DRAGEN™

DRAGEN RNA Pipeline

Ultra-Rapid Transcriptome Data Analysis

3min	97.8%	83.3%	94.7%	91.5%
FASQ > BAM	Specificity	Sensitivity	Specificity	Sensitivity
80 Million Reads	Read Alignment		Splice Junction	

Overview

The DRAGEN RNA Pipeline performs Next Generation Sequencing (NGS) secondary analysis of RNA transcripts. The RNA Pipeline offers multiple operating modes, including reference-only alignment and annotation-assisted alignment with gene fusion detection. The DRAGEN Gene Fusion module leverages the DRAGEN RNA spliced aligner to perform split-read analysis on supplementary (chimeric) alignments to detect potential breakpoints, while adding minimal processing time to the overall pipeline.

DRAGEN transcriptome alignments are compatible with downstream transcript assembly tools, novel transcript discovery, differential gene expression, and other third party applications.

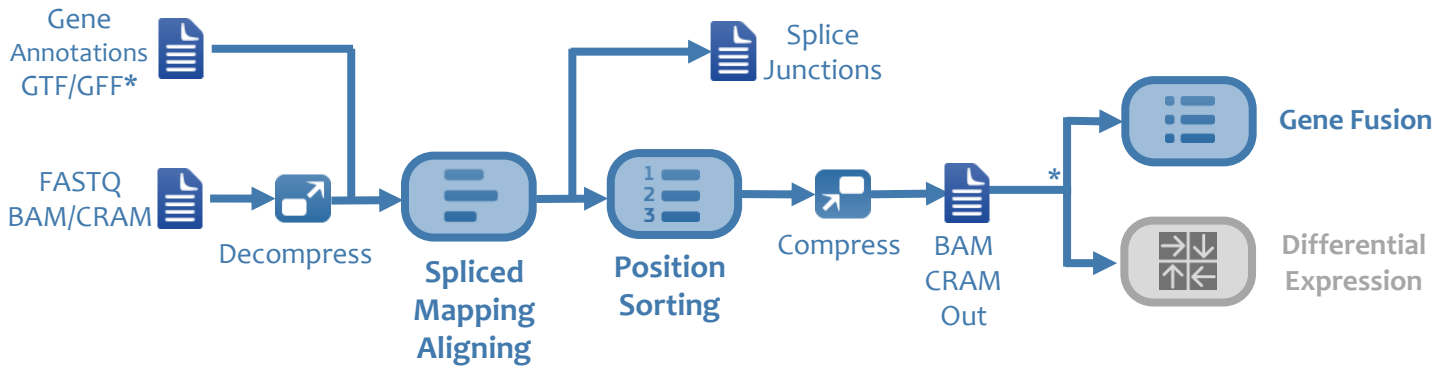
Highlights

- **Highly Accurate**—detects more true positives and less false positives
- **Sensitive**—demonstrates higher sensitivity than leading RNA analysis tools
- **Rapid**—analyzes genomic data orders of magnitude faster than other software on the market
- **Flexible**—compatible with multiple downstream analysis tools
- **Easy to Use**—fully functional API, CLI, and GUI
- **Accessible**—can be used onsite, in the cloud, or as a hybrid cloud solution

