DRAGEN™

DRAGEN Transcriptome Pipeline

Ultra-Rapid RNA-seq Data Analysis

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

Overview

The DRAGEN Transcriptome (RNA-Seq) Pipeline performs Next Generation Sequencing (NGS) secondary analysis of RNA transcripts. The Transcriptome Pipeline offers multiple operating modes, including reference-only alignment and annotation-assisted alignment.

DRAGEN transcriptome alignments are compatible with downstream transcript assembly tools, novel transcript discovery, differential gene expression, gene fusion detection, and other RNA-Seq applications.

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.

In addition to improved speed, the DRAGEN Transcriptome Pipeline achieves greater accuracy as verified on standard benchmark datasets.

How Does it Work?

DRAGEN takes raw read data produced by a sequencing instrument, such as Illumina's HiSeq 2500. For the Transcriptome Pipeline, DRAGEN outputs a Splice Junction file and an aligned BAM/CRAM file that is compatible with downstream analysis tools.

The DRAGEN platform includes a fully functional and easy to use graphical user interface (GUI) and Workflow Management System, enabling customers to easily schedule multiple workflow runs, analyze results, compare different pipelines, monitor multiple networked DRAGEN cards and receive update software releases.

Comprehensive Set of Pipelines

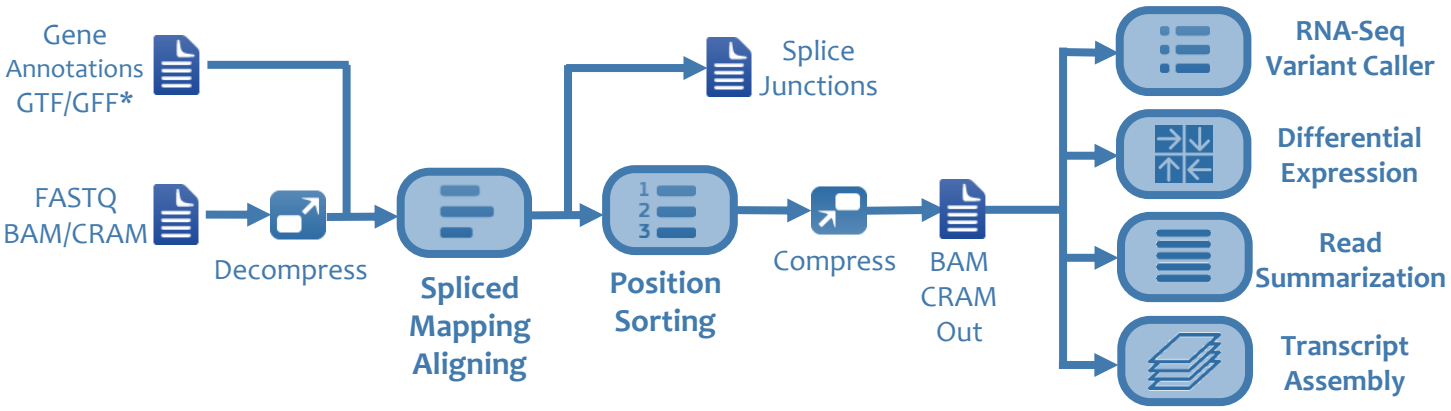
DRAGEN supports pipelines for Whole Genome, Whole Exomes, Targeted Panels, Epigenome / Methylome, RNA-seq / Transcriptome, Microbiome and Cancer Tumor / Normal.

