

# Integrating Illumina DRAGEN secondary analysis within your infrastructure

An overview of how to  
integrate DRAGEN software  
into different genomic data  
analysis infrastructures



## Introduction

With advances in next-generation sequencing (NGS) technology, the volume of generated sequencing data continues to grow exponentially. This growth requires fast and efficient analytical methods that maintain high standards in accuracy for variant calling. Customers may struggle with updating their existing bioinformatics infrastructure to keep up with the vast amount of data and increases in computing power necessary for analysis of new applications that require deeper sequencing. 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, ultra-rapid secondary analysis of sequencing data. This is achieved by using highly reconfigurable fieldprogrammable gate array technology (FPGA) to provide hardwareaccelerated implementations of common genomic analysis algorithms. Some of these applications include BCL conversion, mapping, alignment, sorting, duplicate marking, and haplotype variant calling. For custom analysis, simply select the settings and produce industry standard alignment files to take advantage of the DRAGEN mapper and variant calling applications.

## Benefits of DRAGEN software integration

- **Increased operational efficiency:** process a whole genome in under 25 minutes and an exome in < 8 minutes, reducing turnaround times.
- **Scalable operations:** run all DRAGEN pipelines with a single DRAGEN server, supporting a vast array of experiment types without needing to invest in new HPC clusters to expand bandwidth and capabilities.
- **Reduced maintenance:** take advantage of quarterly software updates, including new features, improvements, and functionality.
- **Considerable cost savings:** Save with a smaller physical footprint, reduced hardware costs, and diminished power and cooling requirements.

“The transition from our existing systems to using DRAGEN was surprisingly easy. We had to make code adjustments, but it wasn’t a major overhaul.”

-Kyle Retterer, Chief Innovation Officer, GeneDx

## HPC DRAGEN server integration

A wide variety of lab types choose to integrate a DRAGEN server into an existing HPC as they look to scale or improve operational efficiency. Whether replacing an existing fleet of compute servers with a DRAGEN server simultaneously, or gradually replacing individual servers as they are retired, the integration process is straightforward. The DRAGEN server is a 2U rackmount server containing an FPGA card pre-programmed with patented, hardware-accelerated genomics analysis pipelines. New DRAGEN versions can be downloaded and installed easily on demand. Enabling DRAGEN acceleration gains within your organization consists of a few simple steps:

- Add one or more DRAGEN servers to your data center
- Update your analysis pipelines to enable use of DRAGEN secondary analysis
- Update your job scheduler software to include DRAGEN secondary analysis (optional)

## Common IT infrastructures

DRAGEN servers can be implemented in a central, regional, or cloudbased infrastructure.

### Centralized DRAGEN on-site deployment

Common academic infrastructures typically consist of individual labs, each with their own fleet of sequencing systems, temporarily storing raw sequencing output data, ie, BCL or cBCL files, on cache storage, such as a Network Attached Storage (NAS), proximal to their instruments (Figure 1). The DRAGEN server nodes are centrally managed as part of an institutional compute cluster and analyses are queued and managed through a workload manager such as SLURM.

### Regional DRAGEN on-site deployment

Common core lab or translational lab infrastructures consist of each sequencing lab purchasing and using their own fleet of DRAGEN servers for rapidly processing

sequencing data from BCL or cBCL files to VCF or gVCF files (Figure 2). Sequencing core labs can use the DRAGEN Platform to perform data quality control (FastQC) and read trimming (poly-G trimming or Trimmomatic) themselves, without transferring FASTQ data to a bioinformatics team or a compute facility. This reduces inefficient back-and-forth communication between teams due to flagged QC failures. These analyzed data can then be uploaded to a central compute facility for downstream analysis, variant annotation, data sharing, and alignment against other genome references.

(API) connectivity between the DRAGEN server and the Illumina Analytics Platform (IAP) customers can customize pipelines in the cloud for sharing across networks. As in the regional DRAGEN deployment scenario, secondary analysis is accomplished with on-premise DRAGEN servers, and the resulting data are transferred to another resource for further processing. In a hybrid cloud scenario, this usually involves a platform-as-a-service provider such as BaseSpace™ Sequence Hub. Cloud services offer a richer experience for data management, such as easy data archiving and integration with thirdparty partners and vendors for downstream tertiary analysis.

