# DRAGEN™ v4.4 – Introducing our most advanced and comprehensive secondary analysis

Shyamal Mehtalia

Director, Product Management, Secondary Analysis

illumina



M-GL-03590

# Agenda

## This webinar will cover:

01 Overview of DRAGEN Secondary Analysis

02 DRAGEN Germline

03 DRAGEN Oncology

04 DRAGEN Multiomics

05 DRAGEN Availability on AWS F2

06 Significant Toolkit and Platform Updates

illumina®

# Overview of DRAGEN Secondary Analysis

# A powerful bioinformatics software suite connecting your entire genomics workflow

PREPARE | SEQUENCE | ANALYZE | INSIGHTS AND REPORTING

**PREPARE**

Clarity LIMS™

Assays

**SEQUENCE**

BaseSpace™ Sequence Hub

Instrument run planning

**ANALYZE**

Connected Analytics™

DRAGEN™

DRAGEN

**INSIGHTS AND REPORTING**

Clinical Research

- Connected Insights — Oncology
- Emedgene™ — Genetic Disease

Discovery

- Connected Analytics - Cohorts — Population Genomics
- Connected Multiomics, Partek™, Correlation Engine — Multiomics

**DRAGEN (Dynamic Read Analysis for GENomics)**

ACCURATE, COMPREHENSIVE, EFFICIENT SECONDARY ANALYSIS

# DRAGEN secondary analysis

## Accurate, comprehensive, efficient NGS secondary analysis

### Accurate

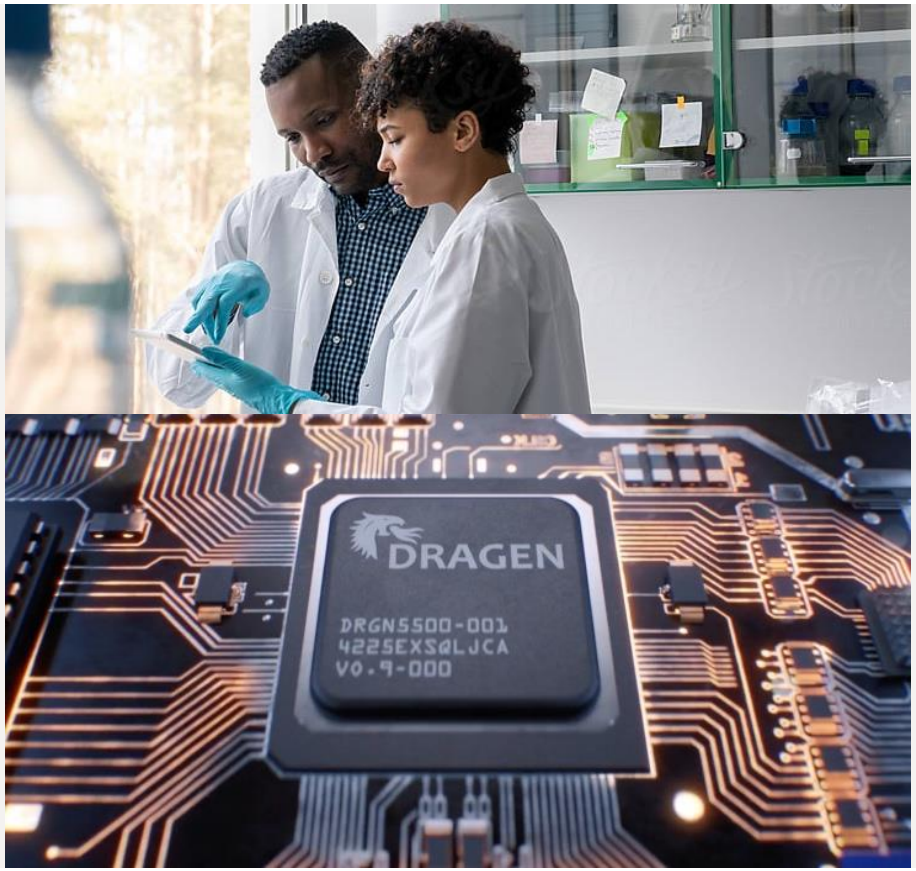
- Winner of 4 Precision FDA challenges[1] for germline and somatic pipelines
- Highly accurate germline variant calling 99.90%[2] F1 score with DRAGEN v4.4

### Comprehensive

- Tools for genomic, epigenomic, transcriptomic and proteomic analysis
- Mapping, UMI QC, Variant calling SNV, SV, CNV, STR, fusion, biomarkers, counting, differential expression, contamination detection and more

### Efficient

- Process a 30X WGS ~ 30mins, with all supported callers[2]
- Reduce FASTQ.GZ file sizes up to 5× with DRAGEN ORA Compression




[1] PrecisionFDA NCTR Indel Calling from Oncopanel Sequencing Data
[2] Illumina internal data on file, 2025

# DRAGEN methods make it a comprehensive analysis choice

- ✓ Pangenome Reference
- ✓ Machine Learning
- ✓ Targeted Callers
- ✓ Mosaic calling

**Read article**

# DRAGEN is available* on different platforms – giving you the flexibility you need

## On premises server

Fast, secure, local analysis in a fraction of time compared to a traditional CPU-based system.

## Illumina cloud

Flexible, secure, scalable managed cloud analysis with Illumina Connected Analytics and BaseSpace Sequence Hub

## Public cloud

Secure cloud analysis with bring your own license to install on AWS, Azure, GCP

## Onboard sequencers

NovaSeq X Series, NextSeq 1000/2000, MiSeq i100

illumina®

* Capabilities vary per deployment options

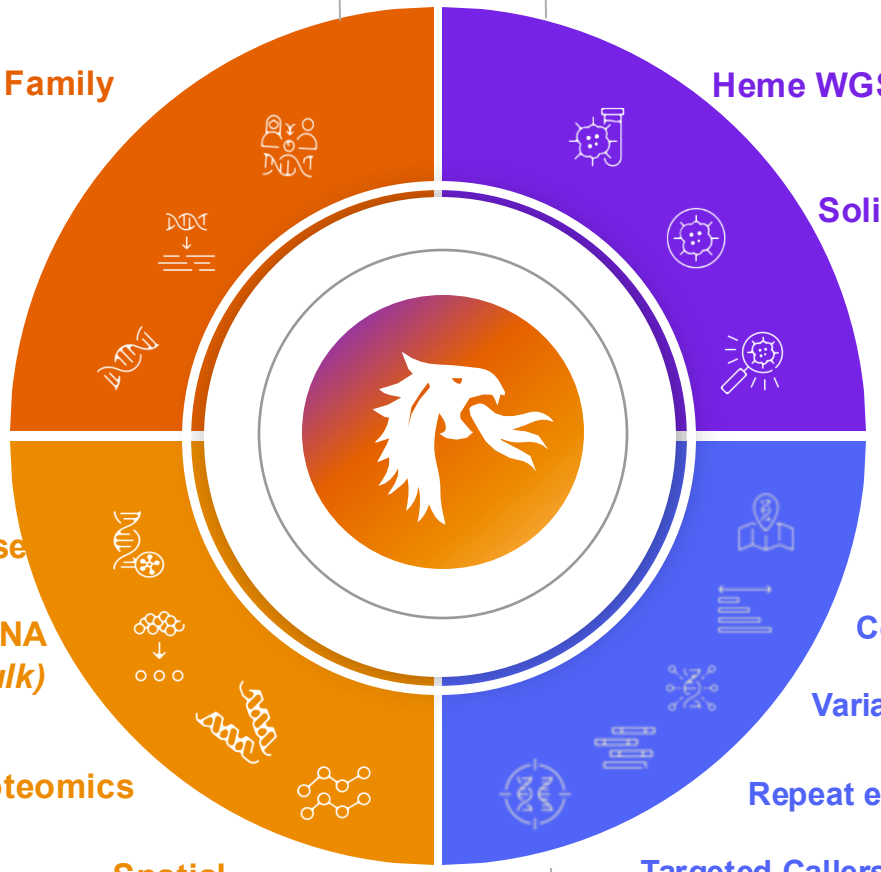
# DRAGEN – Genetic Disease, Oncology, Multiomics, Toolkit

**Genetic Disease Research**
**Emedgene**

- Trio/Joint Family
- WES Germline
- WGS Germline *(Including Constellation)*

**Oncology Research**
**Connected Insights**

- Heme WGS
- Solid WGS T/O and T/N
- Minimal Residual Disease



**Multiomics**
**Connected Multiomics**

- 5 Base
- RNA *(Single cell 3'RNA Prep & Bulk)*
- Proteomics
- Spatial

**Toolkit**

- Multigenome mapper *(with pangenome reference)*
- Compression
- Variant callers
- Repeat expansions
- Targeted Callers

illumina®

# The most comprehensive secondary analysis solution[1]

| General Tools | Genetic Disease | Oncology | Multiomics |
|---|---|---|---|
| Graph Capable Mapping | Single Nucleotide Variant | Single Nucleotide Variant | Bulk RNA [2] |
| Adapter Trimming | Indel | Structural Variant | Single-cell RNA |
| Poly G Trimming | Structural Variant | Copy Number Variant | Single-cell ATAC |
| FastQC Metrics | Copy Number Variant | Tumor Only | NanoString GeoMx |
| BCL Conversion | Regions of Homozygosity | Tumor/Normal | Protein Quantification |
| ORA Compression | Repeat Expansion | Tumor Mutational Burden | Spatial Transcriptomics |
| UMI Collapsing | Targeted Callers | Microsatellite Instability | **Infectious Disease** |
| | Star Allele Caller | MRD for WGS | COVID Lineage |
| | Spinal Muscular Atrophy | Liquid Biopsy | Metagenomics |
| | Trio Calling | Gene Fusion Detection | RPIP / UPIP Pipelines |
| | Joint Calling | Methylation | **PopGen** |
| | Imputation | Heme WGS | gVCF Caller |
| | Specialized callers (MRJD) | | Joint Calling |

- Analyze whole genomes, exomes, panels, transcriptomes, and methylomes in a single platform.
- Call single-nucleotide variants (SNVs), structural variants (SV), copy number variants (CNVs), repeat expansions, targeted callers, and biomarkers.
- Oncology – Supports tumor only and tumor/normal workflows with and without UMIs

Table shows applications available with DRAGEN on-premises server. Cloud and Onboard availability vary. Details here

illumina®

For Research Use Only. Not for use in diagnostic procedures.



1 – When compared against leading secondary analysis players who have submitted to PrecisionFDA v2 Truth Challenge Benchmark Data,
2 – DRAGEN RNA pipeline performs RNA quantification, fusion gene calling, and RNA variant calling.

# DRAGEN v4.4 with easy-to-use oncology apps, enhanced multiomics, AWS F2 support, and more

## Genetic Diseases

**30% boost to SV calling accuracy** with multigenome mapper with pangenome reference

**Personalized pangenome w**ith significant accuracy gains and improved run time to WGS germline analysis

**WGS Cytogenetics** - DEL/DUPs | AOH | mosaic CNV | Autosomal and sex chr aneuploidy | Allele specific copy number

**Targeted Callers Updates -** Streamlined WGS Analysis, SMN1, HBA1/2 supported on ILMN exome v2.5 + spike-in probes kit

## Oncology

**Easy-to-use oncology apps** with push-button analysis for Heme WGS and Solid WGS tumor-normal

**Enhanced RNA fusion Calling** with improved accuracy, breakpoint resolution for in/out-of-frame determination, and more

**Somatic CNV enhancements** with improved purity/ploidy estimation and model detection in low-quality or challenging tumor samples, WGS PON support and more

## Multiomics & more

**Support for new multiomic assays** – SIngle Cell 3'RNA Prep, Proteomics, 5-Base (Methylation)

**AWS F2 Instances -** DRAGEN v4.4 will support AWS F2 instances

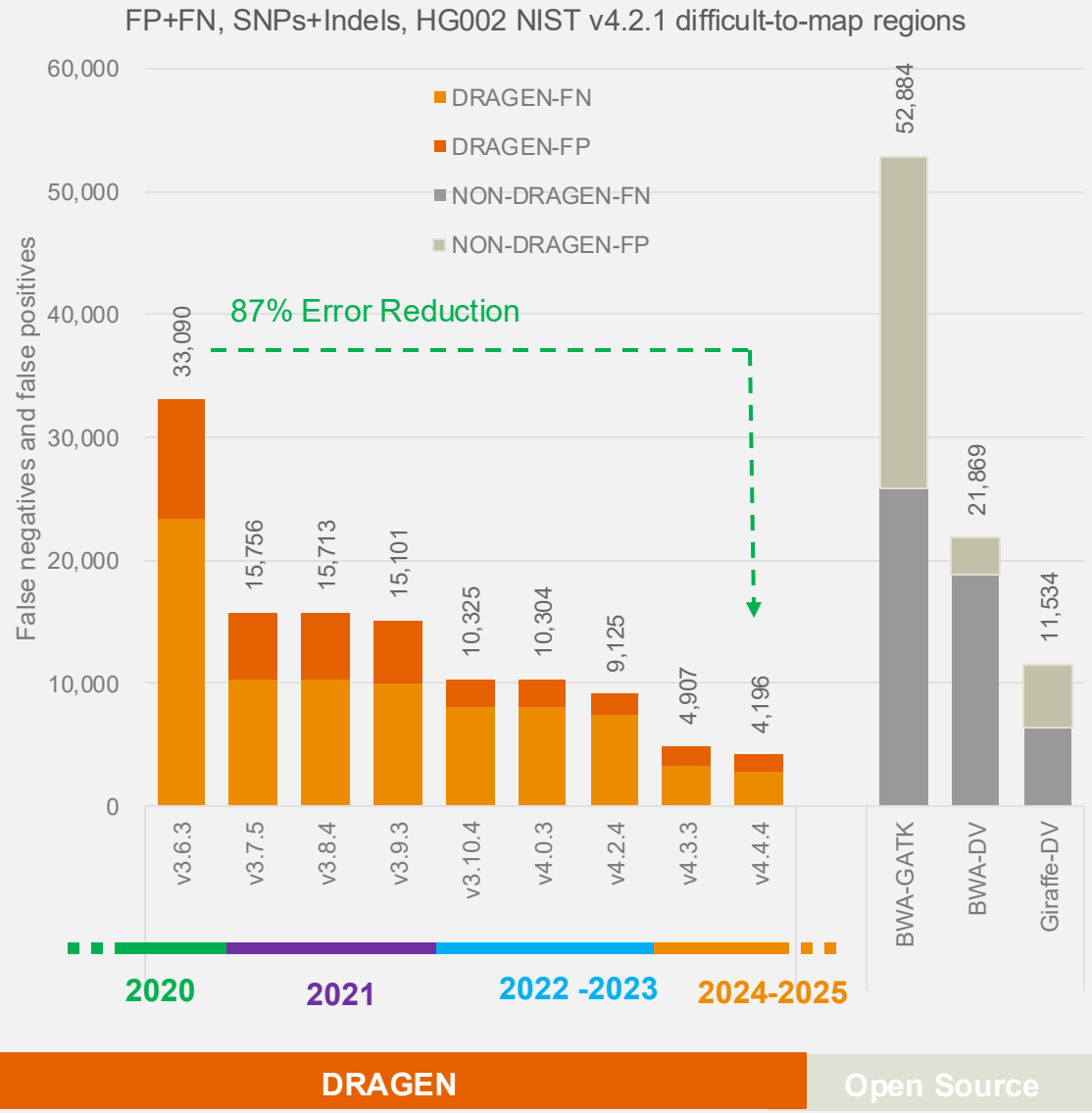
**30% speed up in FASTQ to ORA** compression.

illumina®

# DRAGEN – A legacy of continuous innovation and improving accuracy in each version

DRAGEN v4.4 continue accuracy advancements, reducing the overall germline variant errors

- **Superior SNPs and INDELs accuracy** in all confident regions and **dark regions of the genome**
- **Biggest leap** in **SV accuracy**
- Personalization for a bigger gain in SNPs and INDELs accuracy




DRAGEN – default settings, no personalization activated
BWA-DV – linear reference approach BWA-0.7.17 (r1188) DV-1.0
Giraffe-DV – pangenome and personalization approach- vg_v1.62.0 DV v1.8.0

# DRAGEN Germline Pipeline

2

illumina®

# DRAGEN multigenome mapper resolves difficult-to-map regions and increases number of variants discovered

**Available since DRAGEN v3.7.5, updated in v4.4.4 with 128 Global pangenome reference* and increased SV support**

**Multigenome Mapping with Population Haplotypes**

| DRAGEN | v3.4 - v3.6 | v3.7 – v3.8 | v3.9 | v3.10 | v4.0 | v4.2 | v4.3 | v4.4 |
|---|---|---|---|---|---|---|---|---|
| Native alt-contigs handling | alt-aware | | alt-masked | alt-masked-v2 | | alt-masked-v3 | alt-masked-v4 | alt-masked-v5 |
| Multigenome mapper | | First-generation | | | | | Second-generation | |
| Pangenome reference* | | 16 European ancestry samples | | 16 European ancestry samples + MHC | | 32 Global ancestry samples | 128 Global ancestry samples | 128 Global ancestry samples + SVs |
| Machine learning | | | v2 | v3.1 | v4 | v7 | v12 | v15 |

Historically referred to as “multigenome (graph) reference”

Allows to map accurately in difficult to map regions of the genome, at no run time cost

Increases the number variants discovered in a genome

Fully compatible with existing hg38 reference and existing BAM/VCF format

Enables SNV/SV/CNV Variant Calling in difficult-to-map regions

1 The quest for accuracy gains in the dark regions of the genomes, Aug. 2024

# Major **leap** in structural variant calling accuracy and improvements in personalized pangenome accuracy

## Redefining SV calling accuracy

- DRAGEN 4.4 improves on F-score by more than **30%***
- Structural variant population haplotypes incorporated in DRAGEN's pangenome reference increase accuracy
- Updates in read alignment, breakpoint detection, population-guided assembly, and genotyping models enhance the accuracy of SV calls.

## Personalized pangenome accuracy

- Personalized **pangenome references now generally available (GA)**
- Innovative downstream variant calling adjustments minimize errors
- Proven mapping and variant calling capabilities, unique against competition.

*On HG002 with both NIST T2TQ100[1] and CMRG[2] benchmarks

# DRAGEN SV with pangenome reference

## SV Accuracy

- DRAGEN 4.4 improves on F-score by more than **30%** on HG002 with both NIST T2TQ100[1] and CMRG[2] benchmarks.
- Significant recall gains in SV insertion detection including MEIs[3].

| Truth Set | Method | Recall | Precision | Fscore |
|---|---|---|---|---|
| NIST T2TQ100 HG002 | DRAGEN 4.4 | 0.807 | 0.963 | **0.878** |
| | DRAGEN 4.3 | 0.511 | 0.953 | 0.665 |
| CMRG HG002 | DRAGEN 4.4 | 0.775 | 0.941 | **0.850** |
| | DRAGEN 4.3 | 0.463 | 0.964 | 0.625 |

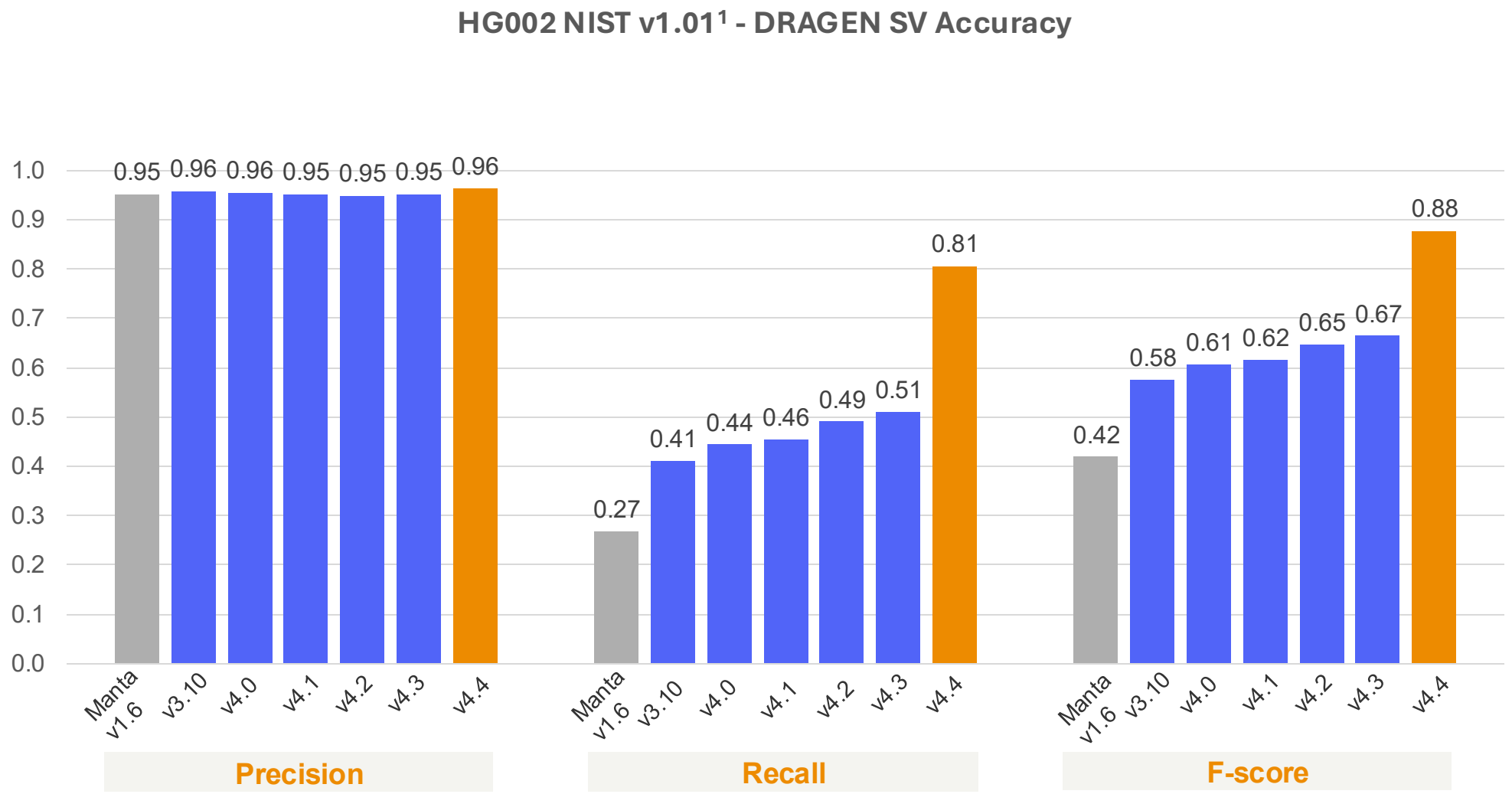
## SV Pangenome Reference

- Structural variant population haplotypes incorporated into DRAGEN's pangenome reference, which includes 128 samples across 26 ancestries.
- Updates in read alignment, breakpoint detection, population-guided assembly, and genotyping models enhance the accuracy of SV calls.

| Method | Type | Recall |
|---|---|---|
| DRAGEN 4.4 | Mobile Element Insertions | **0.906** |
| DRAGEN 4.3 | Mobile Element Insertions | 0.885 |
| Mobstr 0.2.4.1 | Mobile Element Insertions | 0.756 |

*1 – Compared against HG002 NIST T2T Q100 v1.1_v0.019 (hg38) with Truvari v.4.2.2*
*2 – Compared against HG002 CMRG v1.0 (hg38) with Truvari v.4.2.2*
*3 – Insertions stratified by annotation from Delage et al on NIST v0.6 (GRCh37)*

# Accuracy improvements of DRAGEN SV calling




illumina®





*1 – Compared against HG002 NIST T2T Q100 v1.1_v0.019 (hg38) with Truvari v.4.2.2*

# Personalized germline small variant caller Official release

**Leverages 128-sample pangenome reference and multigenome-mapper output to compute personalized 2-haplotypes reference**

- ✓ Use the **2-hap personalized reference** to impute variants, used as priors in the Variant Caller, create new personalized ML model
- ✓ Easy to use end-to end workflow with single flag
- ✓ ~ 25% SNPs FP+FN reduction and ~ 9 % Indels FP+FN reduction in 4.2.1 high confidence regions, compared to germline small variant caller
- ✓ Adds less than 4 minutes to default small variant calling runs on a DRAGEN P4 server

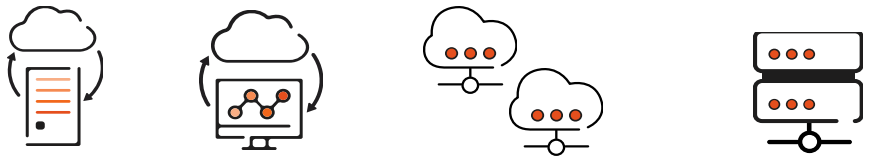
Activated with
*--enable-personalization=true*

Run time on 35x WGS approx. 22 minutes for alignment and small variant caller on a Phase4 server

Requires a pangenome “v5” hash table

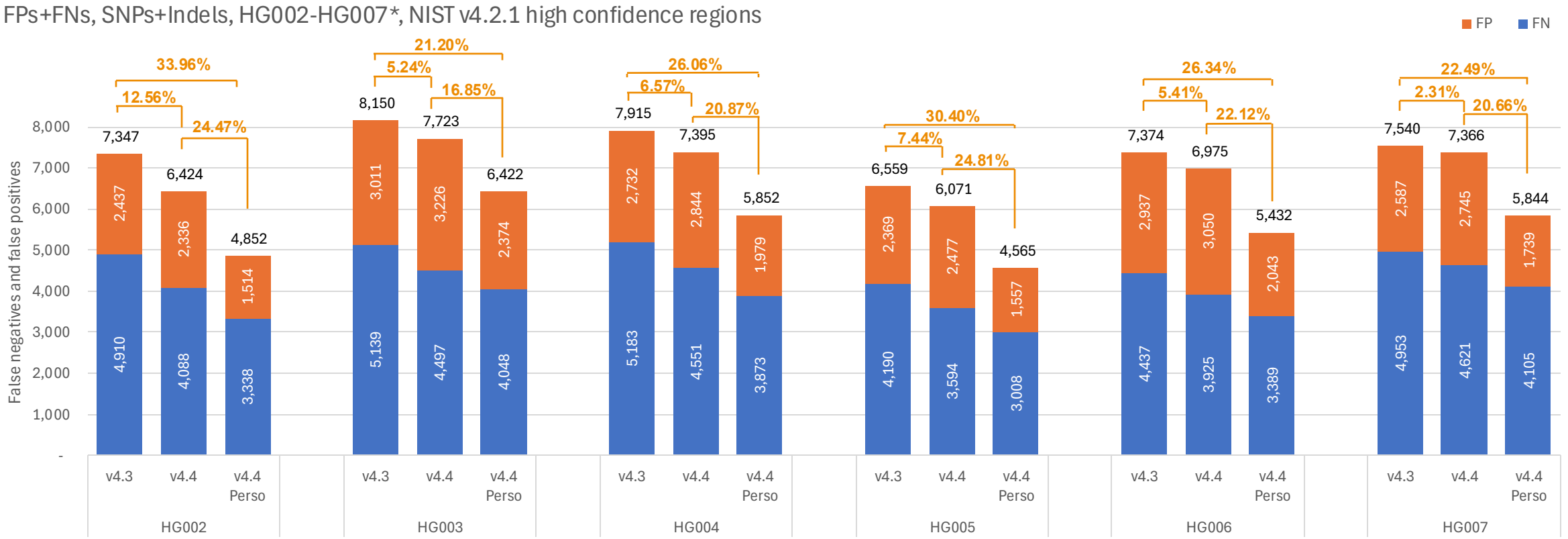
Available only for germline small variant caller

**Availability**



# Personalized germline small variant caller – accuracy

DRAGEN v4.4 adds further small VC gain (5.35% in SNP, 10.30% in Indel).
Personalization adds a further gain of 21.63% for very little added run time (3 minutes and 40 seconds).



