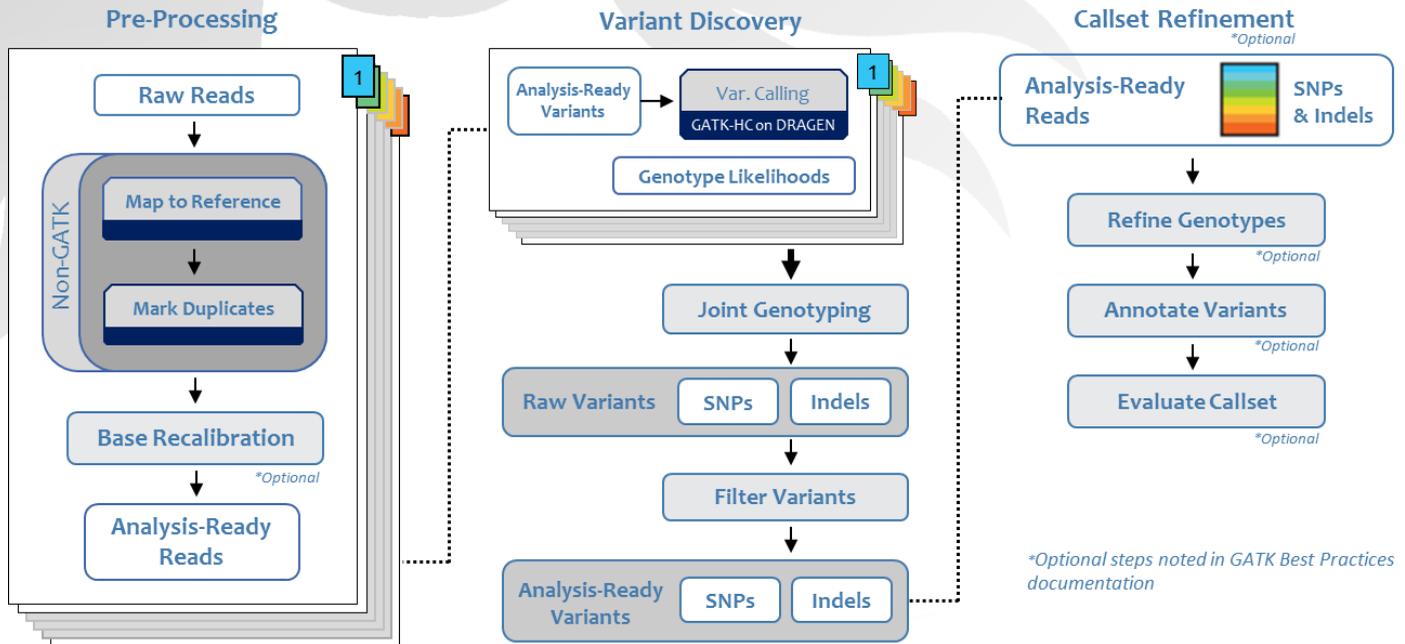


GATK Best Practices Workflow on DRAGEN

Accelerate your GATK Best Practices Workflow on the DRAGEN Platform



GATK Best Practices for Germline SNPs and Indels in Whole Genomes and Exomes on the DRAGEN Platform

Accelerated and Cost-Effective

The GATK Best Practices Workflow for Germline SNPs and Indels in Whole Genomes and Exomes is available on the DRAGEN Platform for customers that have a valid GATK License from the Broad Institute. Harnessing the tremendous processing power of the DRAGEN Bio-IT Platform, the GATK Best Practices Workflow on DRAGEN reduces the time required to analyze a whole genome from FASTQ to VCF at 30x coverage to ~22 minutes. This time saving translates directly to significant cost savings both onsite and in the cloud. DRAGEN supports versions 3.1 and 3.6 of the GATK-HC Variant Caller.

Complete End-to-End Solution

The GATK Best Practices Workflow on DRAGEN is a complete end-to-end solution that includes all required analysis phases as specified by the Broad Institute. The workflow is fully configurable and ready for use out-of-the-box. The workflow ingests BCL or FASTQ files and produces BAM, VCF and/or gVCF files as well as depth of coverage metrics. GATK Best Practices on DRAGEN can be utilized onsite, in the cloud or as a seamless hybrid cloud that includes a fully functional and easy to use Web Portal and Workflow Management System.



GATK Best Practices on DRAGEN

GATK Best Practices describes the key principles of the processing and analysis steps required to go from raw reads coming off a sequencing instrument through to an appropriately filtered variant callset that can be used in downstream analyses. GATK Best Practices breaks the workflow into two required analysis phases and one optional phase that describes ways to handle or fine-tune the output VCF file.

Analysis Phases

1. Pre-Processing

Pre-processing starts from raw sequence data and produces analysis-ready BAM files. This involves alignment to a reference genome and data cleanup to correct for technical biases and make the data suitable for analysis.

