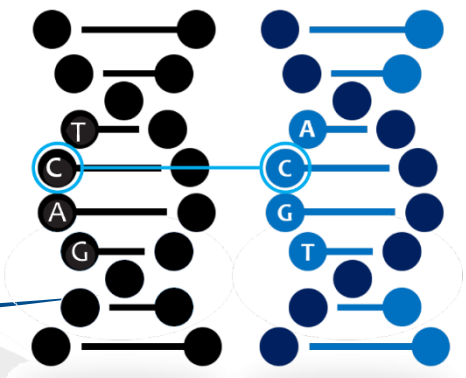
DRAGEN™



DRAGEN Somatic Pipeline V2

Ultra-Rapid, Highly Accurate Analysis of Tumor Sequence Data

Ultra Rapid

Highly Accurate

Low Cost

Cloud or Onsite

Overview

The DRAGEN™ (Dynamic Read Analysis for GENomics) V2 Somatic Pipeline allows ultra-rapid analysis of next-generation sequencing (NGS) data to identify cancer-associated mutations. DRAGEN is able to call SNPs and INDELs from both matched tumor/normal pairs and tumor-only samples. DRAGEN produces rapid results while achieving a level of accuracy greater than top somatic variant callers.

How it works

For the tumor/normal pipeline, both samples are analyzed jointly such that germline variants are excluded, generating an output specific to tumor mutations. The tumor-only pipeline produces a VCF file that can be further analyzed to identify tumor mutations. Both pipelines make no ploidy assumptions, enabling detection of low-frequency alleles. New features of DRAGEN V2 are a sample-specific calibration algorithm that improves accuracy, and refined mapper and aligner algorithms.

Highlights

- **Highly Accurate**—detects more somatic variants with greater precision than other variant callers
- **Improved INDEL calling**—DRAGEN V2 features a novel sample-specific calibration algorithm that improves INDEL calling accuracy
- **Sensitive**—detects low-frequency alleles with high precision
- **Rapid**—analyzes genomic data orders of magnitude faster than other variant callers
- **Flexible**—supports both tumor/normal and tumor-only analysis
- **Easy to Use**—fully functional API, CLI, graphical user interface (GUI) and Workflow Management System
- **Accessible**—can be used onsite, in the cloud, or as a hybrid cloud solution

