

# Illumina DRAGEN™ Bio-IT Platform

Accurate, ultra-rapid secondary analysis.

## Highlights

- Delivers accurate data**  
 Detects small variants with high analytical sensitivity and specificity
- Increases lab efficiency**  
 Processes a whole genome at 30x coverage in ~25 minutes, and a whole exome at 100x coverage in ~8 minutes
- Implements easily and cost-effectively**  
 Reduces hardware investments and consumption costs, with push-button or command-line options
- Supports multiple applications**  
 Supports a variety of applications both on-premise and in the cloud through BaseSpace™ Sequence Hub

As we continue to unlock the power of the genome with new and advanced applications, the amount of data generated from next-generation sequencing (NGS) rapidly expands. In 2018, more than 100 petabytes of data were generated by Illumina systems. To keep up with the vast amount of data, customers require data analysis tools that can efficiently translate raw sequencing data into meaningful results without compromising accuracy or cost. Furthermore, to harness the benefits of NGS, organizations that are new to this technology require easy-to-use solutions that reduce financial and expertise barriers to adoption.

The Illumina DRAGEN Bio-IT Platform is engineered with tight customer collaboration to address the key pain points associated with analysis of NGS data, developing a highly accurate, ultra-rapid secondary analysis solution that meets the needs of both small research labs and population-scale genomic projects.

## About the DRAGEN Platform

The Illumina DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform enables labs of all sizes and disciplines to do more with their genomic data. It provides secondary analysis of NGS data from genomes, exomes, and transcriptomes. Fundamental features of the DRAGEN Platform address common challenges in genomic analysis, such as lengthy compute times and massive volumes of data. Without compromising accuracy, the DRAGEN Platform delivers quickness, flexibility, and cost efficiency.

The DRAGEN Platform is a combined hardware and software solution. It offers a variety of secondary analysis pipelines that are engineered to run on field-programmable gate array technology (FPGAs). FPGAs provide hardware-accelerated implementations of genomic analysis algorithms, including BCL conversion, mapping and alignment, sorting, duplicate marking, and haplotype variant calling.

The DRAGEN Platform produces valuable metrics, including:

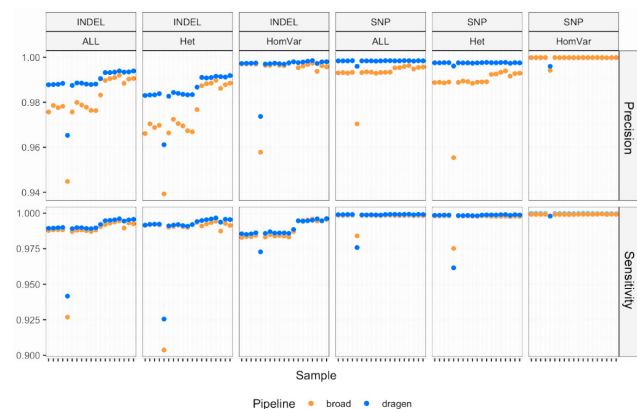
- Library prep quality control (QC)
- Analysis QC
- Demultiplexing
- Duplicate reads
- Raw data processing tools similar to SAM/PICARD

The reprogrammable nature of the DRAGEN Platform enables Illumina to develop an extensive suite of pipelines. Multiple pipelines can be run on a single DRAGEN Server. Engineered from the ground up using optimized software algorithms with hardware acceleration, DRAGEN pipelines are continually improved and additional pipelines are released to deliver added functionality, enhanced accuracy, and speed improvements.

The DRAGEN Platform is available both on-premise and in the cloud through BaseSpace Sequence Hub, and a subset of DRAGEN pipelines are now available on the NextSeq™ 1000 and NextSeq 2000 Sequencing Systems. All DRAGEN pipelines can be version controlled.

## Accurate Results

DRAGEN Platform implementations are built upon world-class algorithms, and stay up to date to adhere to evolving industry standards and best practices. Exceptional analytical sensitivity and specificity are provided for genomic applications and workflows. The DRAGEN Platform is engineered to remove biases and other sources of error to maintain accuracy. Improved algorithms in each new version of the DRAGEN Platform are designed to increase accuracy. A comparison of DRAGEN v3.4 vs GATK 4 by the Broad Institute, published in early 2020, saw overall improvements in sensitivity and specificity (Figure 1).<sup>1</sup>



**Figure 1: High accuracy in single nucleotide variant detection**—For variant calling comparison with a popular variant calling platform, reference genomic DNA was sequenced using three Illumina library preparation kits and separately analyzed through DRAGEN v3.4 or GATK4 for detection of insertions/deletions (indels) and single nucleotide polymorphisms (SNPs).