### Hybrid cloud infrastructures

Some customers currently split analysis and storage functions between on-premise and cloud-based functions (Figure 3). Using Application Programming Interface

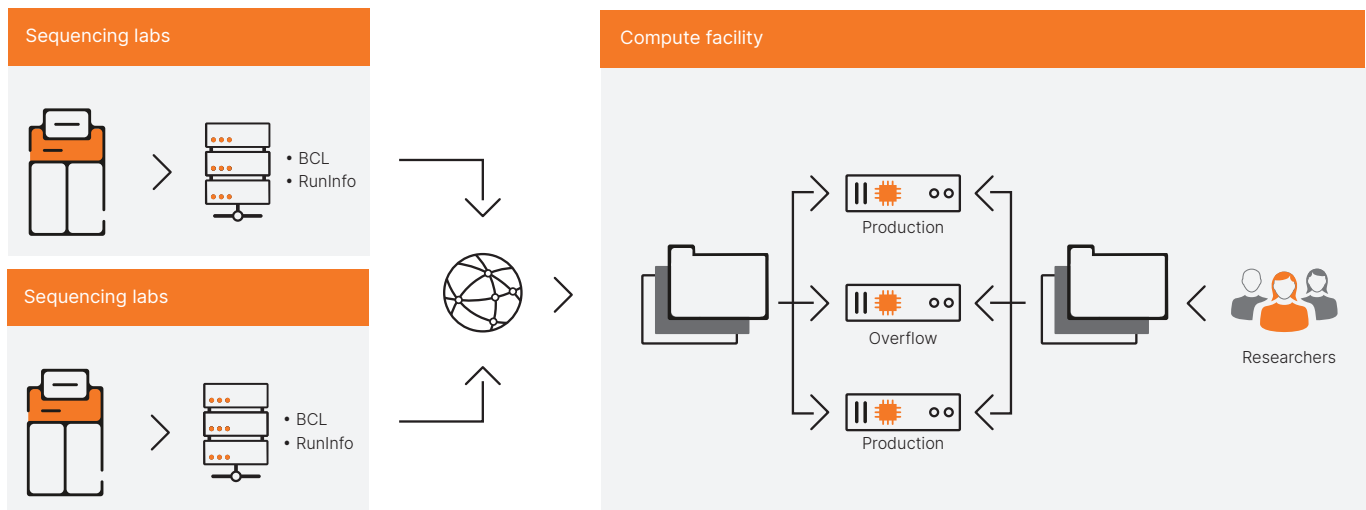


Figure 1: Centralized DRAGEN secondary analysis onsite deployment —Individual labs, each with their own fleet of sequencers, can temporarily store raw sequencer output data on cache storage proximal to their instruments. Data are transferred to a centrally-managed compute and storage infrastructure with DRAGEN Servers. Work queues are managed to maintain production service levels with fail-over capabilities.

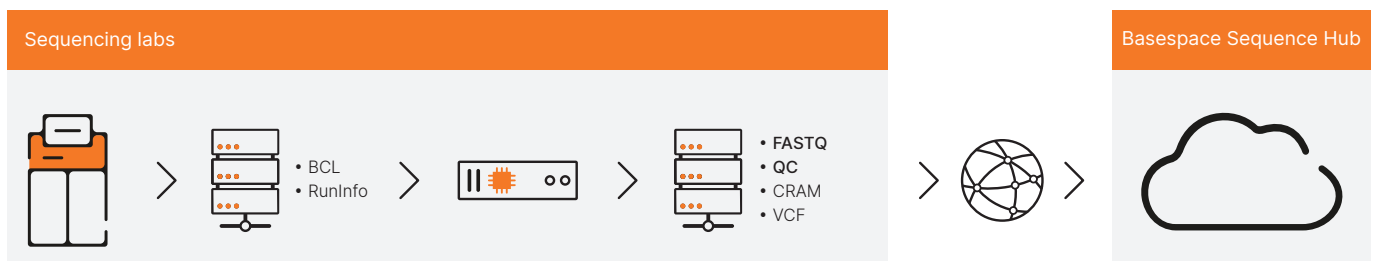


Figure 2: Regional DRAGEN secondary analysis onsite deployment —Core labs can locate DRAGEN servers within the sequencing facility to process data rapidly and minimize network file transfers.

