22-FEB-2021 CAP #8205

Illumina DRAGEN™ Bio-IT Platform

Raul Torrieri

Field Bioinformatics Scientist – Illumina LATAM

illumina®

Setting the Stage

1

illumina®

Investing in Secondary Analysis

Business Considerations for Investing in Secondary Analysis

01

Upfront Investments

Key things to consider when implementing secondary analysis include:

- Upfront hardware investments
- Bioinformatics & IT teams
- Compute and storage costs
- Direct instrument integration
- Security and compliance
- Ensuring accuracy
- Turnaround time
- Out of the box / Plug and play

Ability to Scale as You Grow

When adopting NGS, it's essential to choose a secondary analysis solution that is built for future expectations of scale.

02

Sequencing at Scale

Conducting NGS at scale poses a unique set of challenges

Cost

Speed

Reproducibility

Scalability

Accuracy

High Performance Secondary Analysis

01

Accurate

Detect SNPs and INDELs with high sensitivity and specificity

02

Ultra-rapid

Reduce analysis to ~25 minutes per 30x whole genome

03

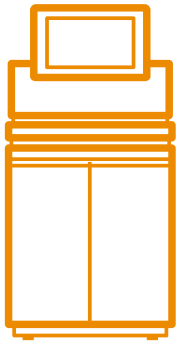
Cost-efficient & scalable

Scale as needed while keeping costs low

Translating Sequencing Data into Insights

Library prep | Sequencing | Secondary analysis | Tertiary analysis

DRAGEN Germline Example



BCL

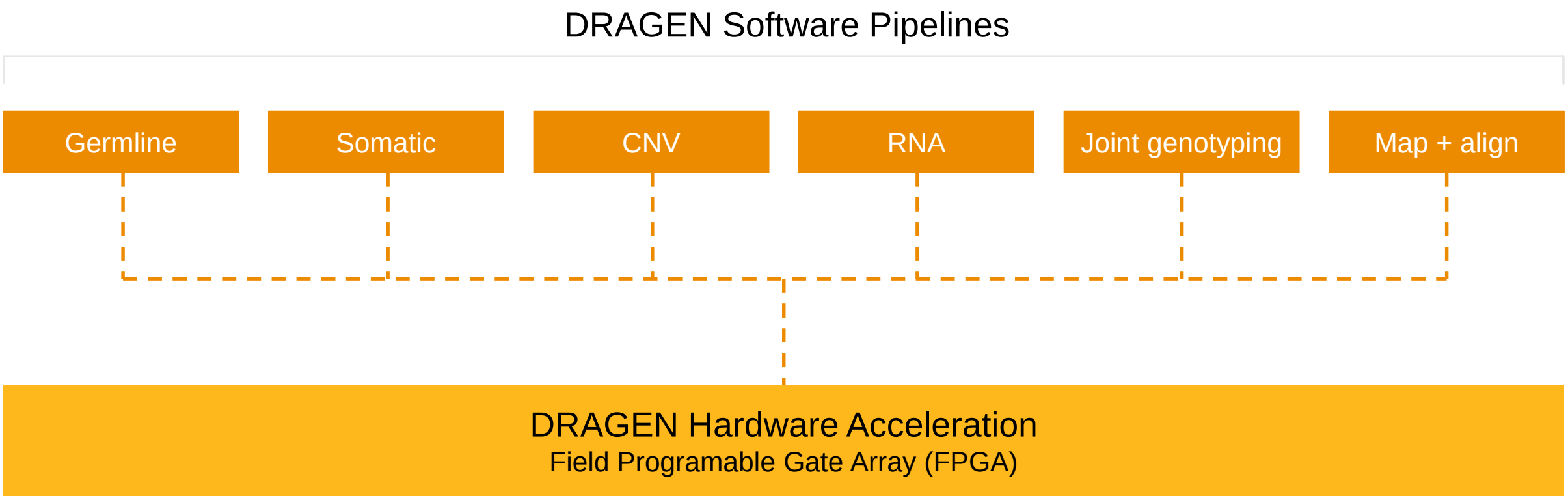
- BCL Conversion
- Mapping + Aligning
- Duplicate Marking
- Variant Calling

VCF

From bases to variants in under 30 minutes

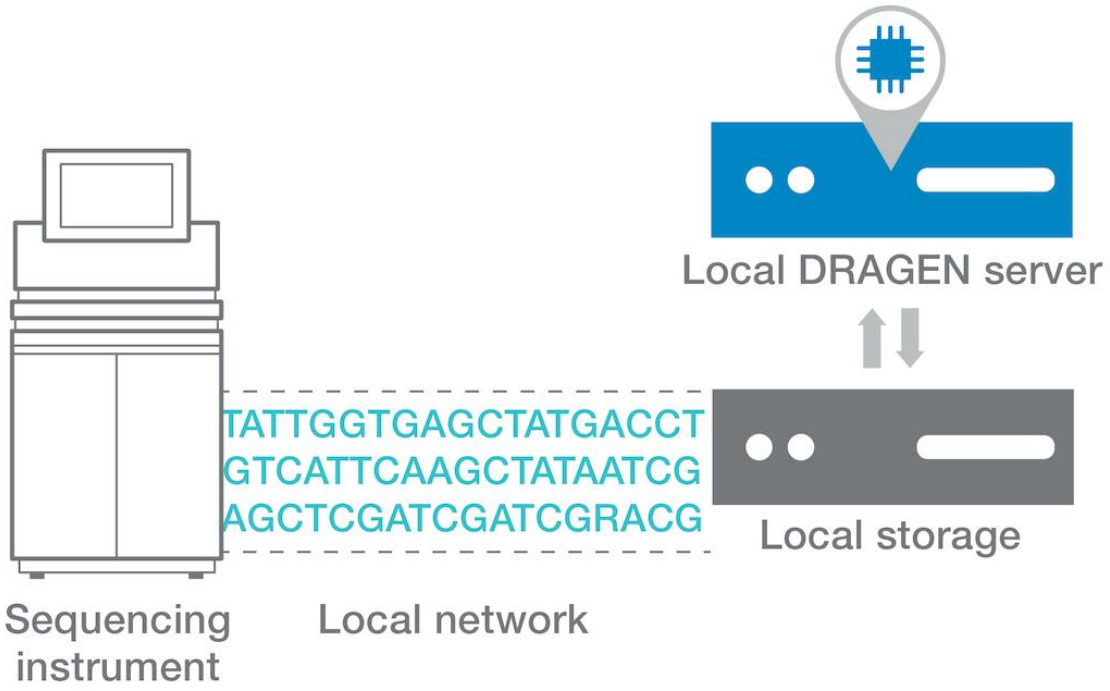
DRAGEN™ is Hardware-Accelerated Secondary Analysis

Dynamic Read Analysis for GENomics



Flexible Data Analysis

On-premise



BaseSpace™ Sequence Hub

Versatile Suite of Secondary Analysis Pipelines

All DRAGEN Pipelines can run on a single DRAGEN Platform

Illumina DRAGEN Germline Pipeline

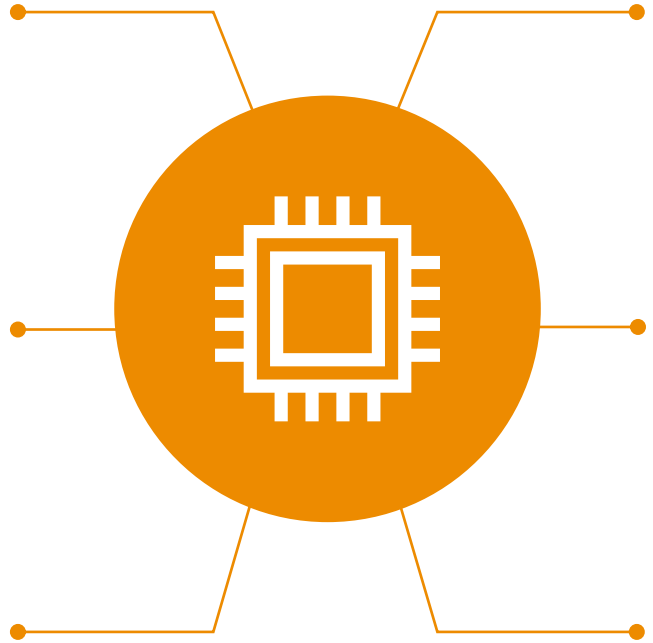
End-to-end (BCL → VCF) NGS analysis, including advanced error model calibration for increased accuracy, and repeat expansion detection and genotyping through Illumina Expansion Hunter.

Illumina DRAGEN Somatic Pipeline

Includes tumor-only and tumor–normal modes, designed for detecting somatic variants in tumor samples.

