DRAGEN™

DRAGEN Copy Number Variant (CNV) Pipeline

Ultra-Rapid, Highly Accurate Analysis of Structural Variants

Overview

The DRAGEN Copy Number Variant (CNV) Pipeline performs ultra-rapid analysis of next-generation sequencing (NGS) data to identify copy number variants. The DRAGEN CNV Pipeline allows for processing of both whole genome and whole exome samples, to detect CNVs arising from germline or somatic mutations, and can be integrated with any existing NGS workflow.

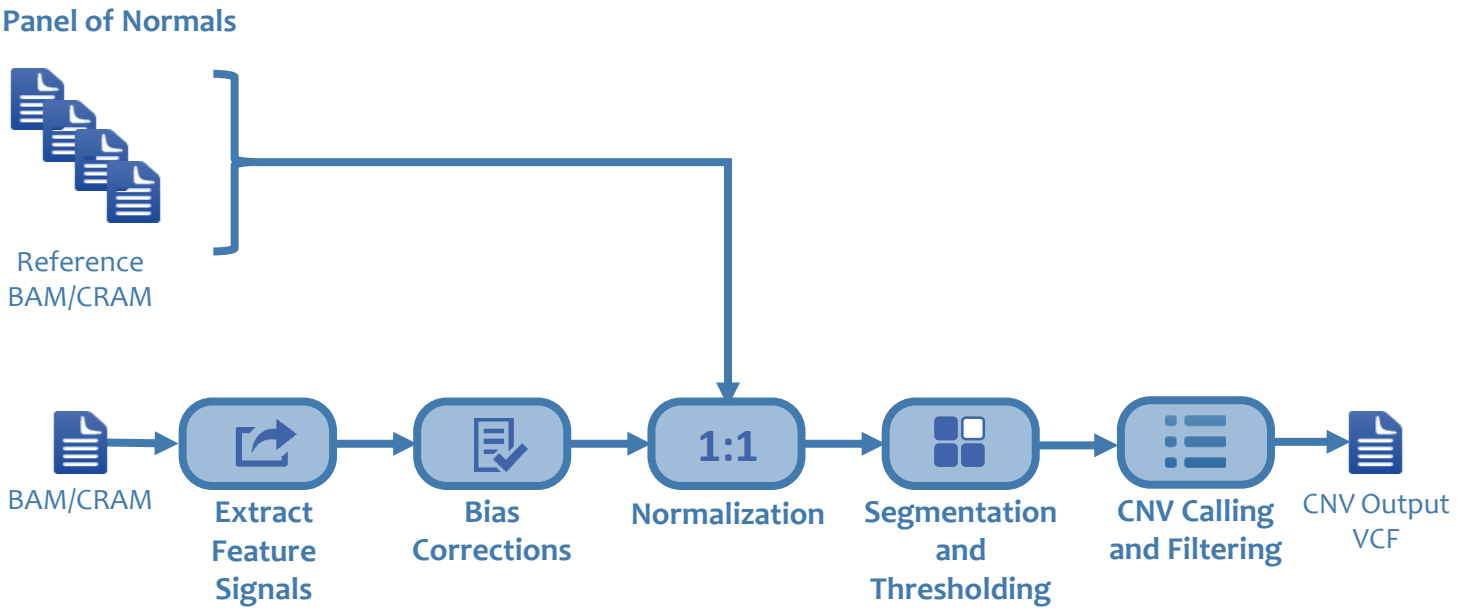
Each stage of the DRAGEN CNV Pipeline is highly configurable making it adaptable to many NGS applications. Selecting samples for a panel of normals allows for flagging samples from different sequencers or library prep kits through correlation metrics. Users may opt to use as few as a single reference sample, or as many as an entire cohort from a population study. A sex genotyper aids in analysis of non-autosomal chromosomes, automatically handling ploidy issues.

Highlights

- **Multiple Applications**—support for whole genome, whole exome, germline, somatic, gene panels, and low pass scenarios
- **Resolution**—detects events down to 100s of bp, such as single exon events in targeted sequencing, or up to whole chromosomal events such as in somatic samples
- **Rapid**—analyzes genomic data significantly faster than other CNV callers. When used in conjunction with DRAGEN mapping and aligning for WGS, CNV analysis takes ~30 minutes
- **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

