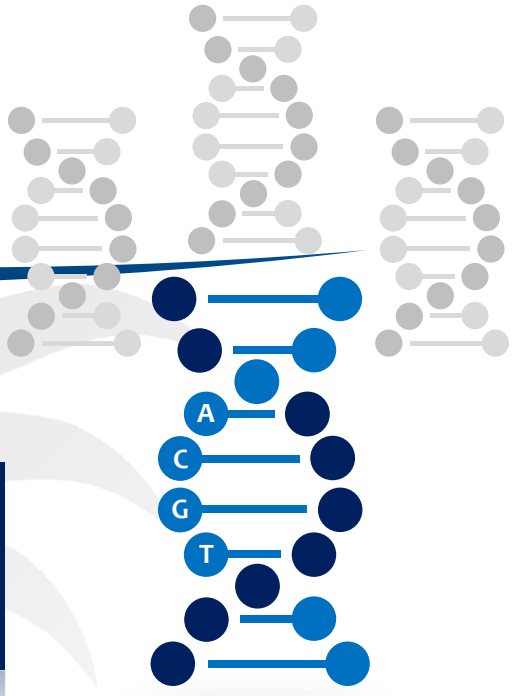
DRAGEN™

DRAGEN Joint Genotyping

Ultra-Rapid Multi Genome Analysis



1.38 hrs	99.5%	99.3%	Hybrid
Trio at 50x	Precision	Sensitivity	Cloud
Joint Analysis from FASTQ → gVCF → VCF	SNP + INDEL Combined	SNP + INDEL Combined	Onsite, Cloud or Hybrid Solution

Overview

The DRAGEN Joint Genotyping Pipeline calls variants from multiple samples at a speed 25x faster than competing pipelines with uncompromising accuracy. The Joint Genotyping pipeline supports pedigree as well as population variant calling from a cohort of samples. The Joint Genotyping pipeline handles up to ten samples at one time. The DRAGEN Population Calling pipeline handles sample sizes of many thousands at once.

The combination of DRAGEN's speed and hierarchical grouping of multiple samples provides the most computationally efficient analysis solution for joint genotyping.

What is DRAGEN?

DRAGEN (Dynamic Read Analysis for Genomics) is a highly reconfigurable Bio-IT Processor which is integrated on a PCIe card and is available in a pre-configured server. DRAGEN can also be integrated directly into sequencing instruments and NGS bioinformatics servers. DRAGEN is offered as a Platform-as-a-Service (PaaS), enabling customers to license various pipelines according to their needs.

How Does it Work?

DRAGEN performs map/align and variant calling on multiple FASTQ files or multi-sample BCL folders produced by a sequencing instrument (such as Illumina's HiSeq X10). The output files are in gVCF format and are fed into DRAGEN's Joint Genotyper to produce a single VCF file for subsequent analysis.

The DRAGEN platform includes a fully functional and easy to use graphical user interface (GUI) and an extensive set of tools, enabling customers to easily schedule multiple workflow runs, analyze results, compare different pipelines, monitor multiple networked DRAGEN cards and manage update releases.