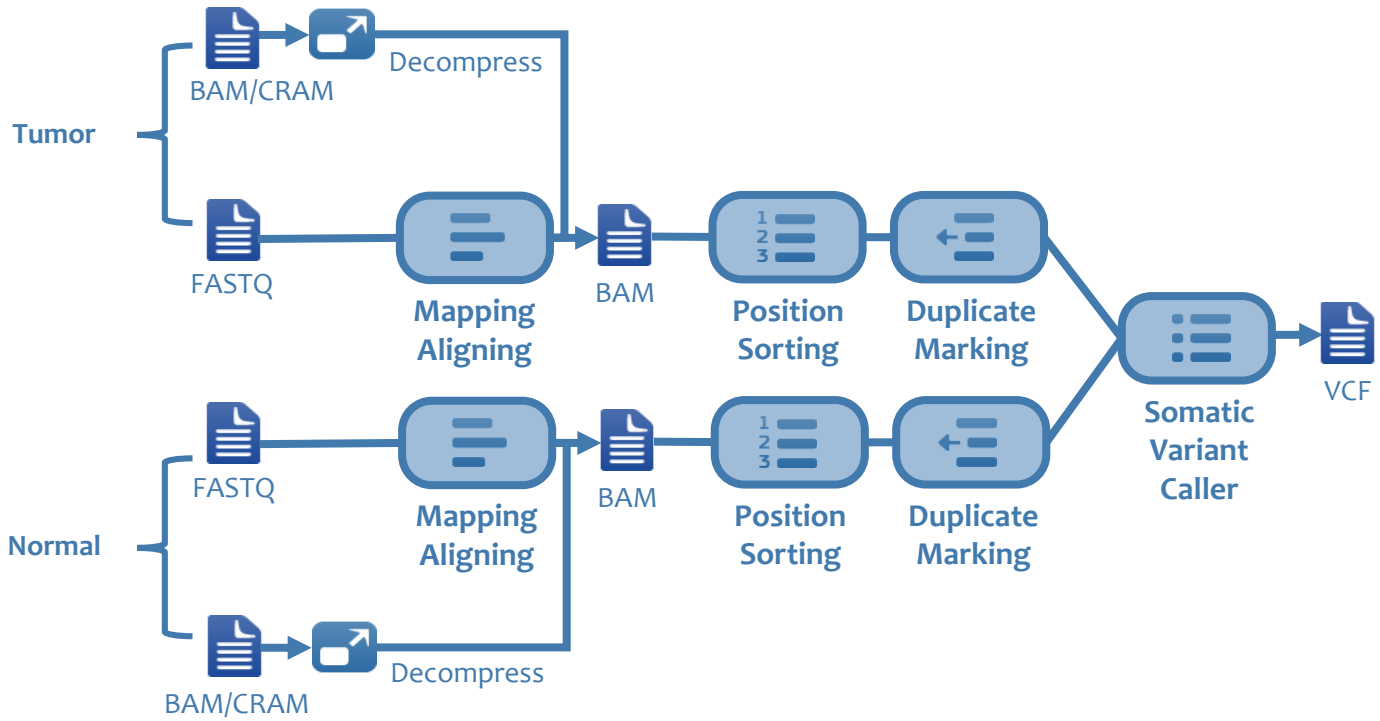
edico genome

DRAGEN V2 Somatic Pipeline

The DRAGEN V2 Somatic Pipeline offers flexible data analysis to suit different needs. It accepts FASTQ, BAM/CRAM, and BCL files and supports NGS input from whole genome, whole exome, and targeted cancer panels. The Somatic Pipeline calls SNPs and INDELs while also reporting allele frequency. The Somatic Pipeline can be automated using the DRAGEN Workflow Management System for ease of use and bulk sample processing. It is available onsite, in the Cloud, or as a hybrid Cloud solution.

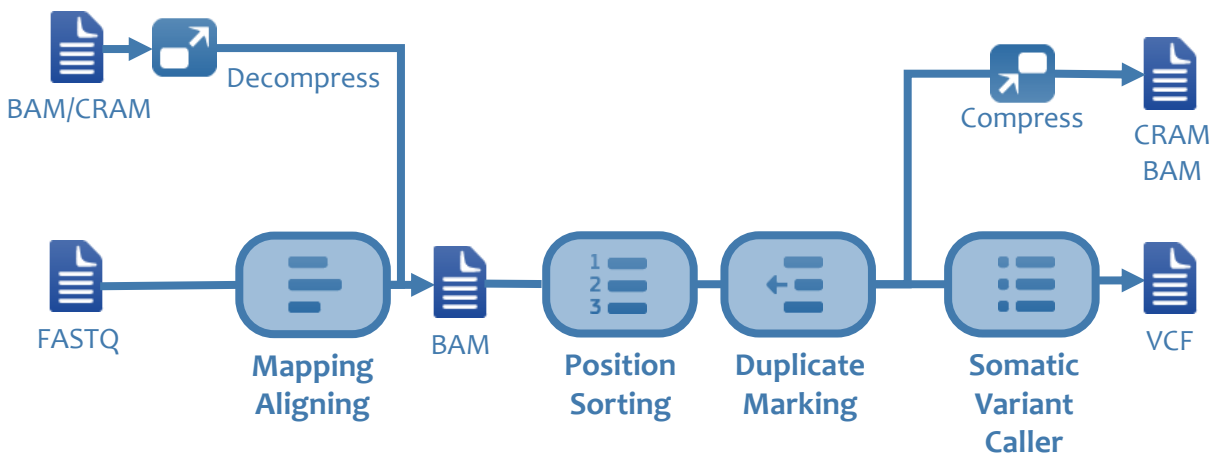
Tumor/Normal Pipeline

In the tumor/normal pipeline, both samples go through identical processing steps and are input into the variant caller, where germline variants are excluded to produce a VCF file specific for tumor mutations.