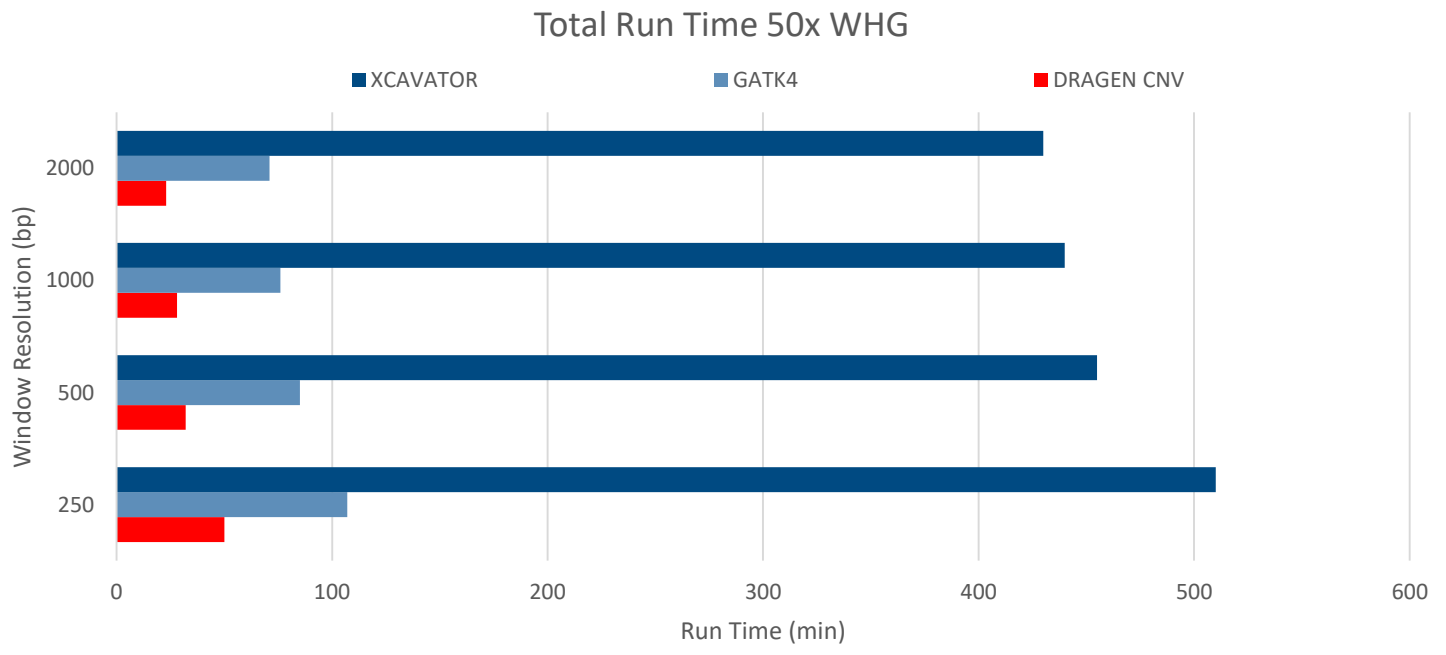
DRAGEN Copy Number Variant (CNV) Pipeline

The DRAGEN CNV Pipeline accepts alignments in BAM/CRAM format. The alignments are processed through the CNV pipeline along with reference samples based on the application. Feature signals such as read counts, paired end, split reads, and allele frequency are calculated per sample. Artifacts such as GC content and capture kit biases are corrected for prior to normalization. Normalization of the case sample via principal component analysis (PCA) and singular value decomposition (SVD) removes noise from the sample's signal. Segmentation is performed on the normalized sample prior to CNV calling and statistical genotyping. Various stages of filtering can be applied to mitigate false positives before emitting the final calls in a standard VCF file as output.



DRAGEN CNV Pipeline Speed

The DRAGEN CNV Pipeline leverages the hardware acceleration of the DRAGEN platform to accelerate the processing of NGS data, allowing for large-scale CNV studies. The detection of CNV involves computationally complex algorithms which often times does not scale well for a large number of samples on general processors. The below table highlights the speed improvement at which DRAGEN is capable of compared to other popular CNV tools.



Pipeline Steps

Input/Output File Formats
- BAM/CRAM
- CNV Output VCF

Extract Feature Signals
- Read Counts, Paired End, Split Reads, Allele Frequency, Sample Variance

Bias Correction
- GC Content, Repeat Regions and Mapability, Target Capture Biases

Normalization
- Data normalized against panel of normals using PCA and SVD

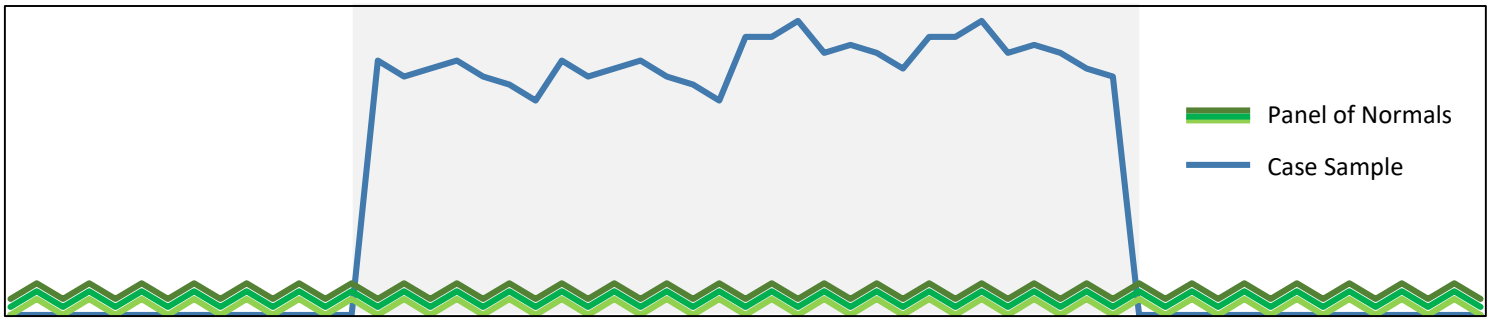
Segmentation and Thresholding
- Segmentation based on statistical changes in signal levels