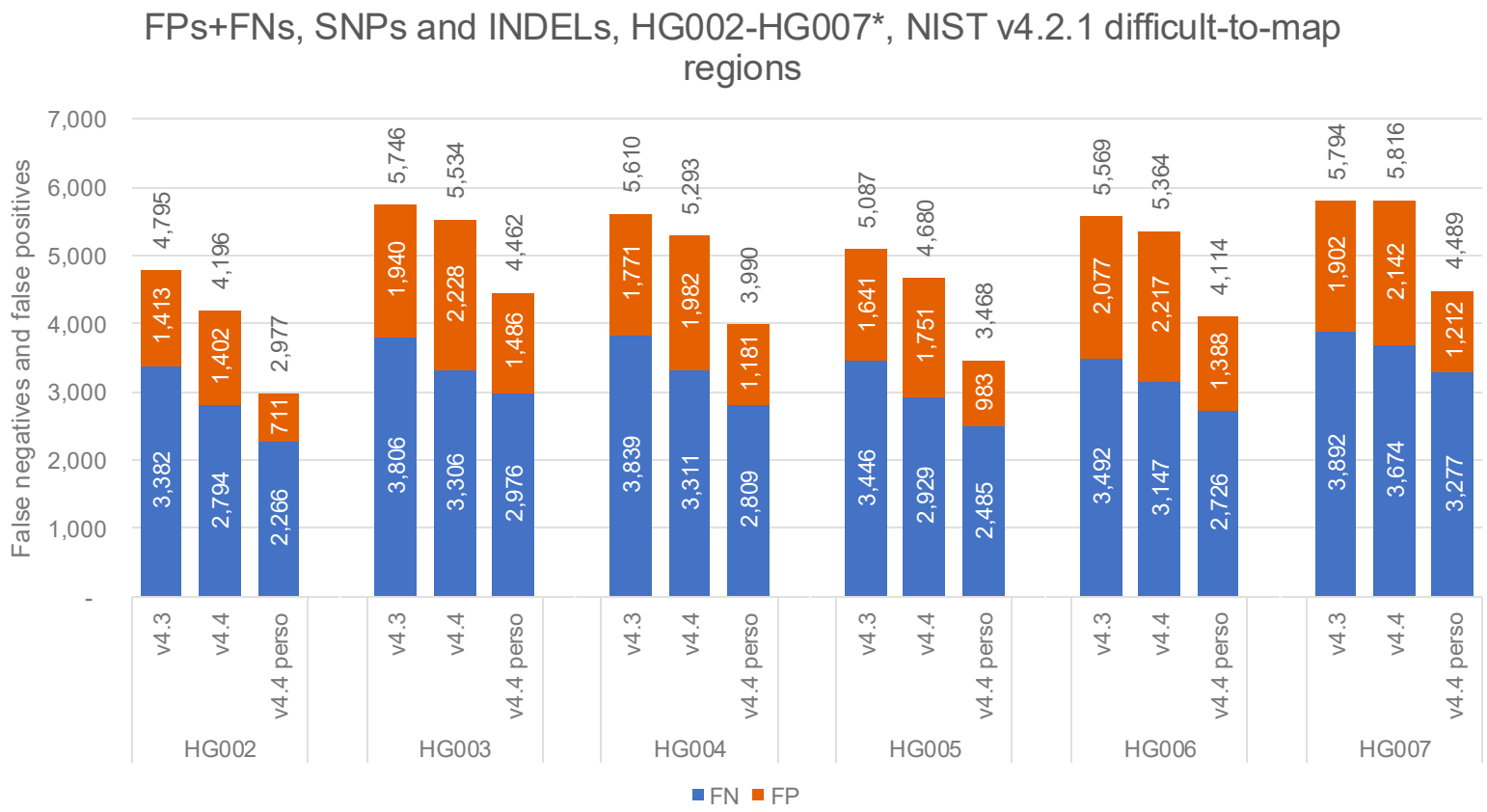

*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Reference

# Personalized small variant caller improvement in the dark regions of the genome

**Personalization adds a further gain of 24.18% for SNPs and INDELS on the dark regions of the genome**

**Difficult-to-map regions[1]**

- regions that have other **homologous regions** in the reference genome for the given read length, number of mismatches, and number of indels.
- **Segmental duplications**




1.Difficult-to-Map Regions BED file. ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/v3.3/GRCh38@all/Union/GRCh38_alllowmapandsegdupregions.bed.gz.

*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Reference

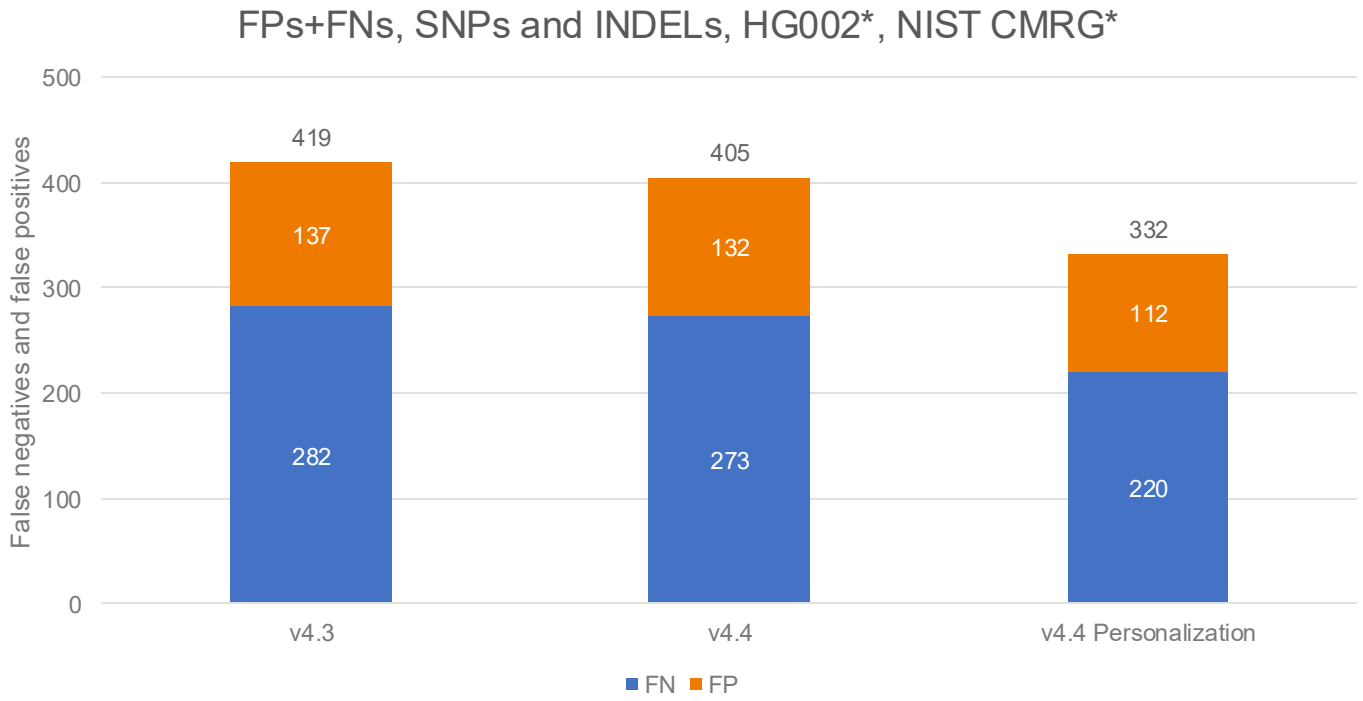
illumina®

# Personalized small variant caller for CMRG

**Personalization reduces by 18% FPs+FNs in Challenging Medically Relevant Genes (CMRG)**

**CMRG[1]**

Curated benchmark of 273 genes with repetitiveness and polymorphic complexity



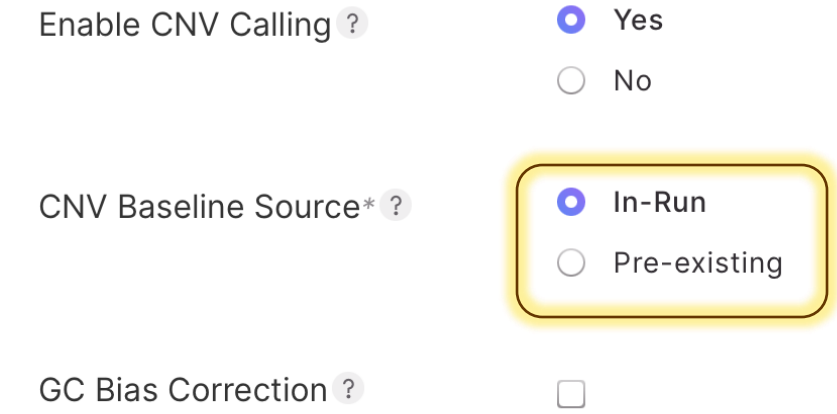

*Sample NS2K P4 XLEAP-advanced recipe 35x Raw coverage

# In-run construction of CNV panel-of-normals for germline enrichment

DRAGEN v4.4 provides workflow-level support for automatically constructing a PoN from a batch of samples.

- Ensures robustness to batch-level effects
- Reduces putative false positives by as much as 65% compared to WGS CNV while maintaining recall
- Eliminates need to maintain and update a pre-built PoN
- End-to-end solution configurable from BSSH run-planning or from existing fastqs/bams/crams on BSSH or ICA

Available via single-click GUI element, or CLI




**Availability** BSSH ICA

# Germline WGS cytogenetics

## B-allele frequency support for Germline CNV

Support for **WGS-based cytogenetics** analysis

- Enables the detection of **AOH/LOH** regions across the genome
- Increases calling **robustness** by considering both **read coverage and minor allele frequency**
- Capable of reporting **mosaic** CNVs

## Concordance with legacy technologies

Processed **98 samples** with 146 CNVs ranging from 40kbp to whole-chromosome aneuploidies

- **> 97% agreement** with results from karyotyping and chromosomal microarray

## Cytogenetics modality reports multiple views of the same sample

- Fine resolution views for shorter alterations (≥ 25kb)
- Coarse resolution views for larger alterations (≥ Mb)

**Example Command Line:**

```
dragen \
-r <HASHTABLE> \
--output-directory <OUTPUT> \
--output-file-prefix <SAMPLE> \
--ref-dir <HASHTABLE> \
--bam-input <BAM> \
--enable-map-align true \
--enable-cnv true \
--cnv-enable-self-normalization true \
--cnv-population-b-allele-vcf <SNP_POP_VCF>
```

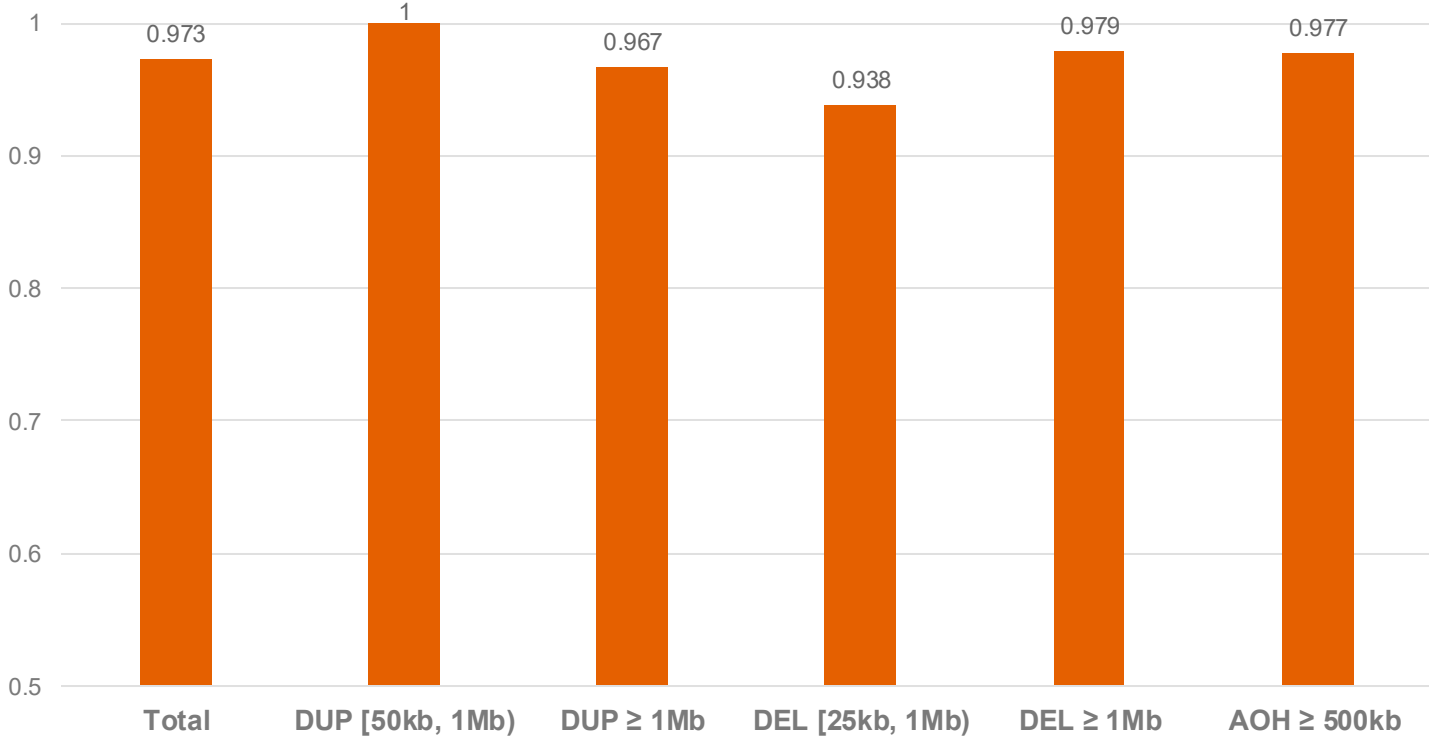
**Availability**

# WGS concordance with Cytogenetics

| Supported events |
| --- |
| DUPs ≥ 50kb, including whole-chromosome |
| DELs ≥ 25kb, including whole-chromosome |
| AOH ≥ 500kb, including whole-chromosome |
| Mosaicism ≥ 20% |

**Coverage:** 30-60x
**Read length:** 2x151bp
**Insert size:** 340-530bp
**N samples:** 98
**Total events:** 146

> 97% Concordance against karyotyping and CMA



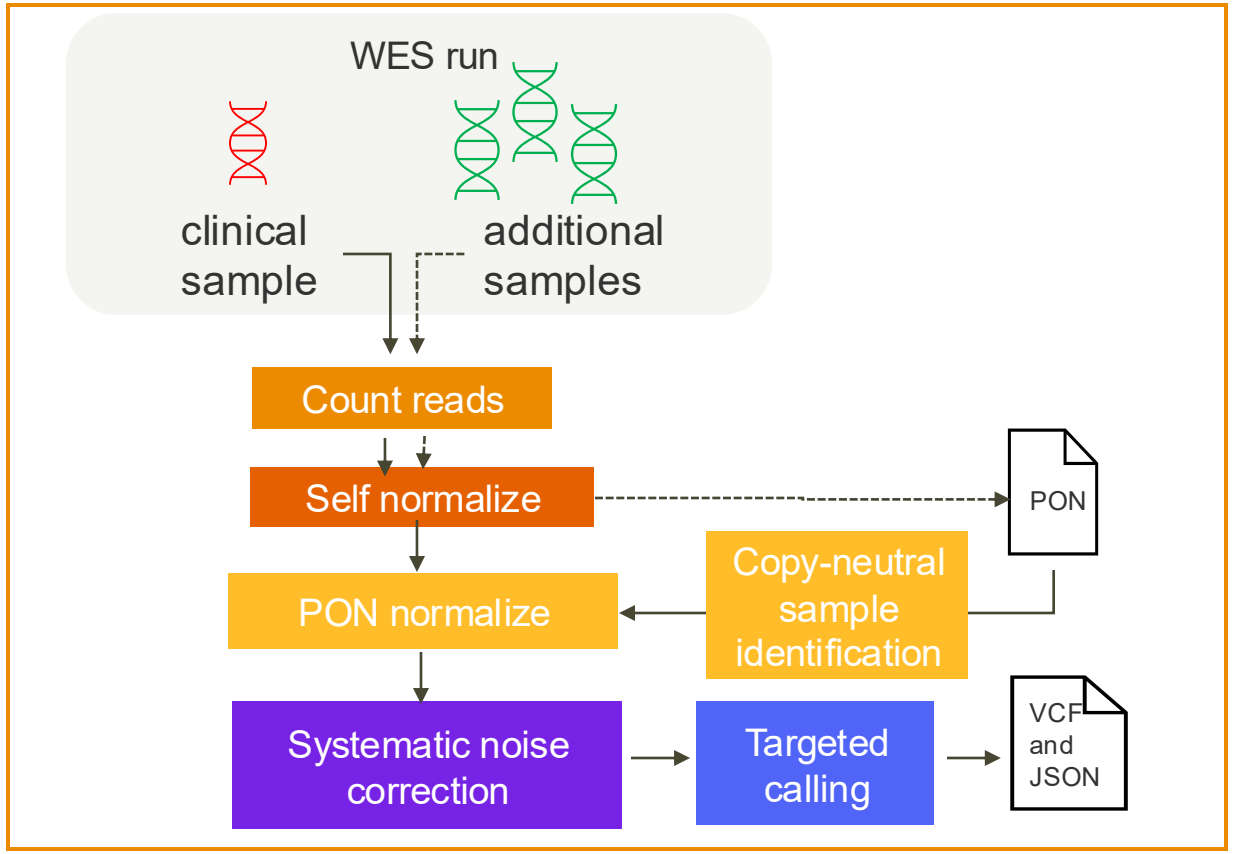

Processed 98 samples with 146 CNVs ranging from 40kbp to whole-chromosome aneuploidies

# New WES Targeted Calling

- Support for HBA and SMN
- Requires Illumina Custom Enrichment Panel v2 specifically designed for DRAGEN Targeted Calling
- Normalization using automatic copy-neutral sample identification from a multiplexed WES run
- Systematic noise correction based on a panel of 576 samples (available for hg19, hs37d5 and hg38)
- 100% benchmark concordance for HBA*
- 99.8% benchmark concordance for SMN *

**Availability**

BSSH | ICA | Multi-Cloud | On-premises server



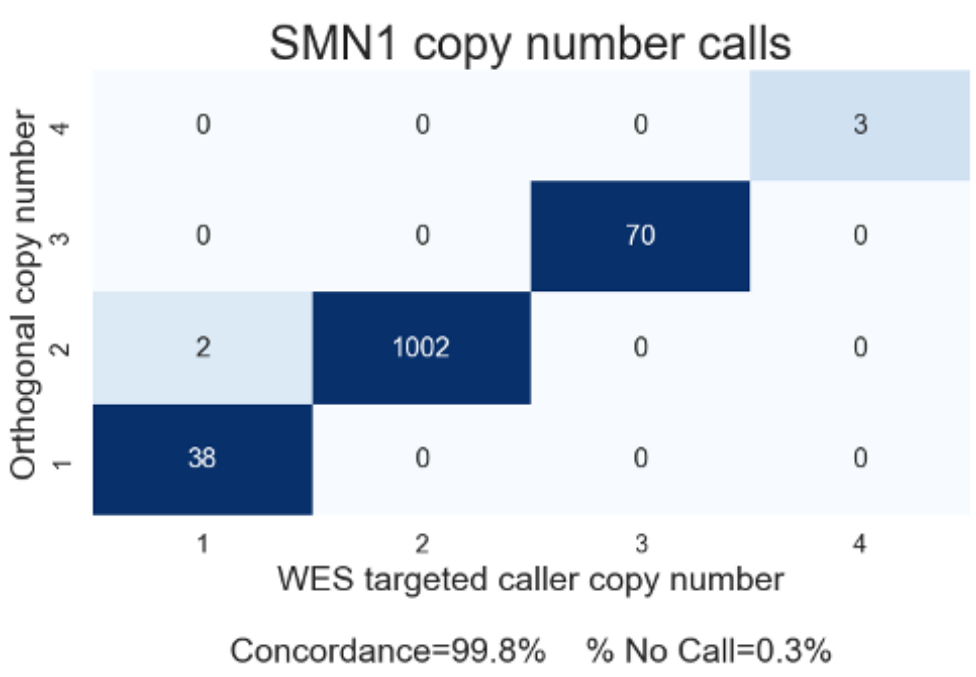

* evaluated against calls derived from (MLPA or qPCR) in a partially blinded and partially consecutive cohort of clinical samples

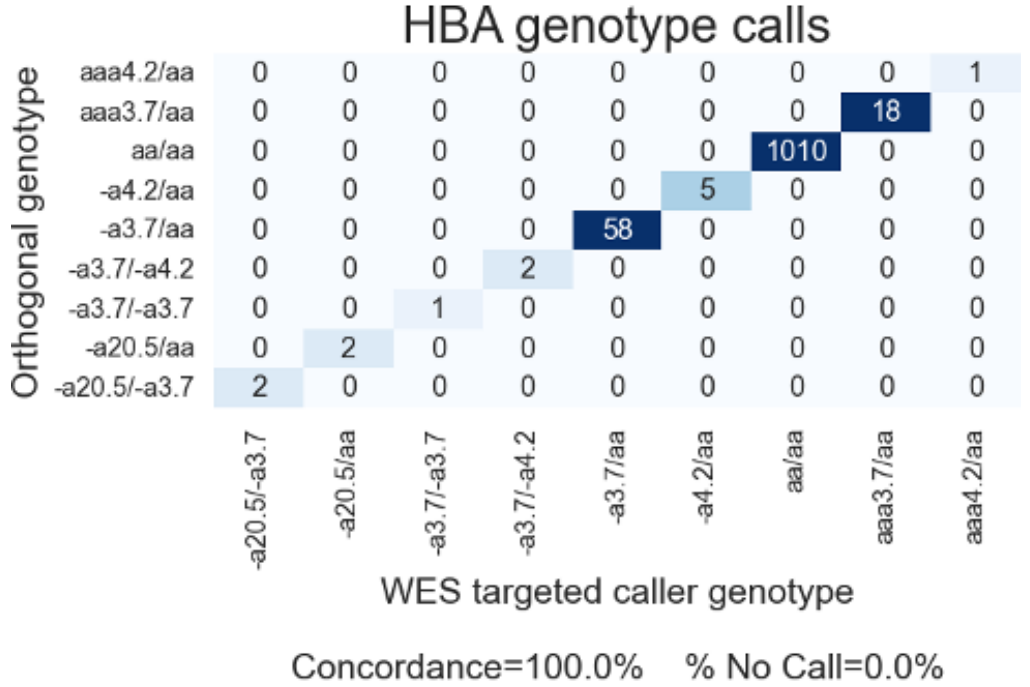
# DRAGEN SMN and HBA targeted callers coming to Exome 2.5 Enrichment in v4.4

## DRAGEN WES Targeted Caller Key Features

- **Spike-in probes:** Additional probes (~359 kbp) to cover current and prospective targeted caller regions.
- **Panel of Normal (PoN):** Specialized Exome 2.5 + spike-in PoN approaches to normalize for system level biases.

- ✓ 99.8% SMN concordance
- ✓ 100% HBA concordance




illumina®





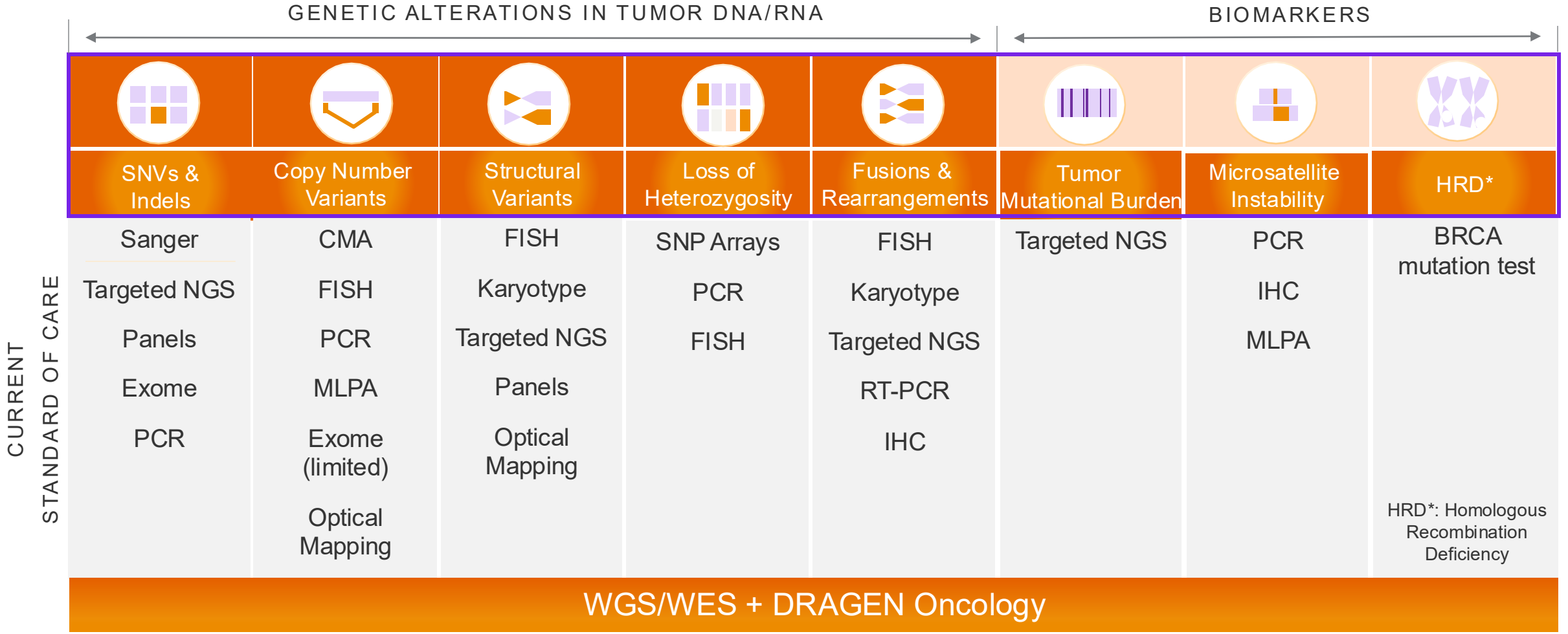
*Concordance matrices shown for 1 unique run and 5 identical runs with different conditions (hybridization time, yield, etc.). Orthogonal validation is comprised of data from Coriell, PacBio, and DRAGEN WGS results.