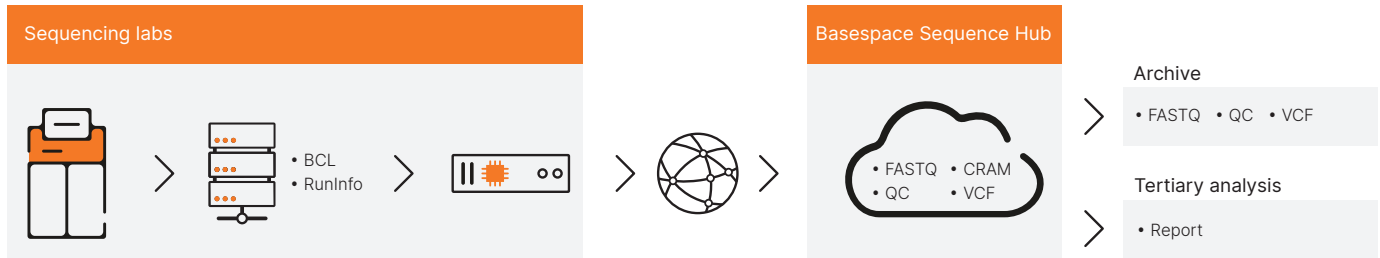


Figure 3: Hybrid cloud DRAGEN secondary analysis deployment —Customers can use a DRAGEN Server to perform primary and secondary analysis, and harness cloud resources, such as BaseSpace Sequence Hub, for multi-sample analysis, data archive, and tertiary analysis.

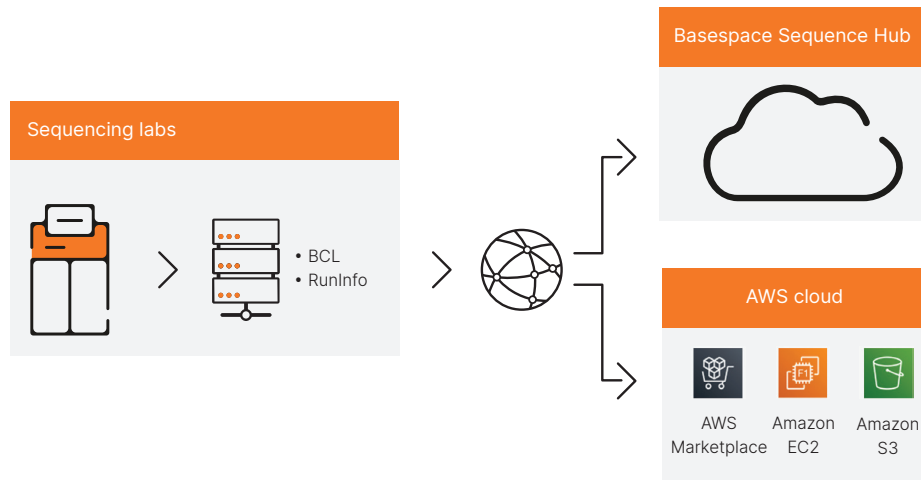


Figure 4: Cloud DRAGEN secondary analysis deployment —Customers can use DRAGEN applications available in the cloud via BaseSpace Sequence Hub or AWS.

### Cloud deployment

Customers can also forego installation and management of a DRAGEN server and choose to leverage DRAGEN applications available in the cloud via BaseSpace Sequence Hub or Amazon Web Services (AWS) Marketplace for primary and secondary data analysis (Figure 4).

### Summary

Integrating DRAGEN secondary analysis into an existing analysis infrastructure is a cost-effective way to gain the important goals of speed, accuracy, cost-savings, compatibility, and scalability. Illumina can guide you in choosing the solution that works best for your needs.

### Learn more

DRAGEN secondary analysis, [illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html](https://illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html)



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

© 2024 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](https://www.illumina.com/company/legal.html).  
M-GL-02547 v1.0