CNV Calling and Filtering
- Segments merged and regions outside of copy-neutral levels are called and annotated

DRAGEN CNV Process

Extract Feature Signals

Read count extraction comparing case sample against a panel of normals



Segmentation

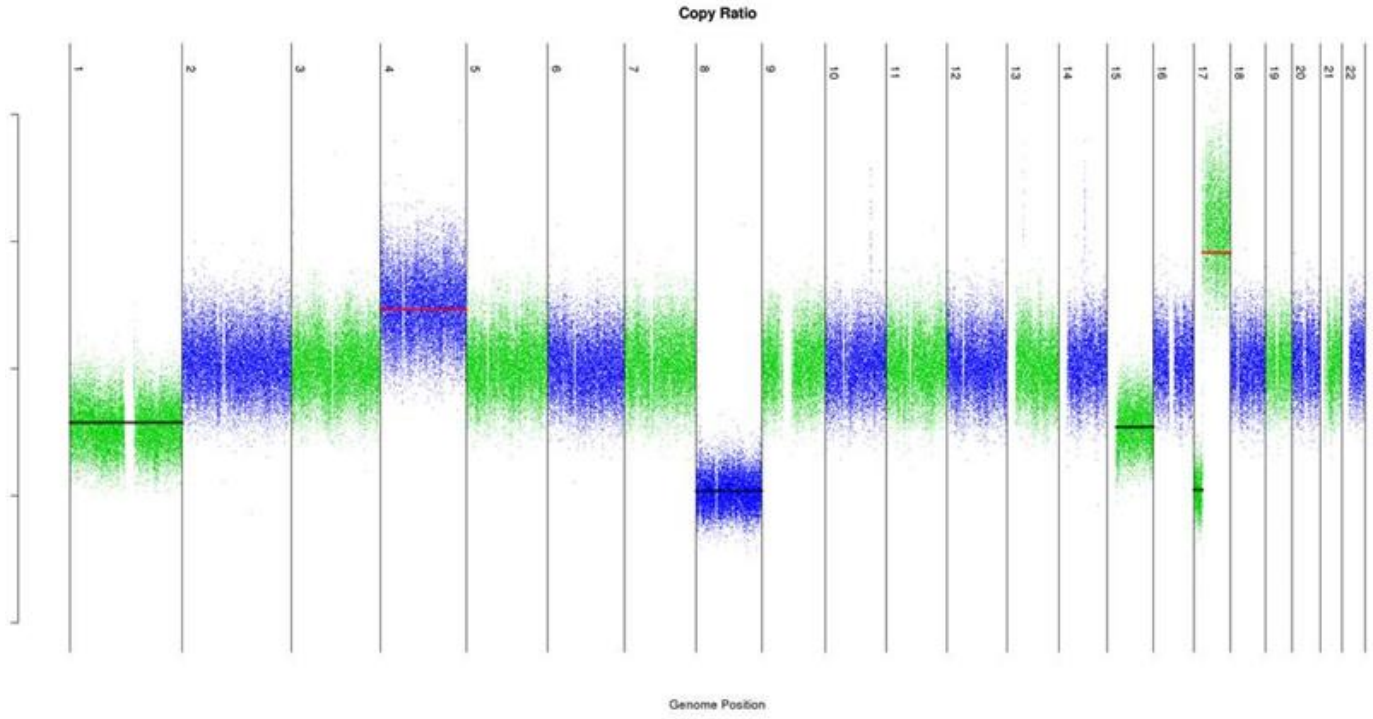
Segmentation of the normalized signal produces candidate events

CNV Calling and Filtering

Adjacent segments with similar means are merged and statistically called

DRAGEN CNV Pipeline – Cancer Application

DRAGEN extracts out the CNV signal from the tumor sample after bias corrections and normalization against a matched normal sample. With large chromosomal aberrations such as those found in cancer samples, the CNV events are visible from coverage plots such as the one shown below. This medulloblastoma cancer sample exhibits both deletions and duplication events, some of which span entire chromosomes.



Clinical Applications – Low Pass Whole Genome

A popular application in clinical studies is the use of low-pass (low coverage, such as 5x) whole genome samples for the detection of pathogenic CNVs. Typically this entails detection of events on the order of >10kb. The resolution of the DRAGEN CNV Pipeline can be controlled to save time and money in this case.

WGS Coverage	Recommended Resolution*
5x	>1000bp
10x	>1000bp
20x	>500bp
30x	>250bp
50x	>250bp

*users can always choose to trade off between resolution and speed

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.

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