# DRAGEN Oncology


illumina®

# WGS/WES provide the most comprehensive analysis among all molecular technologies

## when powered by accurate and scalable informatics



# Simplifying oncology biomarker detection with pre-configured pipelines on premises

## DRAGEN TOOLKIT (key features)

| | |
|---|---|
| Map/ Align | QC |
| SNV | CNV |
| SV | RNA |
| TMB | MSI |
| HRD | HLA |
| MRD detection | Sample matching |

Modular. Experienced bioinformaticians can configure the commands

## STREAMLINED PIPELINES (EARLY ACCESS)

| | |
|---|---|
| Tumor/ Normal WGS | Heme WGS |
| cfDNA WGS* | MRD WGS* |
| Solid Tumor Panel* | Methylation WGS* |
| Solid Tumor RNA Panel* | Solid Tumor/ Normal Panel* |
| Heme Panel* | MRD Panel* |
| cfDNA Panel* | cfDNA Methylation Panel* |
| Solid Tumor WTS* | |

Coming Soon

Easy-to-use, application-specific oncology pipelines

* Planned for future releases

# Illumina products provide all components for Heme WGS

**Library Prep**

Highly open platform compatible library prep kits from Illumina™
Illumina DNA PCR-Free Prep
Illumina DNA Prep

**Sequencing**

Powerful, high throughput sequencing with
NovaSeq™ 6000 or
NovaSeq™ X/X+


**Secondary Analysis**

Accurate, comprehensive and ultra-efficient, easy to use and pre-packaged application for secondary analysis with DRAGEN™ for malignancies like AML, MDS etc.
DRAGEN Somatic
WGS Heme pipeline

**Connected Insights***

**Insights**

Enables streamlined variant interpretation and research reporting from integrated knowledge bases and automation
Illumina Connected Insights™

[1] See full lists of compatible 3rd party automation on illumina.com

# DRAGEN Heme WGS Analysis

### SNPs & INDELs
Best in class variant calling detection for SNPs and INDELs in genes such as NPM1.

### Structural Variants
Large structural variants including recurrent gene fusions such as PLM::RARA and BCR::ABL1.

### DUX4 Rearrangements
Detection of rearrangements such as DUX4::IGH in difficult to map regions.

### Internal Tandem Duplications
Increased sensitivity for ITDs within genes such as FLT3, KMT2A.

### Copy Number Variants
CNVs including arm level, whole chromosome, and sub-clonal events.

### Purity/Ploidy Aware
Tumor purity estimation enables accurate calling of chromosomal ploidy states.

### Loss of Heterozygosity
Allele specific copy number and LOH calling.

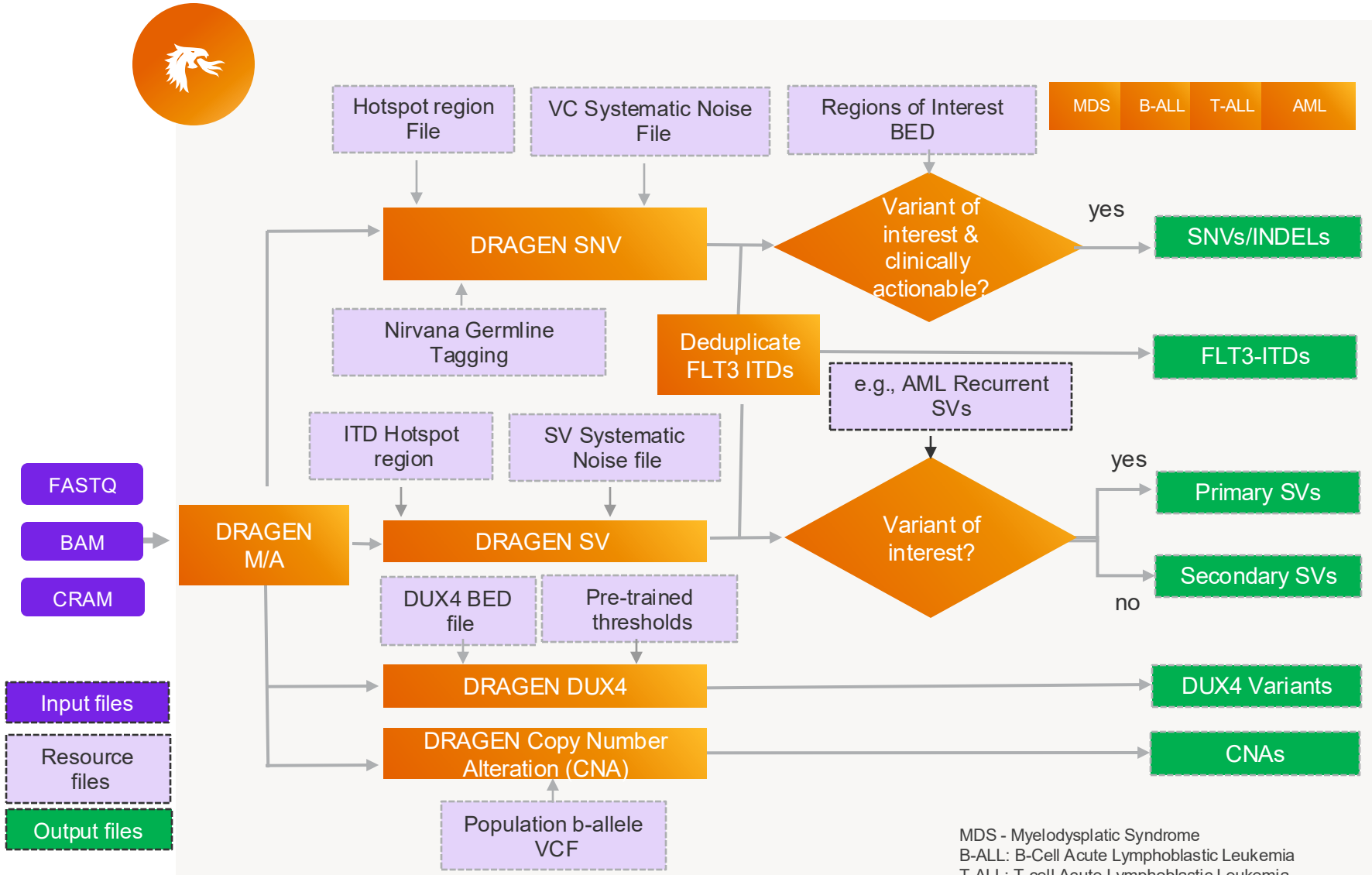
### Fast Streamlined Analysis
Single execution to analyze all variant types reducing total analysis time.

### Simplified Interpretation
Single assay and analysis prevents conflicting results.

1. **Replacement for conventional methods**
2. **Accurate Variant Detection**
3. **Enhanced efficiency & accessibility**

# Out of the box Heme WGS application (early access)

- **First and only commercially available**[1] Heme WGS application
- WGS **provides a comprehensive** genomic **characterization for hematologic malignancies**, well beyond what cytogenetics applications can achieve[2]
- An **easy-to-use prepackaged** application on the **DRAGEN server** abstracts the complexities of setting up own pipeline, integrated with **Connected Insights** for **streamlined workflow**




illumina®

For Research Use Only. Not for use in diagnostic procedures.

1 As compared to other commercial secondary analysis vendors
2 https://www.illumina.com/company/news-center/feature-articles/WGS-for-heme.html

# DRAGEN WGS Tumor-Normal Analysis

### SNPs & INDELs
Including the ability to call phased complex/ variants.

### HLA
Four-digit Human Leukocyte Antigen genotyping.

### HRD

### CNA
CNV and CNAs, as well as arm level & whole chromosome.

Optional PON for improved calling on FFPE samples.

### Biomarkers
Biomarkers including TMB and MSI.

### Simplified Interpretation
Single assay and analysis prevents conflicting results.

### Structural Variants
Large structural variants including recurrent gene fusions such as PLM::RARA and BCR::ABL1.

### Somatic and Germline results
Germline results based on the matched normal.

### QC
Optional coverage QC over multiple regions.

### Germline
Targeted callers e.g. CYP* StarAllele (PGx)
Variable Number Tandem Repeat

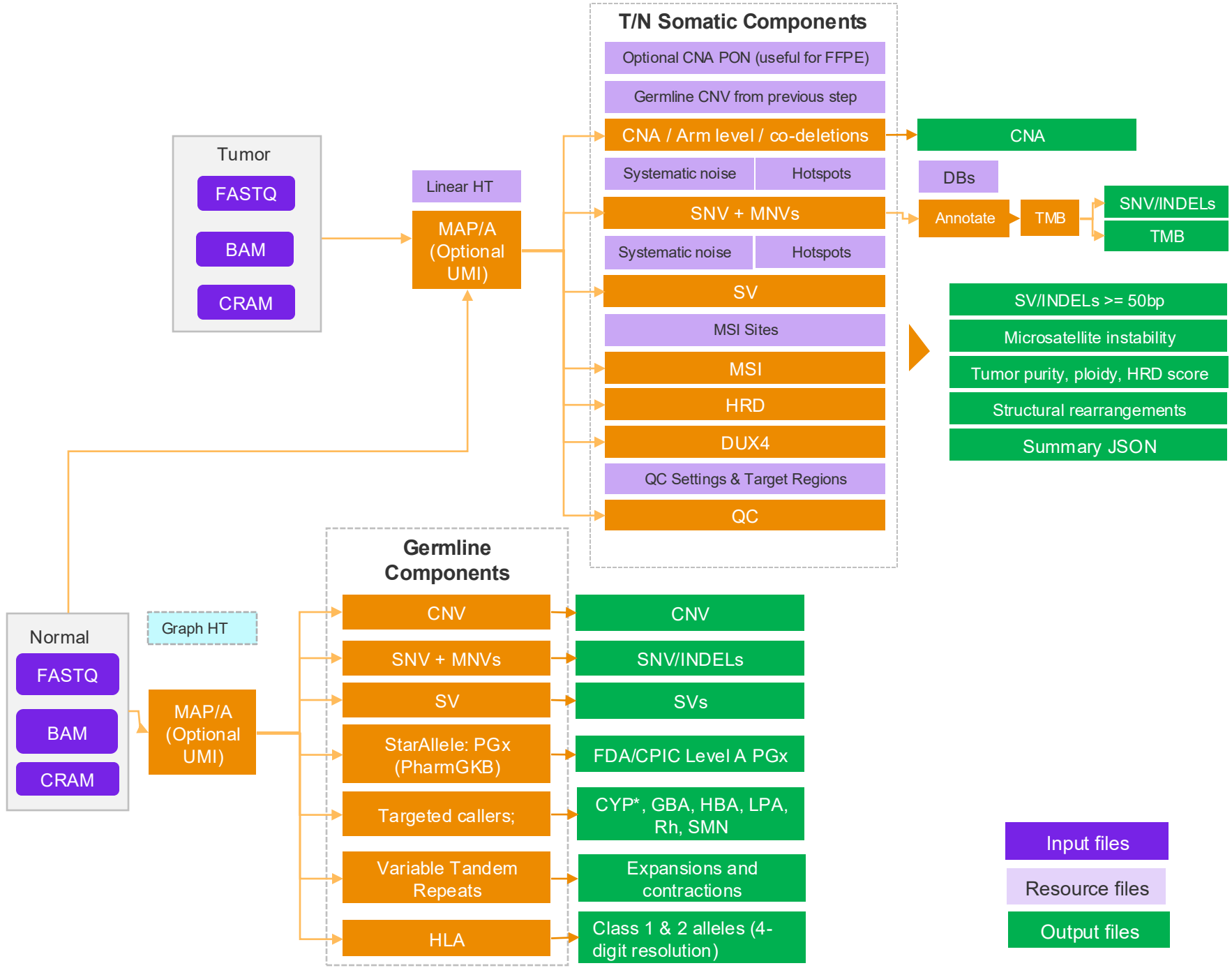
### DUX4
Structural rearrangements between DUX4 and other genes (including IGH)

1. **Replacement for conventional methods**
2. **Accurate Variant Detection**
3. **Enhanced efficiency & accessibility**

illumina®

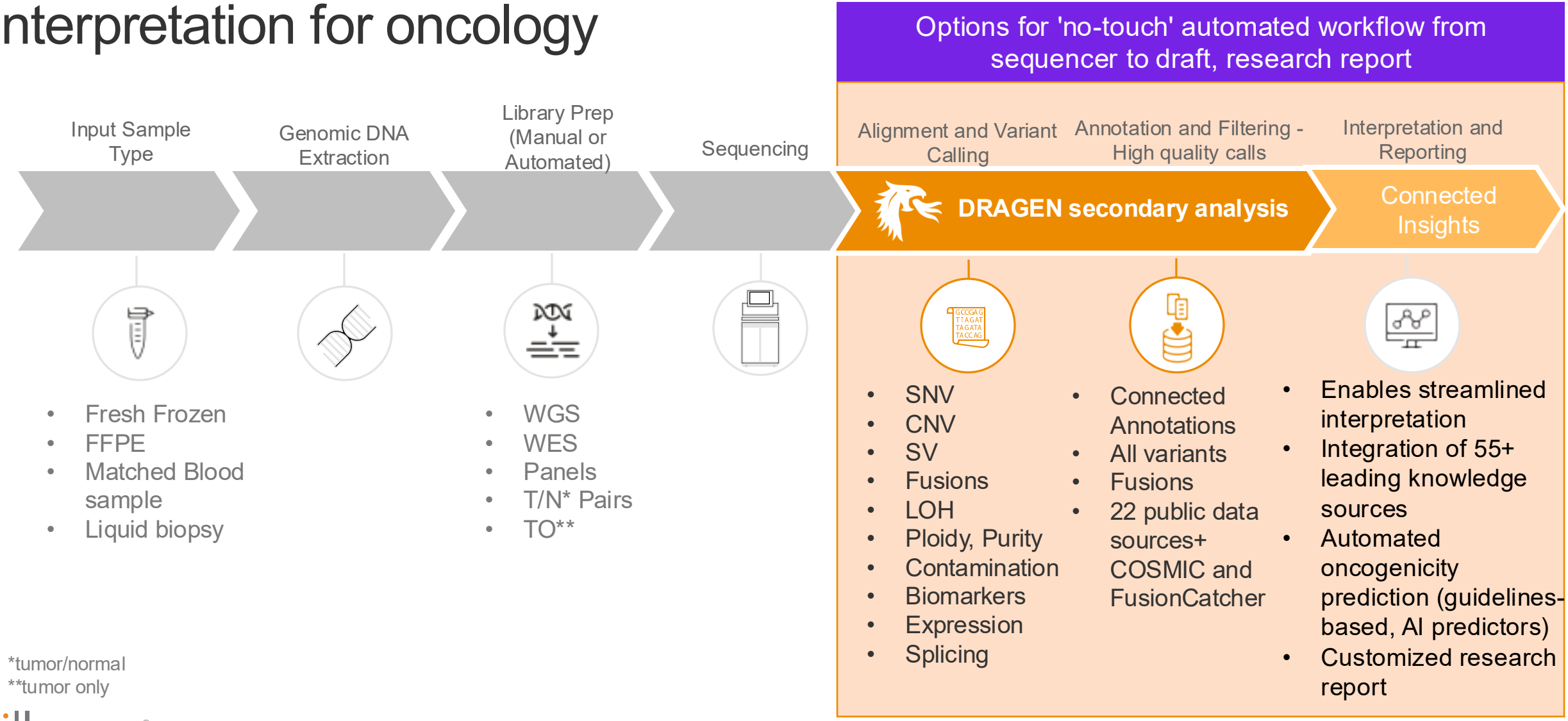
# Out of the box WGS DNA Tumor-Normal Application (early access)

- Simplifies WGS tumor-normal workflow
- **Prepackaged** application **on DRAGEN server** that abstracts with recommended settings and resource files for WGS solid samples, with both somatic and germine results in a single workflow
- Support for various files types – FastQ, BAM, CRAM
- Integration with Illumina Connected Insights, streamlines samples to insights workflow




**For Research Use Only. Not for use in diagnostic procedures.**

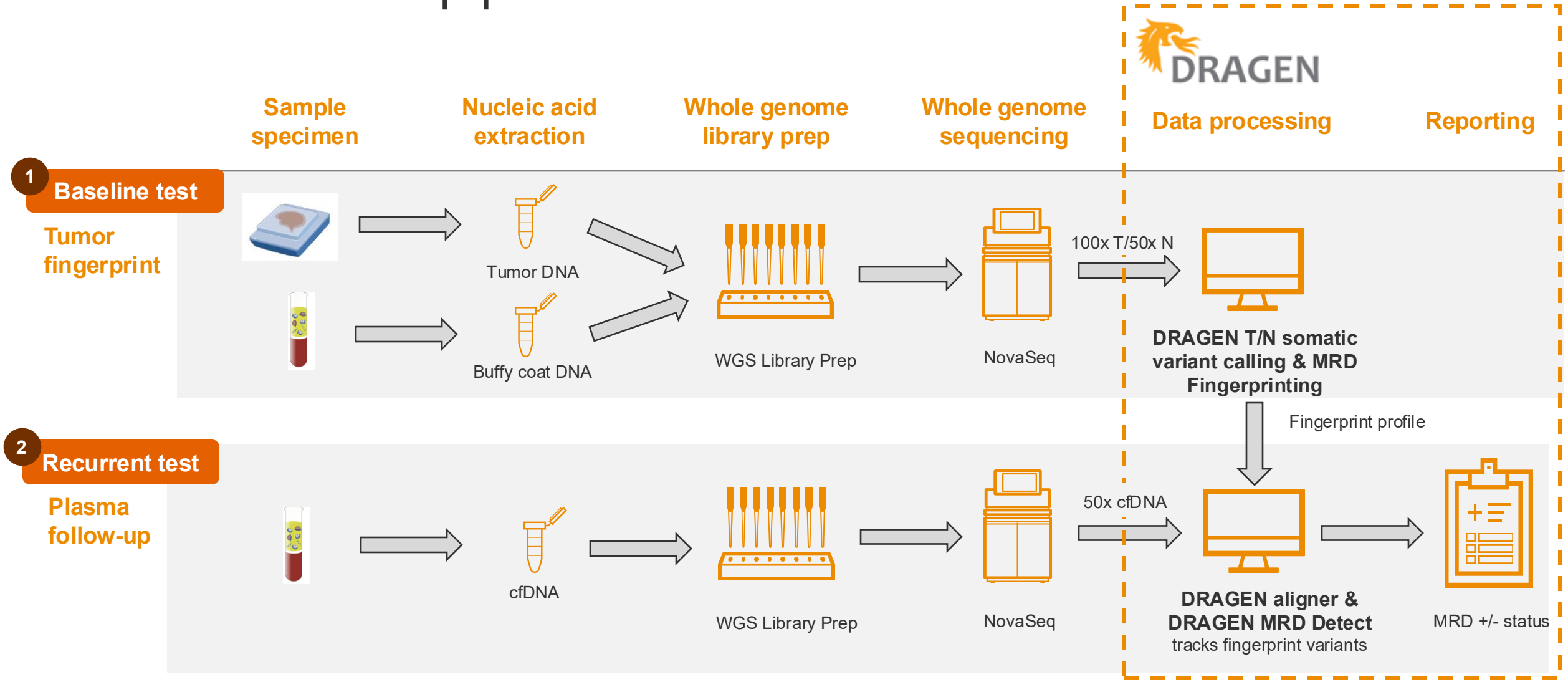
# DRAGEN with Connected Insights enables streamlined analysis & interpretation for oncology




*tumor/normal
**tumor only

illumina®

# DRAGEN MRD pipeline for tumor-informed WGS MRD



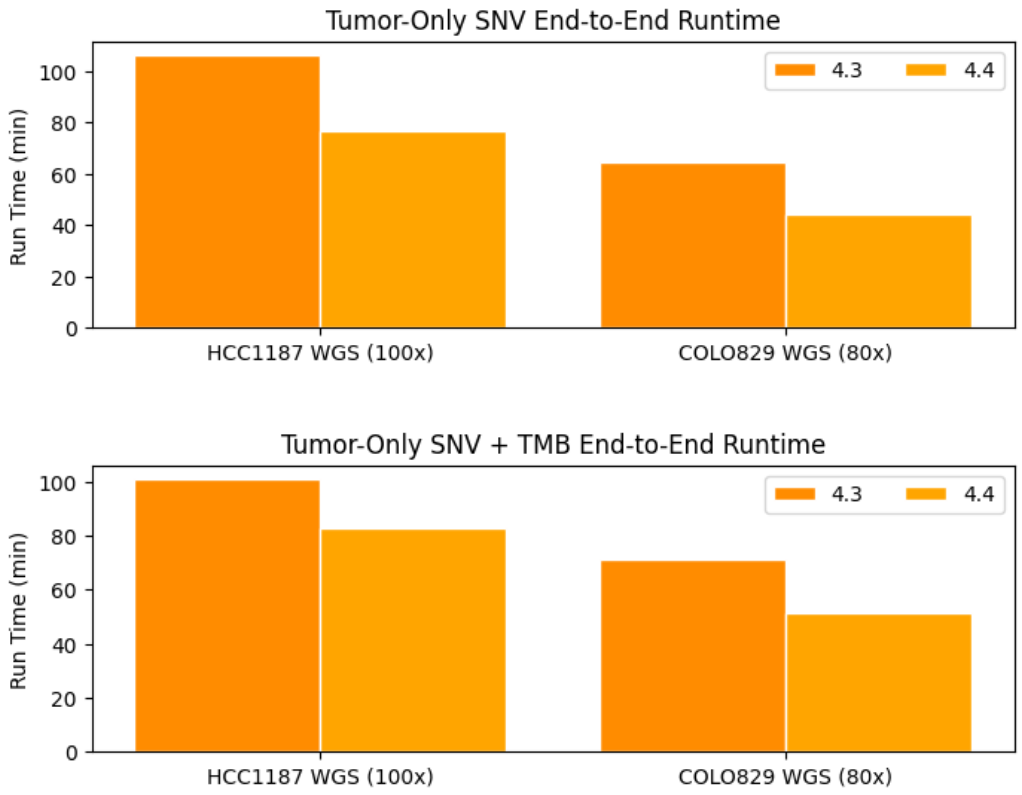
# Faster tumor-only SNV and TMB run time by ~25%

## **Significant speed up in** VC Germline Tagging and TMB

- No longer required to run full variant annotation
- Tumor-only run time improvements up to 25%

## User Interface changes

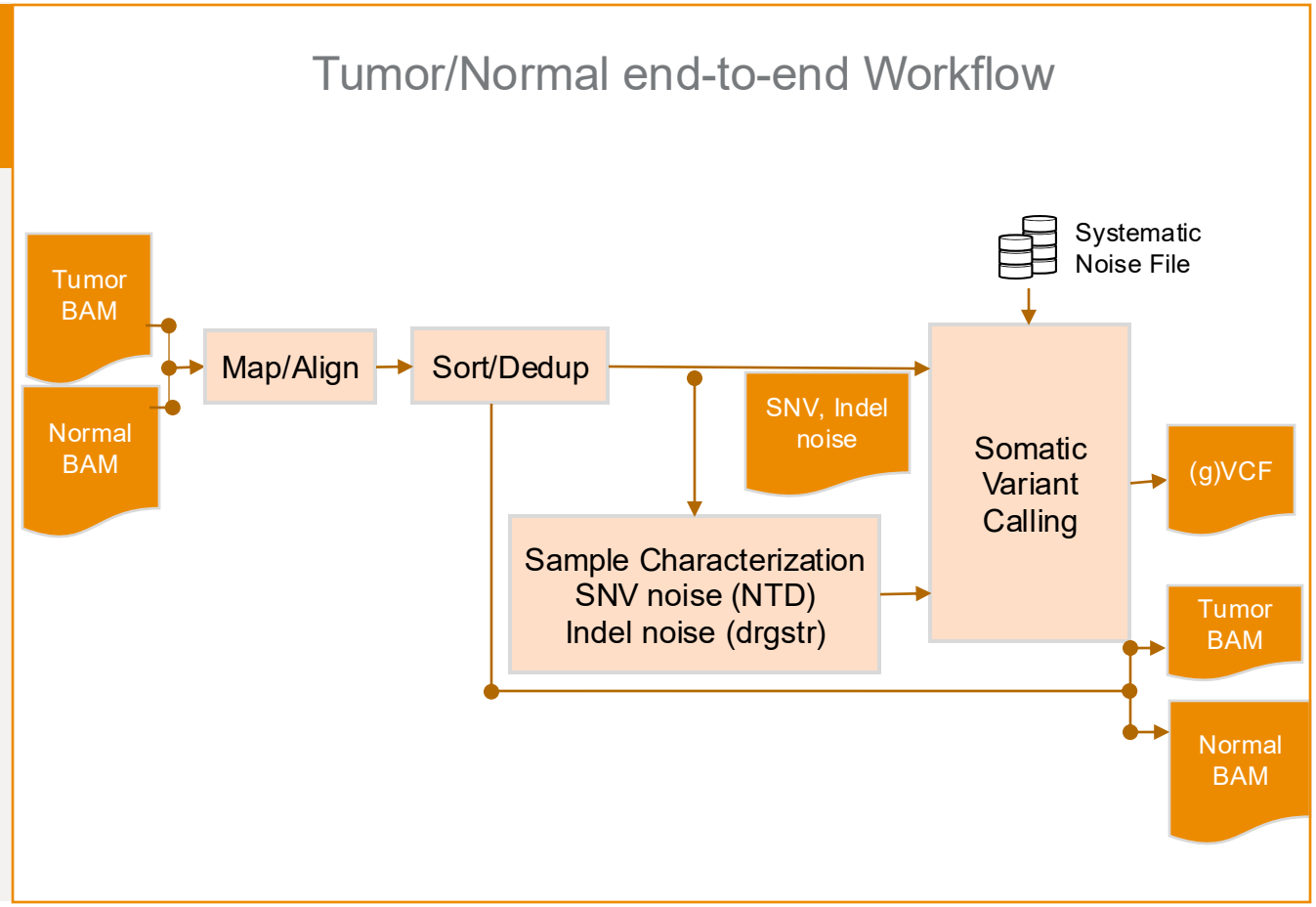
- Pass “--enable-variant-annotation true” and it will do full annotation on all DRAGEN (G)VCF outputs
- Omit “--enable-variant-annotation” and it will automatically only annotate the VCF file with reduced set of annotations needed for TMB/tagging



# Support for Tumor + Normal BAM/CRAM input to mapper

## Enables tumor/normal end-to-end analyses from BAM/CRAM input

- End-to-end analysis for tumor/normal is now supported for all input types
- No longer need to remap tumor and normal BAMs/CRAMs separately before VC
- Cuts what used to be a 3-execution workflow down to a single DRAGEN run
- Greater flexibility for users to design their preferred workflows

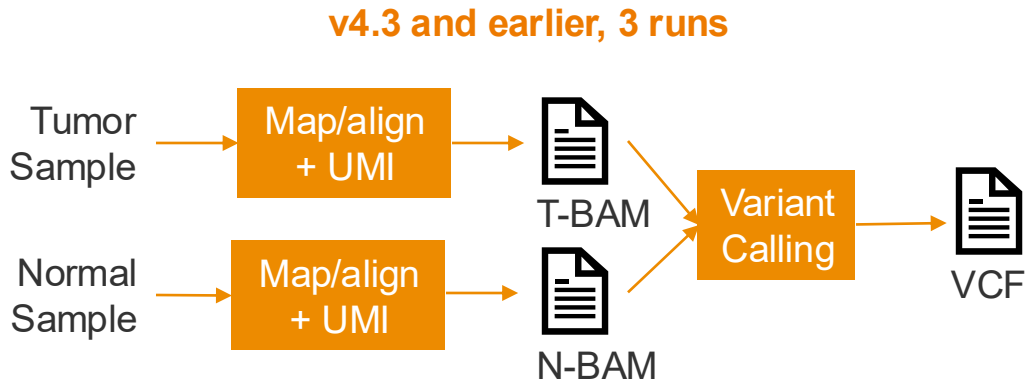


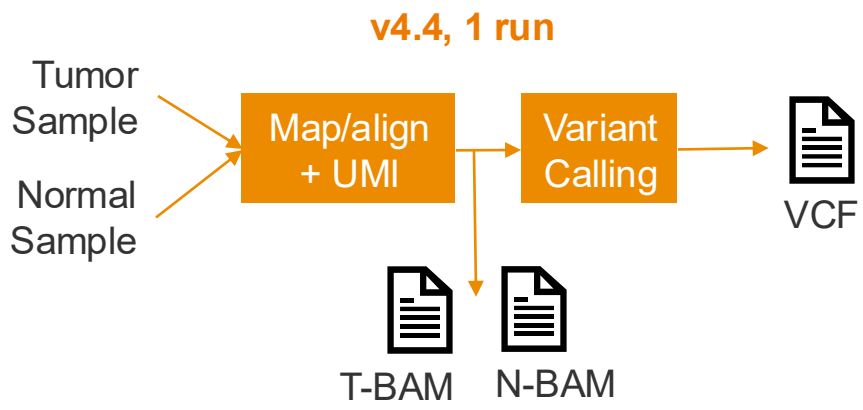
# Support for Tumor/Normal workflow with UMI

## DRAGEN now supports Tumor/Normal input with UMI in a single run

- Somatic workflow in Tumor/Normal mode may now include UMI collapsing for Tumor, or for both Tumor and Normal samples.
- Eliminates the need for 3 separate runs to do alignment + UMI collapsing per sample, and Variant Calling from the BAMs
- Feature enabled via

  --tumor-normal-has-umi={both, tumor}
- Supported for
  - WGS, Exome or Panel samples types
  - All UMI types