DRAGEN RNA Pipeline

Performs transcriptome analysis starting with splice junction discovery and alignment, followed by gene fusion detection.



DRAGEN Joint Genotyping Pipeline

Calls variants across the genomes of multiple individuals, and scales to large cohorts of samples at expedited speeds with uncompromising accuracy.

DRAGEN CNV Pipeline

Performs copy number variant (CNV) analysis for germline and somatic exomes and genomes. Various levels of filtering can be applied to mitigate false positives before emitting the final calls

DRAGEN Map + Align Pipeline

Capable of ultra-rapid mapping and aligning DNA and RNA for both exomes and genomes.

New and enhanced DRAGEN pipelines are released periodically. For a comprehensive list, visit www.Illumina.com/DRAGEN

The Values of DRAGEN

2

illumina®

Easy to Implement

From command line to push-button

Bioinformatics Expertise

Push-button

- Simple Graphical User Interface (GUI)
- Managed service
- HIPAA* and GDPR compliance
- Workgroup capabilities
- Easy data sharing

BaseSpace SEQUENCE HUB

Single command launch

- Easy-to-learn Linux based Command Line Interface (CLI)
- Simple command line execution

Advanced command line

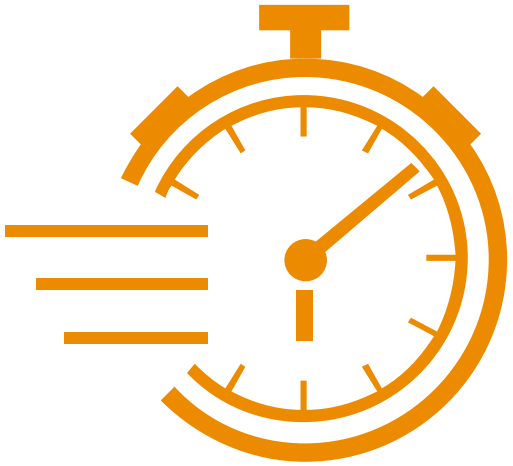
- Script back-to-back jobs
- Make different configuration files for different applications

On-premise

*HIPAA compatibility applies in the US only with BaseSpace Enterprise

Record-Breaking Analysis Speed

Guinness World Records®
Fastest Genetic Diagnosis



Sample > Answer

19.5 Hours

In 2018, Rady Children's Institute for Genomic Medicine set the Guinness World Records® for Fastest Genetic Diagnosis leveraging the Illumina DRAGEN Bio-IT Platform.

Guinness World Records®
Fastest Analysis of 1,000 Genomes

1000WHG

2hr25min

In 2017, Children's Hospital of Philadelphia (CHOP) set the Guinness World Records® for Fastest Analysis of 1,000 Genomes using the Illumina DRAGEN Bio-IT Platform in the cloud.