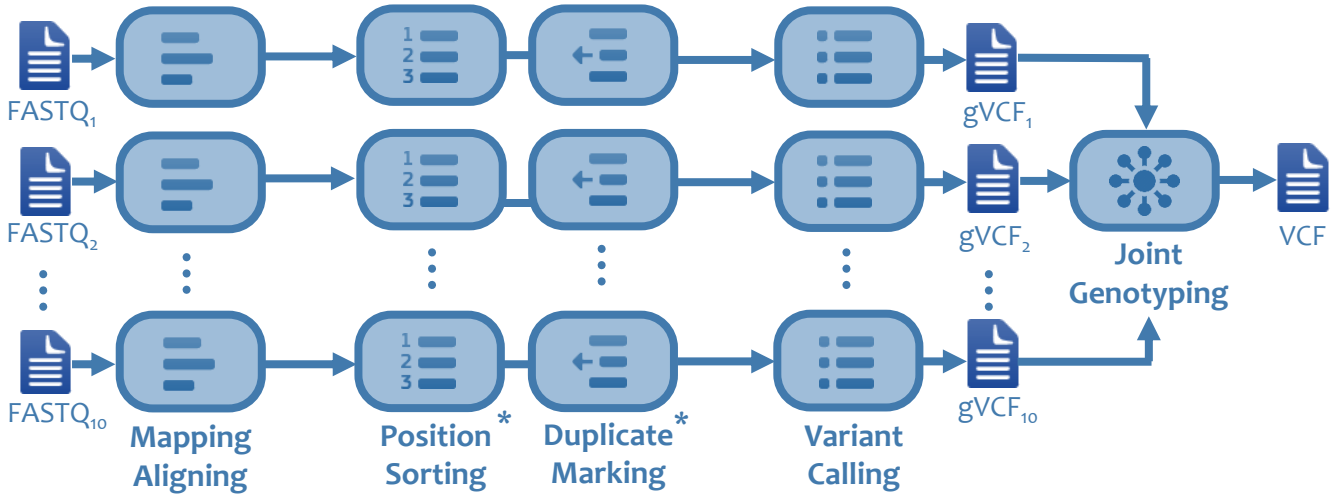
Comprehensive Set of Pipelines

DRAGEN supports pipelines for Whole Genome, Whole Exomes, Targeted Panels, Epigenome / Methylome, RNA-seq / Transcriptome, Microbiome and Cancer Tumor / Normal. All pipelines are available both onsite and in the cloud via a pay-per-use or subscription service. Data usage based pricing tiers are available to cater for customers performing benchtop sequencing all the way up to population scale.

edico genome

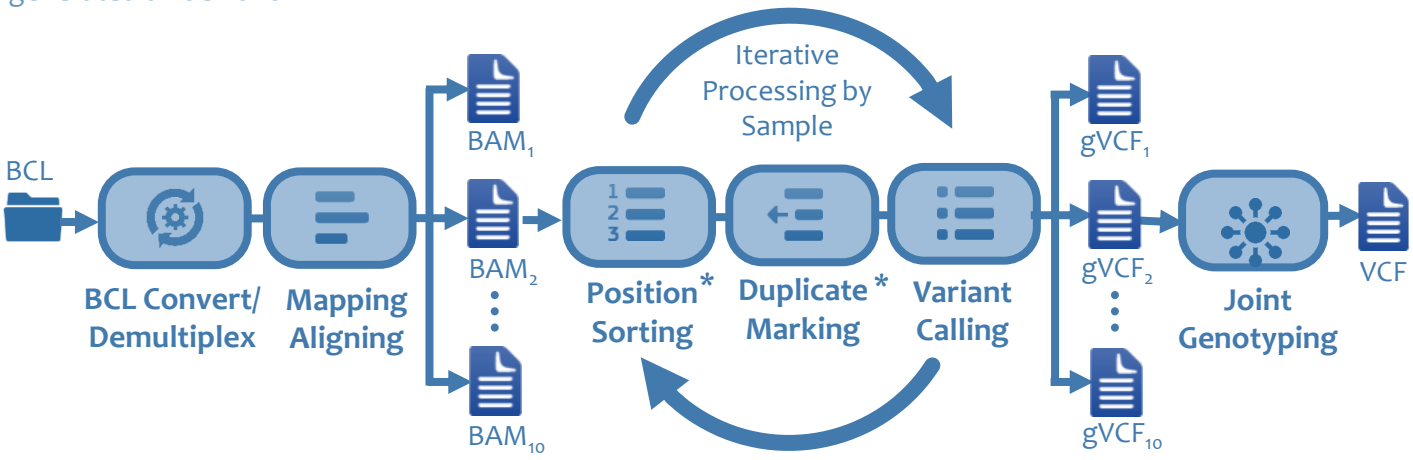
DRAGEN Joint Genotyping Pipeline

The DRAGEN Joint Genotyping pipeline enables variant calls to be made with information from multiple samples. DRAGEN produces an output gVCF file for each of the individual samples. Each gVCF file provides a comprehensive record of every position in the genome. The gVCF files are fed into the DRAGEN Joint Genotyper to produce a single VCF for subsequent joint or family analysis. The Joint Genotyping pipeline handles ten samples at one time. The DRAGEN Population Calling pipeline handles sample sizes of many thousands at once.



Joint Calling from BCL

In the event the user is joint calling samples sequenced on the same flow cell, he can take advantage of the capability of DRAGEN to simultaneously map/align multi-sample inputs to speed up the overall process of joint calling. DRAGEN is capable of processing BCL data directly, eliminating any FASTQ conversion step. The BCL data is fed directly to the pipeline to produce unique gVCF files for each sample. Intermediate BAM/CRAM files can be generated on demand.



