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.

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.

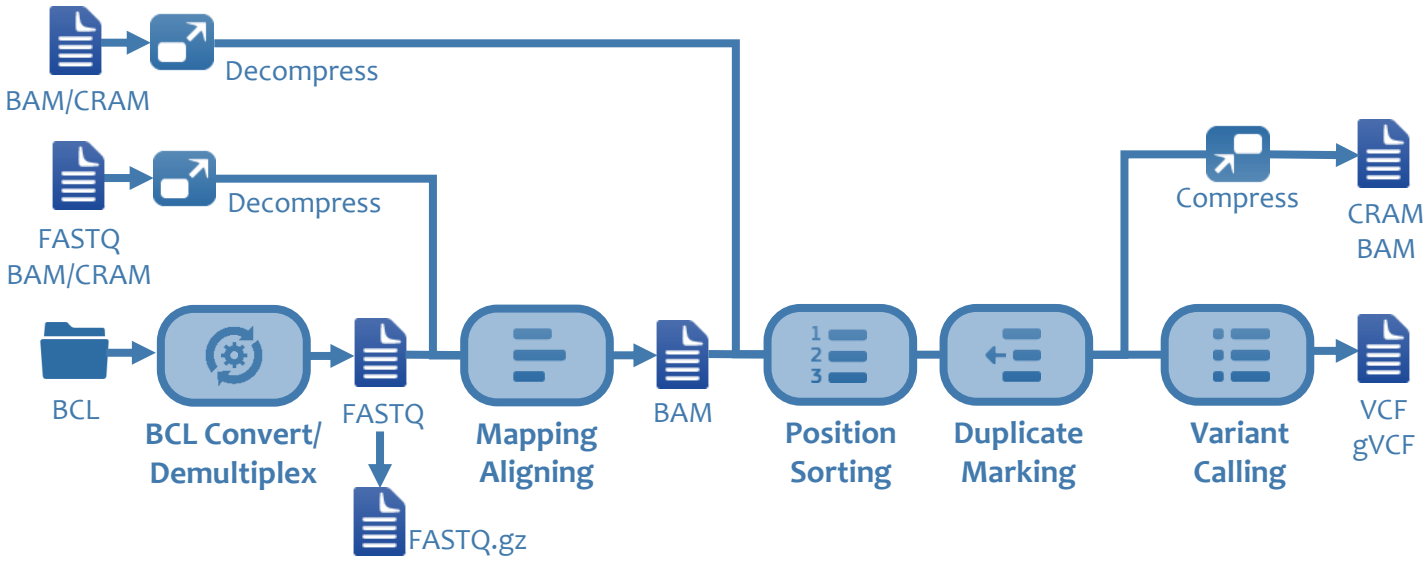
Highlights

- **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. Aside from algorithmic improvements, the DRAGEN Germline V2+ Pipeline improves on the DRAGEN Germline V2 Pipeline, which 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.



Other features of DRAGEN Germline V2+ Pipeline

Enhanced Quality Metrics

The DRAGEN Germline V2+ Pipeline now includes additional quality metrics for the Mapper and Variant Caller. The genome coverage histogram now shows percentage of bases up to 100x, and also coverage over a target region when a BED file is provided. Additionally, alignment coverage and SNP counts are now available for the X & Y chromosomes including coverage and SNP ratios over those chromosomes. The DRAGEN Germline V2+ Pipeline also provides average and median mitochondrial and autosomal coverages.

Improved Variant Calling in Sex Chromosomes

The DRAGEN Germline V2+ Pipeline software supports accurate processing of the ploidy in pseudoautosomal regions (PAR). Mainly during variant calling for males, the PAR regions on ChrX will be processed using the diploid model and the rest using haploid model. For females, since ChrX is already diploid, there is no PAR region related change. This extends to gVCF and jointly called VCF files and is important with Trio analysis.

