DRAGEN™

DRAGEN Germline V2 Pipeline

Ultra-Rapid Genome and Exome Data Analysis

24min
FASTQ → VCF
WHG @ 30x Coverage

Industry-best sensitivity
SNP and INDEL variant detection

Highly accurate
Detects true variants; low false positive rate

Hybrid Solution
Available Onsite and in the Cloud

Overview

The DRAGEN Germline V2 Pipeline provides ultra-rapid analysis of NGS data, and reduces the time required to analyze a whole human genome from ~20 hours (BWA-GATK) to ~24 minutes onsite and ~34 minutes in the cloud (AWS F1.16xlarge instance), while achieving industry leading variant calling accuracy. New features in the DRAGEN Germline V2 Pipeline include improved mapper and alignment algorithms, Alt-Aware mapping for Graph-based references, and a sample-specific error modeling algorithm that greatly improves variant calling accuracy for INDELs.

How It Works

The DRAGEN platform performs secondary analysis of NGS data using optimized algorithms for alignment, mapping, sorting, duplicate marking, haplotype variant calling, and compression. DRAGEN accepts FASTQ, BAM, and BCL file formats and outputs BAM/CRAM, VCF, and gVCF files suitable for downstream analysis.

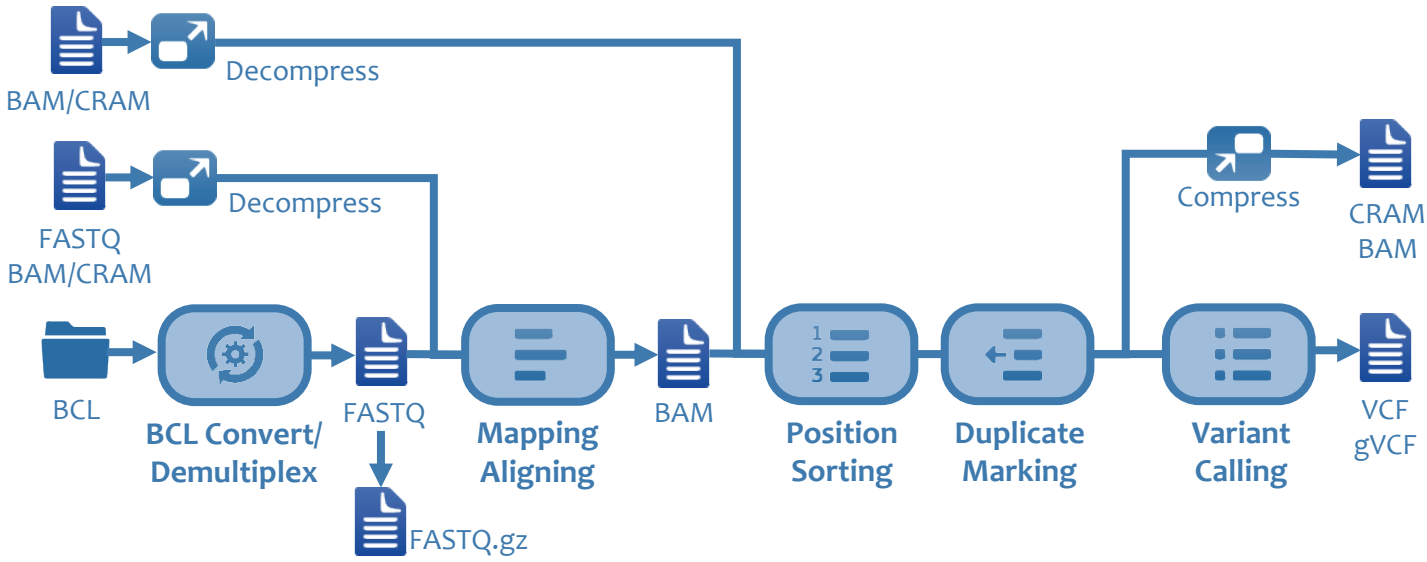
Highlights

- **Comprehensive**—Analyzes sequencing data from whole genome, whole exome, and targeted panels.
- **Ultra-rapid**—50x faster than BWA-GATK-HC (whole human genome, 30x coverage, DRAGEN onsite).
- **Accurate**—Detects SNPs and INDELs with industry leading sensitivity and specificity.
- **Easy to Use**—Fully functional API, CLI, and GUI. Can stream BCL files directly from sequencing instrument.

