illumina®

# Illumina DRAGEN™ Bio-IT Platform

Accurate, ultra-rapid secondary analysis-without compromise.

## Highlights

- **Accurate data**
  Detects small variants with high analytical sensitivity and specificity
- **Ultra-rapid speeds**
  Process a whole genome at 30× coverage in ~25 minutes, and a whole exome at 100× coverage in ~8 minutes
- **Cost-effective and easy to implement solution**
  Reduces hardware investments and cloud-based costs, with push-button or command-line options
- **Robust applications**
  Supports a variety of applications and methods both on-premise and in 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 the 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 will require easy-to-use solutions that reduce the 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 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, enabling labs of all sizes and disciplines to do more with their genomic data.

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

The DRAGEN Platform produces a robust portfolio of metrics, including:

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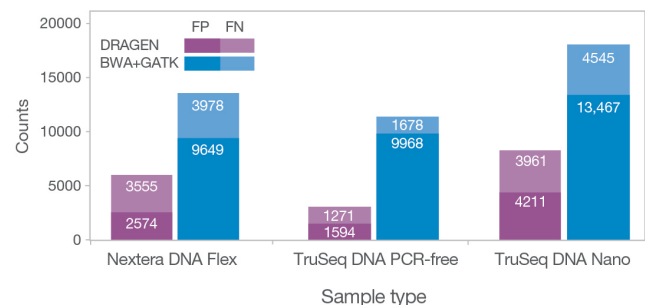- Raw data processing tools similar to SAM/PICARD

The reprogrammable nature of the DRAGEN Platform enables Illumina to develop a robust suite of pipelines, and is beneficial for allowing customers to run multiple pipelines 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 on-premise and in the cloud through Illumina's BaseSpace Sequence Hub, and 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, ensuring that accuracy is maintained across the board. In the 2017 PrecisionFDA Hidden Treasures – Warm Up Challenge, the DRAGEN Platform received the highest score in five out of six accuracy measures for whole-genome variant calling among platforms that recognized all 50 variants.[1] Improved algorithms in each new version of the DRAGEN Platform are designed to increase accuracy (Figure 1).



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 the DRAGEN Platform or BWA+GATK. Resulting variant calls were compared to a reference genome truth set to determine which variants were falsely detected (false positives, FP) or not detected (false negatives, FN).

## Ultra-rapid analysis

The DRAGEN Platform achieves accelerated turnaround times throught 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.

The DRAGEN Platform can process NGS data for an entire human genome at 30× coverage in about 25 minutes and a human exome at 100× 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 (Figure 2)[2,3]



**Figure 2: Record-holding speed for genomic analysis**—The DRAGEN Platform was used by two independent institutions to set speed records in genomic analysis.[1,2]

## 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 $5/genome and $3/exome.*

DRAGEN pipelines include native compression of aligned reads in the CRAM file format, reducing data footprint by 50% over standard BAM files.

\* Approximate cost. Varies based on input sample used
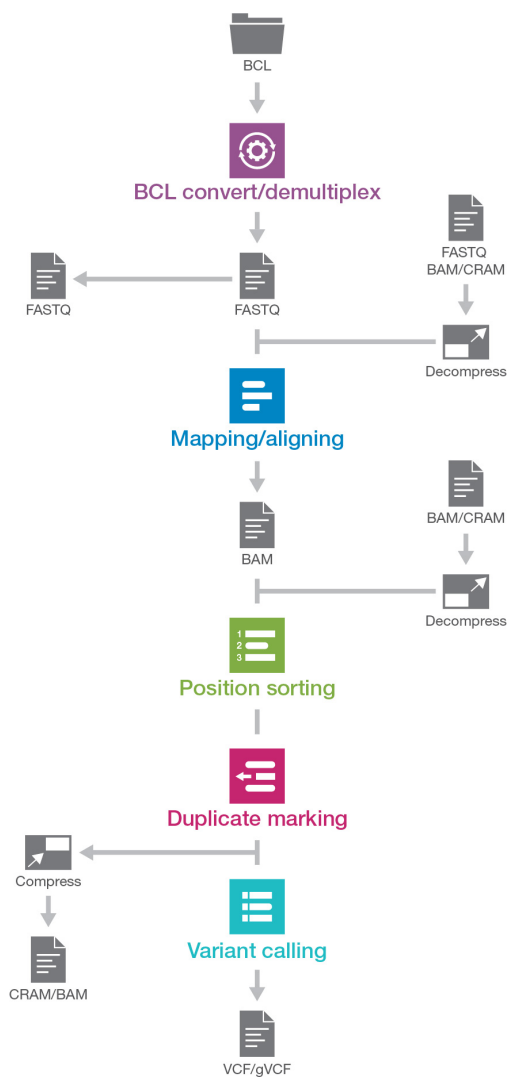
## Robust applications

The DRAGEN Platform features a robust suite of secondary analysis pipelines (Table 1) that support genome, exome, and RNA analysis. Pipelines can accept input files or create output files as different stages of the pipeline (Figure 3), and many 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.