## Increase lab efficiency

The DRAGEN Platform achieves accelerated turnaround times through its FPGA backbone. In contrast to conventional CPU-based systems that execute lines of software code to perform an algorithmic function, FPGAs implement these algorithms as logic circuits, providing an output almost instantaneously, resulting in significant gains in lab efficiency.

The DRAGEN Platform can process NGS data for an entire human genome at 30x coverage in about 25 minutes and a human exome at 100x coverage in ~8 minutes on-premise, compared with more than 10 hours using a traditional CPU-based system. The DRAGEN Platform was also used to set two world speed records for genomic data analysis.<sup>2,3</sup>

"It is faster, more cost-effective, and more accurate. That's a nice trifecta."

-Alexander Bisignano, CEO, Phosphorous

## Cost-effective solutions

The DRAGEN Platform can reduce on-premise investments in server clusters and utilization of cloud-computing resources. A single on-premise DRAGEN Platform can replace up to 30 traditional compute instances, reducing hardware, maintenance costs, and other expenses, including power consumption and cooling. DRAGEN on BaseSpace Sequence Hub offers the same high-quality DRAGEN pipelines coupled with the flexibility and security of BaseSpace for approximately \$5/genome and \$3/exome (varies based on input sample). DRAGEN pipelines include native compression of aligned reads in the CRAM file format, reducing data footprint by 50% over standard BAM files.

## Supports multiple applications

The DRAGEN Platform features a robust suite of secondary analysis pipelines (Table 1) that support a wide variety of experiment types, including genome, exome, and RNA analysis. Pipelines can accept input files or create output files at different stages of the pipeline (Figure 2) and can be run on a single on-premise server or on BaseSpace Sequence Hub.

For a full listing of DRAGEN pipelines, visit [www.illumina.com/DRAGEN](http://www.illumina.com/DRAGEN).

## Custom References

The DRAGEN Reference Builder, also referred to as a hash table, can be used by the customer to generate a non-human or non-standard reference. Created references can be used as an 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.

Table 1: DRAGEN offerings support a variety of applications

Feature	DRAGEN Server	DRAGEN BaseSpace	NextSeq 1000/2000 System
Demultiplexing (BCL convert).	✓	✓	✓
Map and align	✓	✓	✓
RNA-Seq (gene fusion and quantification)	✓	✓	✓
Exome enrichment (germline and somatic)	✓	✓	Germline only
Whole genome (germline and somatic)	✓	✓	Germline only
Methylation	✓	✓	—
Joint genotyping	✓	✓	—
TruSight Oncology 500 liquid biopsy	✓	—	—
DRAGEN Metagenomics	✓	✓	—

DRAGEN is also available via an Amazon Machine Image (AMI) on AWS Marketplace.

## Easy-to-implement

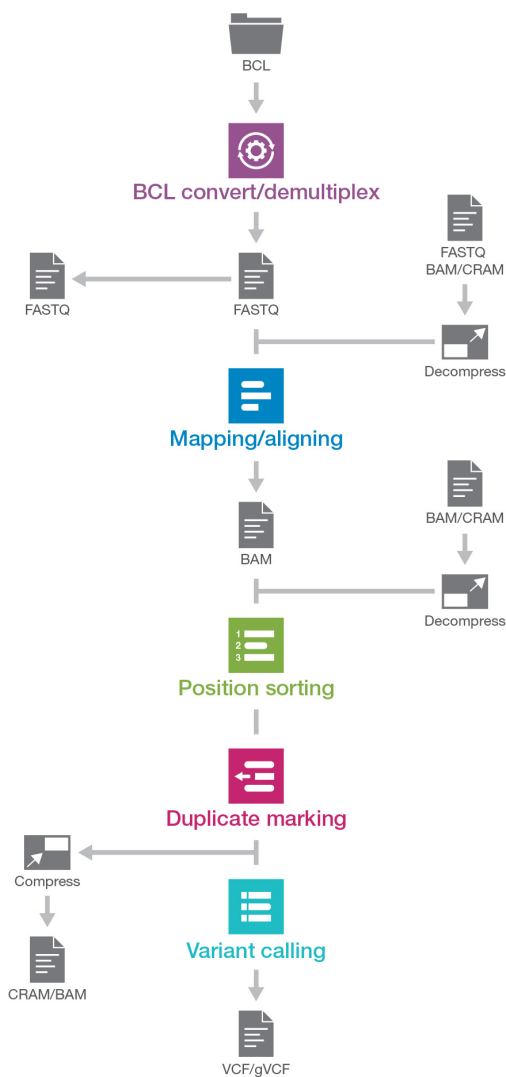
DRAGEN on-premise and cloud offerings via BaseSpace Sequence Hub provide solutions for labs with varying levels of bioinformatics expertise, from simplified push-button analysis to command-line programming (Figure 3). The DRAGEN Server can be easily integrated into new or existing on-premise setups. The DRAGEN Platform does not require additional configurations, and is ready out of the box.