edico genome

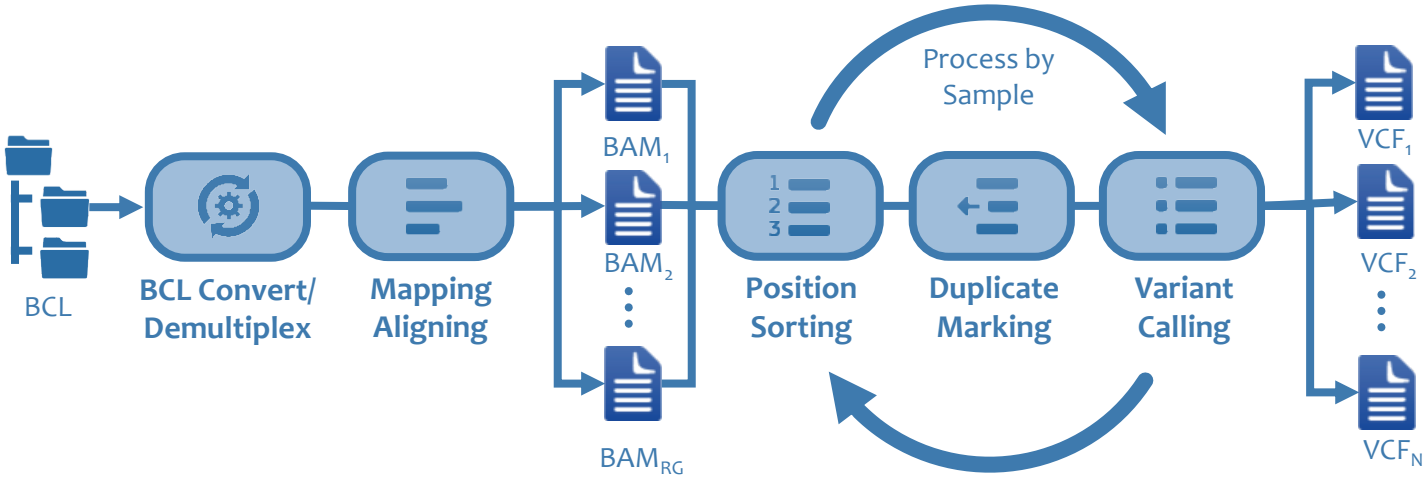
DRAGEN Germline V2 Pipeline

The DRAGEN Germline V2 Pipeline provides ultra-rapid genome sequence analysis with industry-leading accuracy. This pipeline features a new sample-specific error modeling algorithm that improves variant calling accuracy, particularly for INDELs. DRAGEN Germline V2 Pipeline was the top performer in 5 out of 6 categories amongst entrants that identified all 50 hidden variants in the 2017 Precision FDA Hidden Treasures Challenge Warm-Up.



Efficient Multi-Sample Processing

DRAGEN is capable of processing BCL data directly, eliminating any FASTQ conversion step. The BCL data is fed directly into the pipeline to produce a unique output VCF file for each sample. Intermediate BAM/CRAM files can be generated on demand. To streamline and automate multi-sample processing, DRAGEN offers a comprehensive Workflow Management System (WMS), which allows users to schedule multiple workflow runs for any pipeline.



