Integrating Illumina DRAGEN™ secondary analysis within your infrastructure

DRAGEN secondary analysis deployment options for end-to-end research workflows

illumına®

Introduction

With advances in next-generation sequencing (NGS) technology, the volume of data generated continues to grow exponentially. This massive influx of data requires fast and efficient analysis methods that do not sacrifice accuracy. Keeping bioinformatics infrastructure current with the vast amount of data and increases in computing power necessary for new applications that require deeper sequencing can be arduous. To address these challenges, Illumina offers DRAGEN secondary analysis. DRAGEN software can be easily integrated into existing high-performance computing (HPC) or cloud-based solutions to enhance operations. This technical note describes some examples of integrating DRAGEN software within typical scenarios.

DRAGEN secondary analysis

DRAGEN software provides accurate, efficient, and comprehensive secondary analysis of sequencing data. This is achieved using highly reconfigurable field-programmable gate array technology (FPGA) to provide hardware-accelerated implementations of specialized genomic analysis algorithms. Some of these applications include BCL conversion, mapping, alignment, sorting, duplicate marking, haplotype variant calling, gene quantification, fusion calling, and variant annotations.

Benefits of DRAGEN software

- Accurate results—DRAGEN achieved a 99.89% accuracy score using the Precision FDA Truth Challenge v2 benchmark data[1]

- Comprehensive solution—analyze whole genomes, exomes, methylomes, and transcriptomes with a single solution that replaces up to 30 open-source tools

- Efficient analysis—process genomes at 40× coverage in ~ 34 min with all supported callers[2]; reduce FASTQ file sizes up to 5× with DRAGEN ORA compression; DRAGEN secondary analysis resulted in two world speed records for genomic data analysis[3,4]

- Cost efficiency—built-in lossless data compression decreases storage costs by up to 80%; preconfigured workflows reduce time and expense for developing analysis pipelines

- Multiplatform accessibility—DRAGEN secondary analysis is available via on-premises server, in the cloud, or directly onboard the NovaSeq™ X Series, NextSeq™ 1000 and NextSeq 2000 Systems, and MiSeq™ i100 Series (Table 1)

- Streamlined integration—easily integrate with Illumina sequencing systems, enabling a streamlined workflow from sequencing to downstream tertiary analysis, including Illumina Connected Insights, Emedgene™, Partek™ Flow™, or Correlation Engine software

Deployment options

Access DRAGEN secondary analysis via an on-premises server, on cloud, or onboard instrument solution. Multiple deployment options support various analysis needs without compromising accuracy, speed, or flexibility.

On-premises DRAGEN server

An on-premises DRAGEN server relies on a local storage solution to collect and store NGS data (Figure 1). The DRAGEN server is a 2U rackmount server containing an FPGA card that is preprogrammed with patented, hardware-accelerated genomics analysis pipelines. New DRAGEN software versions can be downloaded and installed easily. On-premises DRAGEN server:

- Supports flexible configuration of DRAGEN features through a command-line interface

- Replaces up to 30 traditional compute instances

- Processes NGS data for an entire human genome at 40× coverage in ~35 minutes

DRAGEN server on HPC systems

A wide variety of lab types choose to integrate a DRAGEN server into an existing HPC infrastructure as they look to scale or improve operational efficiency. Whether simultaneously replacing an existing fleet of compute servers with a DRAGEN server, or gradually replacing individual servers as they are retired, the integration process is straightforward. Enabling DRAGEN acceleration gains within your organization consists of a few simple steps:

- Add one or more DRAGEN servers to your data center

- Update analysis pipelines to enable use of DRAGEN secondary analysis

Table 1: DRAGEN secondary analysis supports an extensive array of secondary analysis applications[a]