*Optional Pipeline Step

Applications

- Genetic Testing for Hereditary Disorders
- Family Studies
- Clinical Diagnostics
- Population Genetics
- Autism Linkage Studies

Pipeline Steps

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

Compression/Decompression
- Decompression of FASTQ, BCL, BAM/CRAM
- Gzip and CRAM in and out

BCL Convert/Demultiplex
- BCL conversion to FASTQ
- BCL can also be processed directly

Mapping/Aligning
- Single end or paired end reads
- Supports read lengths from 26 bp to 10k bp

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

Duplicate marking
- Based on starting position & CIGAR string
- Highest quality duplicate report

Variant Calling
- Haplotype variant caller with reassembly
- Uses Hidden Markov Model and Smith-Waterman Alignment

Joint Genotyper
- Single Up to 10 input gVCF files
- Jointly called VCF output

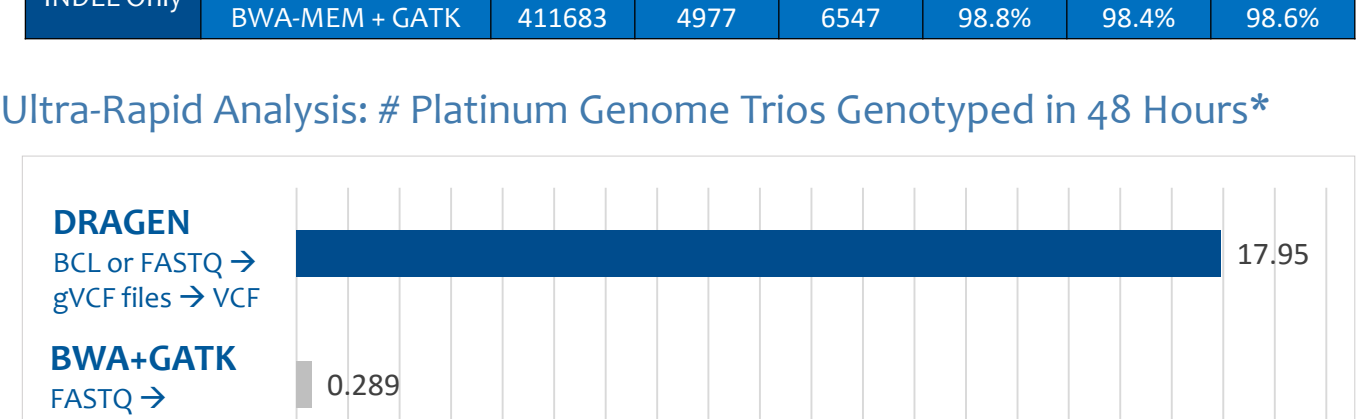
Speeds: Joint Genotyping Pipeline*

Pipeline	DRAGEN	BWA-MEM + GATK-HC	DRAGEN Speed Up
Platinum Genomes Trio @ 50x coverage + Joint Genotyping	2:40:46	166:45:00	69X

Accuracy: Joint Genotyping Pipeline*

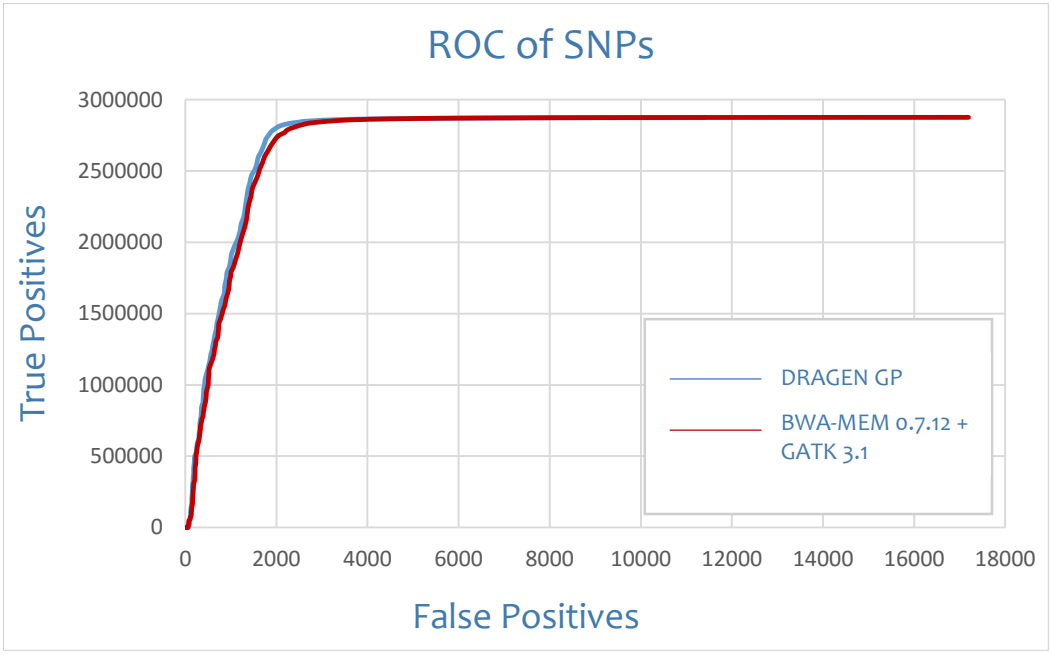
Accuracy	SNP+INDEL Combined	True-Pos	False-Pos	False-Neg	Precision	Sensitivity	F-Measure
SNP + INDEL Combined	DRAGEN	3284846	15741	22590	99.5%	99.3%	99.4%
	BWA-MEM + GATK	3288002	22168	22867	99.3%	99.3%	99.3%
SNP Only	DRAGEN	2874110	11407	16078	99.6%	99.4%	99.5%
	BWA-MEM + GATK	2876319	17191	16320	99.4%	99.4%	99.4%
INDEL Only	DRAGEN	410736	4334	6512	99.0%	98.4%	98.7%
	BWA-MEM + GATK	411683	4977	6547	98.8%	98.4%	98.6%

