DRAGEN™ Bio-IT Platform

Accurate, comprehensive, and efficient secondary analysis for NGS data

- Accurate results
- Comprehensive applications
- Efficient analysis

illumına®

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

M-GL-00680 v2.0 | 1

Introduction

Unlocking the power of the genome through next-generation sequencing (NGS) is critical to advances in research and medicine. To maximize genetic insights from NGS, researchers require data analysis tools that can accurately and efficiently translate raw sequencing data into meaningful results. Furthermore, to harness the benefits of NGS, organizations require easy-to-use solutions that accommodate a range of users and have lower financial and technical barriers to adoption.

The Illumina DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform was developed to address important challenges associated with analyzing NGS data for a range of applications, including genome, exome, transcriptome, and methylome studies. The DRAGEN platform is a secondary analysis software suite that processes NGS data and enables tertiary analysis to drive insights. The available tools make up a highly accurate, comprehensive, and efficient solution that enables labs of all sizes and disciplines to do more with their genomic data.

Accurate results

The DRAGEN Bio-IT Platform generates exceptionally accurate results. In the 2020 Precision FDA Truth Challenge V2 (PrecisionFDA V2), DRAGEN v3.7 won most accurate in All Benchmark Regions and Difficult to Map regions for Illumina sequencing data.[1,2] Innovations in Graph Genomes and Illumina Machine Learning with DRAGEN v3.10 set a new standard for data accuracy across all sequencing technologies, achieving a 99.83% F1 score in All Benchmark Regions (Figure 1).[1,2] 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) regions.
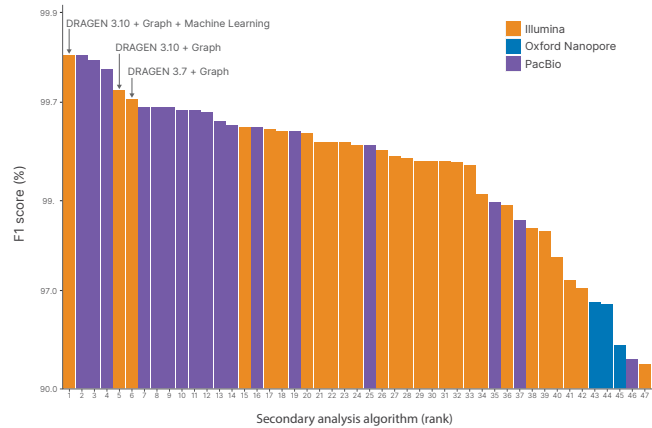

Figure 1: Accuracy of DRAGEN 3.10, Graph and Machine Learning compared to the PrecisionFDA Truth Challenge v2 submissions in the All Benchmark Regions data set— The DRAGEN 3.10 + Graph + Machine Learning analysis method was awarded first place, tied with the PacBio read submission. DRAGEN 3.10 + Graph shows improvement over the DRAGEN 3.7 + Graph due to graph and reference/alt-contig handling improvements. The Y-axis, F1 score (%) is a calculation of true positive and true negative results as a proportion of total results.

Comprehensive coverage

With the included suite of variant callers—repeat expansion, structural variation (SV), copy number variation (CNV), ExpansionHunter, and targeted callers such as *SMN*, *CYP2D6,* and *HLA*)—DRAGEN software offers broad genomic coverage. In addition, DRAGEN Graph with parallel compute architecture effectively lengthens Illumina reads and reaches low-complexity regions, resolving areas of the genome that are difficult to assess due to repeat sequences. This improves coverage of potentially medically relevant genes and enables single nucleotide, copy number, and structural variant calling in difficult-to-map regions.

Efficient analysis

The DRAGEN Bio-IT Platform is specifically designed to give labs the data analysis speed and file options they need to get the greatest benefits from NGS data sets. DRAGEN software is hardware accelerated and uses field-programmable gate array (FPGA) architechure to

achieve rapid turnaround times. The efficiency of DRAGEN analysis algorithms resulted in two world speed records for genomic data analysis.[3,4] In practical applications, the on-premise DRAGEN Bio-IT Platform can process NGS data for a whole genome equivalent at 34× coverage in ~36 min vs. > ~8 hr with a traditional CPU-based system.

To deal with the storage demands of large NGS data files, DRAGEN Original Read Archive (ORA) technology provides lossless 5× compression of FASTQ files. The lossless compression of DRAGEN ORA maintains the details of FASTQ files and is remarkably fast, requiring ~8 minutes for compression of 50 GB to 70 GB FASTQ files.