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

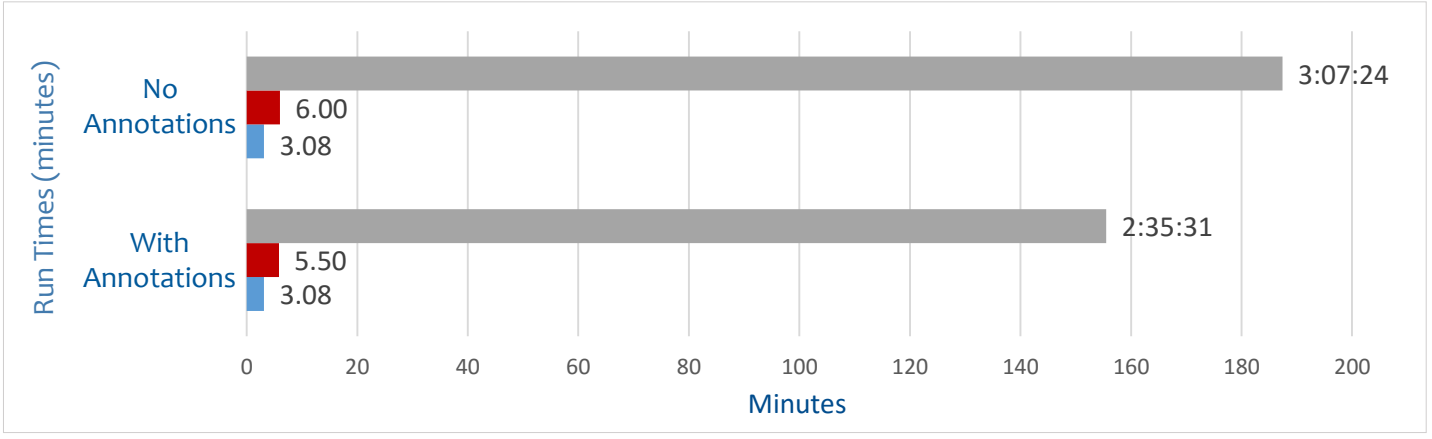
DRAGEN Transcriptome Pipeline

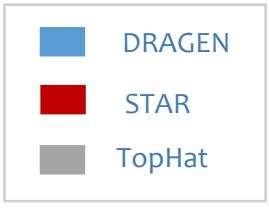
The DRAGEN Transcriptome pipeline accepts input FASTQ/BAM/CRAM and produces an output aligned BAM/CRAM. DRAGEN offers the option to input a gene annotations file (GTF) to guide the spliced alignments. DRAGEN is also capable of running in a "2-pass" mode which uses novel splice junctions, as detected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



Transcriptome Pipeline Speed

The DRAGEN Transcriptome 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

- **Developmental Studies**
- **Cancer Testing**
- **Food Supply Safety**
- **Drug Discovery**
- **Differential Expression Research**

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
- BCL can also be processed directly

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

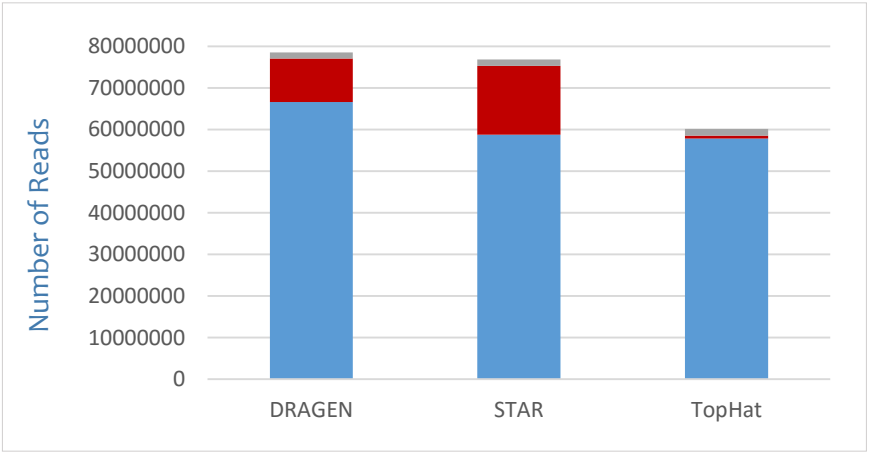
Position Sorting
- Binning by reference range
- Sorting of bins by reference position

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

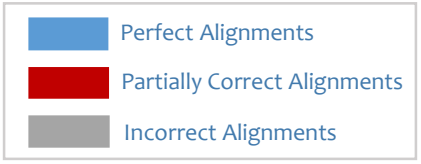
Downstream Analysis Tools
- Outputs compatible with downstream tools
- Tools include featureCounts and DESeq