2. Variant Discovery

Variant discovery starts from analysis-ready BAM files and produces a callset in VCF format. This involves identifying genomic variation in one or more individuals and applying filtering methods appropriate to the experimental design.

3. Callset Refinement*

After a callset has been generated and filtered according to GATK Best Practices recommendations, there are several options to evaluate and refine the variant and genotype calls further.

**Optional*

➔ The GATK Best Practices documentation: software.broadinstitute.org/gatk/best-practices/

Pre-Processing

The first phase of the GATK Best Practices Workflow involves recommendations to properly prepare raw data files for analysis.

Map to Reference: The first step is mapping the sequenced reads to the reference genome to produce a file in SAM/BAM format sorted by coordinate. The workflow offers supreme flexibility of data analysis and can handle any raw data input format and can stream BCL data directly from sequencer storage, a solution unique to the DRAGEN Platform, enabling the customer to go directly from raw sequencing data to an output VCF. DRAGEN can also convert BCL to FASTQ or BAM/CRAM then proceed with the GATK Best Practices Workflow.

Mark Duplicates: Once data has been mapped to the reference genome, the workflow marks duplicate reads from DNA fragments that may have been sequenced multiple times because of PCR amplification or optical duplicate artifacts. Marking duplicates mitigates biases introduced by data generation steps by flagging the duplicates but the process does not remove the reads.

Variant Discovery

GATK Best Practices divides variant discovery into two separate steps: variant calling and variant filtering. The first step is designed to maximize sensitivity while the filtering step aims to deliver a level of specificity that can be customized for each project.

Call Variants: The workflow is capable of calling SNPs and Indels simultaneously via local de-novo assembly of haplotypes in an active region. The workflow uses hardware-accelerated implementations of the Smith-Waterman and PairHMM algorithms for variant calling.

Joint Genotype: The workflow produces an output gVCF file for each individual sample providing a comprehensive record of each position in the genome along with genotype likelihoods. The gVCF files are fed into the joint genotyping tool to produce a multi-sample VCF for subsequent joint or family analysis.

Filter Variants: GATK Best Practices recommends filtering the raw variant callset by using a variant quality score recalibration (VQSR), which uses machine learning to identify annotation profiles of variants that are likely to be real, and assigns a VQSLOD score to each variant.



Comprehensive Platform for Genomic Data Analysis and Storage

The GATK Best Practices Workflow on DRAGEN plugs into the larger DRAGEN eco-system, a comprehensive platform for analyzing, reanalyzing, storing and archiving genomic data at the lowest cost and highest fidelity. The DRAGEN Platform is based on the highly reconfigurable DRAGEN Bio-IT Processor which uses a field-programmable-gate-array (FPGA) to provide hardware-accelerated implementations of genome pipeline algorithms, such as BCL conversion, compression, mapping, alignment, sorting, duplicate marking and haplotype variant calling. The highly flexible DRAGEN Platform allows users to effortlessly set-up and manage highly complex workflows and can be utilized onsite, in the cloud or as a seamless hybrid cloud that includes a fully functional and easy-to-use Web Portal and Workflow Management System. The Hybrid Cloud configuration provides users with the flexibility to scale up to the cloud during times of high capacity and return to onsite analysis when demand is reduced.

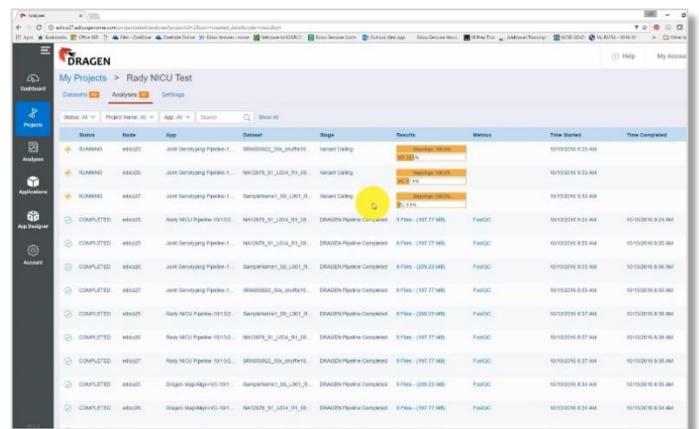
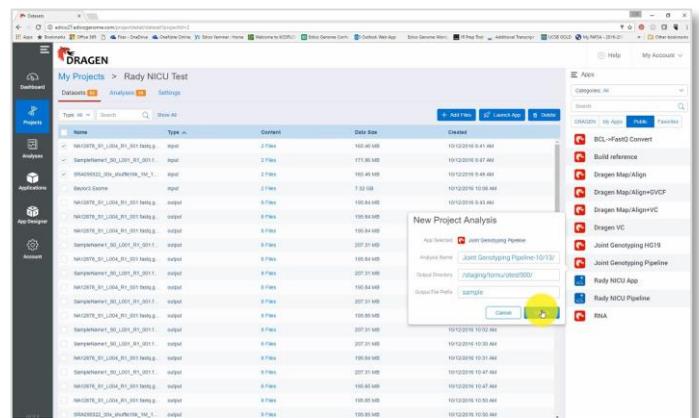
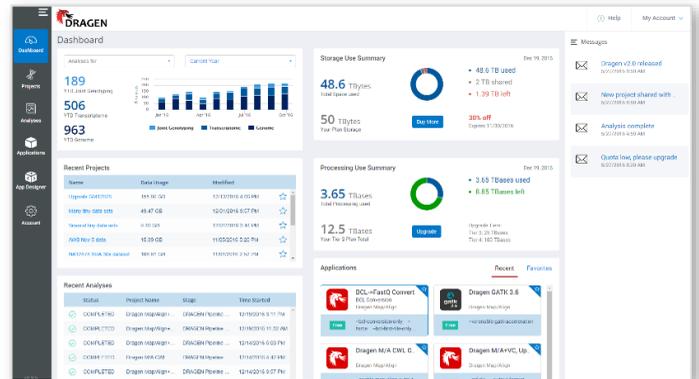
Workflow Management System