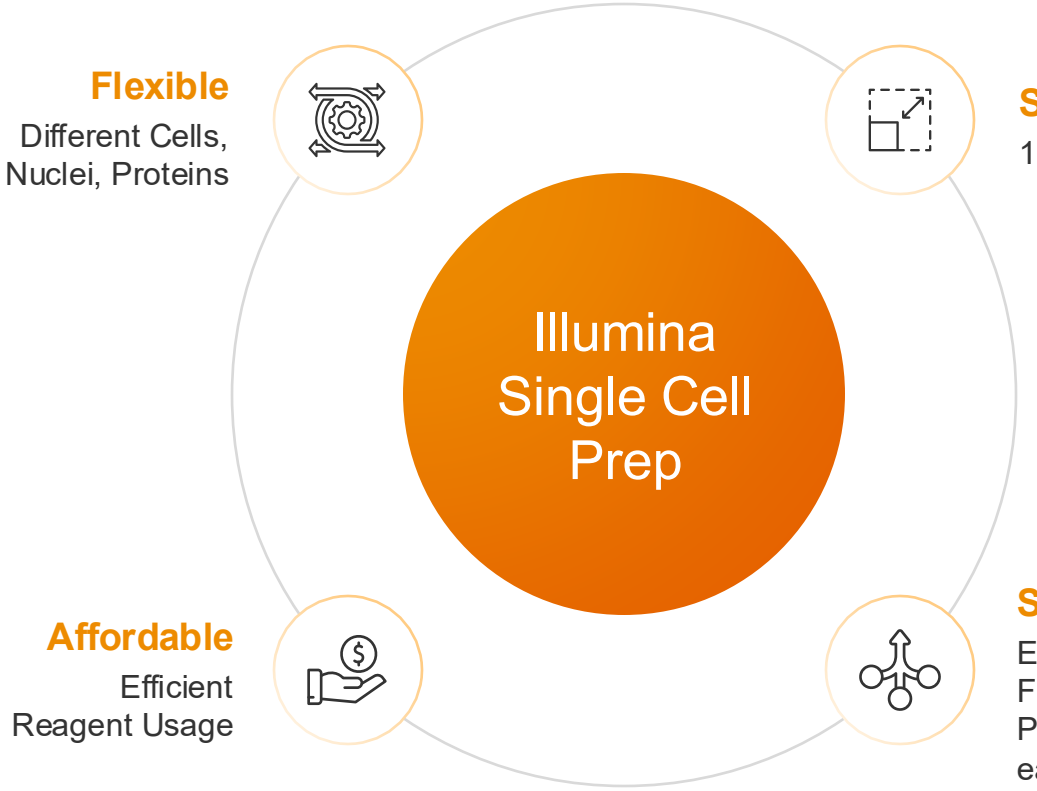
# DRAGEN Multiomics


illumina®

# Eliminating limitations in single-cell analysis

## Current Industry Constraints

- Expensive instrumentation and consumables
- Limited scale
- Rigid workflows
- Significant expertise required
- Analysis requires bioinformatics capabilities




**Flexible**
Different Cells, Nuclei, Proteins

**Scalable**
100 to 1M Cells

**Affordable**
Efficient Reagent Usage

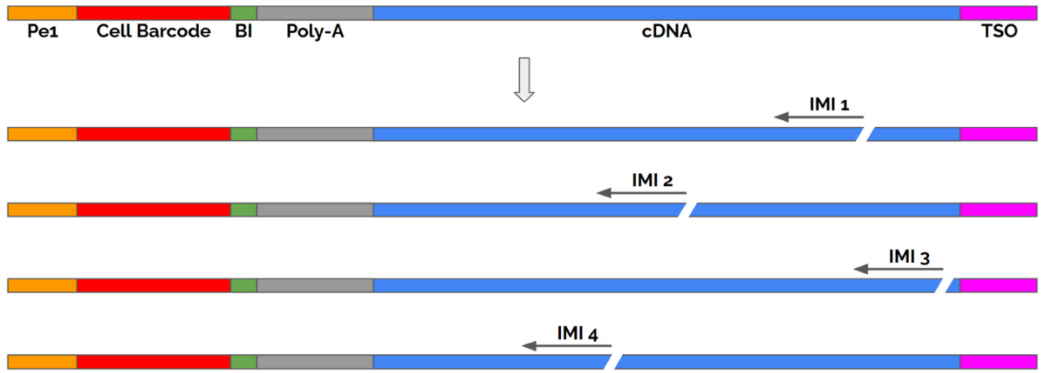
**Simple**
Easy Workflow, Flexible Stopping Points, intuitive, easy-to-use analysis

# Illumina Single Cell 3’ RNA Analysis available on DRAGEN

**DRAGEN now supports scRNA analysis from Illumina Single Cell Prep**

- Output: Raw matrix, filtered matrix of cells, and full set of metrics output for each sample
- New molecular counting strategy utilizing binning indexes (Bis) and intrinsic molecular identifiers (IMIs)
- Clustering and UMAP now supported by DRAGEN reports
- Barcode and feature counting and correction with new IMIs-per-molecule (IPM) algorithm automatically corrects IMI counts

## Structure of Illumina Prep FASTQ files




**Availability**

BSSH | ICA | Multi-Cloud | On-premises server

# Analyze multiple scRNA preps simultaneously with state-of-the-art methods and visualizations

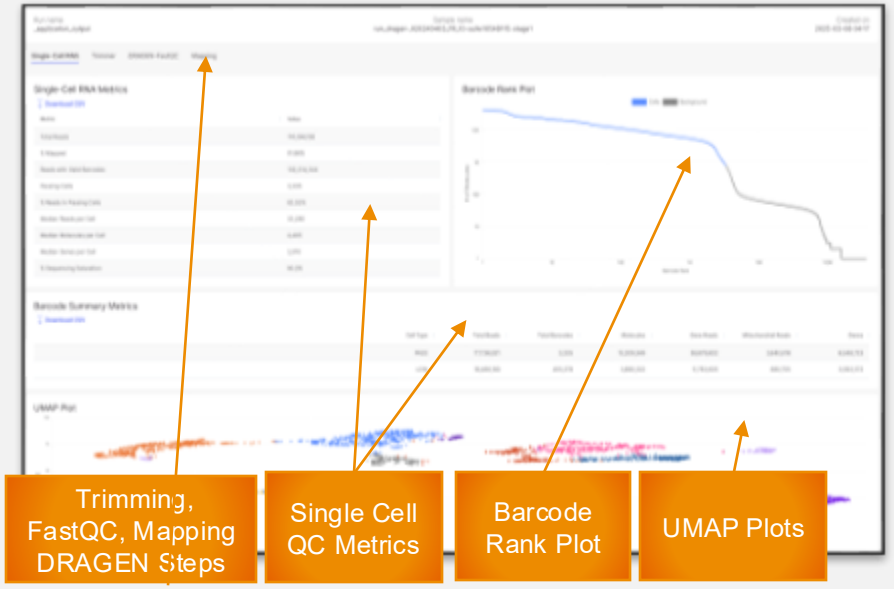
## DRAGEN Single-Cell RNA Pipeline

Hardware acceleration provides efficient secondary analysis of large kits

| Kit | # Cells | Run Time* | |
|---|---|---|---|
| | | Open-Source | DRAGEN |
| T2 | 2k | 30m | 6m |
| T20 | 20k | 2.5h | 0.5h |
| T100 | 125k | 28h | 2.5h |

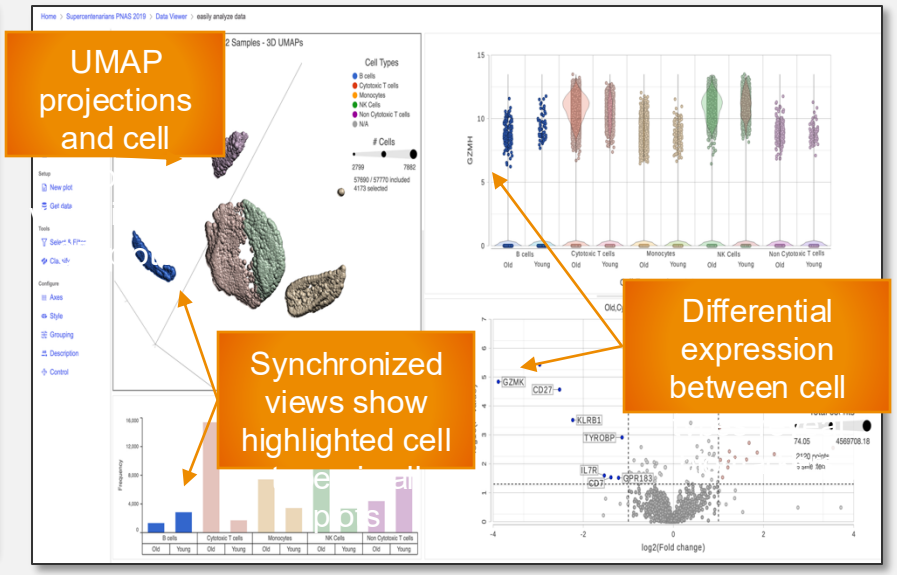
*Illumina internal data on file, 2025, DRAGEN, open-source data from Pipseeker

DRAGEN reports highly accurate and visual reports for secondary analysis QC




## Connected Multiomics / Partek Flow

DRAGEN outputs are available in Connected Multiomics to explore the data further



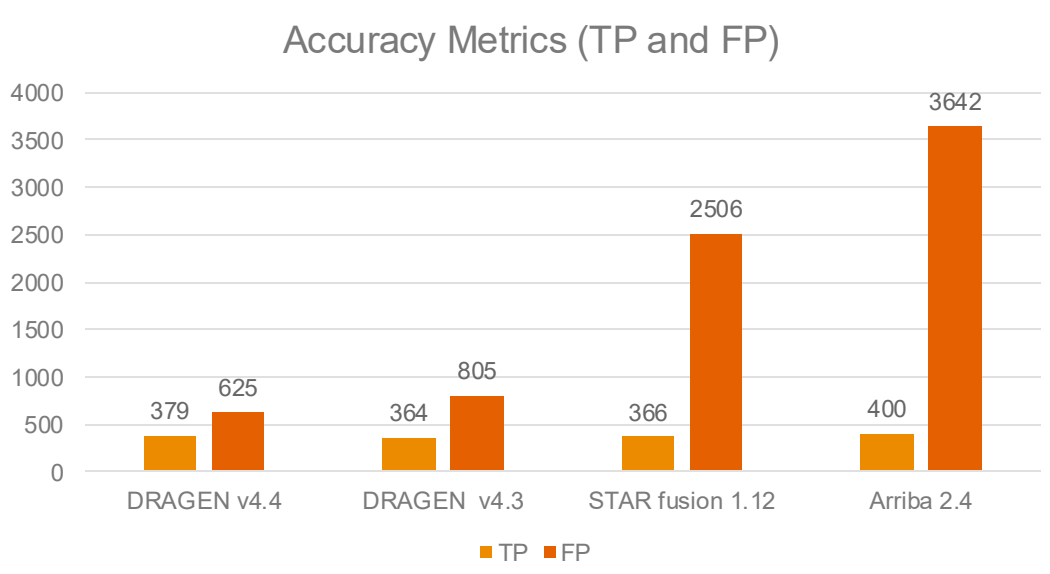

**DRAGEN scRNA license is included with the purchase of the Illumina Single Cell 3'RNA assay and will be available for free to use on the server with DRAGEN v4.4**

# DRAGEN Bulk RNA

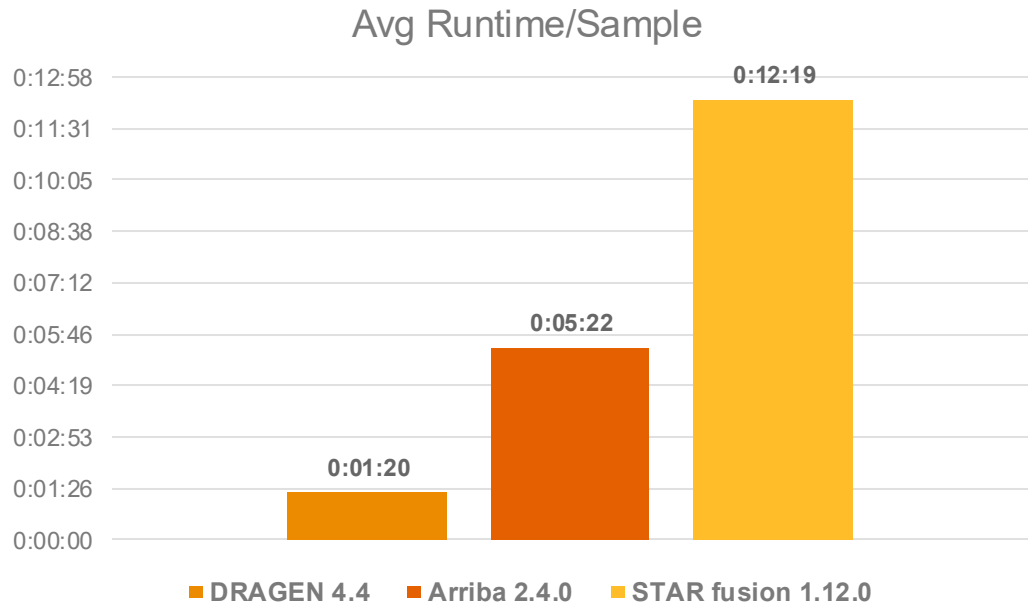
illumina

# Improved accuracy of RNA gene fusion detection

## Improvement across versions and vs 3rd Parties




## With ultra-rapid analysis time




- ✓ Total of 179 samples with known validated & “expected” fusions from five different sources
- ✓ 400 of the 531 “expected” fusions are detectable across all callers.

# Assembled fusion sequences and precise breakpoints for in-frame/out-of-frame determination

**Gene fusion sequence is assembled and aligned to determine precise breakpoints**

- Fusion sequence assembly is enabled using "**--rna-gf-output-fusion-sequence**" (set to **true** by default)
- De novo assembly is performed on all evidential split reads to produce fusion sequence
- Assembled sequence are aligned against left and right breakpoint reference sequences to determine more precise breakpoints
- Fusion candidates VCF and *.final files automatically report more precise breakpoints based on alignments of assembled sequences

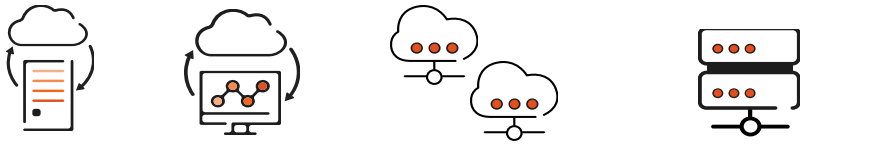
## Output of <*prefix*>.*fusion_candidates.final*

| … LeftBreakpoint | RightBreakpoint | … | **FusionSequence** | BreakpointLeeway … |
|---|---|---|---|---|
| … chr2:134120290:+ | chr22:22922721:+ | … | GCGACCTCGCG… | -1\|+2 … |
| … chr22:23290413:+ | chr9:130854064:+ | … | GGGTTTCTGAAT… | -0\|+0 … |
| … chr20:50795173:+ | chr17:61368327:+ | … | GGCGCAACCAC… | -1\|+2 … |

*Corrected* *New Fields*

- "LeftBreakpoint" and "RightBreakpoint" updated to more precise values based on alignment of assembled sequence
- Full fusion sequence reported under "FusionSequence"
- "BreakpointLeeway" indicates allowable leeway for top alignment. For example, -1|+2 can be shifted to the left by 1 or to the right by 2 bp but maintain maximal alignment score

**Availability**



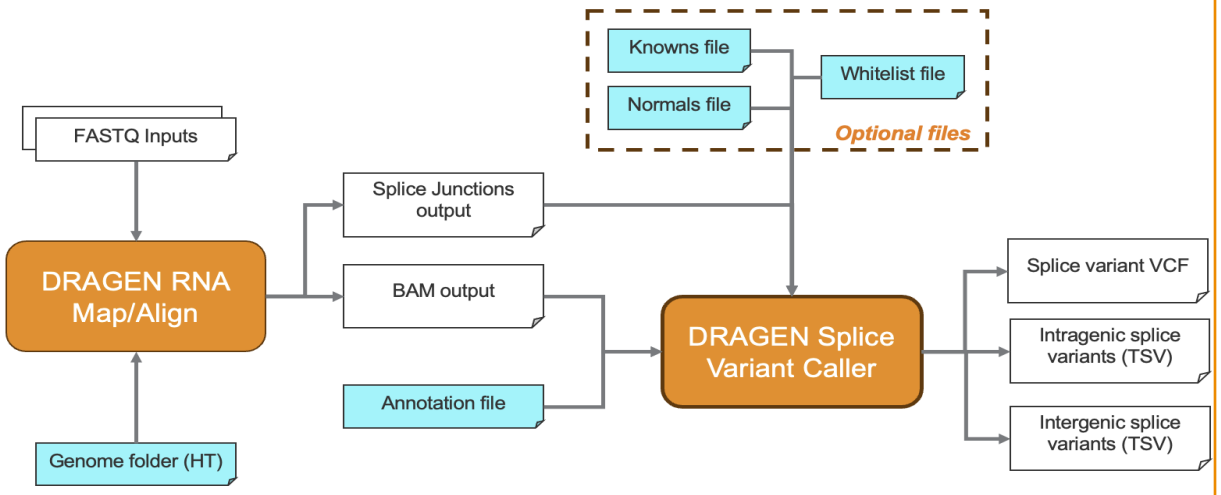
# RNA splice variant caller – increased sensitivity with new ML model and mapper updates

## Enhanced RNA splice variant caller

- ✓ **Updated mapper tuned for better splice sensitivity**
- ✓ **New ML model with scoring and filters**
- ✓ **Trained on WTS NIST HG005 with long read data**
- ✓ **Outputs:**
  - Intragenic splice variants: *splice_variants.tsv*
  - Intergenic splice variants: *splice_variant_fusions.tsv*

## New features

- Uniquely mapping read counts
- Unique and multi mapping read counts
- PCR-duplicated read counts
- Maximum and average mapQ
- Overhang as a ratio of the read length
- Confidence score (0 to 1) and filter (PASS/FAIL)




**Availability**

BSSH | ICA | Multi-Cloud | On-premises server

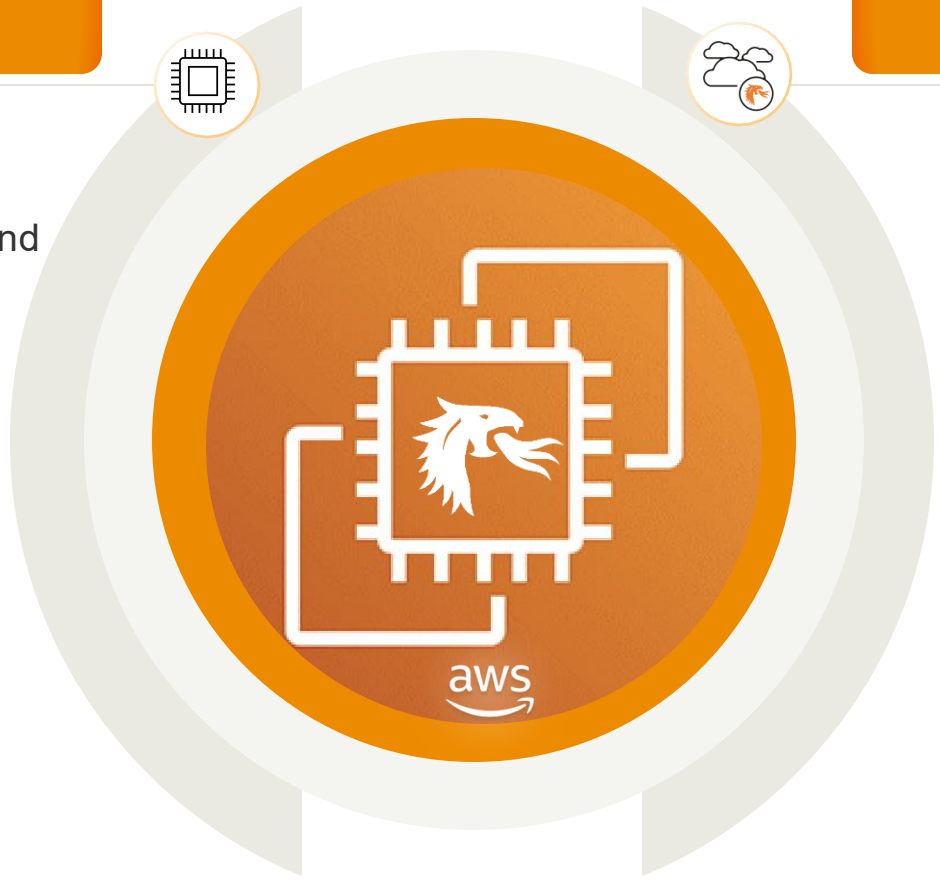
# DRAGEN on AWS F2

illumina®

# New AWS EC2 F2 FPGA instance types deliver faster TAT and improved scalability

## Second-generation FPGA-powered instance types

- New CPUs with more vCPU cores and faster clock speed
- More instance RAM
- Larger FPGA
- Improved networking and EBS bandwidth

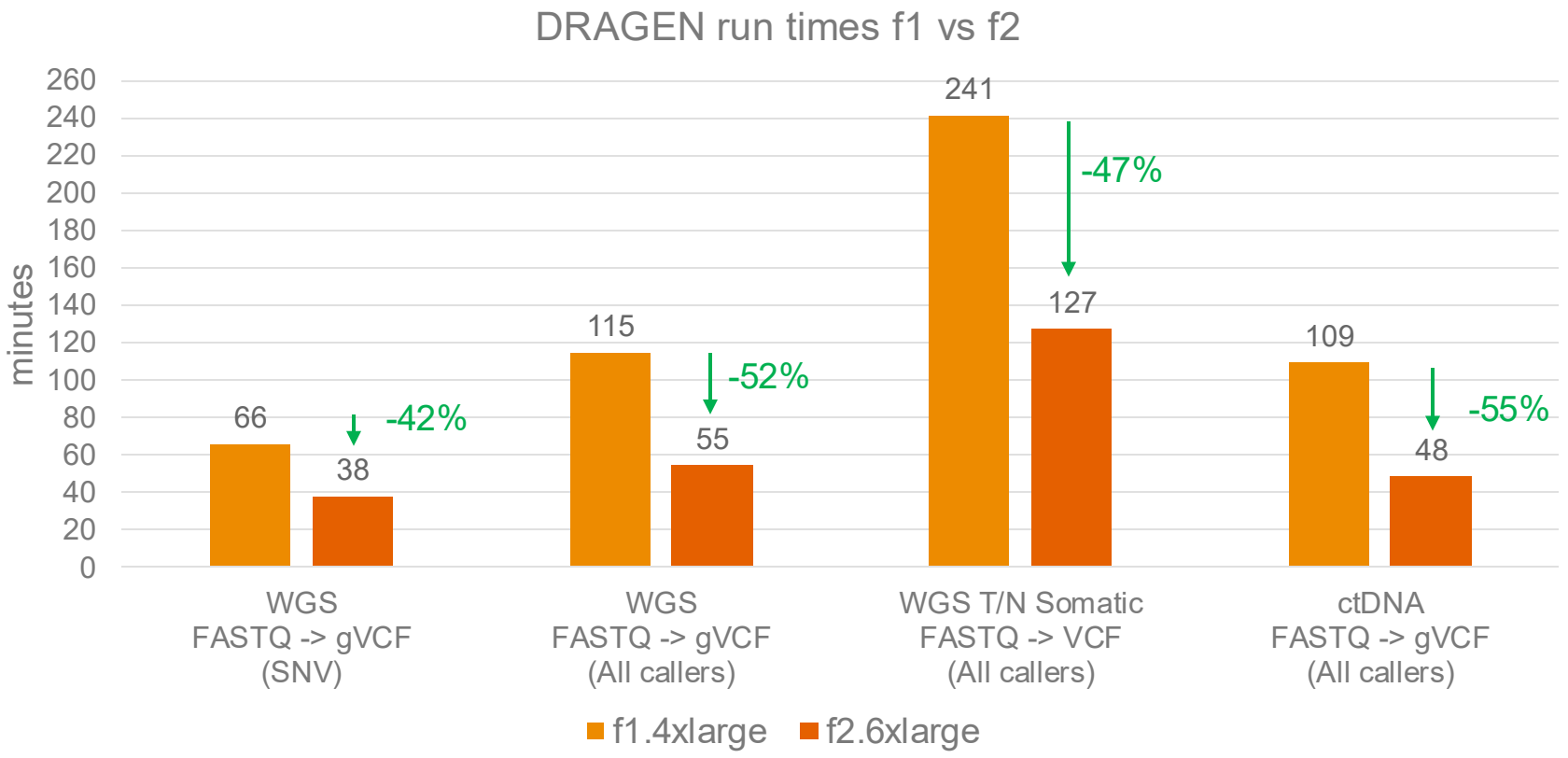



## DRAGEN v4.4 with native support for F2

- All BSSH/ICA apps with v4.4 prioritizes execution on f2 instances
- Improved turn-around times and capacity on the BSSH/ICA platform
- Lower cost for BYOL customers
- Available for BSSH/ICA and BYOL customers

# DRAGEN on AWS EC2 F2 performance

**Significant reduction in turn-around time across all workflow types**




| | |
|---|---|
| Reference | hg38 alt-masked graph |
| WGS | HG002 Novaseq PCR free 35x |
| | SNV: FQ > MA, SNV VC, Targeted > BAM, gVCF out |
| | All Callers: FQ > MA, SNV VC, CNV, SV, Repeats, Targeted > BAM, gVCF out |
| WGS T/N Somatic | HCC1187 T/N 110x/40x |
| | All callers: T/N FQ > MA, SNV VC, CNV, SV, HLA, MSI, TMB, HRD > BAM, VCF out |
| ctDNA | 961M reads |
| | All callers: Tumor FQ > MA, UMI collapse, CNV, SV, MSI, HLA, TMB, QC, Nirvana Annotation -> BAM,VCF |

illumina®

# Toolkit and platform updates


illumina®

# BCL Convert updates for greater flexibility and efficiency

## **Gain flexibility** with new sample sheet settings

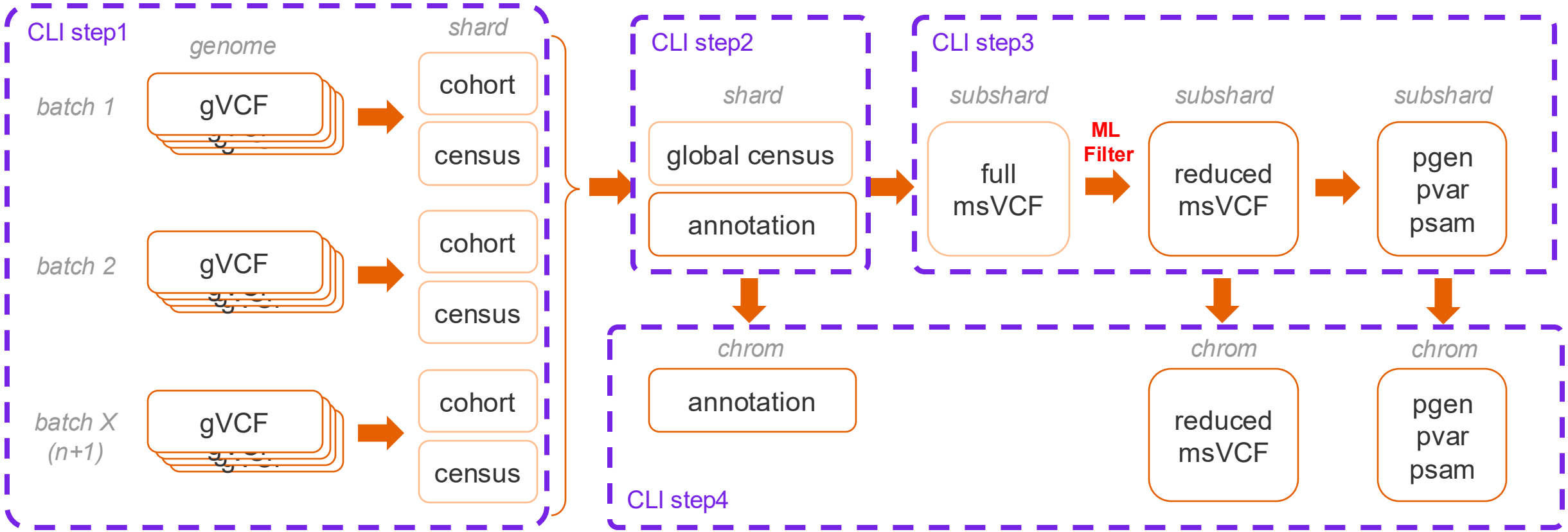
- **OverrideReads**: Redefine read structure by overriding RunInfo (read cycle boundaries, genomic/index status)
- **Index Orientation**: Specify index2 orientation as forward or reverse
- **Custom columns support**:
  - Use `custom_*` fields for per-readgroup customer data
  - v2 sample sheets ignore these fields without error – helpful for **sample management**

## **Optimize memory** with updated outputs

- **Memory optimization:** New settings reduce memory usage for large sample batches
- New `Demultiplex_Detailed_Stats.csv` file provides cycle & transition error details for mismatch tolerant demultiplexing
- Extended columns to per-tile `Demultiplex_Tile_Metrics.csv` file, to include derived stats identical to those in the aggregate `Demultiplex_Stats.csv` file

# gVCF Genotyper in ICA: aggregating variants at scale

From gVCF to msVCF, variant annotation and pgen output using the PopGen CLI




illumina®

# DRAGEN ORA compression - BCL to FASTQ.ORA up to 30% faster

- Convert BCL into ORA skipping intermediate FASTQ.GZ files, and decreasing the runtime up to 30%
- When using compression during BCL convert, simply add the command *--bcl-ora-direct true*.
- Not set as default. Limited to a max number of samples per lane of 40

Note: this option cannot be used if output requested is interleaved FASTQ.ORA files

**Example DRAGEN CMD line**

```
dragen
   --bcl-compression-only-true
   --bcl-input-directory <DIR>
   --sample-sheet <DIR>
   --fastq-compression dragen #cannot be used with
   'dragen-interleaved'
   --ora-reference <./oradata/path to reference files>
   --bcl-ora-direct true
   --output-directory <DIR>
```

**Availability**

(Not available on instrument)

**On-premises server**

# Operating System support

**Out with the old, in with the new**

- CentOS 7 support is deprecated
  On-premise builds, AWS AMIs and Azure VMs for el7 still exist for v4.4 but no longer distributed.
  Customers in need may ask for the el7 builds.
  Guidance that v4.5 will not have the ability to generate el7 anymore.