Accelerated Speeds

Analyze more in less time

22:1

Efficiency ratio

Analyze up to 57 genomes/day
vs ~2.5/day with standard solutions

2 hours

S4 Flow Cell Demultiplexing

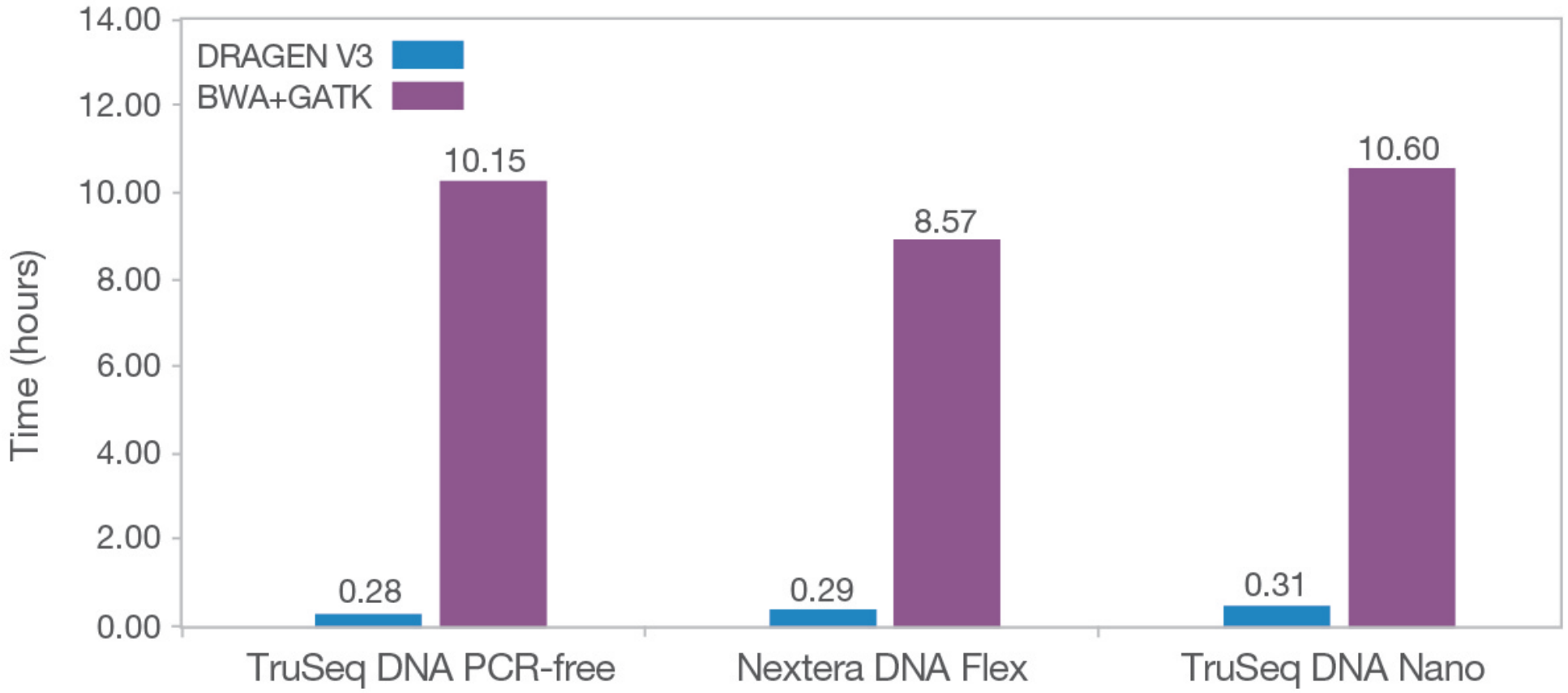
As fast as **25 minutes**

FASTQ > VCF

WHG @ 30x coverage

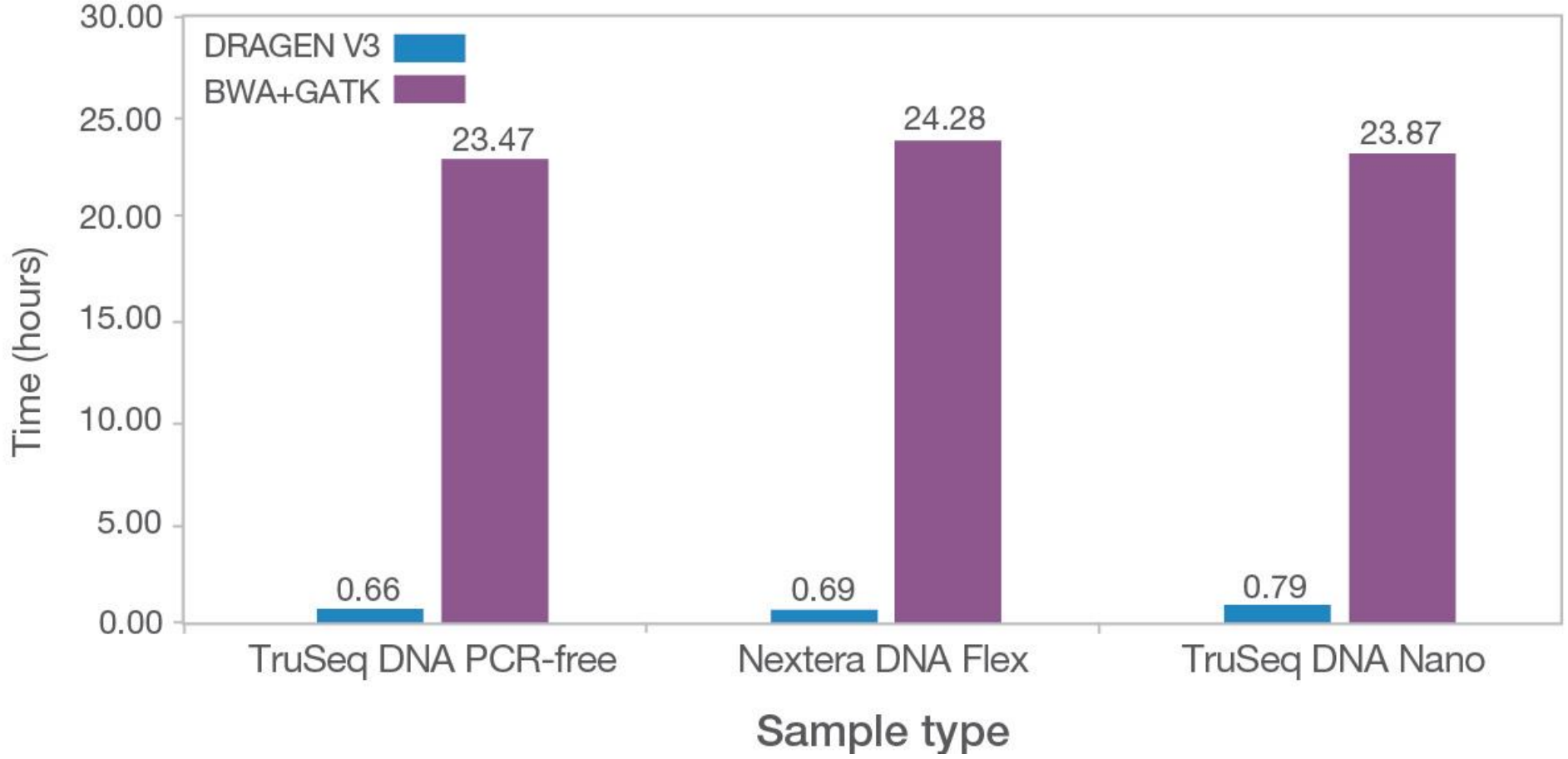
Accelerated Speeds

DRAGEN on-premise v. BWA+GATK



Accelerated Speeds

DRAGEN on BaseSpace Sequence Hub v. BWA+GATK



Accurate Data

Detects small variants, copy number variants and structural variants
with high analytical **sensitivity** and **specificity**

precisionFDA

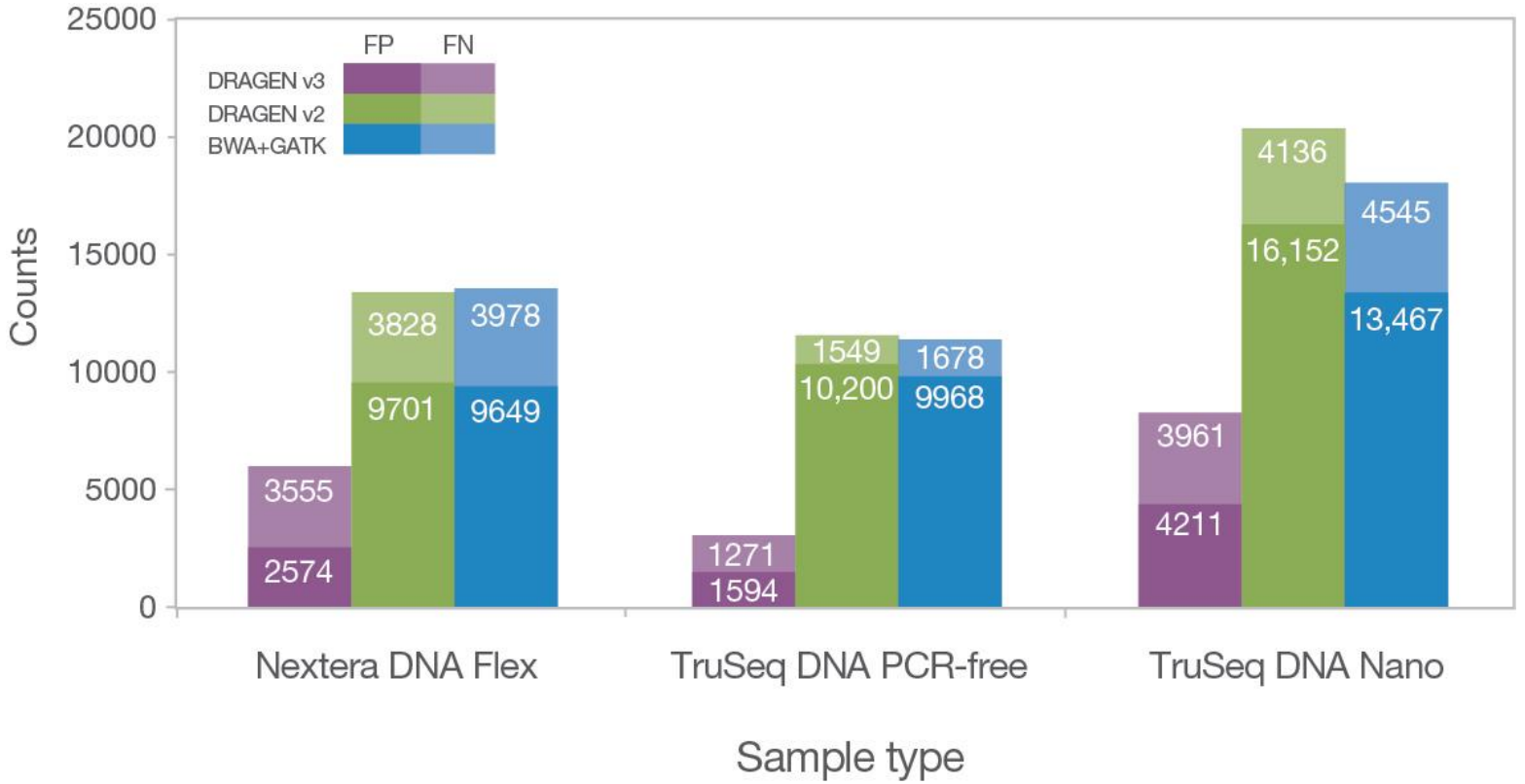
DRAGEN identified all 50 hidden variants
and ranked 1st in the following categories*

Hidden Variants	Indel Precision	Indel Recall	Indel F-Score	SNP F-score	SNP Recall
50/50	1st	1st	1st	1st	1st