FPGA and hardware-acceleration

The highly configurable FPGA is essential for the efficiency of the DRAGEN Bio-IT Platform. FPGA programing allows for ultraefficient hardware-accelerated implementations of genomic analysis algorithms, such as base call (BCL) file conversion, mapping, alignment, sorting, duplicate marking, and haplotype variant calling. The flexible nature of FPGAs enables Illumina to develop an extensive suite of DRAGEN application pipelines, with continual updates and additions to deliver the best possible accuracy, comprehensiveness, and efficiency.

Support for multiple applications

The DRAGEN Bio-IT Platform meets the needs of labs performing a wide range of NGS applications. Features include versatile pipelines that can accept input data files and create output files at different stages of the pipeline (Figure 2). The selection of DRAGEN secondary analysis pipelines supports a wide variety of experiment types, including whole-genome sequencing, enrichment panels, single-cell RNA, bulk RNA, and methylation analysis (Table 1).

Custom references

The DRAGEN Reference Builder, also referred to as a hash table, can be used to generate a nonhuman or nonstandard reference. Created references can be used as input for all DRAGEN apps that support customer reference files. The DRAGEN Reference Builder requires a FASTA file. Most DRAGEN pipelines include built-in support for hg19, hg238 (with or without HLA*), GRCh36, and Hs37d5.
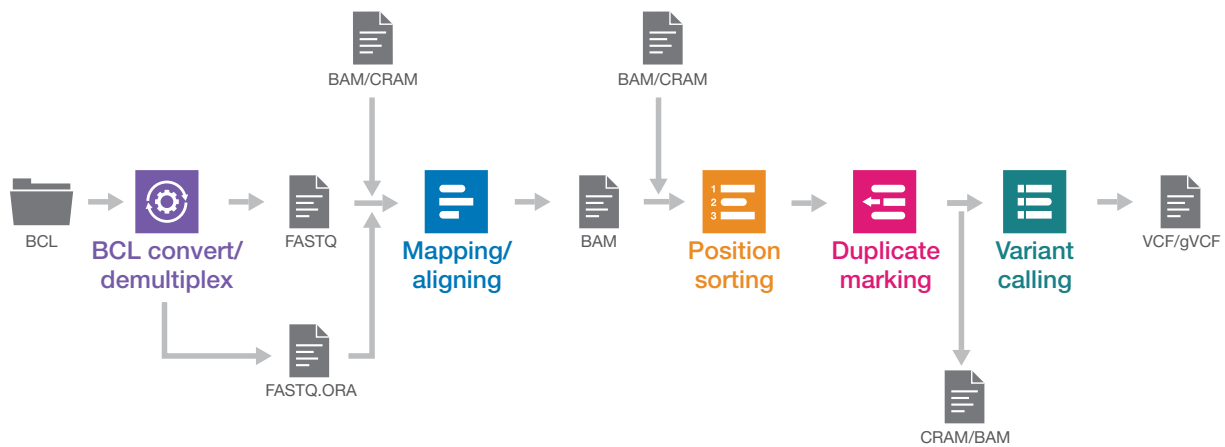
---

* HLA: human leukocyte antigen

Figure 2: Flexibility of DRAGEN pipelines—Each DRAGEN pipeline contains a specific set of steps to support accurate and efficient analysis. The DRAGEN pipeline provides the flexibility to accept various input files and produce a range of output types, enabling users to customize their experience and produce their desired file format.

Table 1: The DRAGEN Bio-IT Platform supports an extensive array of secondary analysis applications