**Table 1: DRAGEN ultra-rapid analysis pipelines support a variety of applications**

| |
|---|
| **Illumina DRAGEN Germline Pipeline**[a] |
| The DRAGEN Germline Pipeline provides end-to-end (BCL → VCF) NGS analysis, including advanced error model calibration for increased accuracy, and repeat expansion detection and genotyping through Illumina Expansion Hunter. |
| **Illumina DRAGEN Somatic Pipeline**[a] |
| The DRAGEN Somatic Pipeline includes tumor-only and tumor–normal modes, designed for detecting somatic variants in tumor samples. |
| **Illumina DRAGEN RNA Pipeline**[a] |
| The DRAGEN RNA Pipeline performs rapid alignment, quantification, splice junction mapping, and fusion detection. |
| **Illumina DRAGEN Joint Genotyping/Population Pipeline**[a] |
| The DRAGEN Joint Genotyping/Population Pipeline calls variants jointly across multiple samples and scales to thousands of samples at expedited speeds with uncompromising accuracy. |
| **Illumina DRAGEN CNV Pipeline**[b] |
| The DRAGEN CNV Pipeline performs copy number variant (CNV) analysis for germline and somatic exomes and genomes. Various levels of filtering can be applied to mitigate false positives before emitting the final calls. |
| **Illumina DRAGEN Methylation Pipeline**[a] |
| The DRAGEN Methylation Pipeline offers multiple operating modes, including reference-only alignment and annotation-assisted. |
| **Illumina DRAGEN Map + Align Pipeline**[c] |
| The DRAGEN Map + Align pipeline is capable of ultra-rapid mapping and aligning DNA and RNA for both exomes and genomes. |
| **Illumina DRAGEN Reference builder**[a] |
| The DRAGEN Reference Builder accepts FASTA files, and builds the proprietary reference used by the DRAGEN apps. |

a. Available on premise and in BaseSpace Sequence Hub.

b. Available only on premise.

c. Available only on premise. To access this function in BaseSpace Sequence Hub, the Germline Pipeline can be used for mapping and alignment, and the exported BAM file can be used for variant calling.

**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.

| BaseSpace Sequence Hub | On Premise | |
|---|---|---|
| **Push button** | **Single command launch** | **Advanced command line** |
| Simple graphical user interface (GUI) | Easy-to-learn Linux based command line interface (CLI) | Script back-to-back jobs |
| Managed service | Simple command line execution | Make different configuration files for different applications |
| HIPAA compliance* | | |
| Workgroup capabilities | | |
| Easy data sharing | | |

**Figure 4: 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 capabilites in several ways:

1) **Keeping up with the NovaSeq™ 6000 System**: A single DRAGEN server can keep up with the output of two NovaSeq 6000 instruments with dual S4 flowcells at full capacity.

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.



**Figure 3: 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.

### 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.

## Easy-to-implement

DRAGEN's on-premise and cloud offerings via BaseSpace Sequence Hub provide solutions for labs with varying levels of bioinformatics expertise - supporting customers from command-line to push-button (Figure 4). The DRAGEN Platform does not require additional configurations, and is ready out of the box.
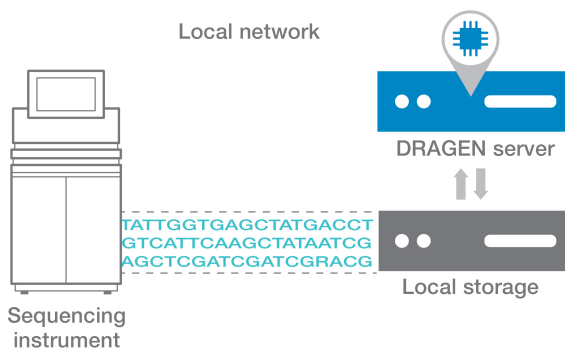
## DRAGEN on-premise

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 5).
DRAGEN on-premise is ideal for:

- **Keeping data local**: For organizations required to keep data on-premise, DRAGEN on-premise is an ideal solution.
- **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 command line interface (CLI) that can be configured for single command launch or advanced command line.



Figure 5: 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.
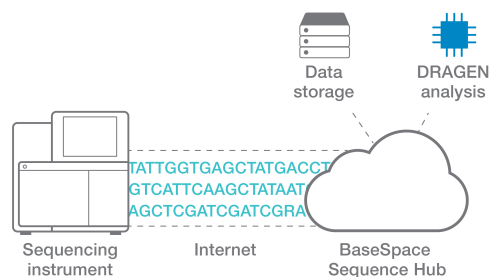
DRAGEN on-premise offers a variety of different licensing tiers, varying from 100,000 to 2,000,000 Gb/year (Table 2. Updated versions and new features are released regularly and can be accessed through the DRAGEN customer portal. DRAGEN on-premise installation service is available.