*Amongst entries that identified all hidden variants

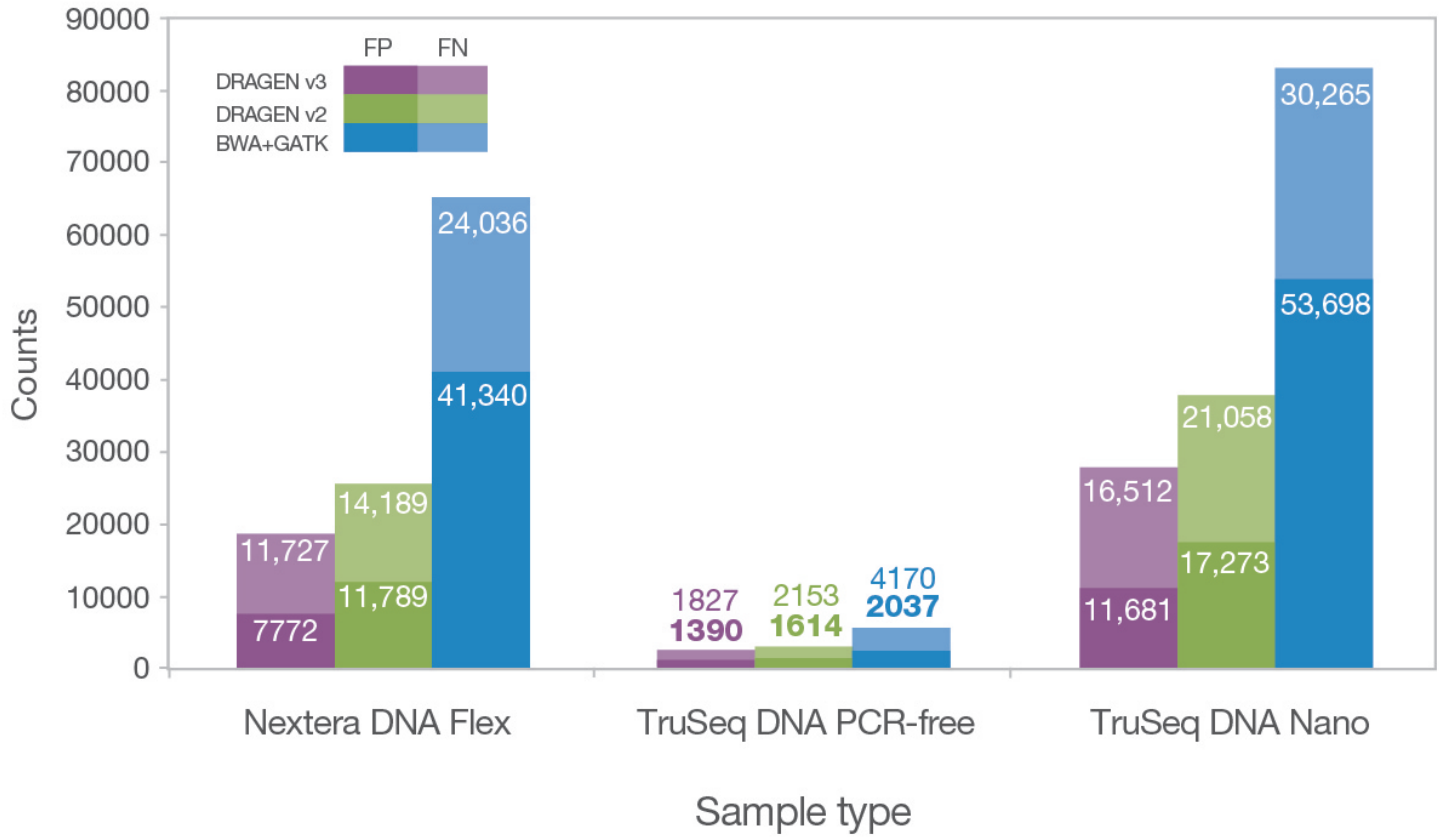
Accurate Data—Small Variant Calling SNV

DRAGEN
detects fewer false positive and false negative SNV calls



Accurate Data—Small Variant Calling Indels

DRAGEN detects fewer false positive and false negative indels



DRAGEN Metrics

The DRAGEN Platform Produces a Robust Portfolio of Metrics

Library Prep QC

Analysis QC

Demultiplexing

Duplicate Reads

Raw Data Processing

Cost-Effective

1 DRAGEN Server can **replace up to 30** traditional compute instances

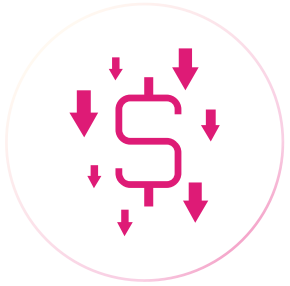
Reduce Hardware Investment

Less compute, storage, power and other associated costs.

Reduce Storage footprint by **50%** using compression (CRAM)

Smaller Storage Footprint

Seamless data compression under-the-hood with CRAM output



$5/Genome* on BaseSpace Sequence Hub

Low Cost Cloud Analysis

Most DRAGEN pipelines run at $5 per sample on BaseSpace Sequence Hub

*Approximate cost. Varies based on input sample used. FASTQ to SNVs.

DRAGEN in Practice

3

illumina®

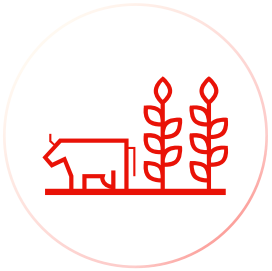
Popular Application Areas

Rare Genetic Disease

Reduces turnaround times required for genomic analysis, when fast results can be a critical factor.

Oncology

Easy-to-implement, cost effective and accelerated secondary analysis of whole exomes and genomes for cancer research.



Agrigenomics

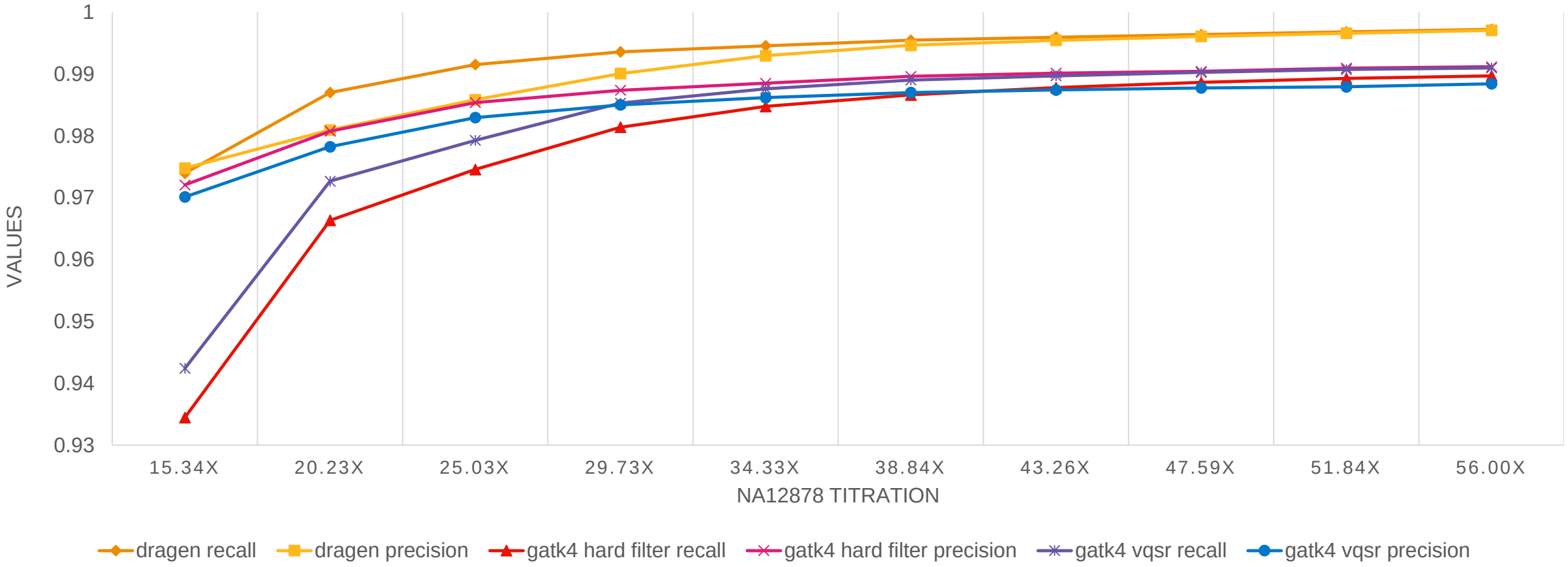
The DRAGEN Platform is well suited to handle large samples and repetitive elements. With a provided reference, the DRAGEN Platform can analyze animals and plants of varying genomic complexities.

Population Genomics

The DRAGEN Platform can rapidly analyze sequenced samples, and accelerate reanalysis as computational tools improve over time.

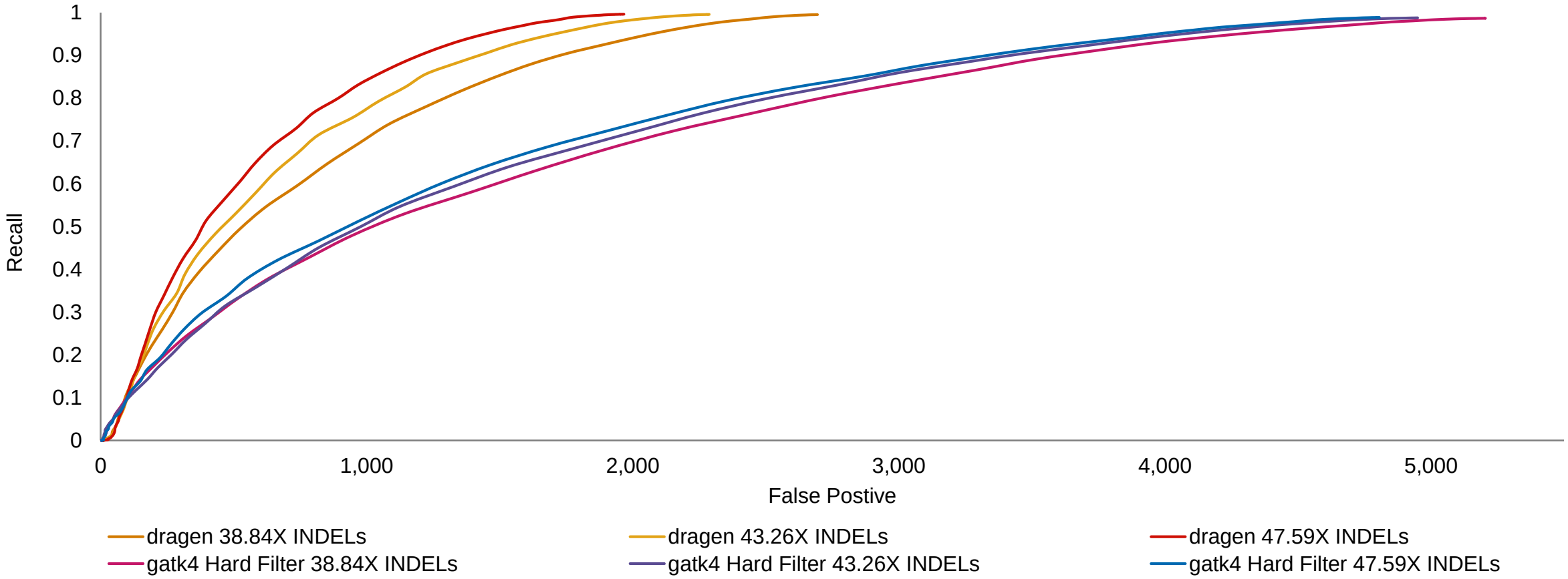
NA12878 Titration INDEL Recall and Precision