Application	DRAGEN On-premise Server	NextSeq 1000 NextSeq 2000	BaseSpace Sequence Hub	Illumina Connected Analytics		Multi-Cloud bring your own license
				Preconfigured	Build your own	
BCL conversion	✔	✔	✔		✔	Not recommended
DRAGEN ORA compression	✔	✔			✔	✔
DRAGEN FASTQ + MultiQC	✔	✔	✔	✔	✔	✔
Whole genome	Germline + somatic	Germline only	Germline + somatic	Germline + somatic	Germline + somatic	Germline + somatic
Enrichment (including exome)	Germline + somatic	Germline + somatic (no CNV)	Germline + somatic	Germline + somatic	Germline + somatic	Germline + somatic
DNA amplicon	✔	✔	✔	✔	✔	✔
RNA	✔	✔	✔	✔	✔	✔
Single-cell RNA	✔	✔	✔	✔	✔	✔
Differential expression		✔	✔			
nanoString GeoMx NGS		✔	✔			
RNA amplicon	✔		✔	Coming soon	✔	✔
Methylation	✔		✔	✔	✔	✔
Metagenomics			✔			
RNA pathogen			✔			
COVID	COVIDSeq, COVID Lineage	Cloud only COVIDSeq	COVIDSeq, COVID Lineage	COVIDSeq, COVID Lineage	COVIDSeq, COVID Lineage	
TruSight Oncology 500 ctDNA	ctDNA available, solid coming soon			ctDNA available, solid coming soon		

Scalability

The DRAGEN Bio-IT Platform enables labs to scale operations as needed while keeping costs and turnaround times low. DRAGEN software can facilitate the expansion of research capabilities in several ways:

1. **Keeping up with the NovaSeq™ 6000 System**: A single DRAGEN server can demultiplex a NovaSeq 6000 System run using an S4 flow cell in less than two hours.

2. **Burst capacity**: During times of increased workloads with high sample volumes, labs can access additional capacity through available parallel DRAGEN platform access options (Figure 3).

3. **Expanding operations**: A single DRAGEN platform can be used to run all DRAGEN pipelines and supported sample types. The accuracy, comprehensiveness, and efficiency of DRAGEN tools enable users to scale up operations without compromising turnaround times or quality of results.

4. **Exomes to genomes**: Ramping from whole-exome sequencing (WES) to whole-genome sequencing (WGS) involves a large increase in generated data. DRAGEN software enables customers to perform analyses from exomes to genomes without large investments in additional hardware infrastructure or cloud-based solutions.
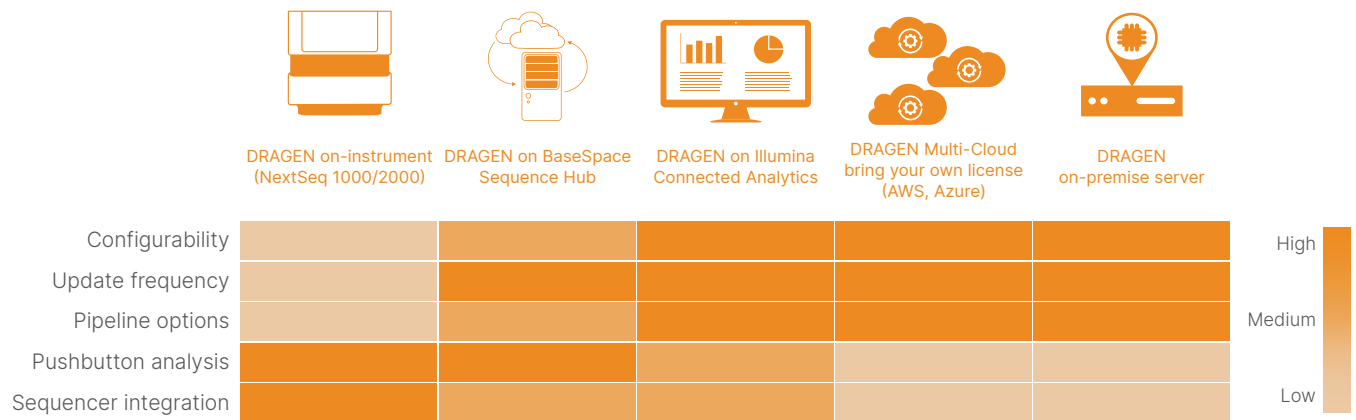
For Research Use Only. Not for use in diagnostic procedures.

Figure 3: DRAGEN Bio-IT pipeline access options with features designed to fit the NGS analysis needs of every lab.

5. **Very large data sets:** The DRAGEN platform offers a simplified workflow for large-scale cohort analysis, featuring multiple pipelines that are used in conjunction to call small and large variants with high accuracy from a cohort sampling. DRAGEN software enables aggregation and genotyping of thousands to millions of genomic variant call format (gVCF) files and aggregates new batches without reprocessing existing batches. The DRAGEN Joint Genotyping Pipeline calls variants jointly across multiple genomes and scales to large cohorts with rapid analysis and uncompromising accuracy.[5] Deployment of the DRAGEN platform on the 1000 Genomes Project data, for example, enabled large scale, accurate variant calling of diverse samples and identification of regions where coverage data are nonuniform or deviate from the assumptions.