- Oracle8, el8 now fully supported on-prem, AWS AMI and Azure VMs
  Azure VMs now based on el8
  AWS Marketplace AMI is now el8

- Oracle 9, el9 now officially supported since v4.4
  Guidance that Illumina DRAGEN server OS image for Oracle 9 to be released in 2025

Terminology
- Red Hat Enterprise Linux (RHEL) is an enterprise Linux operating system developed for the business market.
- el7, el8, el9 are abbreviations for Enterprise Linux major version 7,8,9 respectively
- CentOS and Oracle are two of many open-source distributions of Linux centered around the RHEL source.

# DRAGEN unified JSON metrics

## Unified JSON output facilitates use of metrics of interest in downstream analysis

### Unified JSON Metrics

- DRAGEN 4.4 generates a unified JSON metrics file <output_prefix>.metrics.json for every run
- Unified JSON metrics for 4.4 can contain the following modules (if enabled):
  - DRAGEN metadata
  - Mapping/Aligning metrics
  - QC Region coverage metrics
  - Variant Caller metrics
  - FASTQC metrics

### Format

- Nested dictionaries by module and by grouping (global metrics vs. per read group metrics, wgs vs. per region)
- Metadata contains version, licensing, run information, etc.
- Metric data matches data in the corresponding CSV files
- Can be easily parsed with standard JSON parsing libraries

### Sample output JSON output:

```
"metadata": {
        "dragenVersion": ”4.4.123",
        "licenseInfo": [ … ],
        "runInfo": { … },
},
"modules": {
        "coverageSummary": {
                ”wgs”: { … },
                “qc-coverage-region-1”: { … },
                …
        },
        "mapAlign": {
                "globalMetrics": { … },
                 “perReadGroupMetrics”: { … }
        },
        ”variantCaller": {
              ”postFilter": { … },
              ”prefilter”: { … },
        }
}
```

# BYOL Cloud – Now you can view your licenses

- Can retrieve license information utilizing the packaged dragen_lic tool for BYOL Cloud credentials.
- Provide the dragen_lic tool the credentials using the same DRAGEN commands, --lic-server or --lic-credentials.
- Can export to a JSON format using the –j option as usual.

```
$ dragen_lic --lic-credentials my_credentials.cfg
User: <username>
DRAGEN Version: <DRAGEN version>
Time: <time>

DRAGEN Core:
    Status: Active
    Used: 68.0 Gbases since 2023-Dec-05
    Quota: 100250 Gbases
    Expiry: 2024-Feb-13
    Includes; Genome, JointGenotype, CNV, Somatic, Transcriptome License(s)

Compression:
    Status: Active - !!! EXPIRING IN LESS THAN 30 DAYS !!!
    Used: 0 Gbases since 2024-Jan-12
    Quota: Unlimited
    Expiry: 2024-Feb-13
```

# New DRAGEN v4.4 BSSH Apps & ICAv2 Pipelines

## BSSH Apps (ICAv2 backend)

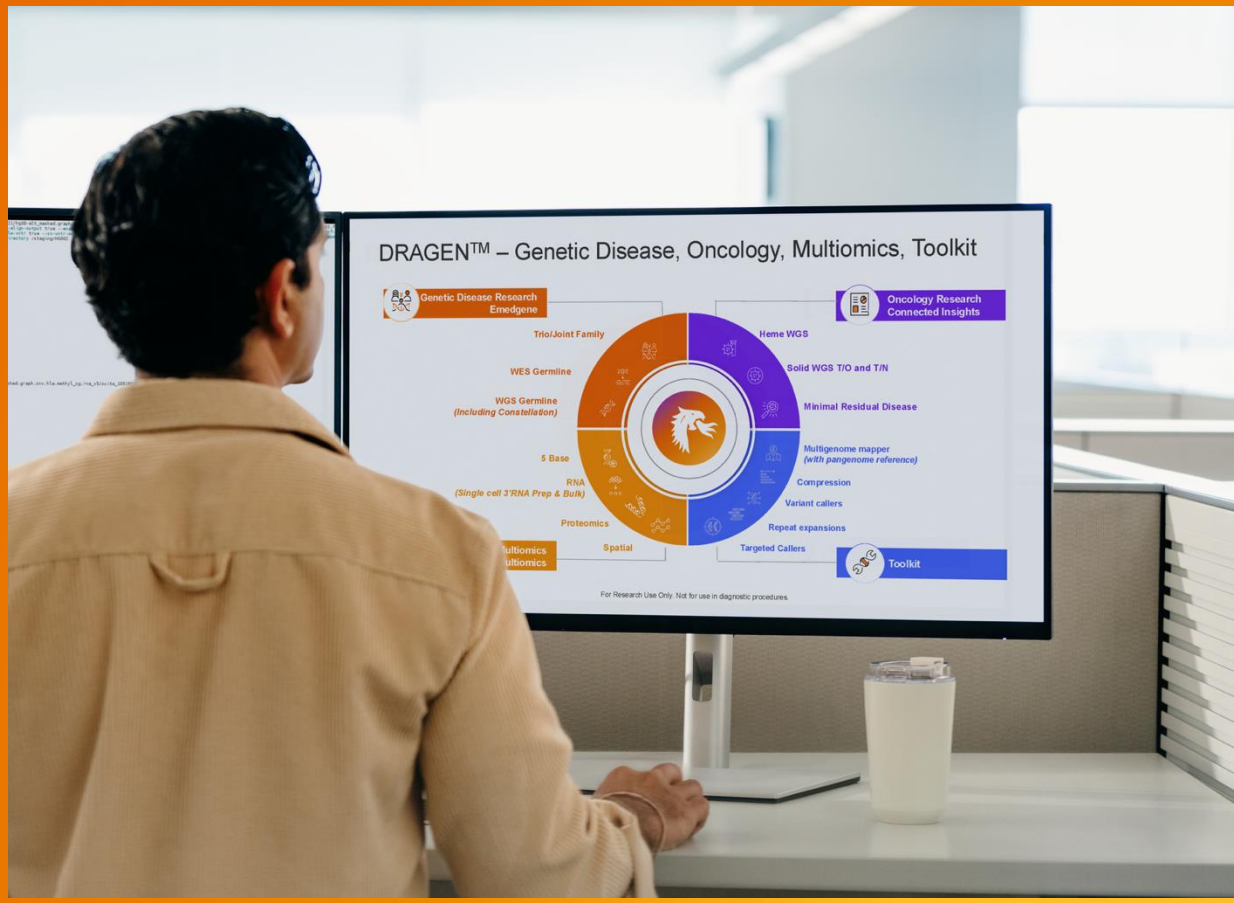
- BCL Convert
- Germline
- Somatic
- Enrichment
- Germline Enrichment from BCLs
- DRAGEN Microbial Amplicon
- Reference Builder (labs)
- Baseline Builder (labs)
- Amplicon (labs)
- RNA (labs)
- Methylation (labs)

## ICAv2 Pipelines (w/ F2 support)

- BCL Convert
- Germline (WGS & Enrichment)
- Somatic (WGS & Enrichment)
- Reference Builder
- Baseline Builder
- Amplicon
- RNA
- Population Haplotyping
- Imputation Reference Panel Builder
- Pedigree (Joint & E2E)
- Methylation
- Illumina Connected Annotations v3.25.1

# Request a DRAGEN free trial

## Accurate, comprehensive, efficient secondary analysis



# Thank you!

illumina®

# Appendix

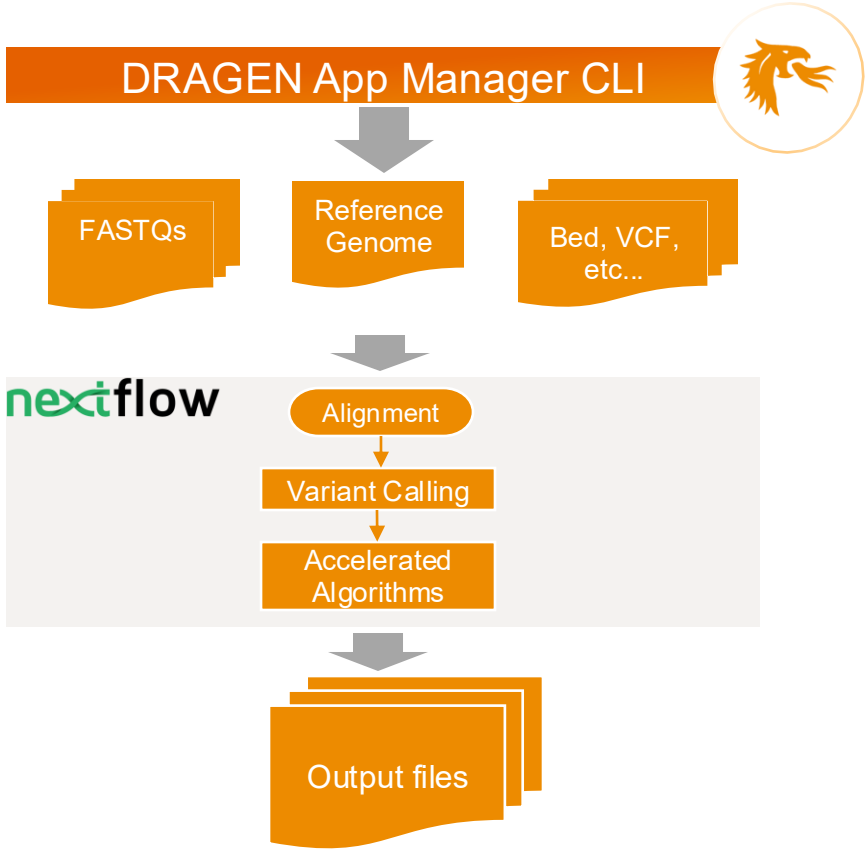
illumina®

# New DRAGEN app manager CLI

## Streamlined DRAGEN interface to install, manage and execute comprehensive DRAGEN apps

### Flexible deployment options with standalone app packages

- Easily install Nextflow DRAGEN workflows as single app packages
- Simple command line interface (CLI) to install and manage DRAGEN workflow resources such as genomes, bed files, etc..
- Single command to execute apps and monitor job status
- Supports current features in DRAGEN v4.4



# DRAGEN CNV updates

## Somatic CNV Model Improvements

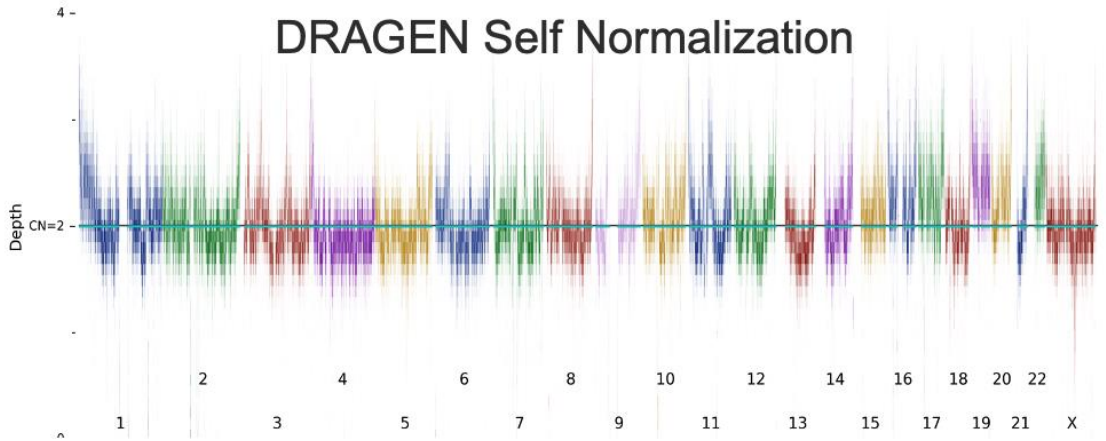
- Updates to purity/ploidy estimation allowing for higher success rate of model detection on challenging tumors.
- Improved detection for samples with low coverage, very few CNVs, noisy BAF profiles, FFPE, or extreme GC bias.

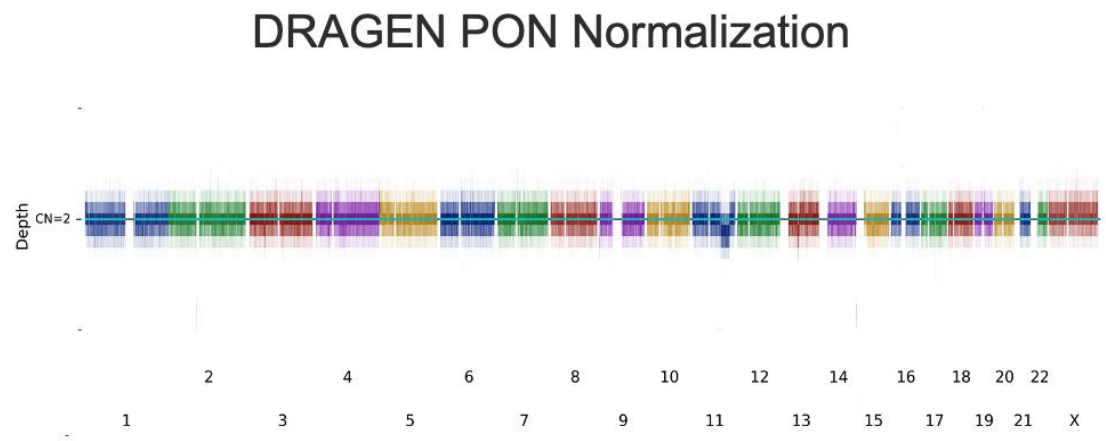
## WGS Panel of Normals Support

- WGS samples with noisy depth profiles can now be normalized with panel of normals to remove systematic biases introduced from cell line replication timing, DNA degradation, or library prep issues.

## WGS Targeted Segmentation BED

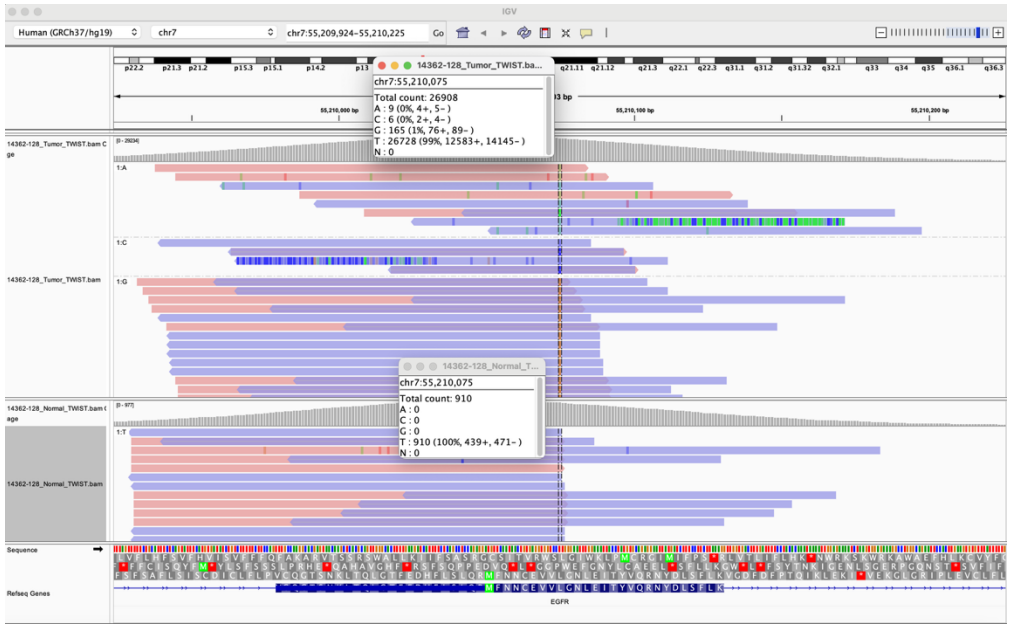
- Forced segmentation of CNV detection allows arm-level and whole chromosome detection, replacing the need for targeted panel tests to detect events such as 1p/19q co-deletions.



# Improved sensitivity for low-VAF variants in amplified regions with sample type auto - detection

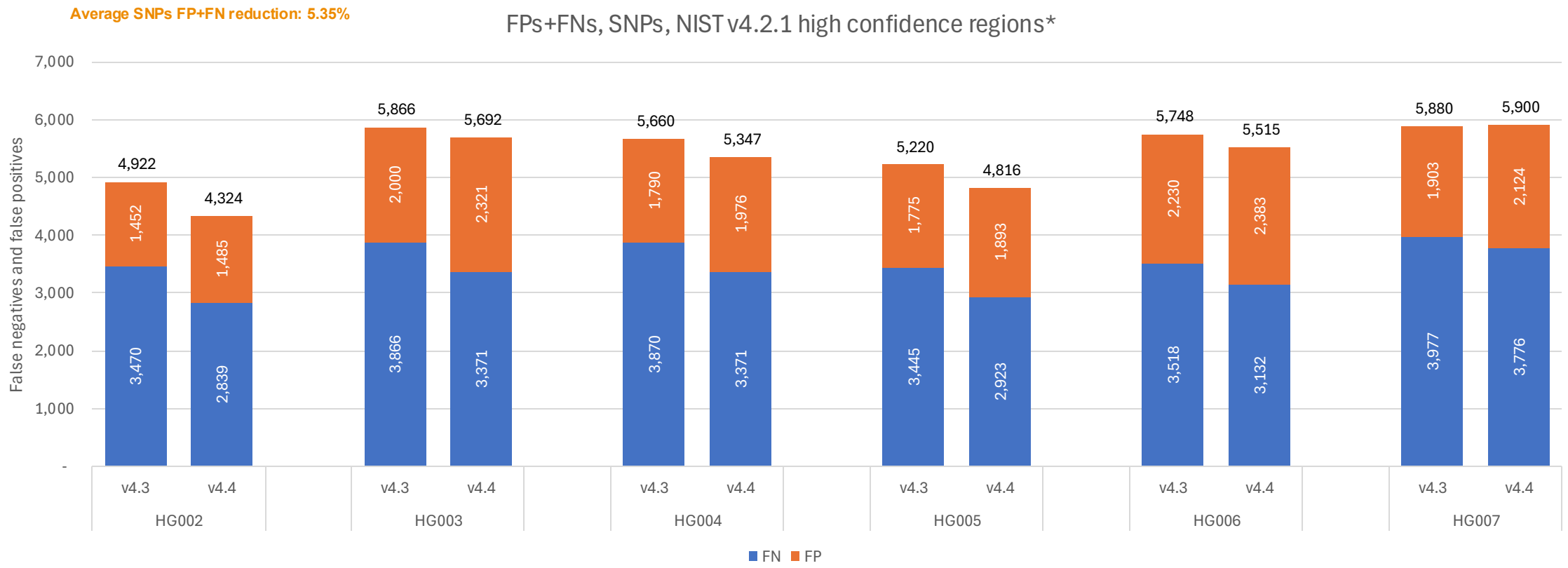
- Column-wise detection update to identify candidate variant events in high-coverage regions
- Improved identification of very low VAF events in copy number amplifications
- With new auto sample type detection, high-coverage support mode enabled automatically for WES/panel samples.
- Systematic noise method for WGS (max) or WES/panel (mean) is applied automatically; no need to set vc-systematic-noise method via the command line




Example SNV with VAF <1% even with >100 supporting reads in EGFR amplification recovered with columnwise detection update

# Germline small variant caller – Accuracy v4.2.1 SNPs

## DRAGEN 4.4 variant caller improvements result in 5.35% reduction in SNPs FP+FN




*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Refeence

# Germline small variant caller in the dark regions of the genome

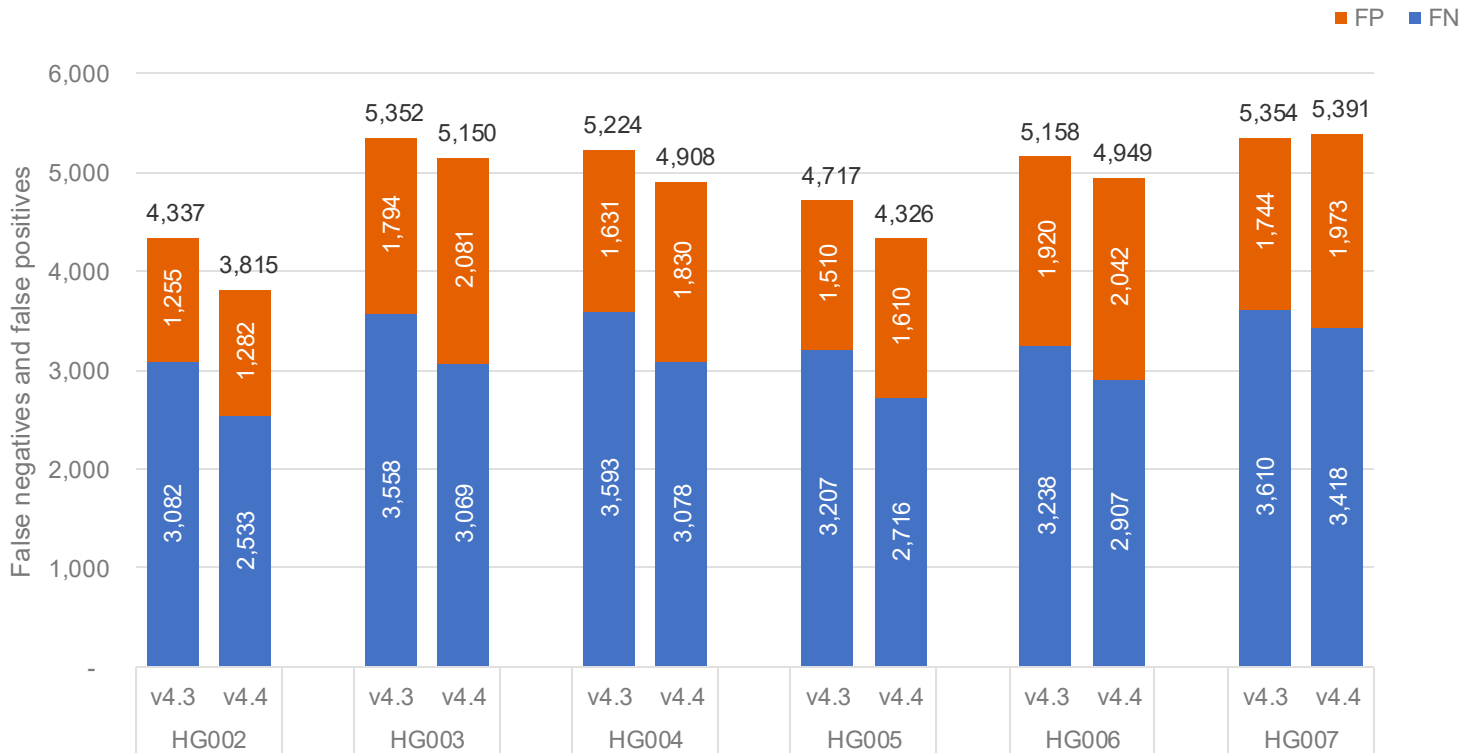
**5.58% reduction in SNPs FP+FN in difficult-to-map regions**

**Difficult-to-map regions**[1]

- regions that have other **homologous regions** in the reference genome for the given read length, number of mismatches, and number of indels.
- **Segmental duplications**

1.Difficult-to-Map Regions BED file. ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/v3.3/GRCh38@all/Union/GRCh38_alllowmapandsegdupregions.bed.gz.