**Push button:** DRAGEN on BaseSpace Sequence Hub makes it easy for labs of varying degrees of informatics expertise to perform secondary analysis in-house at a low cost.

**Command line:** DRAGEN on-premise offers a command line interface which can be used for single-command launch with an easy-to-learn, Linux-based command line interface (CLI) or advanced command line.

"The transition from our existing systems to using the DRAGEN Platform was surprisingly easy. We had to make code adjustments, but it wasn't a major overhaul."

-Kyle Retterer, CIO, GeneDx



**Figure 2: Flexibility of DRAGEN pipelines**—Each DRAGEN pipeline contains a unique set of steps in accordance with its function. Demonstrated by the DRAGEN Germline Pipeline above, DRAGEN provides the flexibility to insert a variety of input files and product a range of output documents, enabling users to customize their experience and produce their desired file format.

BaseSpace Sequence Hub	On Premise	
<b>Push button</b>	<b>Single command launch</b>	<b>Advanced command line</b>
<ul style="list-style-type: none"> <li>Simple graphical user interface (GUI)</li> <li>Managed service</li> <li>HIPAA compliance*</li> <li>Workgroup capabilities</li> <li>Easy data sharing</li> </ul>	<ul style="list-style-type: none"> <li>Easy-to-learn Linux based command line interface (CLI)</li> <li>Simple command line execution</li> </ul>	<ul style="list-style-type: none"> <li>Script back-to-back jobs</li> <li>Make different configuration files for different applications</li> </ul>

**Figure 3: Options for DRAGEN implementation**—In BaseSpace Sequence Hub, users can simply select the app, input info, and start a run. DRAGEN on-premise uses a command line interface. For novice users, an easy-to-learn and operate command line interface can be used. For more advanced users, an advanced command line interface allows for added customization. \*HIPAA compatibility applies in the US only with BaseSpace Enterprise.

## Scalability

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

- 1. Keeping up with the NovaSeq™ 6000 System:** A single DRAGEN server can demultiplex in under two hours with a NovaSeq 6000 instrument with an S4 flow cell.
- 2. Burst capacity:** During times of high capacity with increased sample volumes, labs can scale to DRAGEN on BaseSpace Sequence Hub for burst capacity. The parallel suite of DRAGEN pipelines makes it possible to transfer analysis into BaseSpace Sequence Hub.
- 3. Expanding operations:** A single DRAGEN Platform can be used to run all DRAGEN pipelines and supported sample types. The speed, accuracy, and cost efficiency of DRAGEN 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 enables customers to easily scale from exomes to genomes without large investments in additional hardware infrastructure or cloud-based solutions.

## Available on-premise or via BaseSpace Sequence Hub

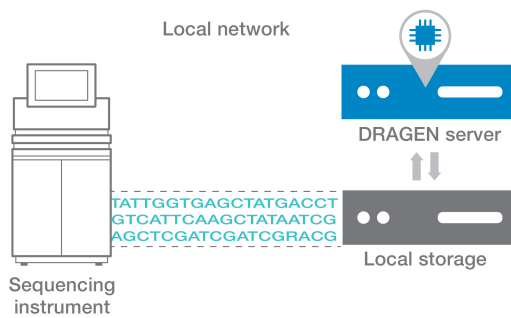
The robust suite of DRAGEN pipelines are available both on-premise and in the cloud via BaseSpace Sequence Hub, enabling labs to use a solution that best suits their needs.

### DRAGEN on-premise

For organizations that wish to keep their analysis local, DRAGEN on-premise offers a robust secondary analysis solution that can be integrated with existing storage solutions (Figure 4). DRAGEN on-premise is ideal for:

- **Keeping data local:** For organizations required to keep data on-premise
- **Limited network connectivity:** In regions with limited or strained connectivity, DRAGEN on-premise can run offline
- **Leveraging existing infrastructure:** DRAGEN on-premise enables labs to utilize their existing storage infrastructure

DRAGEN on-premise relies on a local storage solution to collect and store NGS data. Once the raw sequencing data has been transferred from the sequencing instrument to the local storage via a local network connection, DRAGEN transfers data from storage into the DRAGEN Server to perform the selected workflow, and then writes the generated analysis output files back to the local storage solution. The DRAGEN Server uses a Linux-based CLI that can be configured for single command launch or advanced command line.