Data generated by SickKids June 2019



GATK4 Hard Filtering vs. DRAGEN ROC InDels

Data generated by SickKids June 2019



DRAGEN on BaseSpace™ Sequence Hub

4

illumina®

Combining the Power of DRAGEN with the Flexibility of the Cloud

DRAGEN on BaseSpace Sequence Hub

Accurate, rapid secondary analysis in an easy-to-use, cloud-based environment

Accurate, Fast Analysis

Simple Workflow

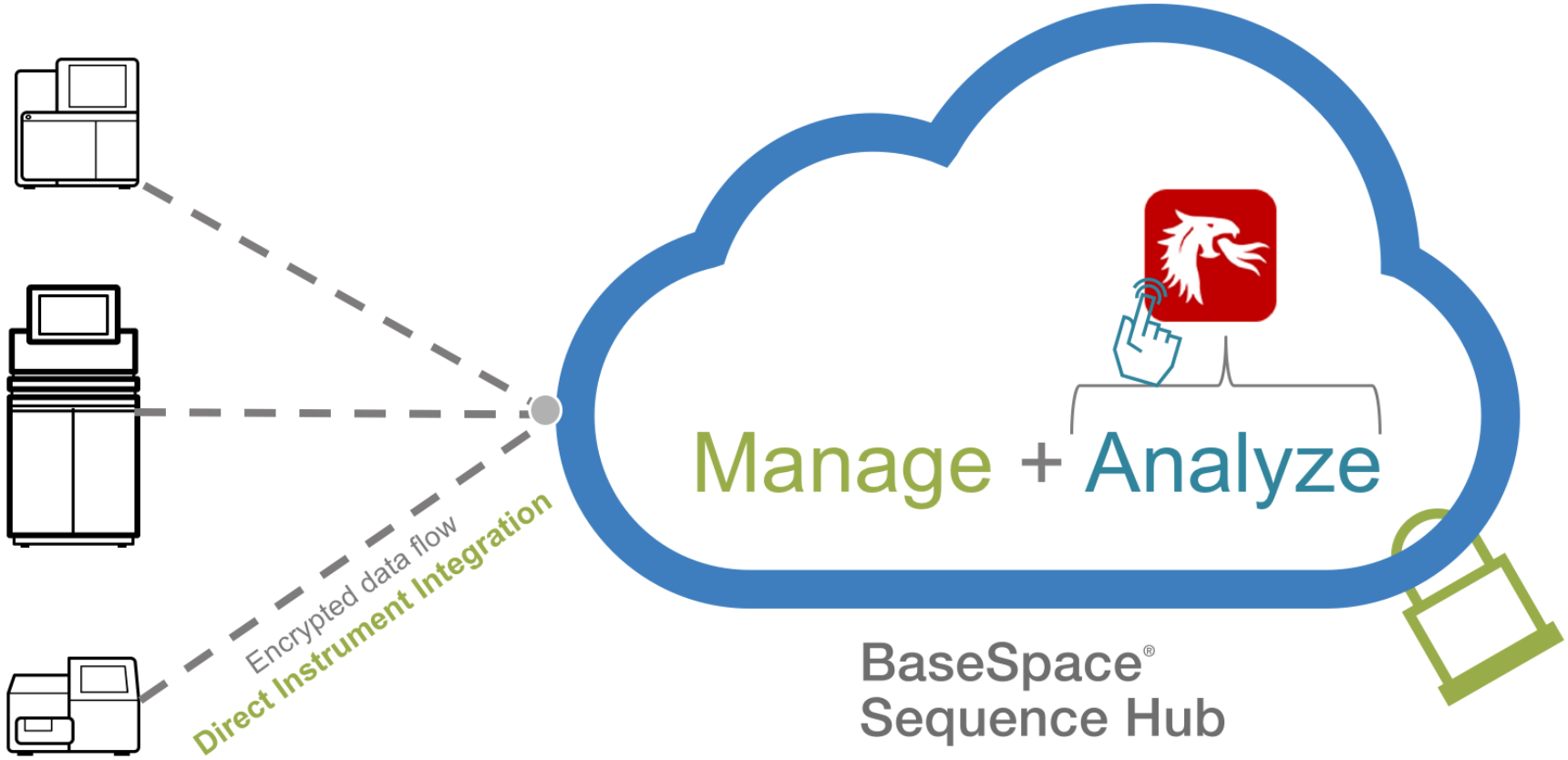
Low-cost, Scalable Platform

Secure, Compliant Environment

Available Pipelines

- DRAGEN Germline Pipeline
- DRAGEN Somatic Pipeline
- DRAGEN Enrichment Pipeline (Q4 19)
- DRAGEN RNA Pipeline
- DRAGEN Joint Genotyping Pipeline
- DRAGEN Methylation Pipeline
- DRAGEN Reference Builder

High Performance Analysis, Simple Workflow



BaseSpace Sequence Hub—Security in the Cloud

Independently audited

HIPAA COMPLIANT

Available with Enterprise Account

bsi. ISO 13485:2016 Medical Device Quality Management

FM 565017

ISO 27001 CERTIFIED by schellman

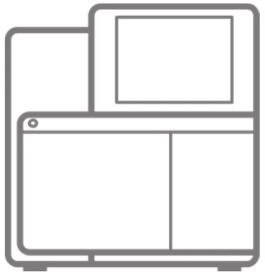
GDPR READY

DRAGEN
Pipeline Overviews

illumina®

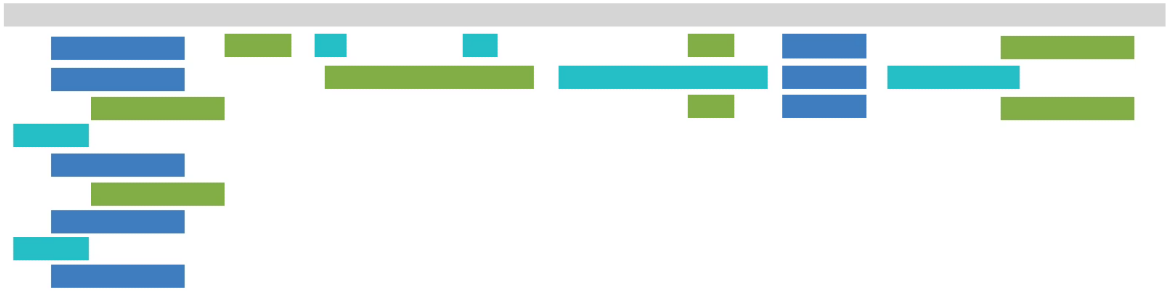
Mapping & Aligning—Germline Pipeline Example