The WMS is comprehensive in terms of functional capabilities: users can effortlessly automate their workflows between any number of sequencing instruments of any type, can automate their pipelines across any number of samples, and can spread these samples across an infinite number of lanes and flow cells. The WMS can integrate with any Laboratory Information Management System (LIMS).

Many of the largest genome sequencing centers in the world use this technology to run thousands of genomes in a year, without the need for widespread resources to facilitate and manage the work. The WMS is scalable in terms of file size, number of files, and/or number of DRAGENs. It enables production-scale, clinical-grade genomic analysis.

Drag and Drop Web Portal

The DRAGEN Web Portal enables customers to use one single interface to manage both the onsite and cloud platform. Customers can manage the complete DRAGEN Platform through the Web portal, including selecting or customizing genomic pipelines, syncing hybrid cloud analysis with onsite DRAGEN server and FPGA platform, automating and scheduling concurrent single, multi-sample and/or population calling pipelines, and compressing data files. The DRAGEN Web Portal features an information-rich dashboard that enables customers to manage everything at a glance, including the complex automation functionality of the Workflow Management System and all DRAGEN ultra-rapid analysis algorithms.



Best Practice Platform Speeds

GATK Best Practices Workflow	DRAGEN Onsite	On CPU Only	DRAGEN Speed Up
BCL / FASTQ -> VCF	0:26:28 / 0:22:11	32:18:20 / 30:20:20	~74X / ~82X
BCL / FASTQ -> gVCF	0:29:27 / 0:28:31	38:51:34 / 36:53:29	~80X / ~78X
Joint Calling (3 gVCF files -> FASTQ -> Joint VCF)	2:24:00	166:45:00	~69X
+ VQSR Recalibration	0:6:58	1:50:00	~15X

DRAGEN Cloud

GATK Best Practices Workflow	DRAGEN Cloud	On CPU Only	DRAGEN Speed Up
BCL / FASTQ -> VCF	0:8:45 / 0:6:12	32:18:20 / 30:20:20	~242X / ~151X
BCL / FASTQ -> gVCF	0:10:20 / 8:16	38:51:34 / 36:53:29	~233X / ~148X
Joint Calling (3 gVCF Files)	1:20:23	166:45:00	~125X

Dataset: SRA056922 NA12878 @ 30x

DRAGEN Tools and Features

In addition to the GATK Best Practices Workflow, the following set of tools are available on the DRAGEN Platform.



VCF Hard Filtering
This tool is designed for hard-filtering variant calls based on specified criteria.



dbSNP Annotation
This tool is designed to annotate variant calls based on their context, as opposed to functional annotation.



Depth of Coverage
This tool processes a set of BAM files to determine coverage at different levels of partitioning and aggregation.



Revolutionary low-cost compute and storage costs

Get DRAGEN Now!

info@edicogenome.com
1-858-260-5234

GATK License from the Broad Institute required for GATK Best Practices on DRAGEN

About Edico Genome

At Edico Genome, we're helping usher in the new era of personalized medicine by enabling a fundamental change in healthcare with customized treatments and data-driven insights tailored to the individual. At the heart of personalized medicine, DNA sequencing technology is advancing at an even more rapid pace than the cell phone revolution. By increasing the speed and accuracy for NGS data analysis, such as whole genome sequencing (WGS), our computing platform makes it easier to discover links between DNA sequence variations and human disease.



info@edicogenome.com
www.edicogenome.com



@edicogenome

edicogenome