Applications

NICU Diagnostics

Clinical Genomics

Sequencing Centers

Agri Genomics

NIPT NIPS

Pipeline Steps

Input/Output File Formats
- Input BCL, FASTQ or BAM/CRAM
- Output BAM/CRAM or VCF/gVCF

Compression / Decompression
- Hardware-accelerated compression and decompression for Gzip and CRAM files

Mapping / Aligning
- Single end or paired end reads
- Read lengths from 26 bp to 10 kbp

Position Sorting
- Binning by reference range
- Sorting of bins by reference position

Duplicate Marking
- Equivalent to Picard tools duplicate marking
- Highest quality duplicate marking reported

Variant Calling
- Includes sample-specific error modeling
- Uses Hidden Markov Model and Smith-Waterman Alignment

Accuracy: PrecisionFDA Hidden Treasures Warm-Up Challenge (Sept 2017)

		SNPs			INDELs			
Entrant	Variant Caller	Specificity	Sensitivity	F-score	Specificity	Sensitivity	F-score	Combined SNP+INDEL F-score
Edico Genome	**DRAGEN Germline V2**	0.997	**0.990**	**0.993**	**0.889**	**0.884**	**0.886**	**0.940**
Aginome	Aginome IMP	0.995	0.986	0.991	0.792	0.859	0.824	0.907
Saphetor	Saphetor 3	0.994	0.971	0.982	0.818	0.821	0.820	0.901
Sentieon	DNAseq	0.992	0.989	0.990	0.752	0.863	0.803	0.897
Samsung Genome Inst.	SGI VC	0.991	0.989	0.990	0.752	0.859	0.802	0.896
New York Univ.	GATK HC	0.986	0.982	0.984	0.75167	0.854	0.7996	0.892
Garvan Inst.	KCCG 1.7.0 HC	0.995	0.989	0.992	0.72771	0.865	0.790	0.891
CNAG	GATK3.6	**0.997**	0.973	0.985	0.73525	0.834	0.782	0.883

Table includes entrants that identified all 50 hidden variants. Top scorers in each category are in bold.

Speed: Single Sample Pipeline

Dataset	Pipeline	DRAGEN Germline V2	BWA + GATK-HC 3.6	DRAGEN Speed Up
SRA056922 NA12878 @ 30x	FASTQ to VCF (onsite)	0:24:24	20:20:20	~50X
	FASTQ to VCF (cloud)	0:33:53	N/A	~36X
	FASTQ to gVCF (onsite)	0:31:22	36:53:29	~71X
	FASTQ to gVCF (cloud)	0:50:22	N/A	~44X

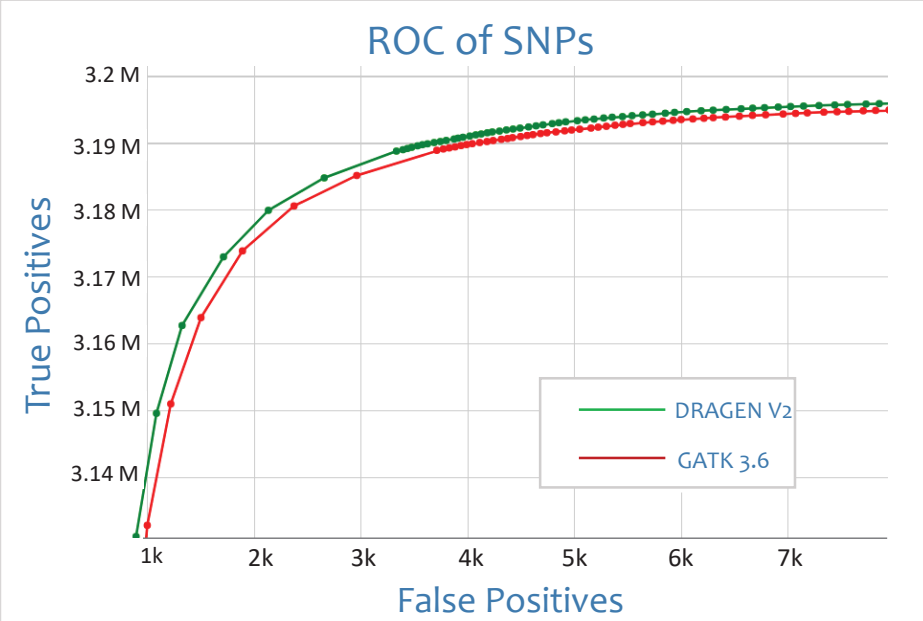
24-core CPU on-site; AWS F1.16x instance in cloud