Mapping Aligning | Position Sorting* | Duplicate Marking* | Variant Calling



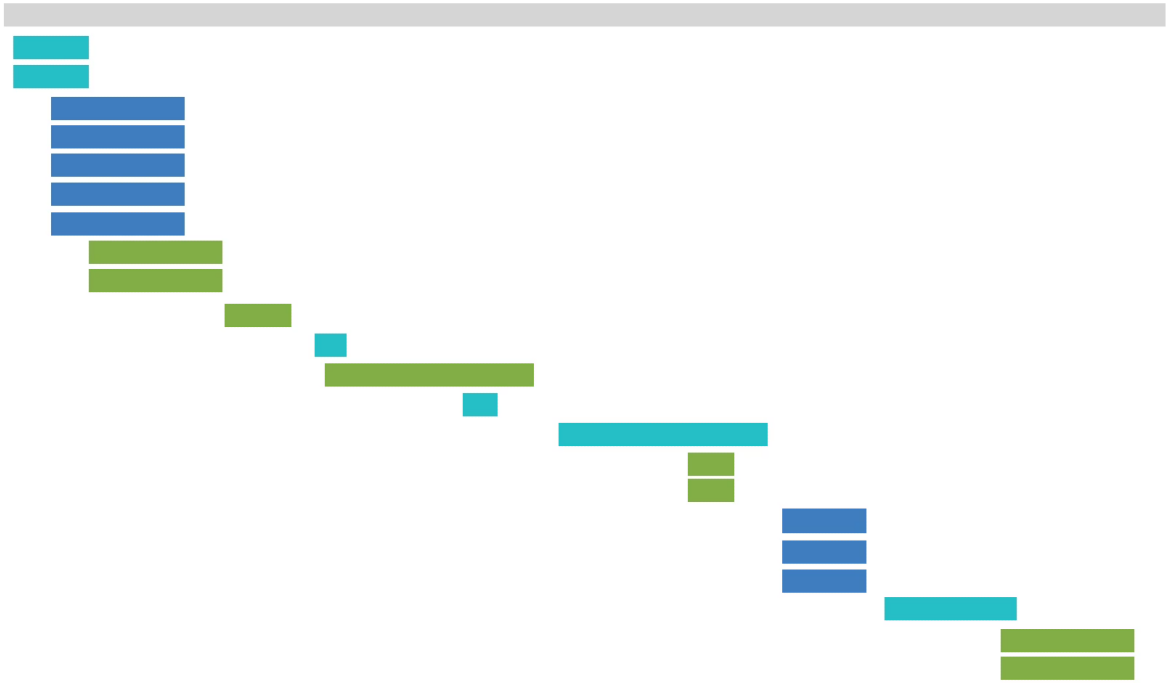
Position Sorting—Germline Pipeline Example

Mapping Aligning | **Position Sorting*** | Duplicate Marking* | Variant Calling



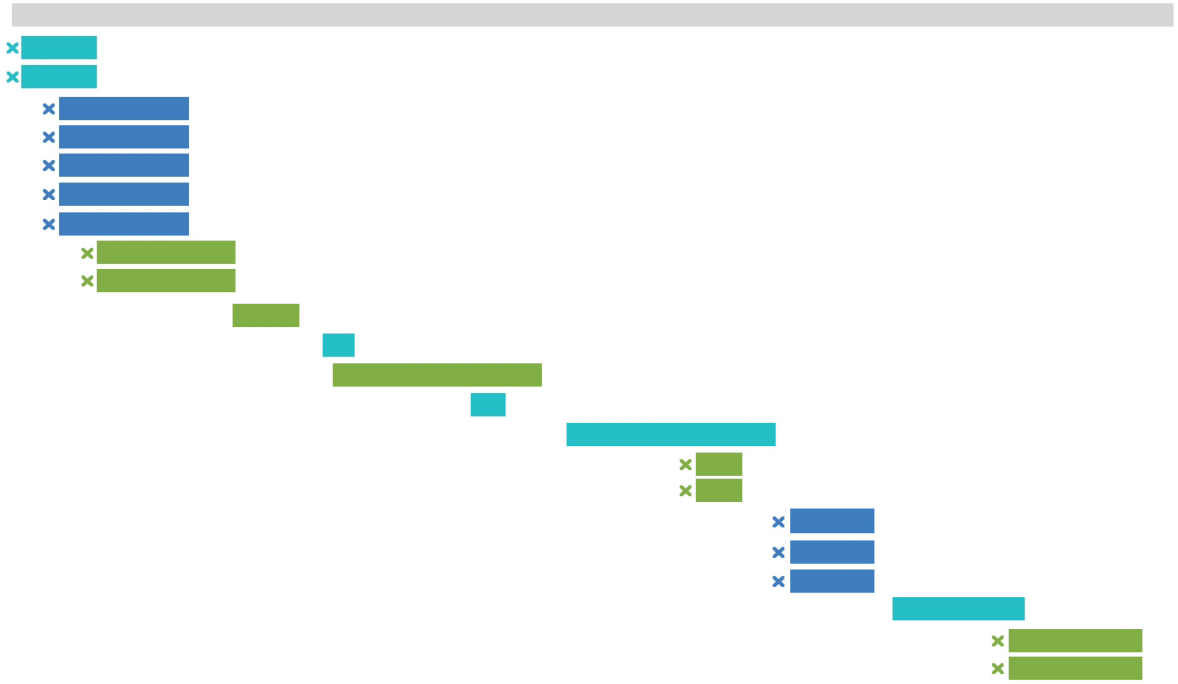
Duplicate Marking—Germline Pipeline Example