edico genome

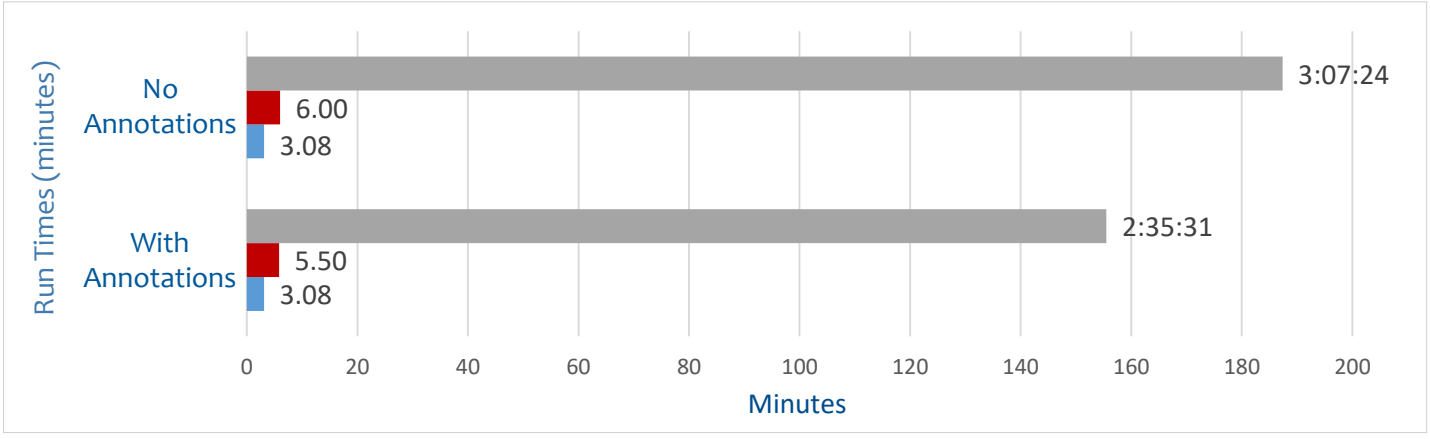
DRAGEN RNA Pipeline

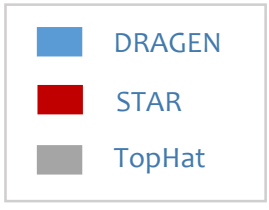
The DRAGEN RNA Pipeline accepts input FASTQ/BAM/CRAM and produces an output aligned BAM/CRAM. DRAGEN offers the option to input a gene annotations file (GTF) to guide the spliced alignments. DRAGEN is also capable of running in a "2-pass" mode which uses novel splice junctions, as detected in the first pass, to guide the second pass mapping / aligning phase.



DRAGEN RNA Pipeline Speed

The DRAGEN RNA Pipeline offers multiple modes, including reference-only alignment and annotation-assisted alignment. The alignment accuracy and splice junction discovery accuracy tables for each mode are shown on the following pages. The reference-only alignment and annotation-assisted alignment pipelines were performed using the Engstrom Sim2 Dataset*.





Dataset*	DRAGEN	STAR 2.5.0a	TopHat 2.0.14
No Annotations	0:03:08	0:05:50	2:35:31
With Annotations	0:03:08	0:06:00	3:07:24

* BEERS Sim 2 datasets obtained from Nature Methods – Systematic evaluation of spliced alignment programs for RNA-seq data. doi:10.1038/nmeth.2722

Applications

Pipeline Steps

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

Spliced Mapping/Aligning
- Single end or paired ends

Splice Junction Output
- Format similar to STAR's SJ.out.tab
- User-configurable junction filters

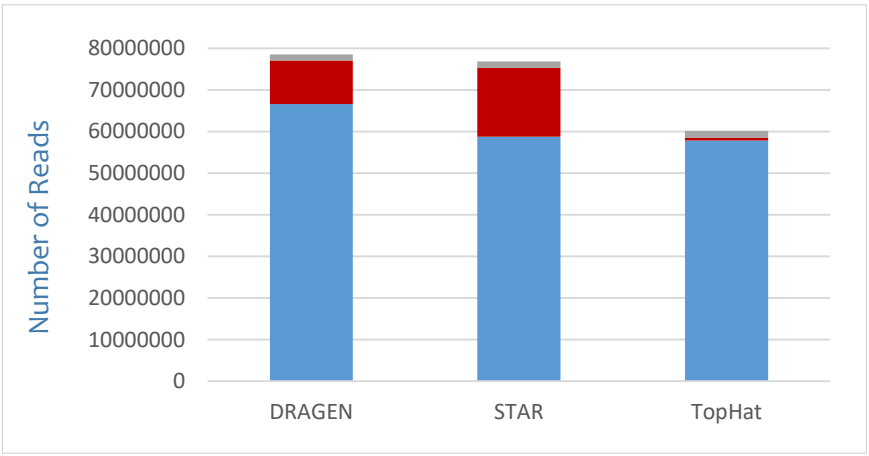
Position Sorting
- Binning by reference range
- Sorting of bins by reference position

Gene Fusion and Downstream Analysis Tools*
- Split-read analysis to detect potential breakpoints
- Outputs compatible with downstream tools
- Tools include featureCounts and DESeq

*Differential expression is underdevelopment

Alignment Accuracy (Reference-Only Alignment)*

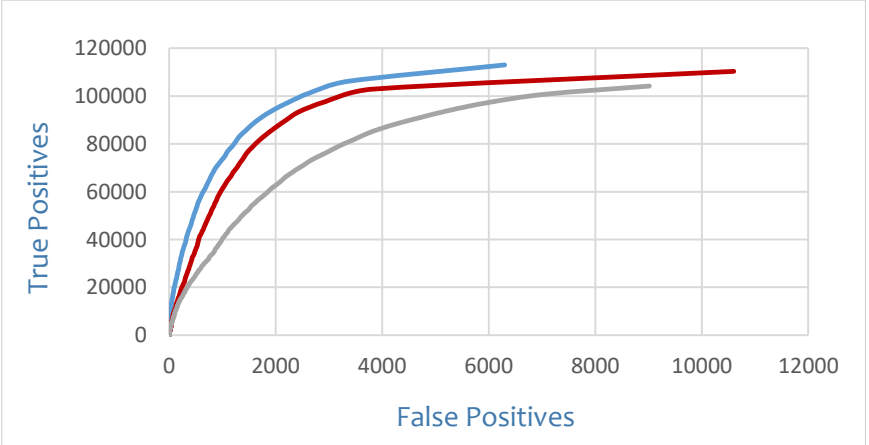
Aligner	True Positive Alignments	False Positive Alignments	Specificity**	Sensitivity
DRAGEN	66641693	1485738	97.81	83.30
STAR	58793482	1578077	97.38	73.49
TopHat	57837349	1597822	97.31	72.29



Read Alignment Accuracy
Each bar plot shows the number of perfect alignments (all bases in read aligned correctly), number of partially correct alignments (at least one base aligned correctly but not all) and totally incorrect alignments.