Application	On-premises server	Onboard Illumina sequencing systems			Illumina cloud platforms	
	DRAGEN server	NovaSeq X Series	NextSeq 1000, NextSeq 2000 Systems	MiSeq i100 Series	BaseSpace Sequence Hub	Illumina Connected Analytics
BCL convert	✔	✔	✔	✔	✔	✔
DRAGEN ORA compression	✔	✔	✔	✔		Coming soon
DRAGEN FASTQ + MultiQC	✔	✔	✔	✔ [b]	✔	✔
Whole genome	Germline + somatic	Germline + somatic	Germline + somatic		Germline + somatic	Germline + somatic
Enrichment (including exome)	Germline + somatic	Germline + somatic	Germline + somatic		Germline + somatic	Germline + somatic
DRAGEN Amplicon	✔		DNA only		✔	✔
RNA	✔	✔	✔		✔	✔
Single-cell RNA	✔		✔		✔	✔
NanoString GeoMx NGS			✔		✔	
Methylation	✔	✔			✔	✔
Metagenomics	✔ [c]				✔	
COVID, IMAP, IMAP-FLU					✔	
TruSight Oncology 500 portfolio	✔				✔ [d]	✔
scATAC-seq	✔				✔	✔
Imputation	✔				✔	✔
PGx Star Allele Caller	✔	✔	✔		✔	✔
Illumina Complete Long Reads					✔	✔
RPIP, RVEK, UPIP, VSP	✔				✔	Beta
Small whole genome				✔	✔	

a. Core DRAGEN software version varies across platforms, speak to a local representative for more information.
b. Supported by Library QC app.
c. Metagenomics applications enabled by Kmer classifier, more tools coming soon.
d. Illumina Connected Analytics subscription required.
Abbreviations: IMAP, Illumina Microbial Amplicon Prep; RPIP, Respiratory Pathogen ID/AMR Panel; RVEK, Respiratory Virus Enrichment Kit; UPIP, Urinary Pathogen ID/AMR Panel; VSP, Viral Surveillance Panel.

- Update job scheduler software (eg, SLURM) to include DRAGEN secondary analysis (optional)
- Mount a network-attached storage solution with support for network file system (NFS) and common internet file system (CIFS) protocols for storage devices

Onboard

The NovaSeq X Series, NextSeq 1000 and NextSeq 2000 Systems, and MiSeq i100 Series include onboard DRAGEN software for rapid, accurate secondary analysis. The software is accessed through a user-friendly graphical interface that allows both expert and nonexpert users to perform analyses and produce results quickly. Onboard DRAGEN software offers a select set of pipelines to cover a range of common applications (Table 1) and includes award-winning machine learning (ML) and multigenome (graph) reference analysis for high-quality variant calling. DRAGEN secondary analysis onboard allows for analysis planning with run planning for a streamlined workflow with reduced touchpoints.

For additional scalability, DRAGEN software onboard can be combined with standalone DRAGEN secondary analysis options, either on premises or on cloud. Benefits of combing standalone DRAGEN software with onboard DRAGEN software include an expanded catalog of workflows, time savings, throughput maximization, and the option to reanalyze samples.

On-cloud

Customers can also choose to forego installation and management of a DRAGEN server and use DRAGEN applications for primary and secondary data analysis (Table 1) through Illumina cloud-based options, including, BaseSpace™ Sequence Hub or Illumina Connected Analytics. Customers can also opt to host DRAGEN software on their private cloud, hosted in either Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (Figure 2).
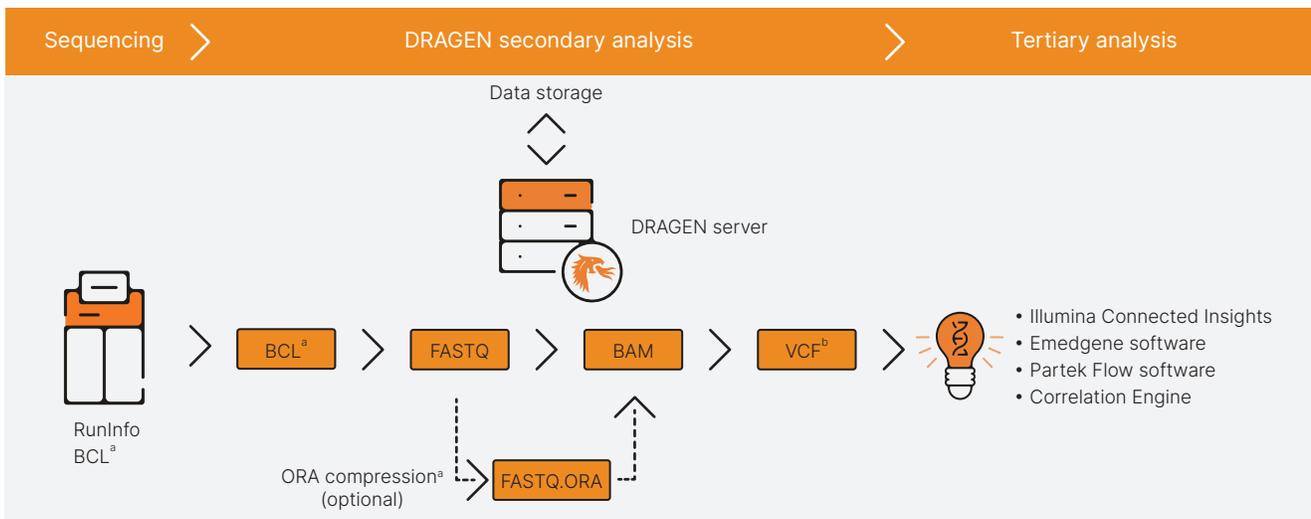
Figure 1: DRAGEN secondary analysis on-premises deployment—DRAGEN servers process data rapidly and minimize network file transfers.