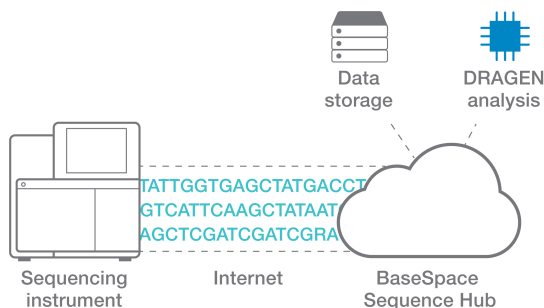
**Figure 4: DRAGEN on-premise solution**—Data is streamed from the sequencing instrument to a local storage solution, and transferred to the DRAGEN server for demultiplexing and secondary analysis. Analysis results are returned back to the local storage solution.

DRAGEN on-premise offers a variety of different licensing tiers, varying from 100,000 to 2,000,000 Gb/year (see Ordering information). Updated versions and new features are released regularly and can be accessed through the DRAGEN customer portal. DRAGEN on-premise installation service is available.

### DRAGEN on BaseSpace Sequence Hub

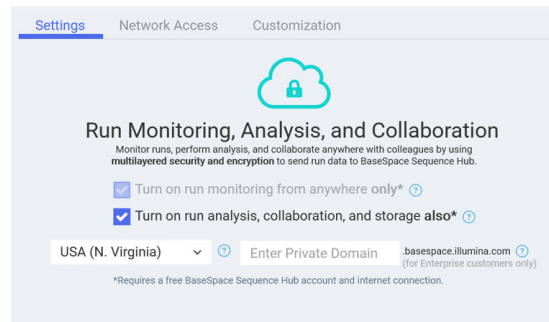
DRAGEN is available in the cloud through BaseSpace Sequence Hub, enabling push-button, rapid, accurate, and cost-effective secondary analysis for labs of all sizes and disciplines. Leveraging Amazon Web Services (AWS) EC2 F1 Instances, DRAGEN on BaseSpace Sequence Hub offers accelerated secondary analysis of genomes, exomes, transcriptomes, and more. DRAGEN on BaseSpace Sequence Hub is ideal for:

- **Ease of use:** Users can stream data directly from their sequencing instrument into BaseSpace Sequence Hub and launch a DRAGEN pipeline with the push of a button
- **Low cost:** With no hardware investments, samples can be analyzed for approximately \$5/genome and \$3/exome
- **Cloud security and compliance:** BaseSpace Sequence Hub is a security-first platform
- **Sharing data:** Projects can be shared securely with collaborators through the cloud
- **Flexibility:** Applications can be used on-demand for small studies or scaled up according to laboratory needs



**Figure 5: DRAGEN in the cloud**—Sequencing data can be transmitted in real time to BaseSpace Sequence Hub, where it is stored and analyzed with selected DRAGEN pipelines.

All DRAGEN pipelines are available in BaseSpace Sequence Hub, combining accurate, accelerated analysis with a secure ecosystem and versatile functionality cloud-based burst scalability. Tight instrument integration enables encrypted data flow directly from the instrument into BaseSpace Sequence Hub for analysis, storage, sharing, and other forms of data management (Figure 5). BaseSpace Sequence Hub connects to the instrument through a wireless internet connection, and can be easily enabled during instrument set-up, or post-setup, through the settings menu on-instrument (Figure 6).



**Figure 6: Easy setup for DRAGEN on BaseSpace Sequence Hub**—Encrypted data flow from instrument to BaseSpace Sequence Hub can be easily setup on-instrument during setup or post-installation via the settings menu.

### Security and compliance in the cloud

BaseSpace Sequence Hub is a security-first platform that has been independently audited and certified for Health Insurance Portability and Accountability Act (HIPAA) compliance, ISO 27001 (Information Security Management System), and ISO 13485 (Quality Management System for Medical Devices), which is available on Enterprise level BaseSpace Sequence Hub accounts. It is built to enable data privacy and is General Data Protection Regulation (GDPR)-ready. BaseSpace Sequence Hub includes end-to-end encryption, auditing, and fine-grained access control. BaseSpace Sequence Hub allows users to update to new versions, revert to older versions, or for labs with a controlled environment, to maintain version consistency (Table 2).