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

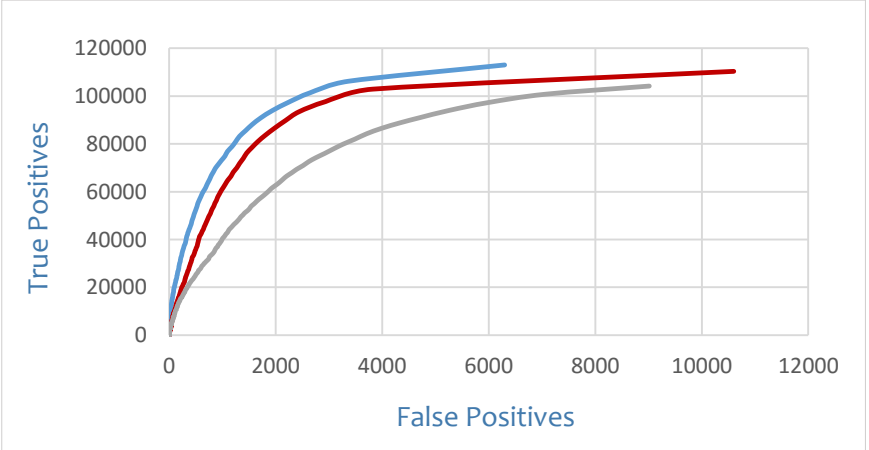


Read Alignment Accuracy
Each bar plot shows the number of perfect alignments (all bases in read aligned correctly), number of partially correct alignments (at least one base aligned correctly but not all) and totally incorrect alignments.

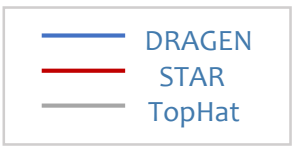


Splice Junction Detection (Reference-Only Alignment)

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



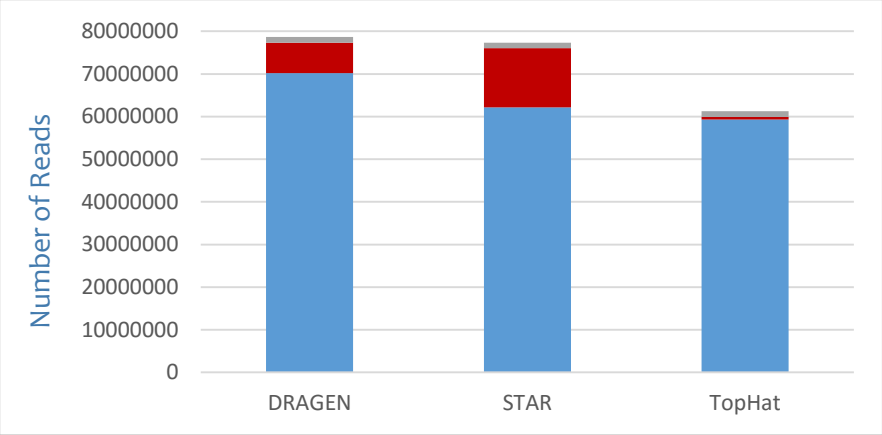
Splice Junction Discovery
Cumulative counts of true and false junctions were computed over a range of thresholds for the number of supporting alignments. A point further to the left on a curve has a higher supporting alignment count threshold than a point to the right.



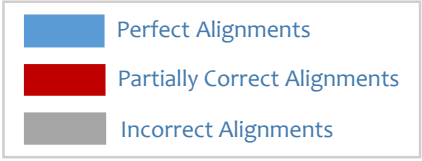
**Sensitivity is defined by True Positives / (Total Read Count). Specificity is defined as True Positives / (True Positives + False Positives)

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



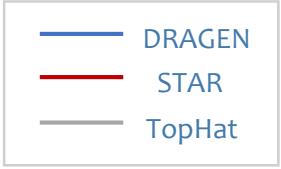
Read Alignment Accuracy: Annotations
With gene annotation input, DRAGEN perfectly aligns at least 10% more reads than STAR or TopHat. The annotation assisted alignment pipelines were also performed using the Engstrom Sim2 Dataset.



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.



**Sensitivity is defined by True Positives / (Total Read Count). Specificity is defined as True Positives / (True Positives + False Positives)

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