Mapping Aligning | Position Sorting* | **Duplicate Marking*** | Variant Calling

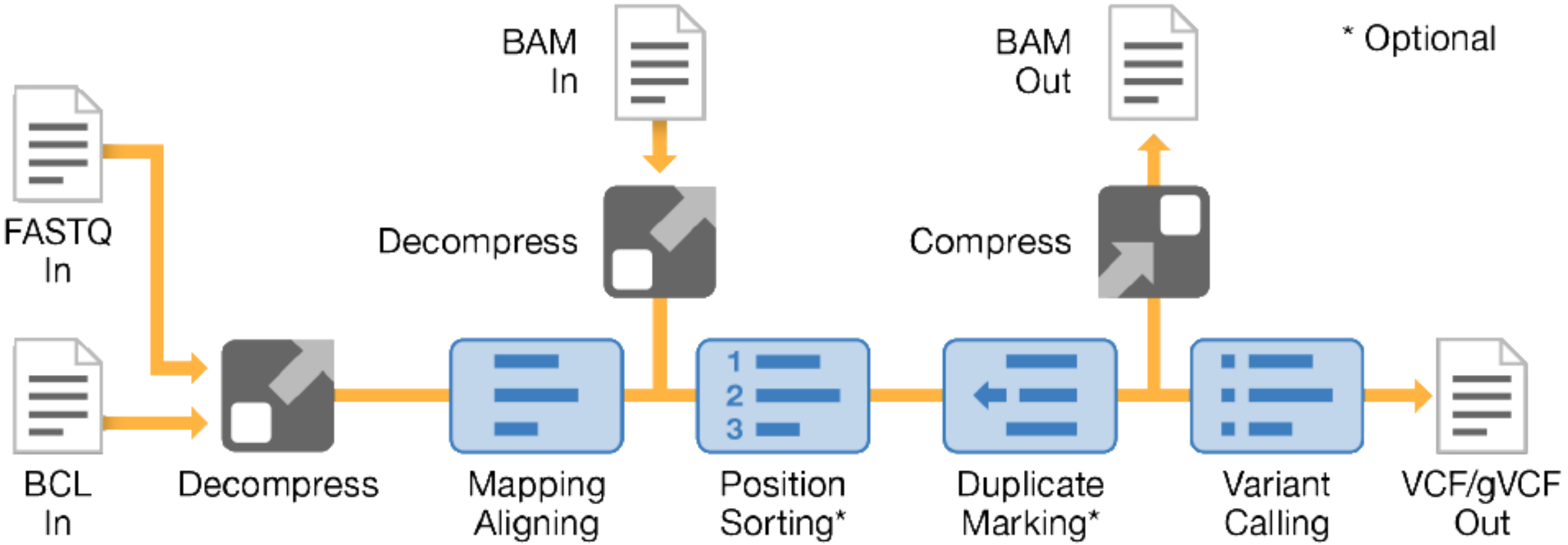


Variant Calling—Germline Pipeline Example

Mapping Aligning | Position Sorting* | Duplicate Marking* | **Variant Calling**

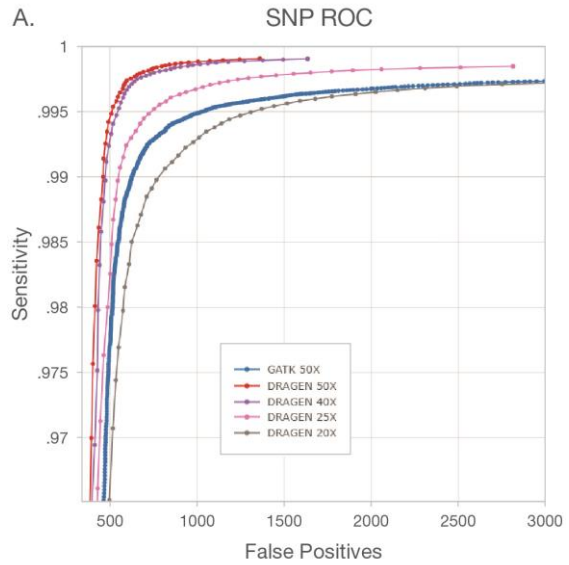


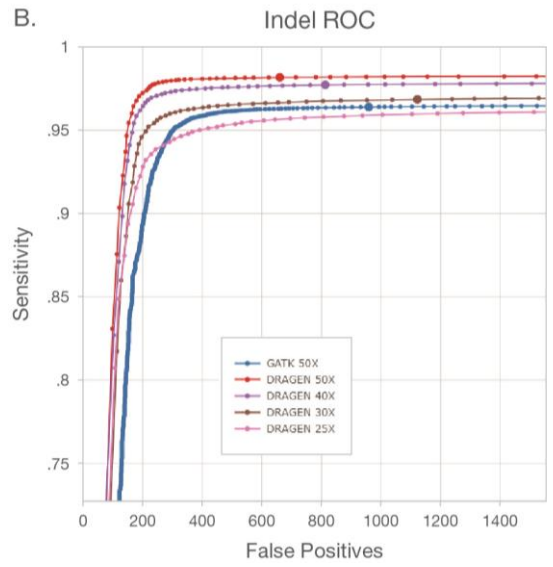
DRAGEN Germline Pipeline



DRAGEN Germline Pipeline—Accuracy

Highly accurate variant calling at low coverage

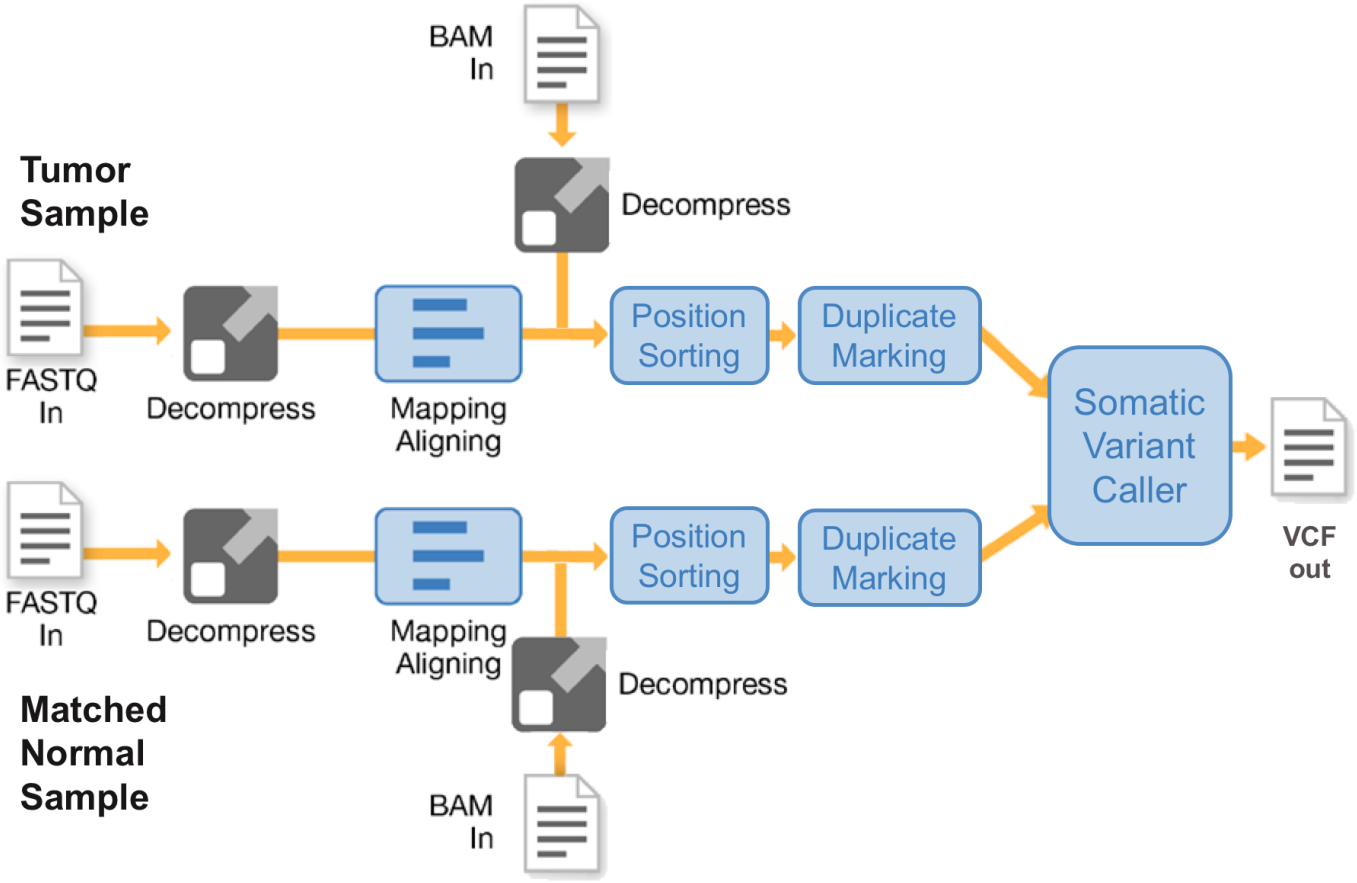




The DRAGEN v3 Germline Pipeline demonstrates high analytical sensitivity with relatively low false positive counts at 50x coverage. Performance comparable to GATK 4.1 at 50x coverage is possible with DRAGEN v3 at coverages lower than 50x.