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

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

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

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

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

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

DRAGEN Germline V2+ Pipeline Algorithmic Improvements

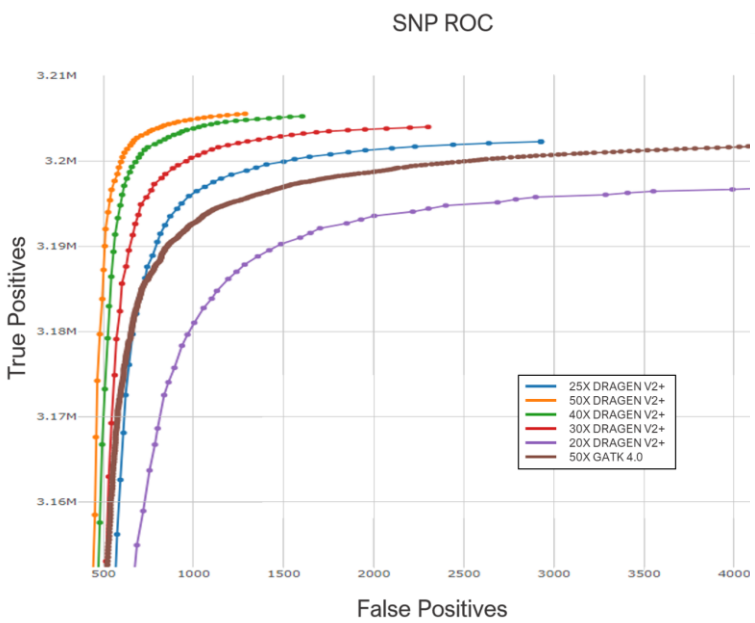
The DRAGEN Germline V2+ Pipeline incorporates enhancements in the algorithm that reduce the false positive rate, produce a better f-measure, and overall greater sensitivity. DRAGEN Germline V2+ reduces false positives without the need to eliminate a large number of reads, leading to its ability to generate higher sensitivity. DRAGEN Germline V2+ achieves false positive reduction through:

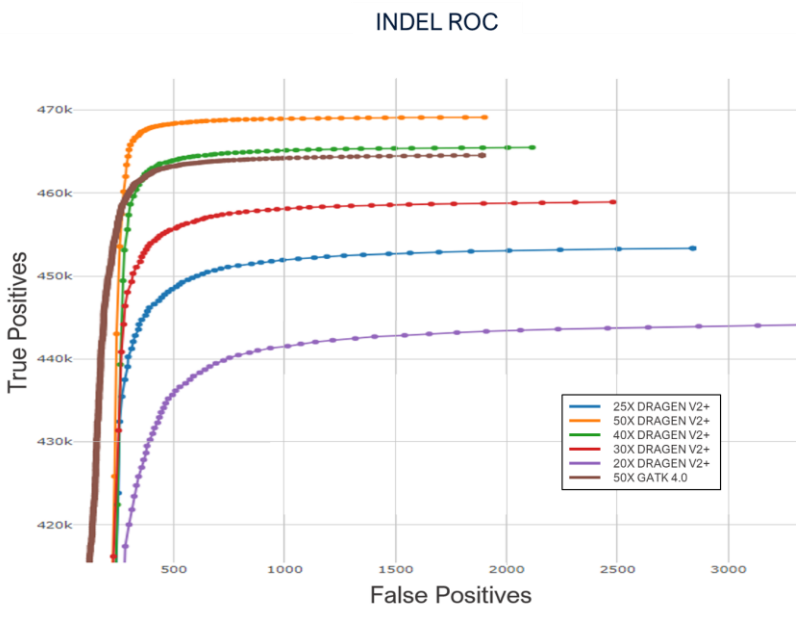
- The identification of contaminant reads and preventing them from contributing to the variant analysis.
- Identification of mis-mapped reads through patterns of reads that are not consistent with properly mapped reads.
- Detecting regions of the genome where the sequencing technology encounters problems during the base incorporation process.

Other algorithmic improvements include a sample-specific error model that improves variant calling accuracy by improving INDEL accuracy for datasets that have undergone polymerase chain reaction (PCR).

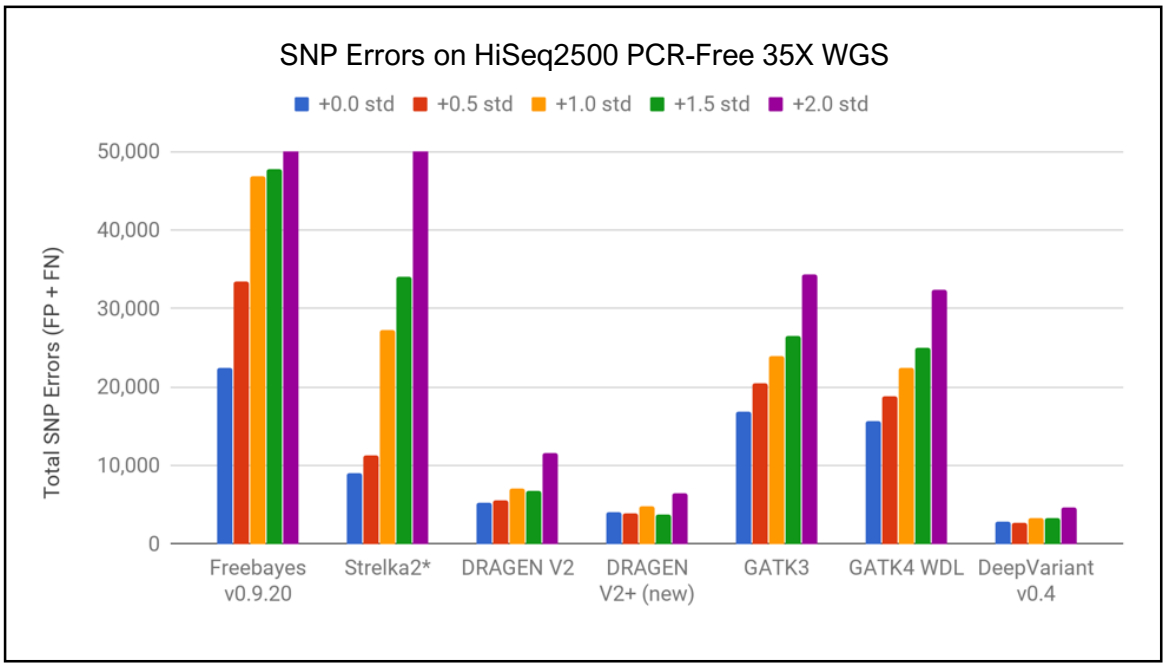
Effects on Sample Coverage: DRAGEN Germline V2+ Pipeline Accuracy