FPs+FNs, SNPs, HG002-HG007*, NIST v4.2.1 difficult-to-map regions




*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Reference

illumina®

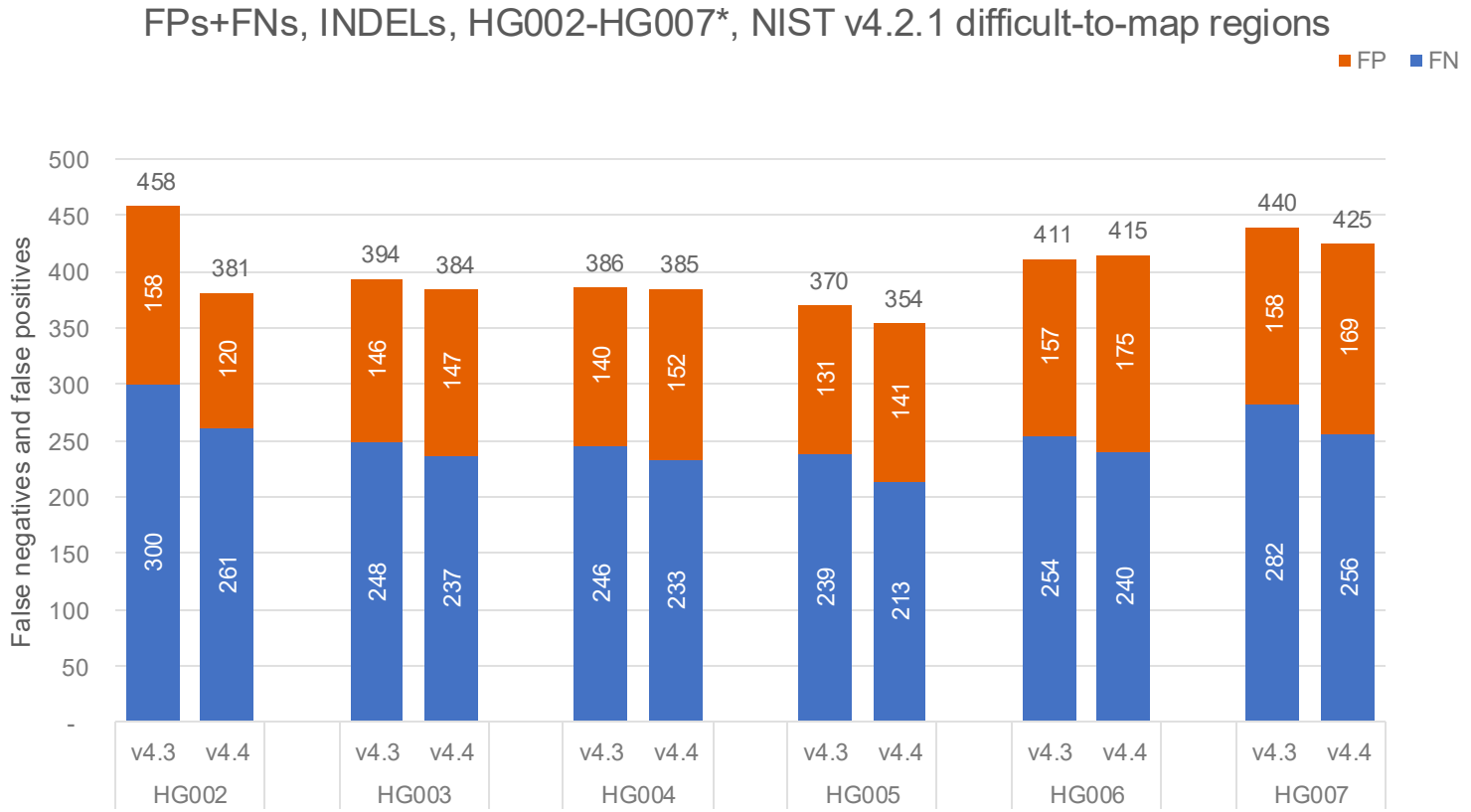
**For Research Use Only. Not for use in diagnostic procedures.**

# Germline small variant caller in the dark regions of the genome

**4.39% reduction in INDELs FP+FN in difficult-to-map regions**

**Difficult-to-map regions[1]**

- regions that have other **homologous regions** in the reference genome for the given read length, number of mismatches, and number of indels.
- **Segmental duplications**




*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Reference

1.Difficult-to-Map Regions BED file. ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/v3.3/GRCh38@all/Union/GRCh38_alllowmapandsegdupregions.bed.gz.

# New cytogenetics output reduces large call fragmentation

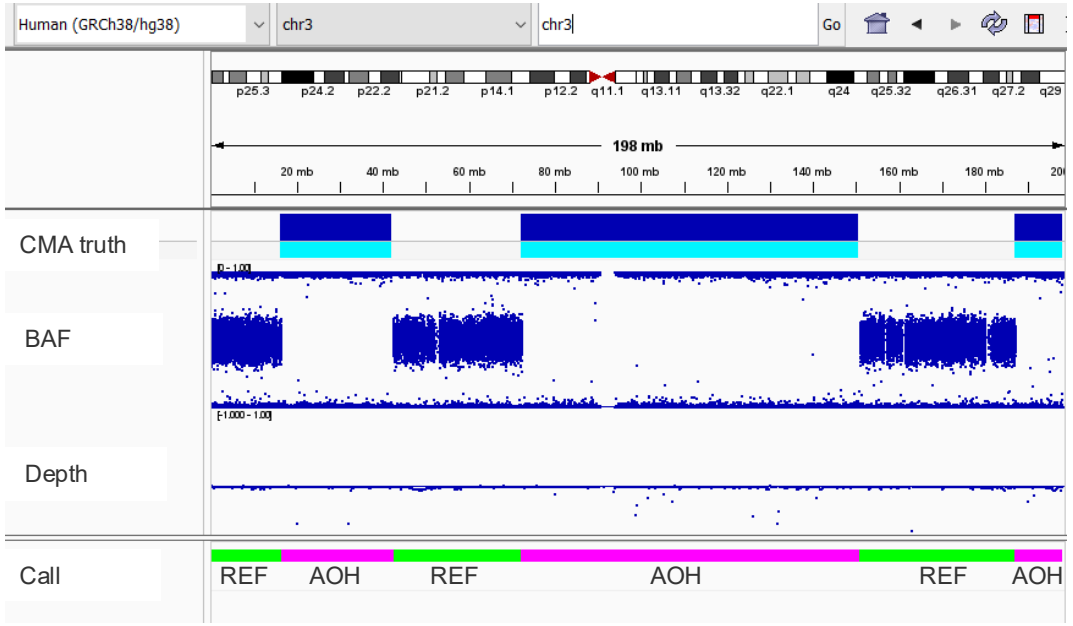
| DUP / DEL / AOH | CNV+SV VCF (*.cnv_sv.vcf.gz) | CYTO VCF (*.cyto.vcf.gz) |
|---|---|---|
| [1kb, 25kb) | Supported | N/A |
| [25kb, 500kb) | Supported | Recommended (RES=25k) |
| [500kb, 1Mb) | Supported | Recommended (RES=500k) |
| ≥ 1Mb | Supported | Recommended (RES=1M) |
| Mosaic Alterations ≥ 100kb | Supported | Supported |

Typical use case:

- Short alterations: CNV+SV VCF recommended for alterations ≥ 1kb
- 25kb to whole-chromosome alterations: CYTO VCF reduces call fragmentation for larger alterations

# Germline CNV can now detect AOH events

> 97% concordance with CMA results on internal AOH validation set (N=34 samples)

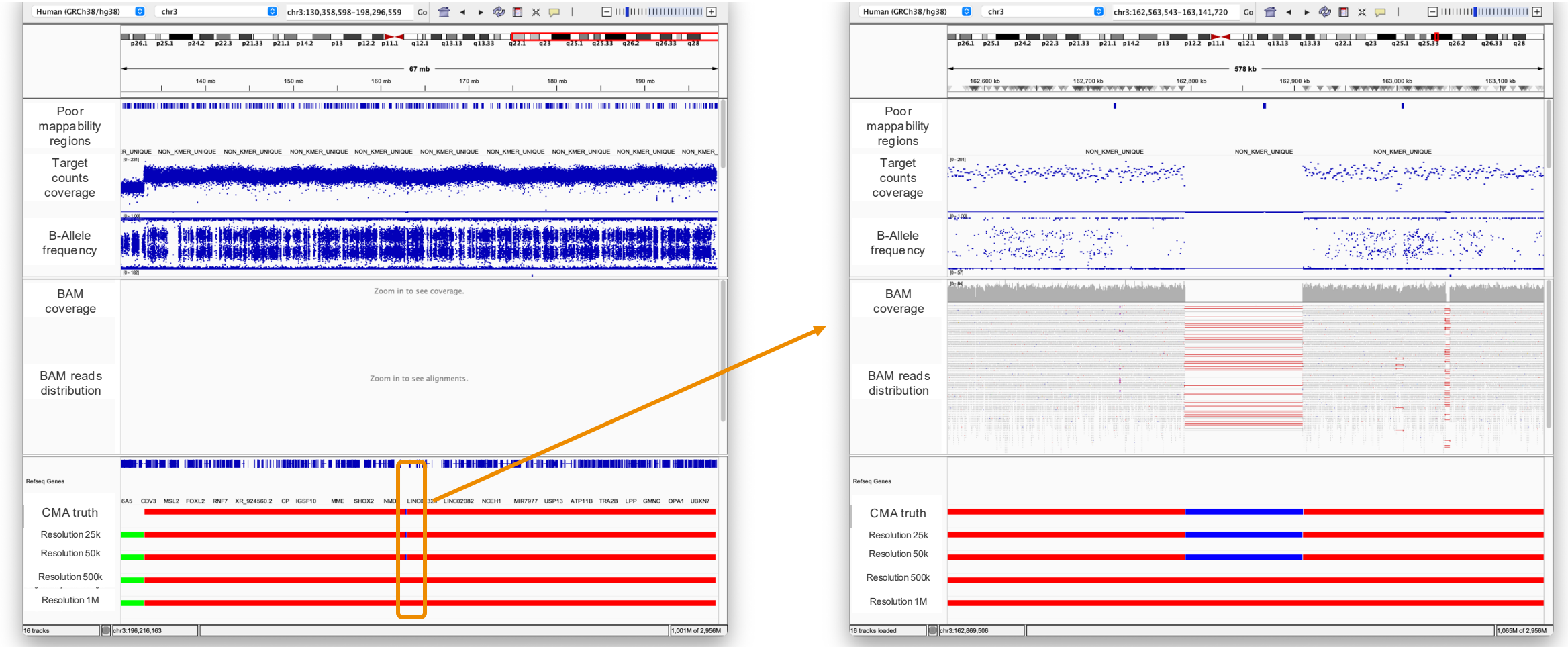



Degree of autosomal homozygosity reported by DRAGEN used as Consanguinity estimate

| Sample | Metric | Expected | Estimated |
|---|---|---|---|
| **HG01970** | Autosomal AOH % (hg38) | 12.2 | 11 |
| **NA19679** | Autosomal AOH % (hg38) | 5.7 | 5 |
| **NA20585** | Autosomal AOH % (hg38) | 4.2 | 4 |
| **NA19648** | Autosomal AOH % (hg38) | 3.3 | 3 |
| **HG01348** | Autosomal AOH % (hg38) | 3.3 | 3 |

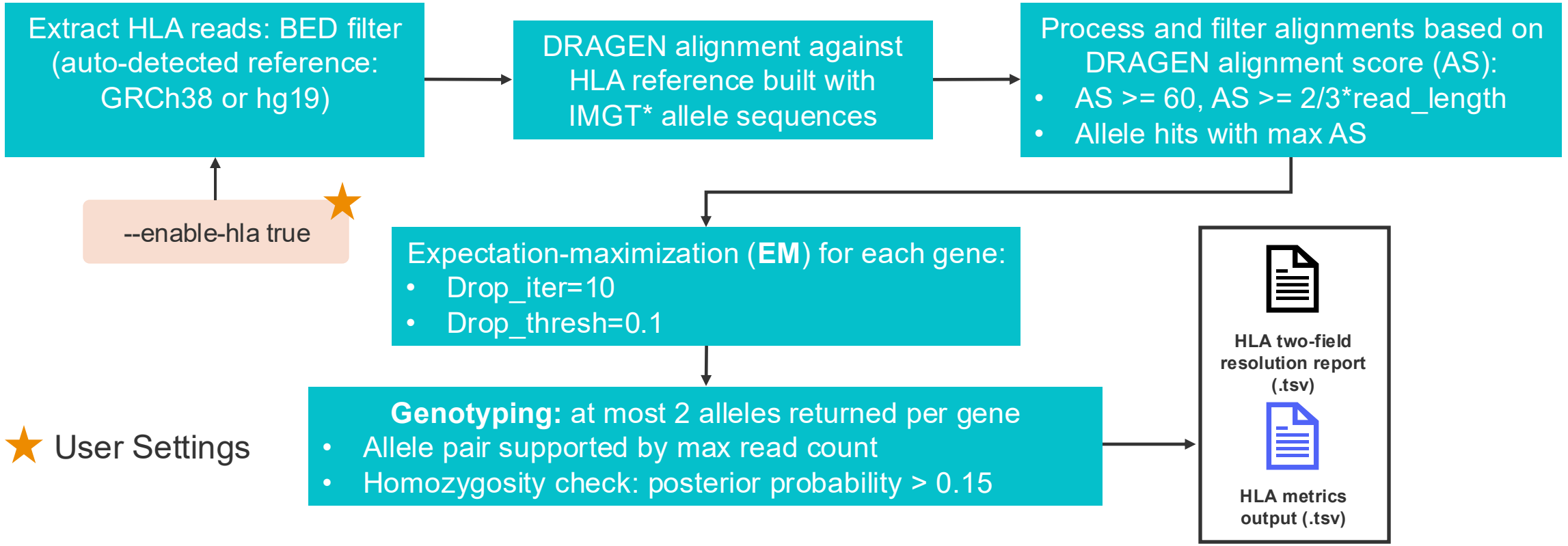
Tested on 5 samples from 1000G with published consanguinity percentage (Dong et al., 2021)

# Cytogenetics modality provides support for multi-resolution views



# DRAGEN v4.4 supports genotyping 41 HLA and HLA-related genes

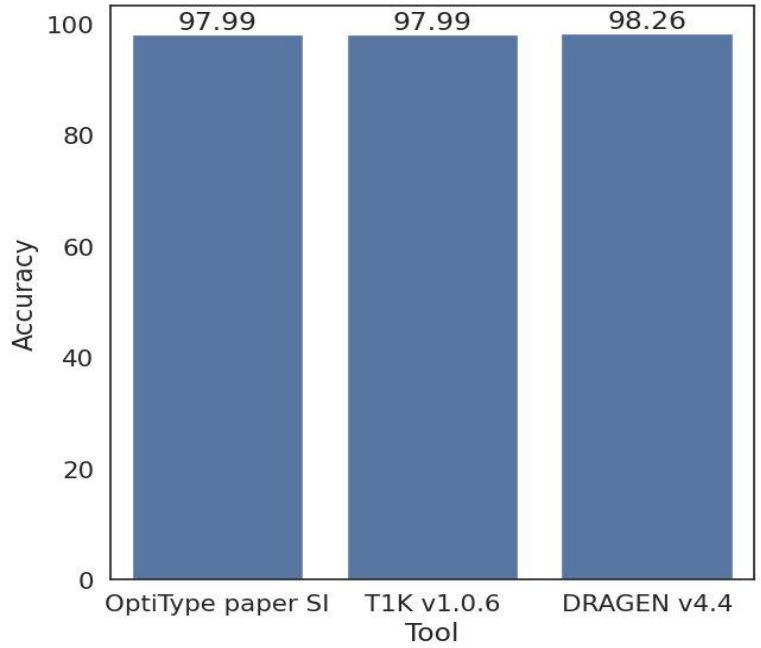
## Nucleotide-based alignments using DRAGEN map-align and expectation-maximization for HLA genotyping




*The ImMunoGeneTics project (IMGT) hosts the IPD-IMGT/HLA database, a comprehensive repository for sequences of the human major histocompatibility complex (MHC) and includes the official sequences named by the WHO Nomenclature Committee For Factors of the HLA System

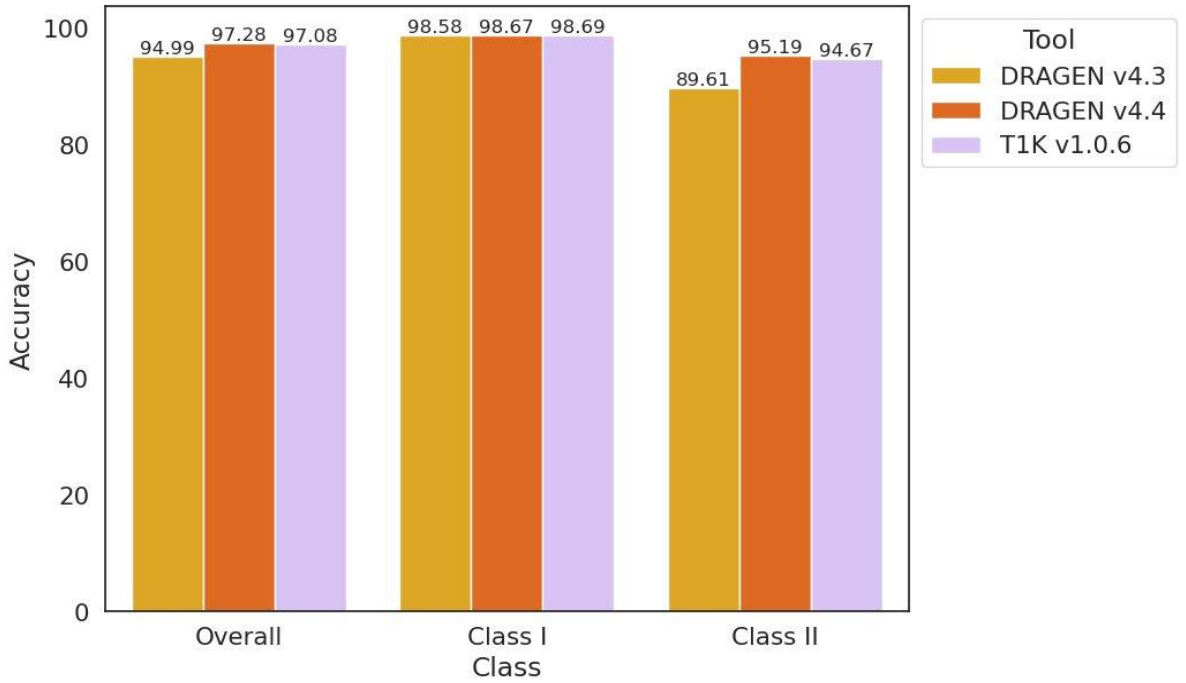
# HLA genotyping accuracy

**Class I genes (A, B, and C), Optitype truth set**




DRAGEN v4.4 HLA caller shows highest accuracy for Class I genes (A, B, and C) compared to OptiType (published data*) and T1K v1.0.6 over 182 1KG samples using Optitype truth set

**Class I genes (A, B, and C), Class II (DQB1, DRB1) 1KG truth set**




DRAGEN v4.4 HLA caller shows higher accuracy overall and higher accuracy for Class II genes (DQB1 and DRB1) compared to T1K v1.0.6 over 1,103 diverse 1KG samples using truth set derived from Sanger sequencing

## HLA caller updates in v4.4

- Extended gene list to all 41 MHC genes: 37 HLA genes and 4 HLA related genes
- Class II accuracy (esp. DRB1) improved by over 10% compared to DRAGEN v4.3
- Support for genotyping HLA starting from Tumor-Normal BAMs (requires map-align)

illumina®

*Szolek et al. OptiType: precision HLA typing from next-generation sequencing data (2014)

# DRAGEN v4.4 HLA supported genes and concordance against T1K

DRAGEN concordance against T1k over 3,202 diverse WGS samples from 1KGP

| gene | gene type | matches | N | %Concordance |
|---|---|---|---|---|
| A | Class I | 6373 | 6404 | 99.52 |
| B | Class I | 6384 | 6404 | 99.69 |
| C | Class I | 6382 | 6404 | 99.66 |
| DMA | Class II | 6362 | 6404 | 99.34 |
| DMB | Class II | 6042 | 6404 | 94.35 |
| DOA | Class II | 6309 | 6404 | 98.52 |
| DOB | Class II | 6194 | 6404 | 96.72 |
| DPA1 | Class II | 6117 | 6404 | 95.52 |
| DPA2 | Class II | 6019 | 6404 | 93.99 |
| DPB1 | Class II | 6162 | 6404 | 96.22 |
| DPB2 | Class II | 5942 | 6404 | 92.79 |
| DQA1 | Class II | 6328 | 6404 | 98.81 |
| DQA2 | Class II | 5958 | 6404 | 93.04 |
| DQB1 | Class II | 6201 | 6404 | 96.83 |
| DQB2 | Class II | 6230 | 6404 | 97.28 |
| DRA | Class II | 6403 | 6404 | 99.98 |
| DRB1 | Class II | 6141 | 6404 | 95.89 |
| DRB3 | Class II | 5541 | 6404 | 86.52 |
| DRB4 | Class II | 3315 | 6404 | 51.76 |
| DRB5 | Class II | 5089 | 6404 | 79.47 |
| E | Class I minor | 6294 | 6404 | 98.28 |
| F | Class I minor | 6368 | 6404 | 99.44 |
| G | Class I minor | 6242 | 6404 | 97.47 |
| H | Class I pseudogene | 6379 | 6404 | 99.61 |
| HFE | Class I like - haemochromatosis | 6314 | 6404 | 98.59 |
| J | Class I pseudogene | 6404 | 6404 | 100 |
| K | Class I pseudogene | 5967 | 6404 | 93.18 |
| L | Class I pseudogene | 6262 | 6404 | 97.78 |
| MICA | HLA class I related | 6175 | 6404 | 96.42 |
| MICB | HLA class I related | 6215 | 6404 | 97.05 |
| N | Class I pseudogene | 6404 | 6404 | 100 |
| P | Class I pseudogene | 6246 | 6404 | 97.53 |
| R | Class I pseudogene | 6404 | 6404 | 100 |
| S | Class I pseudogene | 6322 | 6404 | 98.72 |
| T | Class I pseudogene | 6346 | 6404 | 99.09 |
| TAP1 | HLA related - Antigen processing | 6093 | 6404 | 95.14 |
| TAP2 | HLA related - Antigen processing | 4202 | 6404 | 65.62 |
| U | Class I pseudogene | 5652 | 6404 | 88.26 |
| V | Class I pseudogene | 6404 | 6404 | 100 |
| W | Class I pseudogene | 5960 | 6404 | 93.07 |
| Y | Class I pseudogene | 4447 | 6404 | 69.44 |
| **Overall** | **ALL** | **246592** | **262564** | **93.92** |

Genes supported in DRAGEN v4.3

# WES targeted calling from command line

- Three steps for targeted calling from a WES run
- Available as a single workflow via ICA/BSSH app
- For correct normalization, each WES run must be processed as a batch and include samples from a single library prep.
  - **Counts generation**: For each sample in a WES run, uses read input to generate a counts file that includes read depth across regions of interest. Other components like CNV target counts generation may be enabled at the same time.
    Output file: *<output-file-prefix>.targeted.exome.counts.json.gz*
  - **PON generation**: A PON file is generated by aggregating the counts files from all samples in the WES run. This is a standalone run of DRAGEN that cannot be combined with other components.
    Output file: *<output-file-prefix>.targeted.pon.json.gz*
  - **Case sample analysis**: Targeted calling is performed from read input along with the PON file and an Illumina-provided systematic noise file*. Can run with other DRAGEN germline enrichment components.
    Output files: *<output-file-prefix>.targeted.json*
    *<output-file-prefix>.targeted.vcf.gz*

**Exome counts generation from prealigned BAM Input Example**

```
dragen \
    -r /staging/human/reference/hg38_alt_aware/DRAGEN/${HASH_TABLE_VERSION} \
    --bam-input /staging/test/data/NA12878.bam \
    --output-directory /staging/test/output \
    --output-file-prefix NA12878_dragen \
    --enable-map-align=false \
    --targeted-generate-exome-counts=true
```

**Exome PON generation from exome counts files from a single batch**

```
dragen \
--output-directory /staging/test/output \
--output-file-prefix run1 \
--targeted-pon-counts-list run1_exome_counts_list.txt
```

**Exome case sample analysis from prealigned BAM Input Example**

```
dragen \
-r /staging/human/reference/hg38_alt_aware/DRAGEN/${HASH_TABLE_VERSION} \
--bam-input /staging/test/data/NA12878.bam \
--output-directory /staging/test/output \
--output-file-prefix NA12878_dragen \
--targeted-pon run1.targeted.pon.json.gz \
--targeted-systematic-noise dragen4.4.targeted.systematic_noise.json.gz \
```

*A reference-specific systematic noise file can be downloaded from the DRAGEN Software Support Site page.

# STR motif composition for *C1* calling (and beyond …)

- Pathogenicity of STR loci like *RFC1* is highly dependent on the motif composition
- DRAGEN 4.4 introduces STR motif composition:
  - **De novo (kmer counting)**
  - **Predefined motif set (HMM decomposition)**
- Motif composition can be enabled on a per locus basis by modifying the STR catalog

Catalog specification example
(HMM decomposition)

```
{
   "LocusId": "RFC1",
   "LocusStructure": "(AARRG)*",
   "ReferenceRegion": "4:39348424-39348479",
   "VariantType": "Repeat",

   "MotifAnalysis": [
    "AAAAG", "AAAGG",
    "AAGGG", "AAGAG",
    "AACGG", "ACGGG",
    "ACAGG", "AAAGGG"] # for HMM
}
```

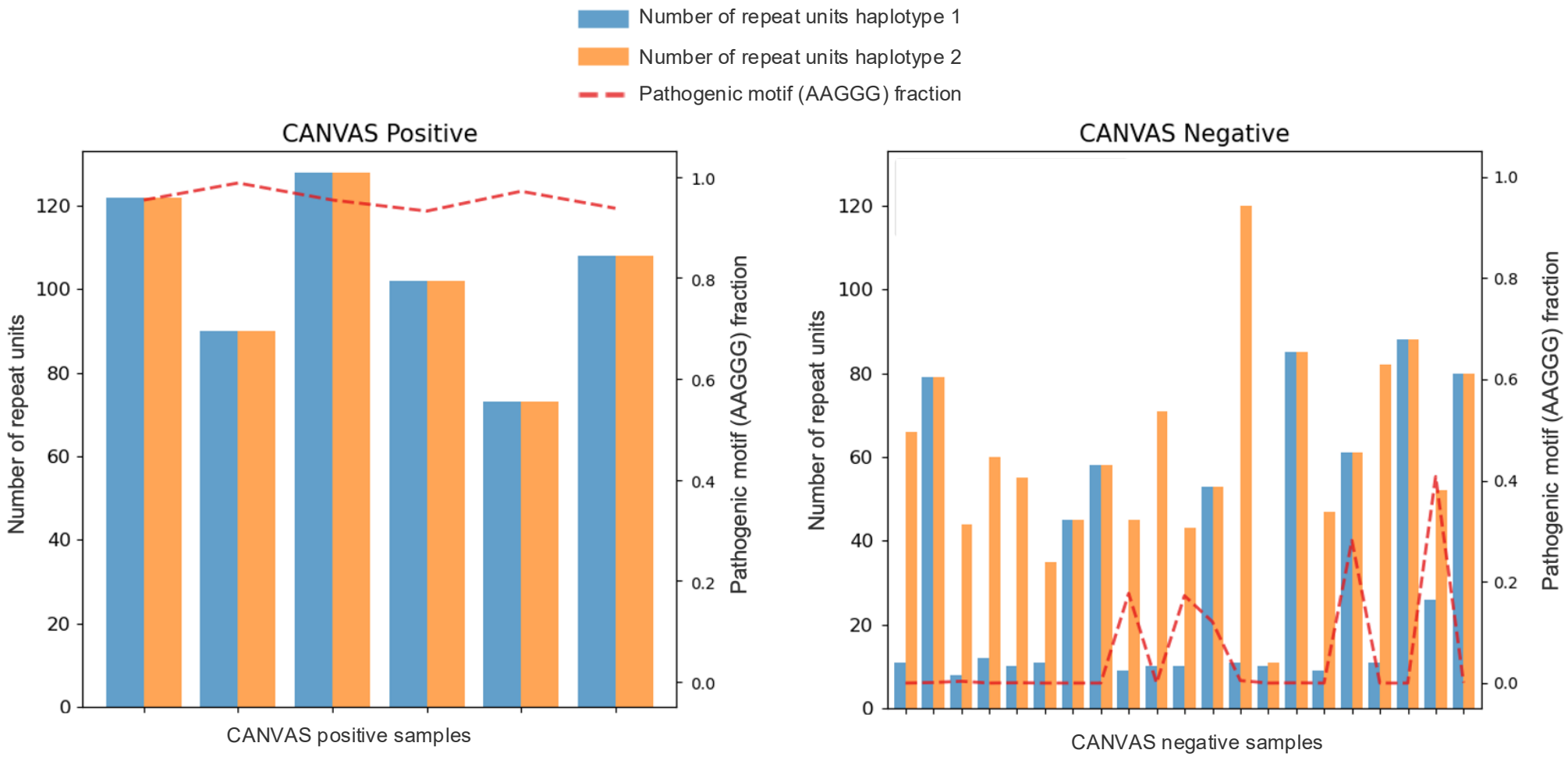
**Enabled by default for *RFC1*!**

New **FORMAT** fields **MOTIFS**, **MF** (motif fractions) and **AMF** (motif fractions per allele):