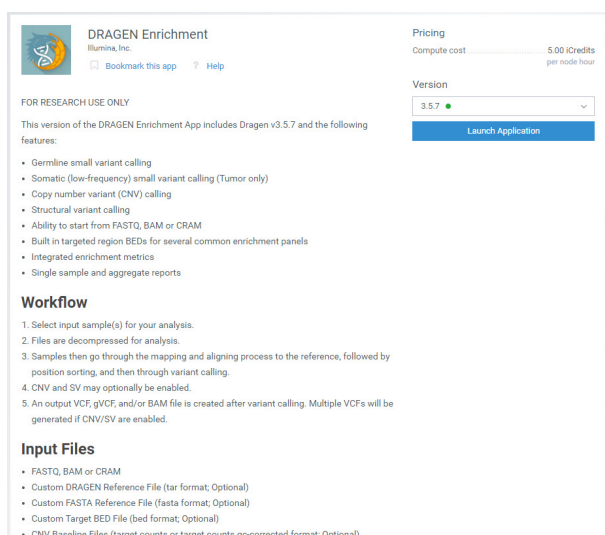
For more details about data security on BaseSpace Sequence Hub, read the [BaseSpace Sequence Hub Security and Privacy white paper](#).

### DRAGEN pipelines in the cloud

All DRAGEN pipelines are made available on BaseSpace Sequence Hub, with new version updates released periodically (Figure 7). DRAGEN can also be deployed via the [Amazon Web Services Marketplace](#).

**Table 2: DRAGEN v3 Server specifications**

Component	DRAGEN v3 appliance
CPU	Dual Intel Xeon Gold 6226 2.7 GHz, 12 cores
Memory	256 GB
Scratch drive	6.4TB NVMe
OS drives	256GB SSD (RAID 1)
FPGA card	DRAGEN
Open PCIe slot	1 x PCIe x16 slot
Form factor	2U
Dimensions	HxWxD 8.8 cm (3.5 in) x 47.8 cm (19 in) x 49.32 cm (19.4 in)
Power supply	1574W 1U Medical grade AC/DC ATX power supply



**Figure 7: DRAGEN pipelines available on BaseSpace Sequence Hub are periodically updated and individual versions are locked down.**

## Learn More

To learn more about the Illumina DRAGEN Bio-IT Platform, email [informatics@illumina.com](mailto:informatics@illumina.com).

Support documentation, including current user guides and install guides, can be accessed through the [Illumina support website](#).

## References

1. DRAGEN-GATK Update: Let's get more specific. [gatk.broadinstitute.org/hc/en-us/articles/360039984151-DRAGEN-GATK-Update-Let-s-get-more-specific](https://gatk.broadinstitute.org/hc/en-us/articles/360039984151-DRAGEN-GATK-Update-Let-s-get-more-specific). Accessed March 16, 2020.
2. Bio IT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. October 23, 2017. [www.bio-itworld.com/2017/10/23/childrens-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed.aspx](http://www.bio-itworld.com/2017/10/23/childrens-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed.aspx). Accessed March 16, 2020.
3. The San Diego Union Tribune. Rady Children's Institute sets Guinness world record. February 12, 2018. [www.sandiegouniontribune.com/news/health/sc-no-rady-record-20180209-story.html](http://www.sandiegouniontribune.com/news/health/sc-no-rady-record-20180209-story.html). Accessed March 19, 2020.

## Ordering information

Product name	Description	Catalog no.	
DRAGEN Server	Includes FPGA chip to accelerate NGS secondary analysis	20040619	
DRAGEN Server Advance Exchange Support Plan	Includes advance exchange for DRAGEN Server; remote technical support (8 x 5)	20032797	
DRAGEN Server Installation		20031995	
Product name	Throughput	Estimated equivalents of 30xWGS	Catalog no.
DRAGEN Level 1 License	100,000 Gb	1000 samples	20027361
DRAGEN Level 2 License	250,000 Gb	2500 samples	20027362
DRAGEN Level 3 License	500,000 Gb	5000 Samples	20027363
DRAGEN Level 4 License	1,000,000 Gb	10,000 Samples	20027364
DRAGEN Level 5 License	2,000,000 Gb	20,000 Samples	20027365

Each license is valid for 1 year.

DRAGEN TruSight Oncology 500 ctdNA requires a separate license.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • [techsupport@illumina.com](mailto:techsupport@illumina.com) • [www.illumina.com](http://www.illumina.com)

© 2020 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](http://www.illumina.com/company/legal.html). Pub. No. 970-2018-002-D QB 6572

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

**illumina®**

970-2018-002-D | 5