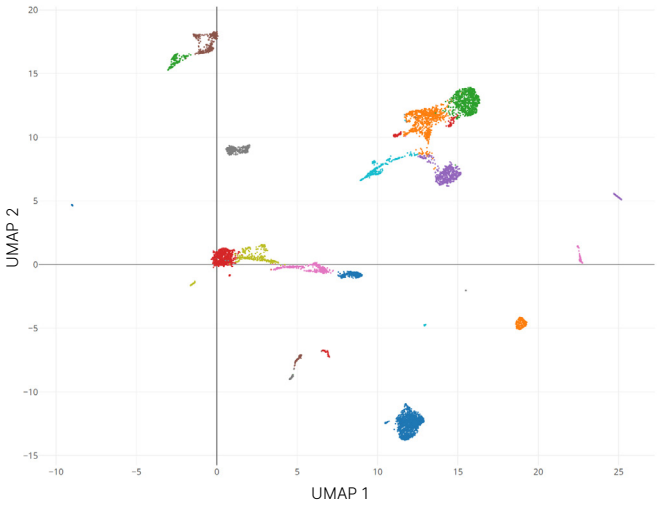


Figure 4: Example cell-type clustering plot from DRAGEN Single Cell RNA pipeline—Uniform Manifold Approximation and Projection (UMAP) analysis allows visualization of individual cells by type or functional states based on complex, multidimensional expression profiles.[4]

Onboard DRAGEN Enrichment pipeline

The onboard DRAGEN Enrichment pipeline provides rapid analysis, including advanced error model calibration for increased accuracy from enrichment and hybridization-based exome panels, and metrics based on the input target. The software can be run in germline or somatic mode, performing germline small variant calling, or somatic (low frequency) variant calling. Copy number variation (CNV) calling and structural variant calling can be enabled in germline mode. The DRAGEN Enrichment pipeline is faster and more accurate than analysis with the Burrows-Wheeler Aligner (BWA) and Genome Analysis Toolkit (GATK) variant caller, particularly in calling indels (Table 2, Figure 5). A flow cell on the NextSeq 1000 and NextSeq 2000 Systems can typically be processed in less than two hours.

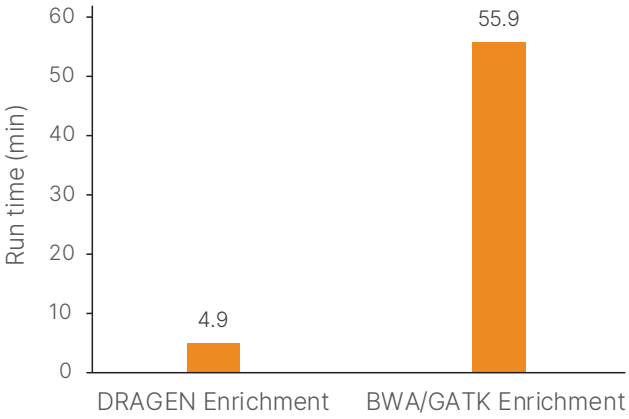


Figure 5: Average run time per sample for the DRAGEN Enrichment pipeline and the BWA Enrichment application on BaseSpace™ Sequence Hub—Germline enrichment analysis of 24 NGS replicates shows faster per sample analysis times for the DRAGEN Enrichment pipeline than the similar analysis using BWA enrichment followed by variant calls with the GATK variant caller.

Onboard DRAGEN RNA pipeline

The onboard DRAGEN RNA pipeline performs secondary analysis of RNA transcripts. The RNA pipeline offers multiple operating modes, including reference-only alignment and annotation-assisted alignment with gene fusion detection. An optional ribosomal RNA (rRNA) filter can be enabled to reduce analysis run time and file size. The gene fusion module leverages the DRAGEN RNA spliced aligner to perform split-read analysis on supplementary (chimeric) alignments in order to detect potential breakpoints, while adding minimal processing time to the overall analysis. Gene fusion detection, transcript quantification, and variant calling can all be enabled using the onboard DRAGEN software features.

Table 2: DRAGEN Enrichment pipeline and BWA/GATK enrichment accuracy results

	SNV			Indel		
	Precision (%)	Recall (%)	F1	Precision (%)	Recall (%)	F1
DRAGEN Enrichment	99.7	94.64	97.11	96.54	83.83	89.74
BWA/GATK Enrichment	99.82	92.77	96.17	96.71	71.32	82.1

Summary

The DRAGEN Bio-IT software onboard NextSeq 1000 and NextSeq 2000 Systems offers easy-to-setup bioinformatics and a simplified user experience in a local implementation. It offers secondary analysis of NGS data generated on the instrument that is faster than DRAGEN pipelines on BaseSpace™ Sequence Hub. Onboard pipelines such as single-cell, enrichment, and RNA offer many speed and accuracy benefits over current offerings.

Ordering information

Product	Catalog no.
NextSeq 1000 Sequencing System	20038898
NextSeq 2000 Sequencing System	20038897
NextSeq 1000 to NextSeq 2000 upgrade	20047256

Learn more

NextSeq 1000 and NextSeq 2000 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

DRAGEN support, support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html

References

1. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Accessed March 14, 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Accessed March 14, 2022.
3. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498. doi:10.1038/ng.806
4. McInnes, L, Healy J, and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv. preprint arXiv:1802.03426. Revised Sep 18, 2020. Accessed May 5, 2022. doi.org/10.48550/arXiv.1802.03426

illumina®

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

© 2022 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html.
M-GL-00729 v1.0