Ultra-Rapid Analysis: # Platinum Genome Trios Genotyped in 48 Hours*

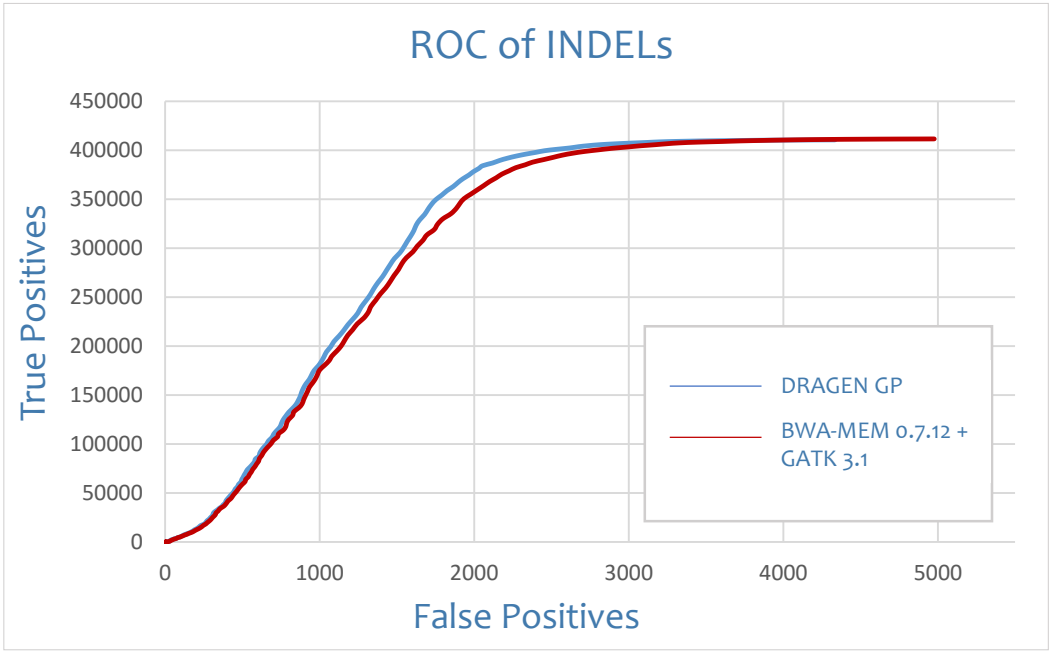


*All DRAGEN results are compared against BWA-MEM 0.7.12 + GATK 3.1 running on comparable servers.

ROC Plots of Variant Calls at 50x Coverage



ROC of SNPs
A SNP (single nucleotide polymorphism) occurs when a single base differs between two genomes, in this case the subject and the reference genome. Use of the NIST Platinum Genome high confidence call set enables performance comparisons between different pipelines. In this ROC plot, a higher count of true positive SNPs and lower count of false positive SNPs is considered better.



ROC of INDELs
An INDEL (insertion or deletion) occurs when bases are inserted or deleted in the subject genome with respect to a reference genome. Use of the NIST Platinum Genome high confidence call set enables performance comparisons between different pipelines. In this ROC plot, a higher count of true positive INDELs and lower count of false positive INDELs is considered better.

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 like 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

edico genome