Speed: Multi-Sample Pipeline

Processing of 12 Samples	Germline V2	Bcl2fastq + BWA + GATK-HC	DRAGEN Speed Up
BCL to FASTQs to VCFs	4:47:56	279:02:36	~58X
BCL to VCFs	3:47:30	N/A**	N/A**

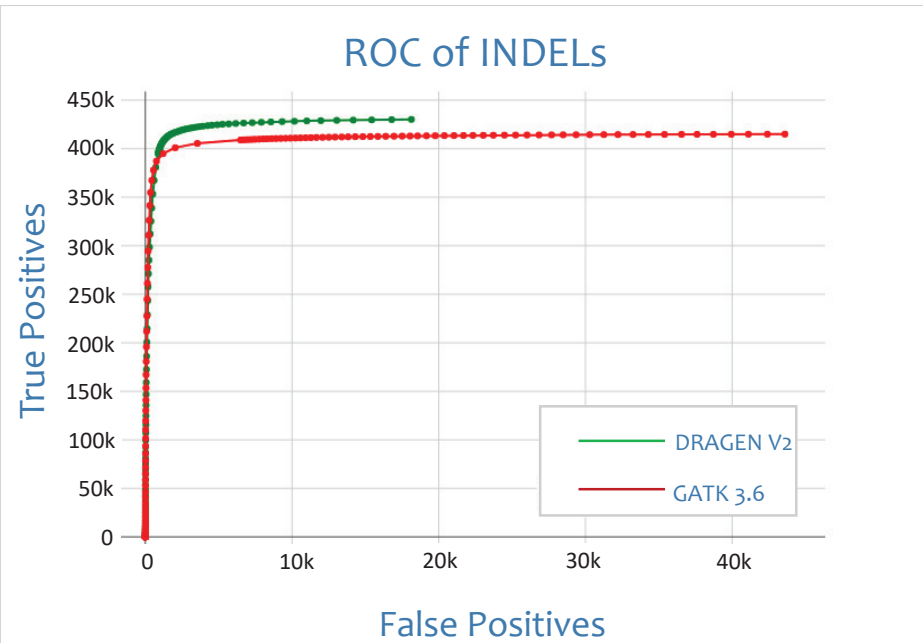
**BCL direct to VCF capability is unique to the DRAGEN pipeline

ROC Plots of Variant Calling Accuracy: DRAGEN Germline V2 vs. GATK 3.6



ROC of SNPs

A SNP (single nucleotide polymorphism) is a single base that differs between two genomes, in this case between the reference human genome hg19 and the NIST Genome in a Bottle sample SRA056922 NA12878 at 30x coverage. DRAGEN Germline V2 had a greater number of true positives relative to false positives compared to GATK 3.6 HC.



ROC of INDELs

An INDEL is the insertion or deletion of one or more base pairs in a sample genome compared to a reference. The sample was SRA056922 NA12878 at 30x coverage. DRAGEN Germline V2 outperformed GATK 3.6 HC in INDEL calling accuracy.

About Edico Genome

Edico Genome is the leading secondary analysis solution provider for next-generation sequencing, delivering its powerful DRAGEN Bio-IT platform to clinical, research and genome centers around the globe. Leveraging field programmable gate array (FPGA) technology, DRAGEN delivers best-in-class accuracy, speeds, scalability and costs, enabling customers of all sizes to focus on what matters most – delivering breakthrough results. The comprehensive set of DRAGEN pipelines can be run onsite, in the Cloud or through a seamless hybrid cloud blend, allowing organizations to scale as their throughput fluctuates. Edico Genome has set two GUINNESS WORLD RECORDS™ titles for its speed, and received top marks for its accuracy in the recent PrecisionFDA Challenge. For more information about DRAGEN, visit www.edicogenome.com.

info@edicogenome.com
www.edicogenome.com
@edicogenome

edico genome