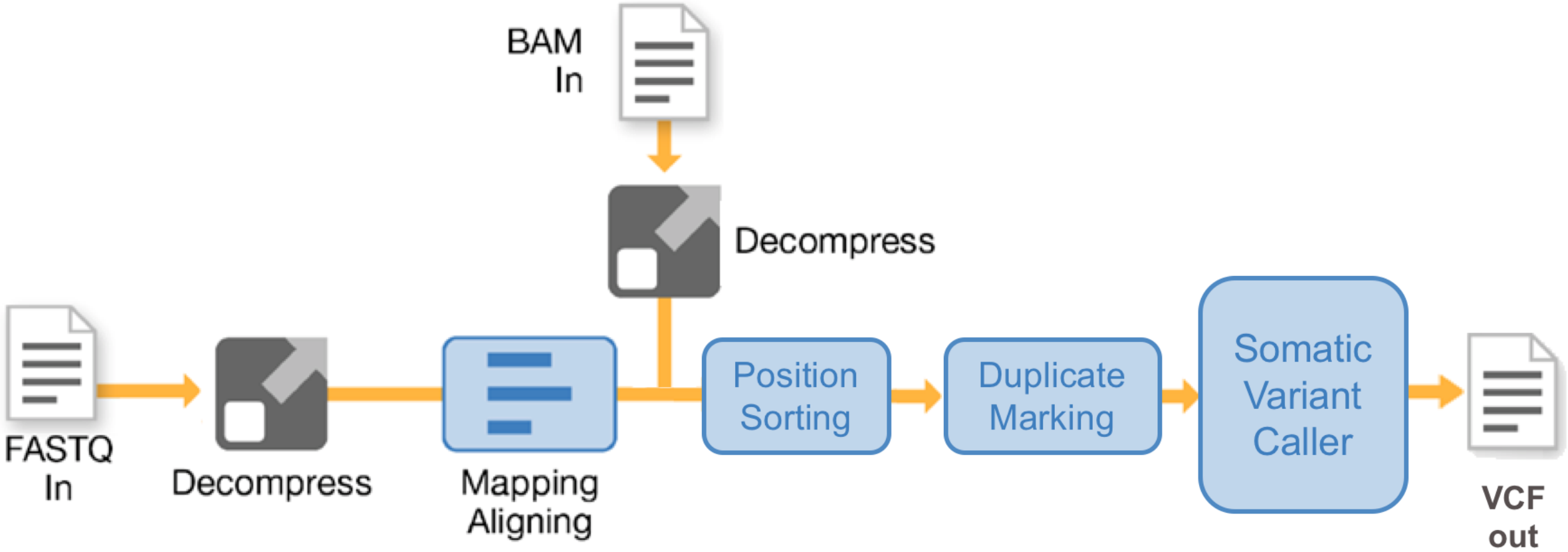
DRAGEN Somatic Pipeline

Tumor/Normal Mode



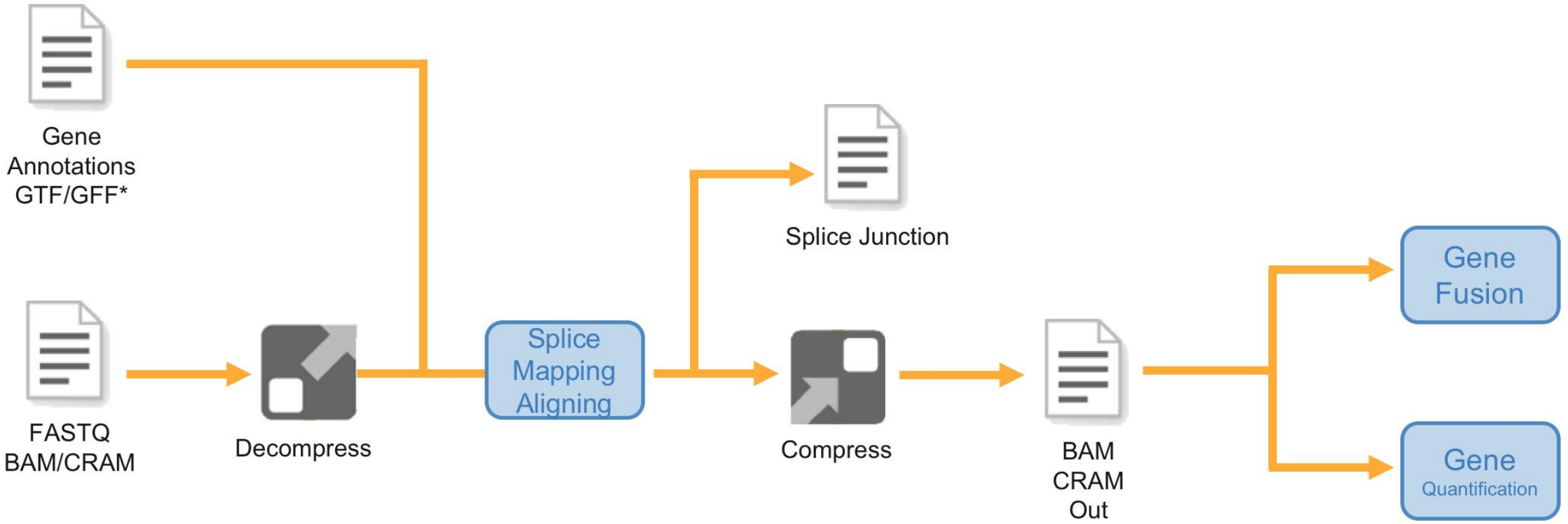
DRAGEN Somatic Pipeline

Tumor-only Mode

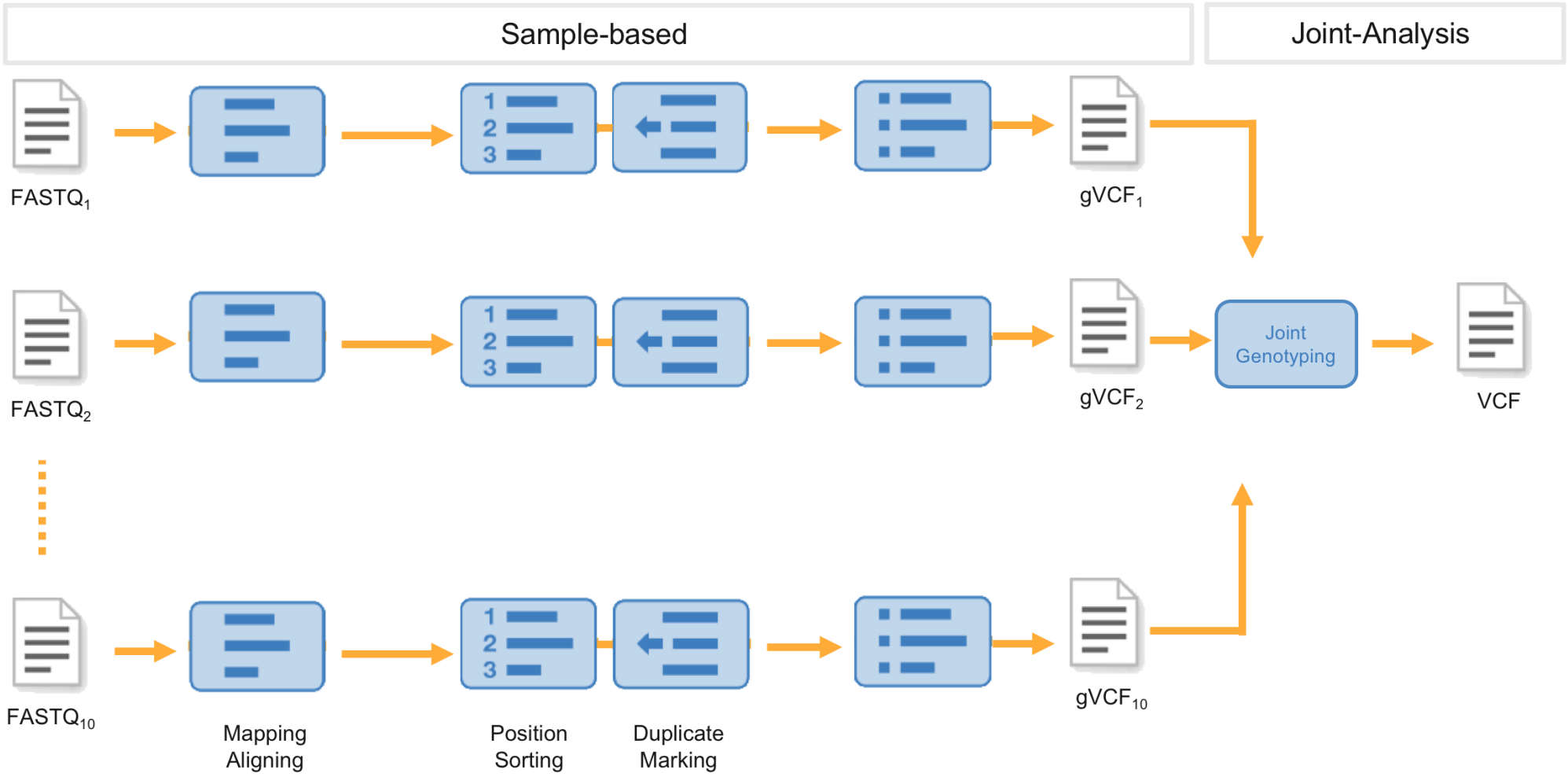


DRAGEN RNA Pipeline

Gene Fusion Detection & Quantification



DRAGEN Joint Genotyping



DRAGEN Automatic QC Metrics Reporting: Mapper

Removes the need to run downstream tools for QC

- Number of samples
- Reads Processed
- Total
- Biallelic
- Multiallelic
- SNPs
- INDELs
- MNPs
- SNP Transitions
- SNP Transversions
- Ti/Tv ratio

- Heterozygous
- Homozygous
- Het/Hom ratio
- In dbSNP
- Novel
- Total
- Biallelic
- Multiallelic
- SNPs
- INDELs
- MNPs

DRAGEN Automatic QC Metrics Reporting: Variant Caller

- Total input reads
- Number of duplicate reads (marked not removed)
- Number of unique reads
- Reads with mate sequenced
- Reads without mate sequenced
- QC-failed reads
- Mapped reads
- Number of unique & mapped reads (excl. dups)
- Unmapped reads
- Singleton reads (itself mapped; mate unmapped)
- Paired reads (itself & mate mapped)
- Properly paired reads
- Not properly paired reads (discordant)
- Reads with MAPQ [40:inf)
- Reads with MAPQ [30:40)
- Reads with MAPQ [20:30)
- Reads with MAPQ [10:20)
- Total reads in RG
- Supplementary (chimeric) alignments
- Average sequenced coverage over genome

- Total alignments
- Secondary alignments
- Supplementary (chimeric) alignments
- Estimated read length
- Bases in reference genome
- Bases in target bed [% of genome]
- Average sequenced coverage over genome
- Average alignment coverage over genome
- PCT of genome with coverage [40x:inf)
- PCT of genome with coverage [30x:40x)
- PCT of genome with coverage [20x:30x)
- PCT of genome with coverage [10x:20x)
- PCT of genome with coverage [5x:10x)
- PCT of genome with coverage [2x: 5x)
- PCT of genome with coverage [1x: 2x)
- PCT of genome with coverage [0x: 1x)
- DRAGEN mapping rate [mil. reads/second]
- Secondary alignments
- Estimated read length
- Insert length: mean

- Number of duplicate reads (marked)
- Number of unique reads
- Reads with mate sequenced
- Reads without mate sequenced
- QC-failed reads
- Mapped reads
- Number of unique & mapped reads (excl. dups)
- Unmapped reads
- Singleton reads (itself mapped; mate unmapped)
- Paired reads (itself & mate mapped)
- Properly paired reads
- Not properly paired reads (discordant)
- Reads with MAPQ [40:inf)
- Reads with MAPQ [30:40)
- Reads with MAPQ [20:30)
- Reads with MAPQ [10:20)
- Reads with MAPQ [0:10)
- Total alignments
- Secondary alignments
- Insert length: standard deviation

Thank You

techsupport@illumina.com | rtorrieri@illumina.com

illumina®