**MOTIFS:MF:AMF** AAAAG,…,AAAGG:0.15,…,0.75:0.454/0.189,…,1.000/0.000

# STR motif composition estimation helps identify pathogenic motif in RFC1 locus in individuals affected by CANVAS

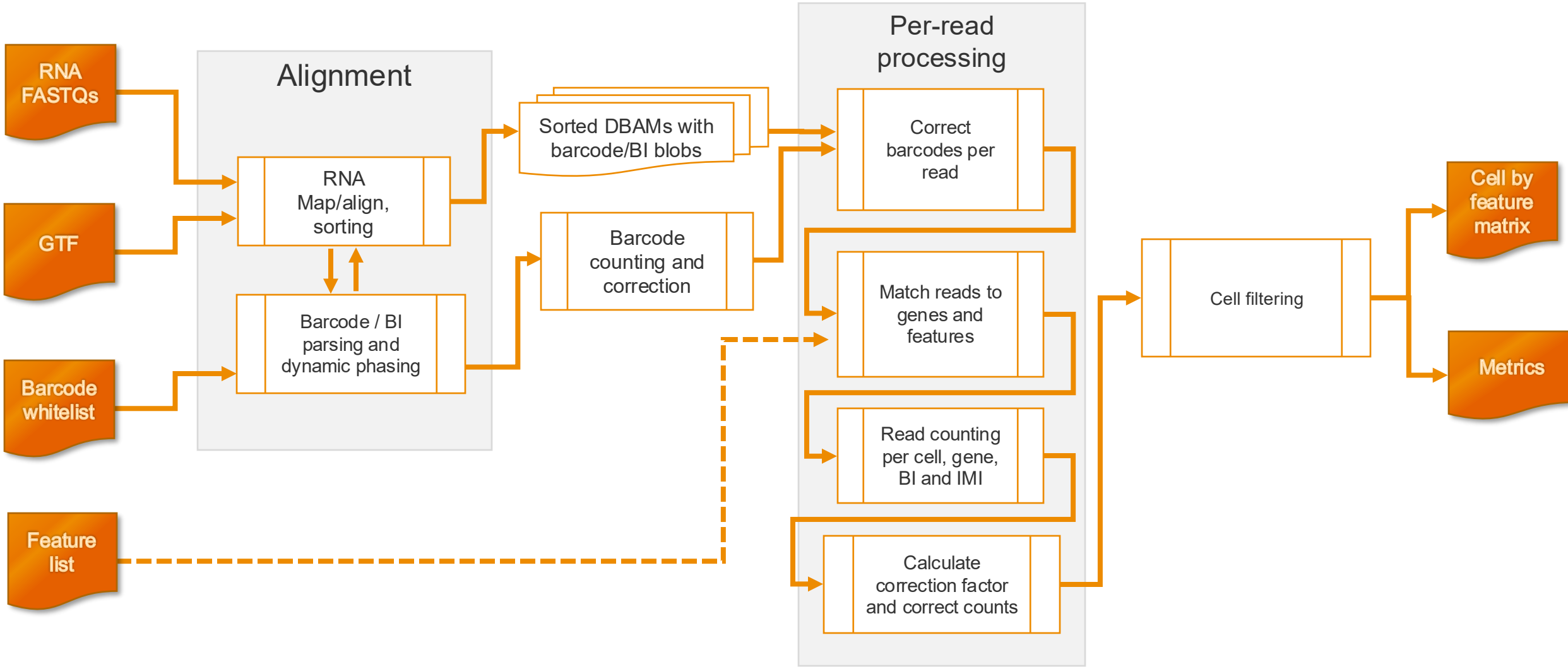
STR sizes and pathogenic motif fractions estimated by DRAGEN STR v4.4




illumina®

* CANVAS (Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome)
** Evaluation performed on a cohort of clinical samples with orthogonally confirmed CANVAS status

# DRAGEN Single-Cell RNA workflow




illumina®

# DRAGEN v4.4 - streamlined multi-region joint detection (MRJD) for de novo germline small variant calling in paralogous regions

**MRJD and DRAGEN Small Variant Caller can now be enabled in the same DRAGEN run**
**Covers six clinically relevant genes -** *NEB*, *TTN*, *SMN1/2*, *PMS2*, *STRC*, and *IKBKG*

- New tags in the VCF INFO column to differentiate different types of variant calls
  - Uniquely placed variants
  - Region-ambiguous variants
  - Potential variants under gene conversion event
  - Alternative locations for uniquely placed variants
- Enabled with *--enable-mrjd=true* option

# DRAGEN MRD: Demonstrated competitive analytical performance

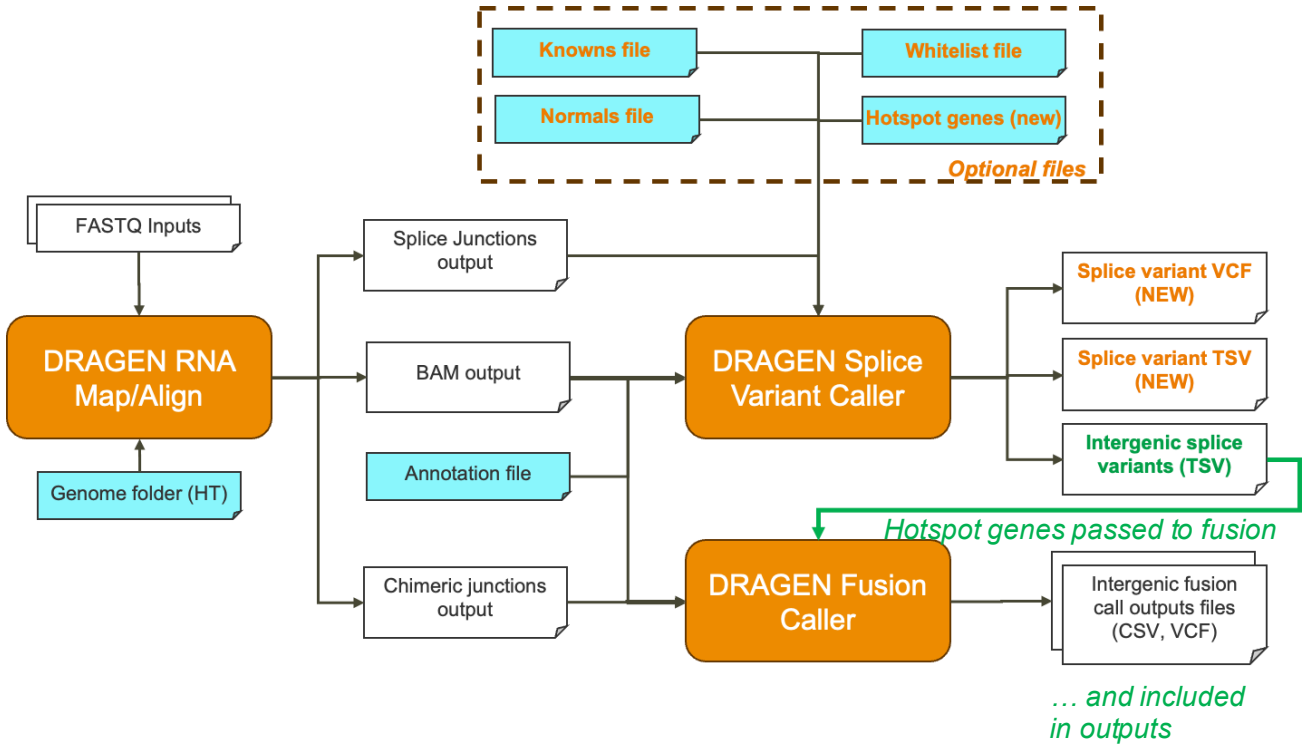
| | Cancer Type | Novaseq 6000 | Novaseq X |
|---|---|---|---|
| LoD95 | NSCLC | 0.003% | 0.003% |
| | CRC | 0.005% | 0.003% |
| | Bladder | 0.005% | 0.005% |
| | Melanoma | 0.006% | 0.006% |
| | Breast | 0.006% | 0.006% |
| Analytical Specificity | | 100%<br>[95% CI: 99.1%-100%] | 100%<br>[95% CI: 99.1%-100%] |

- Analytical sensitivity and specificity are comparable between NovaSeq 6000 and NovaSeq X
- Data generated using Illumina's R&D library prep kit compatible with both FFPE tissue (50-100ng), normal (buffy coat, 50ng), and cfDNA from plasma (2-5ng) samples

# Updated gene fusion + splice variant caller interfaces

- Normally only passing intergenic splice variants are passed to the fusion caller to be scored by the fusion ML model (skips filtered or readthrough genes)
- Option added to specify a list of hotspot genes that may contain spliced fusions: **--rna-splice-variant-fusion-genes=<file>**
- These genes will always be passed to the fusion caller if found (regardless of pass/fail)

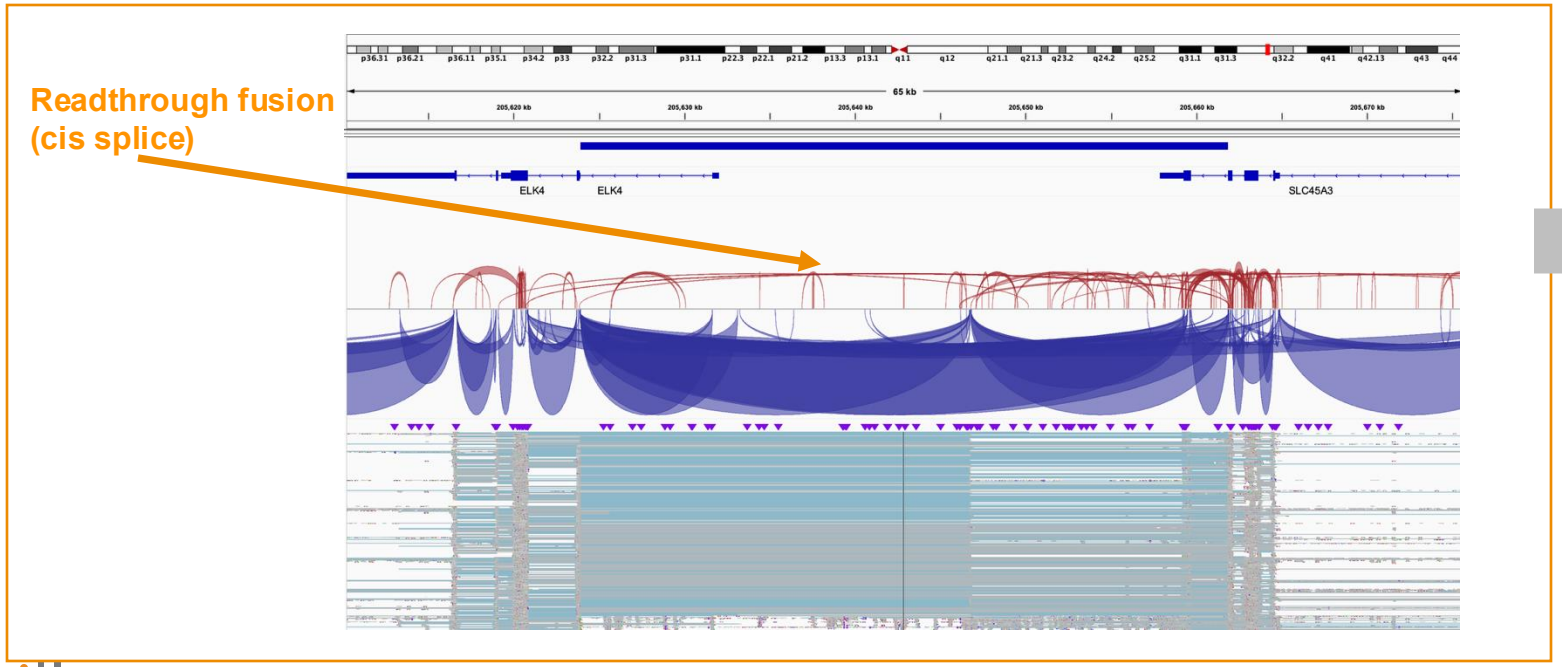
- **Read-through fusions** are fusions of adjacent genes on the same strand (normally filtered out)
- New option added to allow fusion calling on readthrough splice variant calls by setting: **rna-splice-variant-enable-readthrough=true**



# Example of read-through fusion calling (**SLC45A3-ELK4**)

- **SLC45A3-ELK4** RNA read-through fusion (cis splicing of neighboring genes) is an onco-marker for prostate cancer(+)
- Called (PASS) in **v4.4** with ***--rna-splice-variant-enable-readthrough=true*** *and* ***--rna-splice-variant-fusion-genes=<user input file>*** containing **SLC45A3,ELK4** genes

+ https://pubmed.ncbi.nlm.nih.gov/28716526/




Entry in intergenic splice variants.csv in **v4.4**

| | |
|---|---|
| gene_start | ELK4 |
| gene_end | SLC45A3 |
| chromosome | chr1 |
| start | 205623892 |
| end | 205661860 |
| filter | PASS |
| strand | - |
| motif | 2 |
| annotated | FALSE |
| unique_dedup_ref_reads | 2845 |
| unique_total_ref_reads | 4390 |
| multi_dedup_ref_reads | 0 |
| multi_total_ref_reads | 0 |
| unique_dedup_alt_reads | 94 |
| unique_total_alt_reads | 111 |
| multi_dedup_alt_reads | 0 |
| multi_total_alt_reads | 0 |
| high_qual_unique_dedup_alt_reads | 90 |
| max_mapQ_ref | 250 |
| max_mapQ_alt | 250 |
| avg_mapQ_ref | 242.65 |
| avg_mapQ_alt | 235.77 |
| max_spliced_alignment_overhang | 75 |
| normalized_overhang | 0.50 |
| score | 0.94 |
| read_through | 1 |

# MNV calling enabled by default in somatic small VC

- By default, phased variants within 2bp of one another will be merged and reported as an MNV/delins
  - Follows HGVS guidelines for variant reporting and enables better downstream annotation of variant consequence
  - 'vc-combine-phased-variants-distance' option (default=2) can be set to custom merging distance threshold [0, 15]. A value of 0 disables merging calls into MNVs.
- Newly introduced 'mnv_component' filter flag is applied to SNVs and INDELs merged into MNV/delins
  - Avoids double representation of variants that only have read support for existence in MNV
  - '--vc-combine-phased-variants-distance-max-vaf-delta` option (default=0.1) controls filtering of component calls based on difference between component call and MNV VAF
- INFO:MNVTAG field links component calls to MNVs

## Sample MNV record and component call records filtered as mnv_component:

chr1 61987 . A G . mnv_component DP=45;MQ=46.10;FractionInformativeReads=1.000;SoftClipRatio=0.02;STR;RU=A;RPA=2;MNVTAG=chr1:61987_AAG->GAC GT:SQ:AD:AF:F1R2:F2R1:DP:SB:MB:PS 0/1:61.74:0,41:1.0000:0,21:0,20:41:0,0,12,29:0,0,24,17:61987
chr1 61987 . AAG GAC . PASS DP=45;MQ=46.10;FractionInformativeReads=1.000;SoftClipRatio=0.02;STR;RU=A;RPA=2;MNVTAG=chr1:61987_AAG->GAC;GermlineStatus=Germline_DB GT:SQ:AD:AF:F1R2:F2R1:DP:SB:MB:PS 0/1:61.74:0,41:1.0000:0,21:0,20:41:0,0,12,29:0,0,24,17:61987
chr1 61989 . G C . mnv_component DP=48;MQ=47.02;FractionInformativeReads=1.000;SoftClipRatio=0.02;MNVTAG=chr1:61987_AAG->GAC GT:SQ:AD:AF:F1R2:F2R1:DP:SB:MB:PS 0/1:61.85:0,43:1.0000:0,23:0,20:43:0,0,14,29:0,0,26,17:61987

[1]Wang, Q., Pierce-Hoffman, E., Cummings, B.B. *et al.* Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nat Commun* **11**, 2539 (2020). https://doi.org/10.1038/s41467-019-12438-5

# Illumina Connected Annotations - Data Manager

- New utility to give the customer complete control over annotation data.
- Available functionalities:
  - View available versions.
  - Create data configuration template.
  - Download data.
  - Validate data versions.

```
--------------------------------------------------------------------------------
DataManager                                              (c) 2024 Illumina, Inc.
                                                                3.25.1-0-g68d01f47
--------------------------------------------------------------------------------

Illumina Connected Annotation data manager

USAGE: dotnet DataManager.dll <command> [options]

COMMAND: list              display all data available for Illumina Connected Annotations
         make-config       create a config file sample for downloader populated with the latest version
         download          download file from the cloud
         version-validate  validate downloaded data version with the provided config
```

https://illumina.github.io/IlluminaConnectedAnnotationsDocumentation/utilities/data-manager

# Illumina Connected Annotations - ISCN like nomenclature

- Provides ISCN-like simple nomenclature to describe karyotype of the input in sample's and variant's level in both VCF and JSON outputs.

| Chromosome | Start Position | End Position | Variant Type | ISCN Notation |
|---|---|---|---|---|
| 8 | 19200001 | 135400001 | deletion | del(8)(p21.3q24.23) |
| 8 | 19200001 | 135400001 | duplication | dup(8)(p21.3q24.23) |
| 8 | 127300001 | 131500000 | copy number gain | dup(8)(q24.21q24.22) |
| 8 | 135400001 | 138900001 | copy number loss | del(8)(q24.23q24.3) |

Sample level notation from Ploidy VCF: `"48,XY,-10,-10,-11,+12,+13,+13,+X,+Y"`

https://illumina.github.io/IlluminaConnectedAnnotationsDocumentation/core-functionality/simple-nomenclature-notation

# Flexible gVCF metric import for iGG

**Pick the metrics you need from the input gVCFs**

- Allows import of metrics beyond the default set (which is unchanged). Almost all gVCF metrics are supported.
- Unwanted metrics can be omitted, saving on intermediate storage space.
- Output customization options, introduced in v4.3, extended to enable output of any imported metric.

To set the fields to be ingested:

--gg-format-to-import=…
--gg-info-to-import=…

INFO fields available for import:

MQ, MQRankSum, ReadPosRankSum, FractionInformativeReads, LOD, R2_5P_bias, MOSAIC

FORMAT fields available for import:

GT, AD, GQ, PL, AF, DP, F1R2, F2R1, GP, MB, PRI, SB, SQ
(N.B. GT is unconditionally imported)

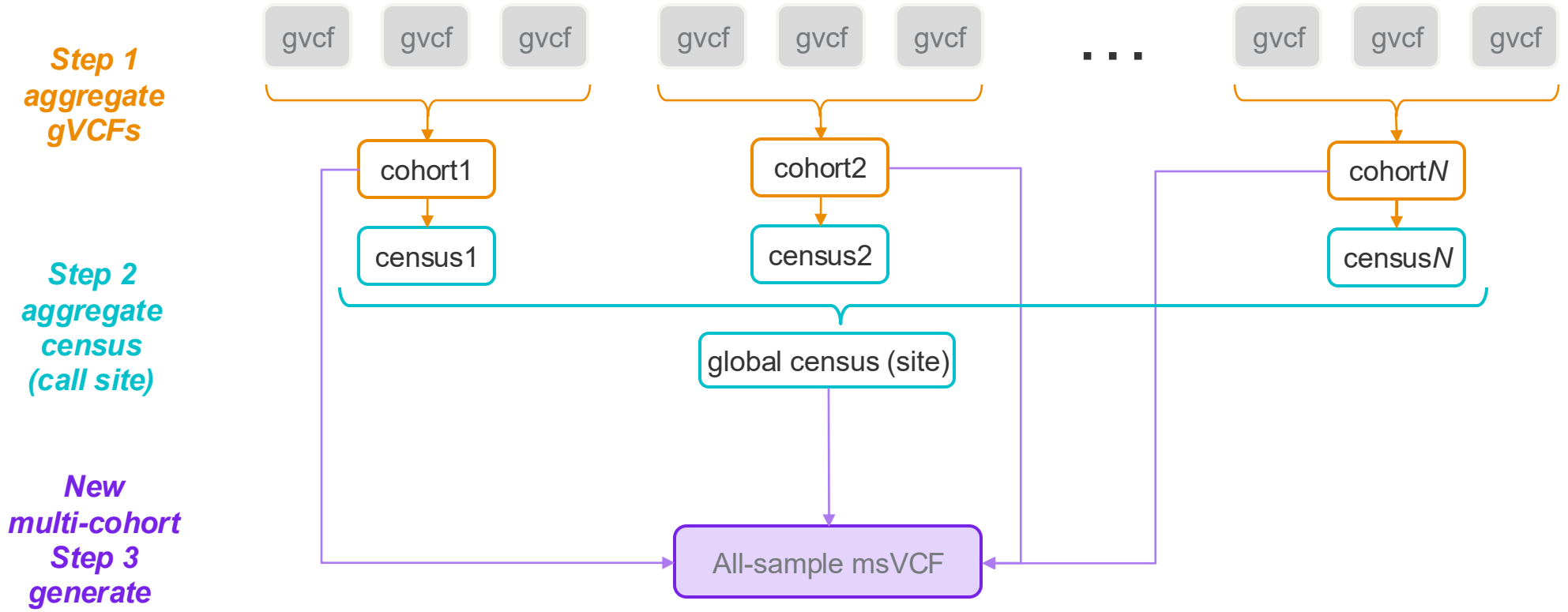
**Availability**

ICA

On-premises server

# Streamlined workflow for multi-batch msVCF file output

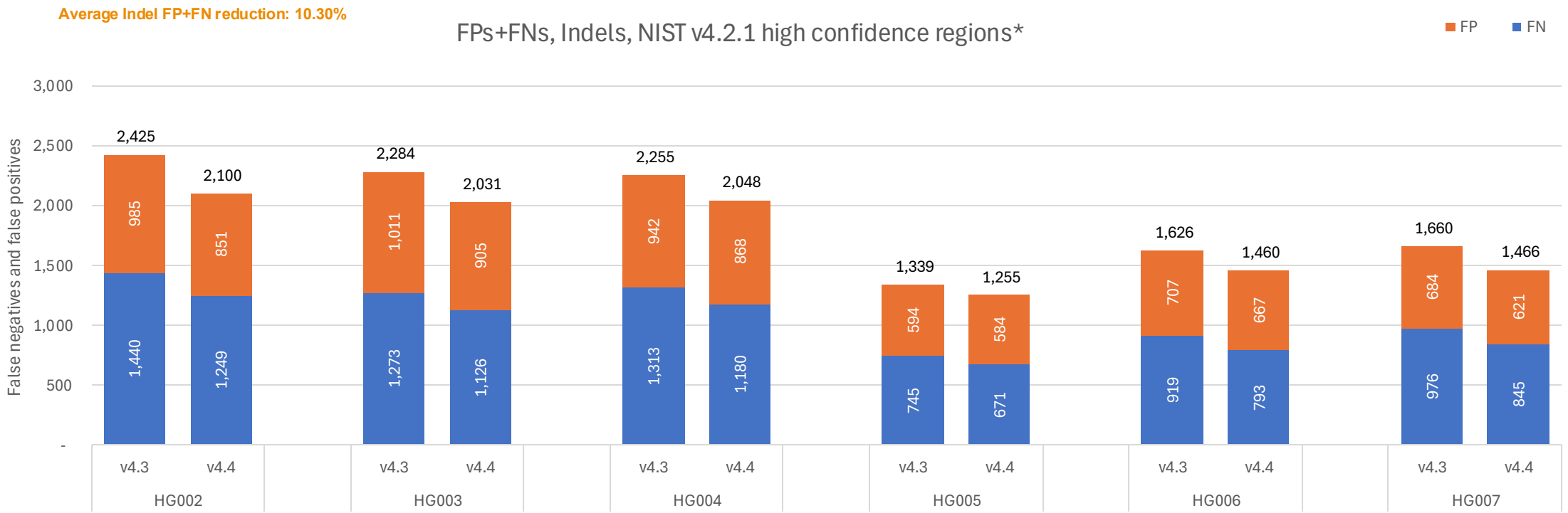
Go directly from cohort files to all-sample msVCF




New step 3 runtime up to ~3 times faster than previous steps 3 + 4, with fewer jobs required.

# Germline small variant valler – Accuracy v4.2.1 indels

## DRAGEN 4.4 variant caller improvements results in 10.30% reduction in Indels FP+FN




*Samples NSX 10B v1.1 35x median autosomal coverage, HG001 removed from analysis since included in the DRAGEN Pangenome Reference

# DRAGEN ORA compression

## Store and easily retrieve information about your sample

- From BCLconvert to FASTQ.ORA, additional metadata are automatically stored to FASTQ.ORA files
- The command *--ora-get-metadata. true*, generates a Json file with all metadata listed
- List of additional metadata:
  - Sequencing platform and flowcell name
  - Run ID
  - Flowcell ID
  - Instrument ID
  - Sample ID
  - First read name
  - index 1
  - index 2

**Example DRAGEN CMD line to generate the metadata json file**

```
dragen
    --enable-map-align false
    --ora-input <fastq.ora>
    --enable-ora=true
    --ora-get-metadata true
```

**Availability**

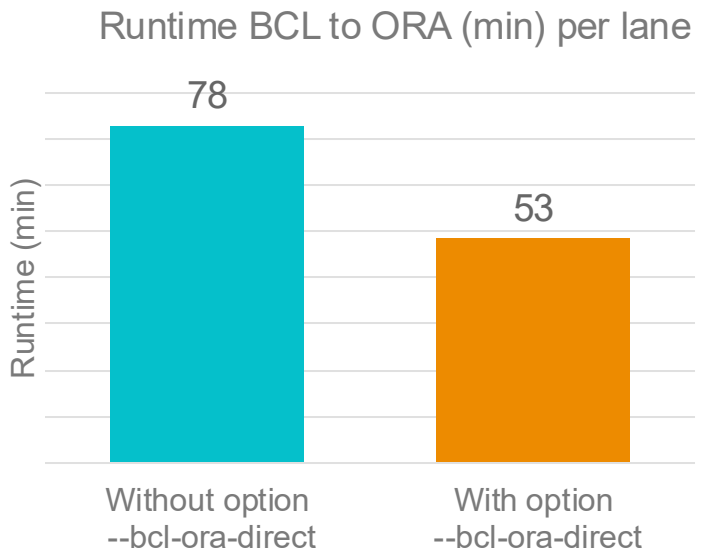
(Not available on instrument)

**On-Premise Server**

# DRAGEN ORA compression - BCL to FASTQ.ORA up to 30% faster

**Runtime BCLconvert to FASTQ.ORA with DRAGEN 4.4**

dataset: 25B flowcell data

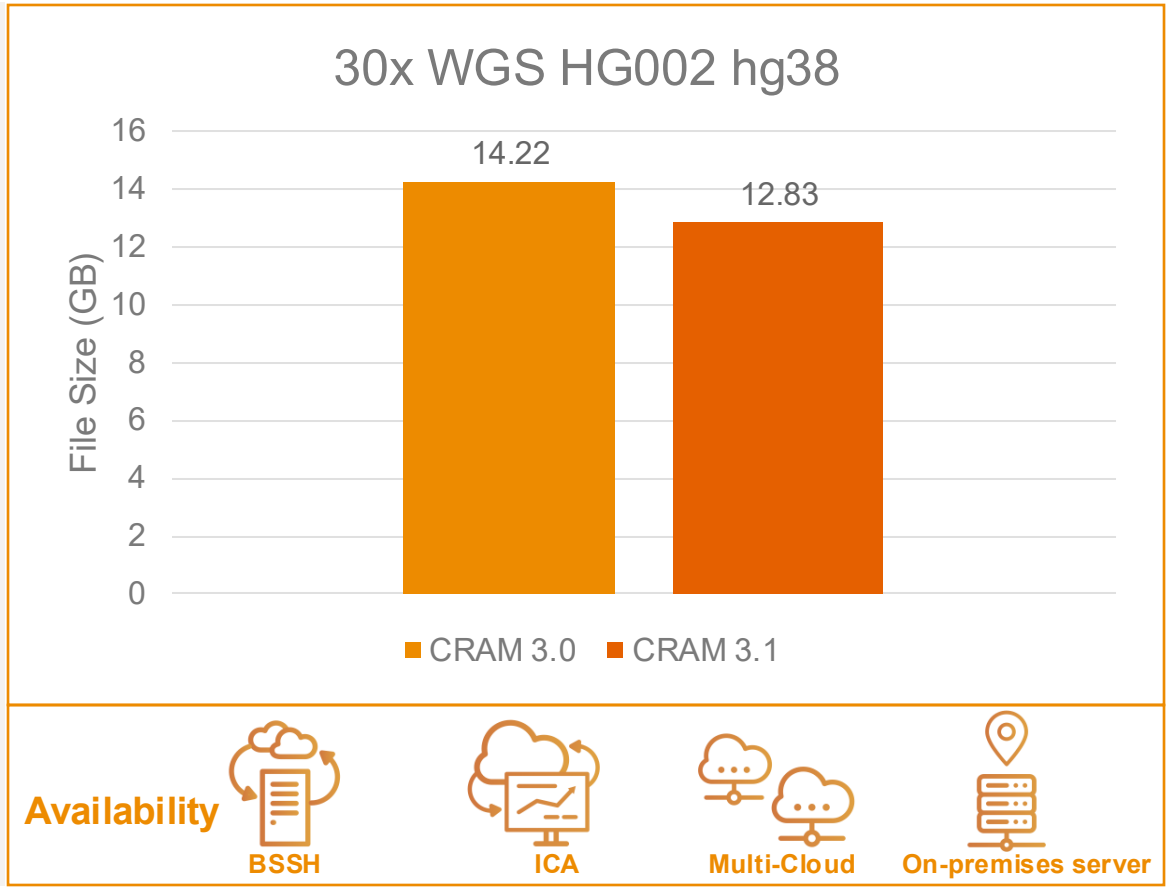



Test conducted on DRAGEN server v4 with all other settings being default and identical across the two benchmarks

# CRAM 3.1 support

**DRAGEN now supports CRAM 3.1**

- Input or generate CRAM files with format 3.1
- CRAM 3.1 files are 7-15% smaller than CRAM 3.0
- No impact on run time
- New option --cram-version=3.1 (default is 3.0)



# New license server domain – license.dragen.illumina.com

## On-premises servers

- HTTP @ lus.edicogenome.com is being replaced with HTTPS @ license.dragen.illumina.com
- Only DRAGEN 4.4+ supports the new domain, previous versions of DRAGEN will need to continue to use HTTP @ lus.edicogenome.com.