Tumor-Only Pipeline

The tumor-only pipeline lacks a matching normal sample and produces a VCF file containing both somatic and germline variants. Users have two options for refining the data to identify somatic variants: 1) Input a panel-of-normals dataset as a germline filter in the variant caller; 2) Compare the output VCF to publicly available databases of germline SNPs and INDELs to remove known germline variants.



Pipeline Steps

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

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 reported

Somatic Variant Calling
- Somatic Variant Caller filters germline variants
- Tumor/Normal analysis filtering is conducted with variant comparison between match-pairs

DRAGEN Delivers Improved Accuracy

The ICGC-TCGA DREAM Mutation Calling Challenge was a contest to find the most accurate tool for detecting variants in synthetically generated mutation datasets. The DRAGEN V2 Somatic Variant Caller was benchmarked against the top-performing variant callers from the DREAM Challenge and outperformed them for both SNP and INDEL accuracy. In the chart below, the winning submission of the DREAM Challenge for synthetic dataset 4 is highlighted in grey; DRAGEN V2 is in blue.

Variant Type	ID	Submission Name	Team	# of Calls	Sensitivity	Precision	F-score
SNP	2559951	MuTect-RSp2	Broad SMC	12272	0.741	0.983	0.845
	2549704	MuTect-RSp	Broad SMC	12282	0.741	0.982	0.845
	2528605	MuTect-R	Broad SMC	12288	0.741	0.981	0.844
		DRAGEN V2 Somatic Variant Caller	Edico Genome	12810	0.767	0.971	0.857
				DRAGEN Improvement over Winner			1.42%
INDEL	2559911	novoBreak indel	Ken Chen Lab	12059	0.788	0.928	0.852
	2556244	EmTooNVDBS	Broad SMC	11907	0.772	0.921	0.840
	2559878	novoBreak_indel	Ken Chen Lab	11341	0.751	0.940	0.835
		DRAGEN V2 Somatic Variant Caller	Edico Genome	11814	0.787	0.953	0.862
				DRAGEN Improvement over Winner			1.15%

DRAGEN Outperforms Top Variant Callers on a Real Tumor Sample

A 2015 study (Alioto et al., *Nature Communications*) compared multiple somatic variant calling tools for accuracy in calling SNPs and INDELs from a medulloblastoma tumor sample. A curated gold set was used to benchmark performance. The DRAGEN V2 Somatic Variant Caller analyzed the same FASTQs as the other tools and produced better measures of accuracy for SNPs and INDELs. The top performing submissions are highlighted in grey; DRAGEN V2 in blue.

Variant Type	Aligner	Variant Caller	True Positive	False Positive	False Negative	Sensitivity	Precision	F-score
SNP	BWA, GEM	Curated Gold Set	1,255	0	0	1	1	1
	Novoalign	MuTect	947	6296	308	0.76	0.13	0.22
	None, BWA	SGA+freebayes	856	62	399	0.68	0.93	0.79
	BWA	MuTect, Strelka	385	3	870	0.31	0.99	0.47
	BWA	qSNP+GATK	842	25	413	0.67	0.97	0.79
	BWA	MuTect	944	272	311	0.75	0.78	0.76
	DRAGEN	DRAGEN V2	950	70	305	0.757	0.931	0.835
				DRAGEN improvement over best performer				5.36%
INDEL	BWA, GEM	Curated Gold Set	337	0	0	1	1	1
	BWA	Platypus	271	224	70	0.79	0.55	0.65
	BWA	GATK, Varscan	167	20	173	0.49	0.89	0.63
	BWA	Pindel	189	82	152	0.55	0.70	0.61
	GEM	samtools, bcftools	103	26	236	0.30	0.80	0.44
	BWA	Strelka	64	3	273	0.19	0.96	0.32
	DRAGEN	DRAGEN V2	253	96	84	0.75	0.72	0.74
				DRAGEN improvement over best performer				13.29%

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

edico genome