Multiplatform accessibility

The suite of DRAGEN Bio-IT pipelines can be accessed through available on-premise, on-instrument, or cloud solutions, enabling labs to select a solution that best suits their needs.

DRAGEN on-premise server

DRAGEN on-premise relies on a local storage solution to collect and store NGS data. After raw sequencing data has been transferred from the sequencing instrument to the local storage via a local network connection, it is transferred to the DRAGEN server to perform the selected workflow. Following analysis, the software writes the generated output files back to the local storage location. DRAGEN on-premise server:

- Supports varying levels of command-line interface
- Replaces up to 30 traditional compute instances
- Processes NGS data for an entire human genome at 30× coverage in < 25 min

Onboard DRAGEN Bio-IT Platform on NextSeq™ 1000 and NextSeq 2000 Systems

NextSeq 1000 and NextSeq 2000 Systems include onboard DRAGEN software for rapid, accurate secondary analysis solution. The onboard DRAGEN software suite offers a select set of pipelines to cover a range of common NGS applications (Table 1) with a user-friendly interface that allows expert and nonexpert users to quickly perform needed analyses and produce results. The onboard DRAGEN Bio-IT Platform:

- Provides access to select DRAGEN informatics pipelines
- Enables users to generate results in as little as two hours
- Uses intuitive pipeline algorithms to reduce reliance on external informatics experts

BaseSpace Sequence Hub

The cloud-based DRAGEN suite available on BaseSpace Sequence Hub, combinines accurate, efficient analysis with a secure ecosystem and versatile scalability. DRAGEN Bio-IT software on BaseSpace Sequence Hub enables pushbutton secondary analysis for labs of all sizes and disciplines. BaseSpace Sequence Hub is a direct extension of your Illumina instruments. Encrypted data flow from the instrument into BaseSpace Sequence Hub, enabling you to manage and analyze your data easily with a curated set of applications. BaseSpace Sequence Hub, powered by Amazon Web Services (AWS):

- Offers a pushbutton, easy-to-use solution for DRAGEN analysis.

- Uses an intuitive graphical user interface for efficient operation by expert and nonexpert users.

- Provides access to powerful computing resources without capital expenditure for additional infrastructure.

Illumina Connected Analytics

The DRAGEN Bio-IT Platform on Illumina Connected Analytics is a comprehensive cloud-based data management and analysis platform that empowers researchers to manage, analyze, and interpret large volumes of multiomic data in a secure, scalable, and flexible environment. Illumina Connected Analytics:

- Provides access to the complete DRAGEN Bio-IT suite for, available as prepackaged pipelines or individual tools for custom pipelines.

- Supports highly automated workflows and custom solutions for optimized high-throughput studies.

- Offers a highly secure environment with guaranteed data residency, single sign-on access, audit logs, and access control supporting Health Insurance Portability and Accountability Act compliance and European Union, General Data Protection Regulation principles.

DRAGEN Multi-Cloud bring your own license (BYOL)

Labs with existing cloud-based services can run DRAGEN analysis on AWS or Microsoft Azure by purchasing the DRAGEN Bio-IT Platform as a virtual image. DRAGEN multi-cloud bring your own license:

- Uses the same software and command line options as DRAGEN On-Premise Server (excluding TruSight™ Oncology 500 pipelines).

- Is suited for advanced users looking to integrate DRAGEN pipelines into their existing cloud environment and workflows.

Summary

The DRAGEN Bio-IT Platform is a powerful suite of software tools that provide accurate, comprehensive, and efficient secondary analysis of NGS data. Multiple DRAGEN Platform options allow labs to select the solution that best suits the type and scale of their projects. As NGS technology continues to make progress, timely updates to the DRAGEN Bio-IT Platform ensure the best possible performance of current pipelines, while new pipelines continue to be added as applications become available.

Learn more

Illumina DRAGEN Bio-IT Platform, Illumina.com/dragen

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

Contact us, illumina.com/company/contact-us.html#/ united-states/customer-care

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. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Accessed March 14, 2022.
4. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. https://www.sandiegouniontribune.com/95899028-132.html. Accessed March 14, 2022.
5. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Accessed March 14, 2022.

illumına®

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