## DRAGEN on BaseSpace Sequence Hub

DRAGEN is available in the cloud through Illumina's 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 offers accelerated secondary analysis of genomes, exomes, transcriptomes, and more.
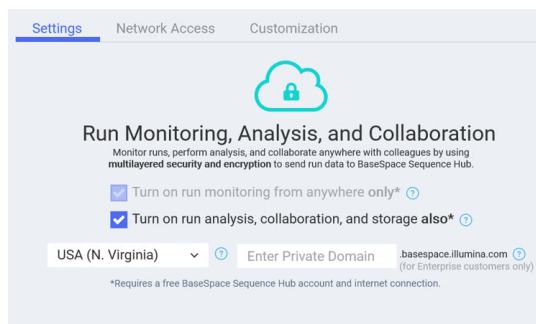DRAGEN on BaseSpace Sequence Hub is ideal for:

- **Ease-of-use**: Users can stream data directly from their sequencing instrument into BaseSpace Sequence Hub and launch the DRAGEN Pipeline with the push of a button
- **Low cost**: With no hardware investments, samples can be analyzed for $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 6: DRAGEN in the cloud—Sequencing Data can be transmitted in real time to BaseSpace Sequence Hub, where it can be stored and analyzed with selected DRAGEN pipelines.

Available within BaseSpace Sequence Hub, DRAGEN in the cloud couples DRAGEN's accurate, accelerated analysis with the secure ecosystem and versatile functionality of BaseSpace Sequence Hub.

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 6). 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 7).



Figure 7: 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 HIPAA compliance, ISO 27001 (Information Security Management System), and ISO 13485 (Quality Management System for Medical Devices)*. It is built to enable data privacy and is 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.

For more details about data security on BaseSpace Sequence Hub, please see the security and compliance white paper.

*Available on Enterprise level BaseSpace Sequence Hub accounts

### DRAGEN Pipelines in the cloud

All DRAGEN pipelines are made available on BaseSpace Sequence Hub, with new version updates released periodically (Figure 8).
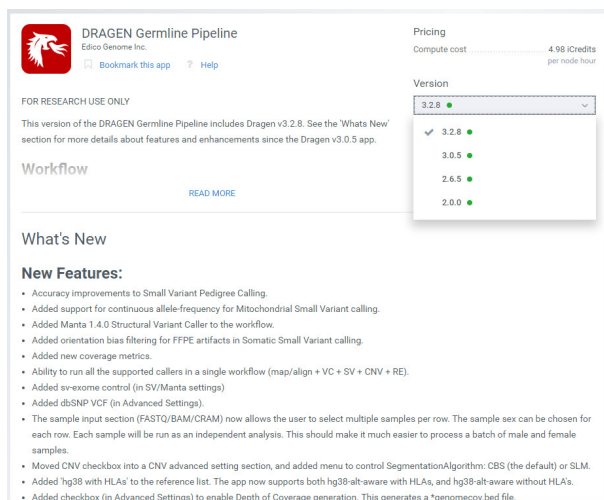


Figure 8: DRAGEN Pipelines available on BaseSpace Sequence Hub are periodically updated and individual versions are locked down.

## Learn More

To learn more about the Illumina DRAGEN™ Bio-IT Platform, email informatics@illumina.com.

Support documentation, including current user guides and install guides, can be accessed through Illumina's support website.

## References

1. Precision FDA Hidden Treasures Warm Up. precision.fda.gov/challenges/1/view/results. Accessed September 14, 2018.
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. Accessed September 19, 2018.
3. The San Diego Union Tribune. Rady Children's Institute sets Guinness world record. February 12, 2018. www.sandiegouniontribune.com/news/health/sd-no-rady-record-20180209-story.html. Accessed September 19, 2018.

Table 2: Ordering information for DRAGEN on-premise

| Product name | Description | Product code |
| --- | --- | --- |
| DRAGEN Server | Includes FPGA chip to accelerate NGS secondary analysis | 20027360 |
| DRAGEN Server Advance Exchange Support Plan | Includes advance exchange for DRAGEN Server; remote technical support ($8 \times 5$) | 20032797 |
| DRAGEN Server Installation | | 20031995 |

| Product name | Throughput | Estimated equivalents of 30×WGS | Product code |
| --- | --- | --- | --- |
| DRAGEN Level 1 License | 100,000 Gb | 1000 samples | 20027361 |
| DRAGEN Level 2 License | 250,000 Gb | 2500 samples | 20027361 |
| DRAGEN Level 3 License | 500,000 Gb | 5000 Samples | 20027361 |
| DRAGEN Level 4 License | 1,000,000 Gb | 10,000 Samples | 20027361 |
| DRAGEN Level 5 License | 2,000,000 Gb | 20,000 Samples | 20027361 |
| Each license is valid for 1 year. | | | |

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