a. BCL convert and ORA compression can be performed onboard the sequencing system or on a DRAGEN server.
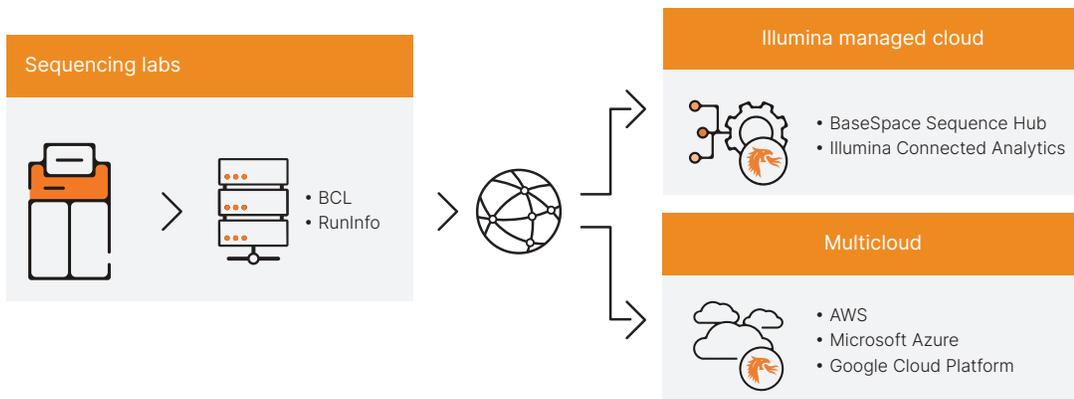b. VCF or other file output option of choice.

Figure 2: DRAGEN secondary analysis cloud deployment—Customers can access DRAGEN applications available in the cloud via BaseSpace Sequence Hub or AWS.

BaseSpace Sequence Hub

Cloud-based DRAGEN secondary analysis applications on BaseSpace Sequence Hub combine accurate, efficient analysis with a secure ecosystem and extensive scalability. DRAGEN 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 flows directly 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):

- Includes a pushbutton, easy-to-use solution for DRAGEN secondary analysis
- Uses a graphical user interface for intuitive and efficient operation by expert and nonexpert users
- Provides access to powerful computing resources without capital expenditure for additional infrastructure

Illumina Connected Analytics

Illumina Connected Analytics is a comprehensive, cloud-based bioinformatics platform that empowers researchers to manage, analyze, and interpret large volumes of multiomic data in a secure, scalable, and flexible environment. Access the DRAGEN secondary analysis suite on Illumina Connected Analytics, available as prepackaged pipelines or select individual tools to incorporate into custom pipelines.

Multicloud

DRAGEN on multicloud is available in regions with FPGA. For details on configuring your instance, refer to the DRAGEN multicloud user guide.

Summary

Integrating DRAGEN secondary analysis into an existing analysis infrastructure allows labs to gain speed, accuracy, efficiency, compatibility, and scalability. There are several options available and capacity can be added to most implementations. Illumina can guide you in choosing the solution that works best for your needs.

Learn more

DRAGEN secondary analysis

References

1. 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. Published January 12, 2022. Accessed October 28, 2024.
2. Data on file. Illumina, Inc. 2024.
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. Published October 23, 2017. Accessed October 28, 2024.
4. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/. Published February 12, 2018. Accessed October 28, 2024.

illumina®

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