The DRAGEN Germline V2+ Pipeline achieves greater accuracy at 25X coverage than GATK 4.0 does at 50X coverage for SNPs. For INDELs, DRAGEN V2+ achieves the same or greater accuracy at 40X coverage that GATK 4.0 achieves at 50X coverage.



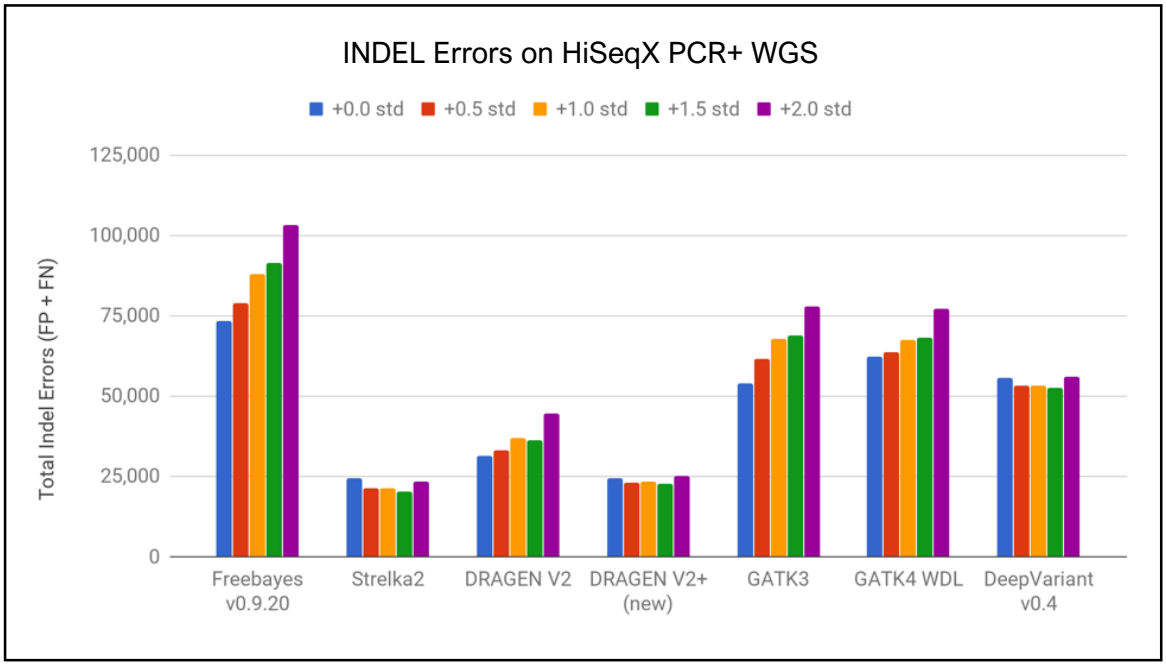


Variant Calling Accuracy: DRAGEN Germline V2+ vs. 3rd Party Analysis Tools



Comparison of SNP Accuracy*

A SNP (single nucleotide polymorphism) is a single base that differs between two genomes. In this comparison using HG001 of DRAGEN Germline V2+ against other third party analysis tools on the market, DRAGEN V2+ demonstrated less SNP errors compared to GATK, Strelka2, and Freebayes.



Comparison of INDEL Accuracy*

An INDEL is the insertion or deletion of one or more base pairs in a sample genome compared to a reference. In this comparison using HG001 of DRAGEN Germline V2+ against other third party tools on the market, DRAGN V2+ demonstrated less INDEL errors than GATK, DeepVariant, and Freebayes.

*Both of these comparisons are independent benchmarks from a trusted third party (DNAnexus) and can be found on the DNAnexus Blog, "How to Train Your DRAGEN – Evaluating and Improving Edico Genome's Rapid WGS Tools." March 2018.

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