Splice Junction Detection (Reference-Only Alignment)*

Aligner	True Positive Alignments	False Positive Alignments	Specificity**	Sensitivity
DRAGEN	113066	6298	94.72	91.49
STAR	110293	10601	91.23	89.25
TopHat	104207	9013	92.04	84.32

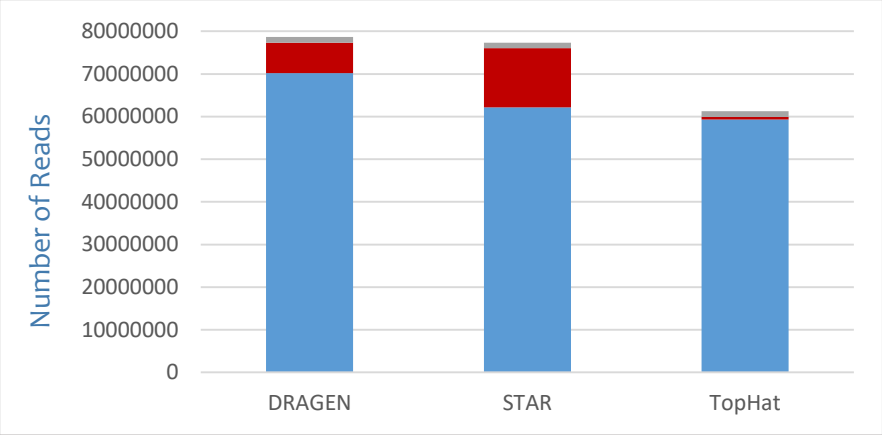


Splice Junction Discovery
Cumulative counts of true and false junctions were computed over a range of thresholds for the number of supporting alignments. A point further to the left on a curve has a higher supporting alignment count threshold than a point to the right.

*BEERS Sim 2 datasets obtained from Nature Methods – Systematic evaluation of spliced alignment programs for RNA-seq data. doi:10.1038/nmeth.2722
**Sensitivity is defined by True Positives / (Total Read Count). Specificity is defined as True Positives / (True Positives + False Positives).

Alignment Accuracy (Gene Annotation Input)*

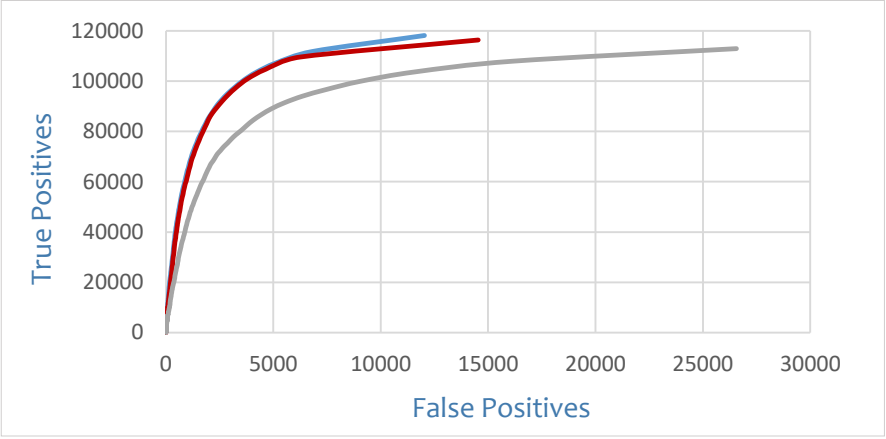
Aligner	True Positive Alignments	False Positive Alignments	Specificity**	Sensitivity
DRAGEN	66641693	1485738	98.17	87.78
STAR	58793482	1578077	97.96	77.66
TopHat	57837349	1597822	97.80	74.10



Read Alignment Accuracy: Annotations
With gene annotation input, DRAGEN perfectly aligns at least 10% more reads than STAR or TopHat. The annotation assisted alignment pipelines were also performed using the Engstrom Sim2 Dataset.

Splice Junction Detection (Gene Annotation Input)*

Aligner	True Positive Alignments	False Positive Alignments	Specificity**	Sensitivity
DRAGEN	118219	12030	90.76	95.66
STAR	116401	14532	88.90	94.19
TopHat	112944	26558	80.96	91.39



Splice Junction Discovery: Annotations
GTF format is used to improve the sensitivity of splice junction discovery. DRAGEN may take a GTF as input, providing the pipeline with the precise locations of known splice junctions for a given species.

* BEERS Sim 2 datasets obtained from Nature Methods – Systematic evaluation of spliced alignment programs for RNA-seq data. doi:10.1038/nmeth.2722
**Sensitivity is defined by True Positives / (Total Read Count). Specificity is defined as True Positives / (True Positives + False Positives).

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 FPGA technology, DRAGEN delivers best-in-class accuracy, speed, 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. For more information about DRAGEN, visit www.edicogenome.com.

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

edico genome