## BYOL Cloud

- HTTPS @ license.edicogenome.com will be replaced with HTTPS @ license.dragen.illumina.com
- All versions of DRAGEN support the new domain, but the default domain will differ.
  - 4.4 and above: license.dragen.illumina.com
  - 4.3 and earlier: license.edicogenome.com
  - You can explicitly define the domain to use with the “credentials-3” configuration option as specified in the user guide.

# Fractional downsampler feature for RNA Pipelines

## Simulates different amount of sequencing for high coverage samples

- Fractional downsampling can be utilized with RNA pipelines
- Subsampling based on user-defined percentage of reads
- Applied to raw reads with no modification (no trimming, no filtering, pre-dedupped)
- Reduce runtime and cost of analysis using high-depth samples
- Any input supported by DRAGEN can be used

To enable the fractional downsampler:

*--enable-fractional-down-sampler=true*

To set percentages of number of reads to keep:

*--down-sampler-normal-subsample=<float>*

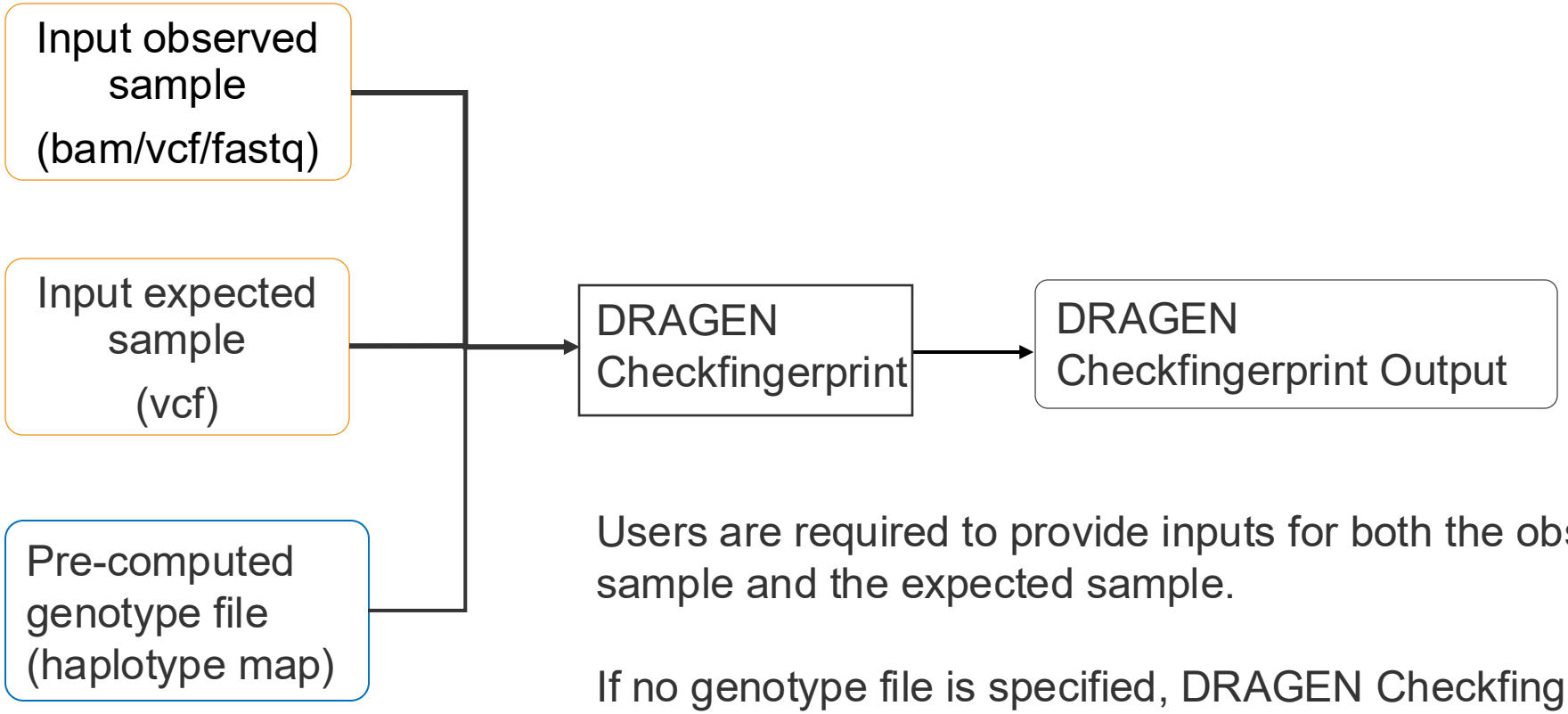
*--down-sampler-tumor-subsample=<float>*

* <float> is the approximate percentage of reads to keep (e.g. 0.05 = 5%)

**Availability**

# DRAGEN Checkfingerprint

Determine if the input sequencing data originates from the same sample




Users are required to provide inputs for both the observed sample and the expected sample.

If no genotype file is specified, DRAGEN Checkfingerprint will automatically utilize the default genotype file.

# DRAGEN Checkfingerprint

- Checkfingerprint compute identity between input samples against a set of pre-computed genotypes.
- The primary output is LOD score, which indicates the relative likelihood that sequencing data originates from the same or different sample.
- A positive LOD value indicates the data originates from the same individual, while a negative values suggests otherwise.
- Checkfingerprint has two mode: read mode and vcf mode
- Vcf mode is recommended for general use case. Read mode is significantly slower.

To enable the Checkfingerprint:

```
--enable-checkfingerprint true
--checkfingerprint-expected-vcf
[path_to_expected_sample_vcf]
```

To enable VCF comparison mode using FASTQ or BAM input, include the additional parameters:

```
--checkfingerprint-enable-vcf-
comparison true
--enable-variant-caller true
```

To enable VCF comparison mode using vcf input:

```
--enable-checkfingerprint true --
checkfingerprint-expected-vcf
[input_expected_vcf] --checkfingerprint-
observed-vcf [input_observed_vcf]
```

# DRAGEN reports

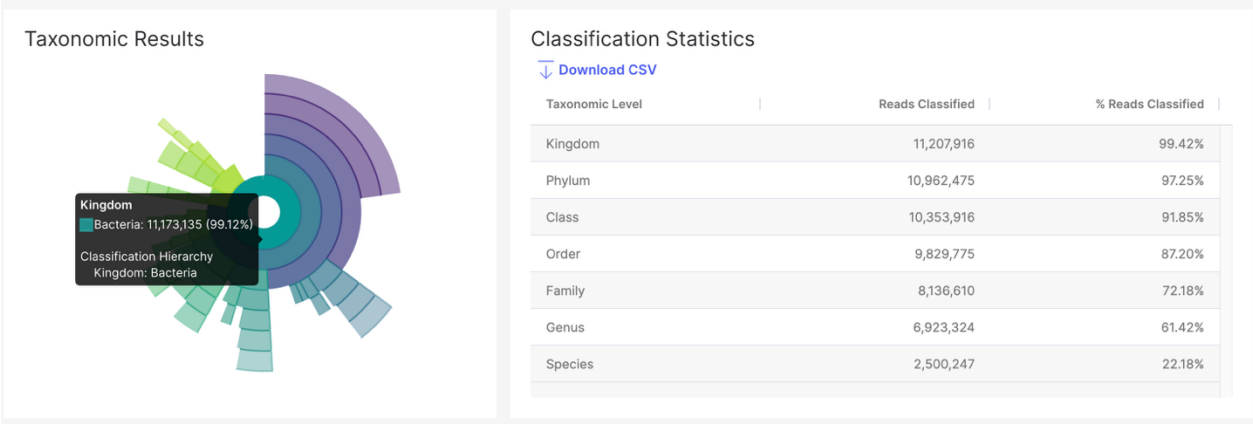
## Key benefit of updates

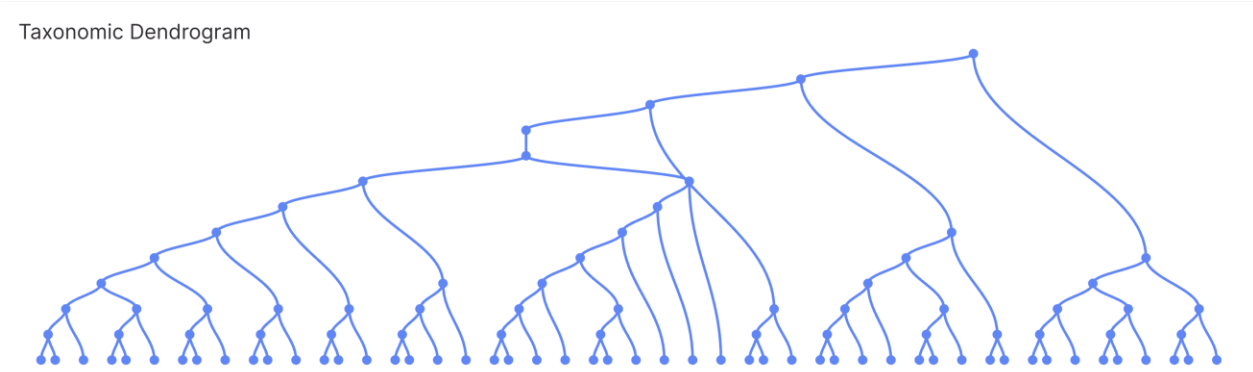
**16S Report**

- Replacement for one of the oldest and most widely used legacy BaseSpace apps
- Added multi-column reports for more complicated designs and smaller plots and tables
- Adds support for multiple new plot types, such as Sunburst plots and Dendrograms

**Addition Features**

- Added a PIPSeq / Single-Cell RNA Report
- Added support for FASTA file downloads
- Added UI testing system to ensure good customer experience with the reports
- New faster CI/CD system for faster updates



# DRAGEN BSSH and ICA Apps Updates

illumina

# BCL Convert

## Key benefit of updates

**Non-DRAGEN features or bug fixes**

- Allow dots and space characters in RunName and ProjectName (consistent with Run Planning)

# DRAGEN Germline

## Key benefit of updates

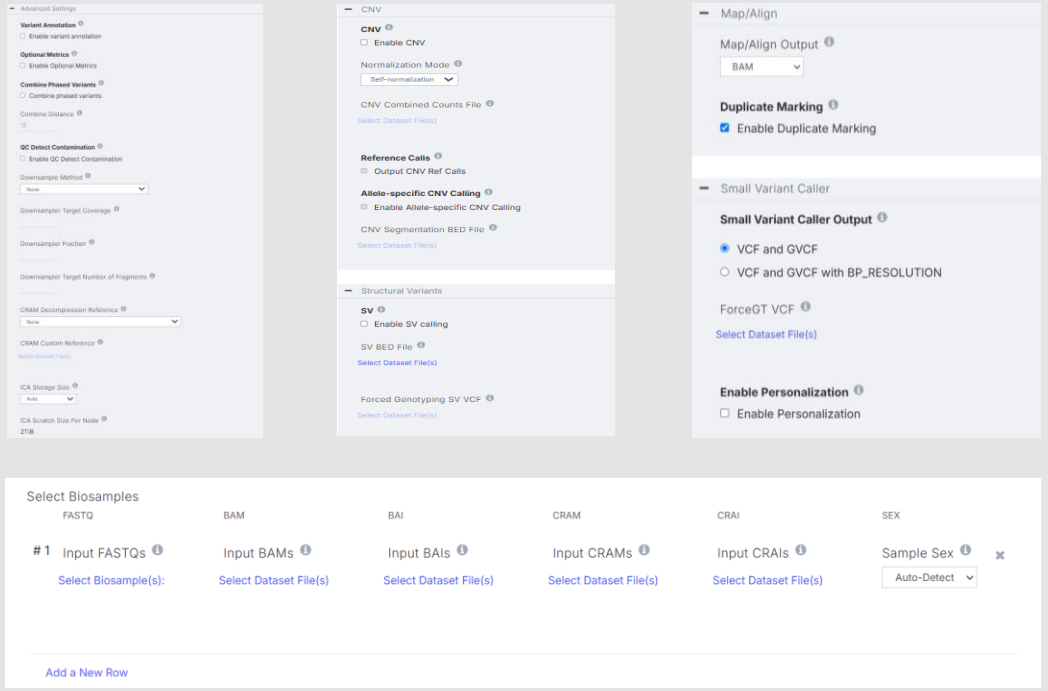
### New DRAGEN features

- Updated to DRAGEN 4.4.4
- Added support for BAIs and CRAIs
- Support v11 reference hash tables (4.4 feature)
- Added additional hs37d5 reference with chr prefix
- Added support for CRAM v3.0 and CRAM v3.1 input
- Cleaned up underutilized inputs
- Removed MRJD as a separate caller
- Removed support for hybrid mode

### Non-DRAGEN features

- Updated ICAv2 backend to use directly mounted input files instead of URLs

### User interface changes



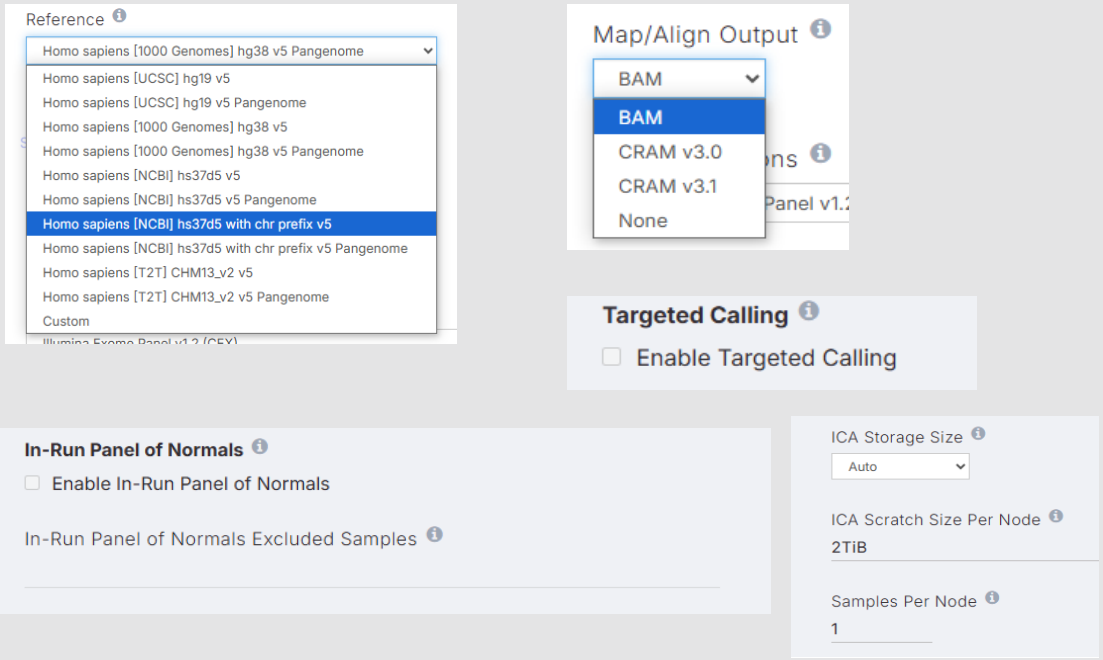
# DRAGEN Enrichment

## Key benefit of updates

### New features

- Upgraded to DRAGEN 4.4.4
- Support v11 reference hash tables (4.4 feature)
- Added additional hs37d5 reference with chr prefix
- Removed under-utilized options
- Added support for CRAM v3.0 and CRAM v3.1 output
- Added support for in-run PON file generation.
- Added support for targeted-caller and building targeted caller baselines files.
- **ICAv2 backend** for the BaseSpace app
- Support setting ICA storage size and scratch size per node

### User interface changes



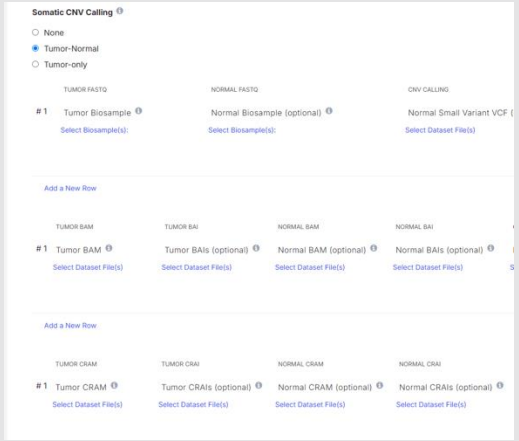
# DRAGEN Somatic

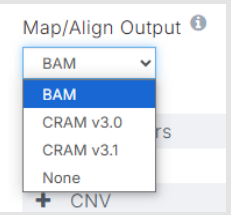
## Key benefit of updates

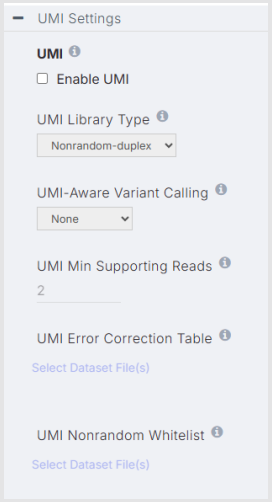
### New features

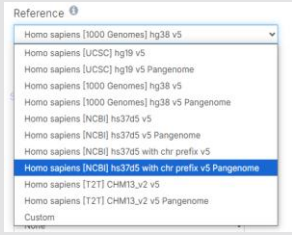
- Upgraded to DRAGEN 4.4.4
- Support v11 reference hash tables (4.4 feature)
- Added additional hs37d5 reference with chr prefix
- Added support for CRAM v3.0 and CRAM v3.1 output
- Enabled UMI support for Tumor-Normal analysis
- Added support for BAM/CRAM input in T/N mode
- Added support for HLA calling in T/N mode from BAM input with realignment
- **ICAv2 backend** for the BaseSpace app
- Support setting ICA storage size and scratch size per node

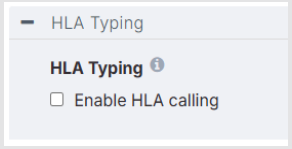
### User interface changes



# DRAGEN RNA

## Key benefit of updates

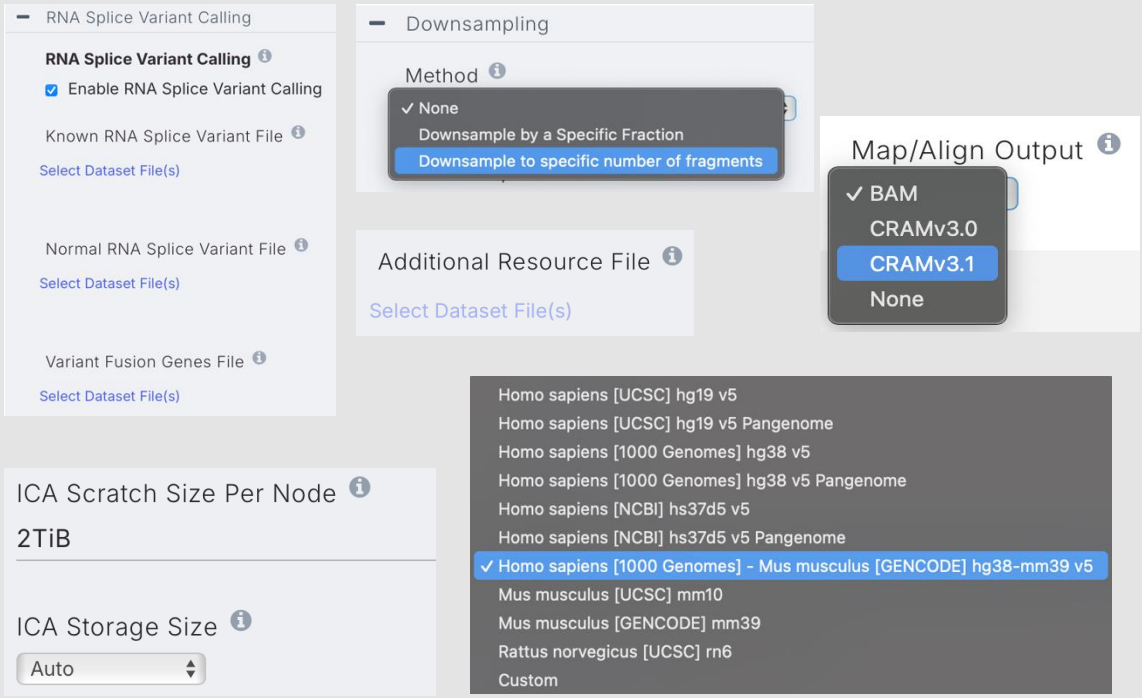
### New DRAGEN features

- Upgraded to 4.4.4
- Support splice variant calling
- Support fractional downsampler
- Support v11 reference hash tables (4.4 feature)
- Support built-in mm39 and hg38-mm39 references
- Set annotation genome automatically
- Added support for CRAM v3.0 and CRAM v3.1 input

### Non-DRAGEN features

- **ICAv2 backend** for the BaseSpace app
- Support additional files
- Support setting ICA storage size and scratch size per node

### User interface changes



# DRAGEN Amplicon

## Key benefit of updates

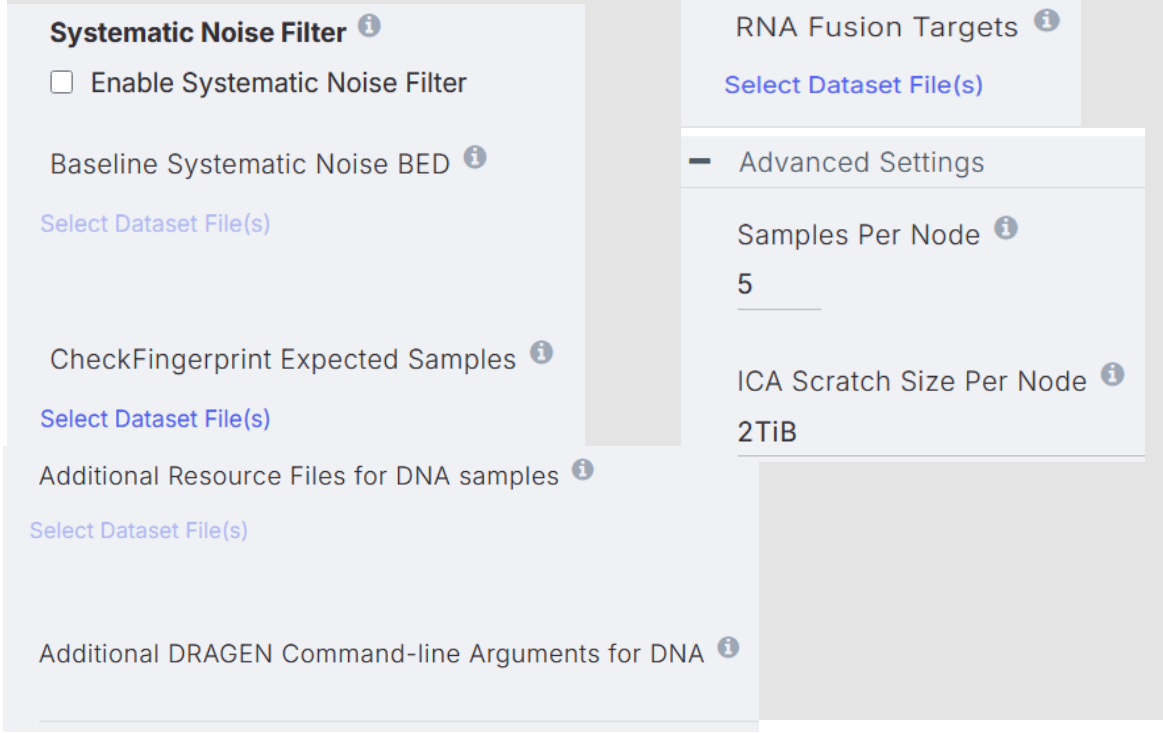
**New DRAGEN features**

- CheckFingerprint always run in VCF comparison mode.
- Support v11 reference hash tables (4.4 feature).
- Support CRAM v3.1.

**Non-DRAGEN features**

- **ICAv2 backend** for the BaseSpace app
- Auto-detect annotation genome.
- Support Systematic Noise Filter for somatic VC.
- Support RNA Fusion Targets.
- Support for additional resource files.
- Support batching.
- Support overriding scratch size for DRAGEN.

**User interface changes**



# DRAGEN Microbial Amplicon (DMA)

Generates consensus sequences from Illumina Microbial Amplicon Prep (iMAP), COVIDSeq, or other amplicon-based libraries from microbial samples

## Key benefit of updates

### v1.1.0 release notes

- Updated DRAGEN to 4.4
- Updated DRAGEN Map/Align and Variant Calling steps to improve concordance with DRAGEN COVID Lineage
- Expanded Influenza reference database to include more sequences and subtypes and improved reference selection logic
    - Added 48 rare A subtypes (e.g. H7N9) in addition to H1N1, H3N2, H5N1
- Enabled downloading analysis-wide consensus sequence FASTA from the report
- Reduced run time
- Minor bug fixes in Pangolin, input config, and report

### DMA replaces BSSH apps near end of life

- DRAGEN Targeted Microbial: May 31, 2025
- DRAGEN COVID Lineage: Dec 1, 2025

**Updated user guide at help.idm.illumina.com**

# DRAGEN joint pedigree and E2E joint pedigree

## Key benefit of updates

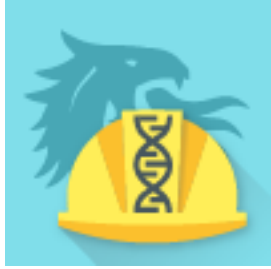
**Overview**

- 2 apps, one starting from GVCFs and another starting from FASTQs
- Processes one family (pedigree) at a time

**Features**

- Update to DRAGEN 4.4.4
- Add REViewer—a viewer for repeat expansions

# DRAGEN Reference Builder



## Key benefit of updates

**New DRAGEN features**

- Generate references in the v11 hash table format

# DRAGEN Population Haplotyping

ICA

## Key benefit of updates

### New features

- Upgraded to DRAGEN 4.4.4
- Added PassOnly filter option during phase common step

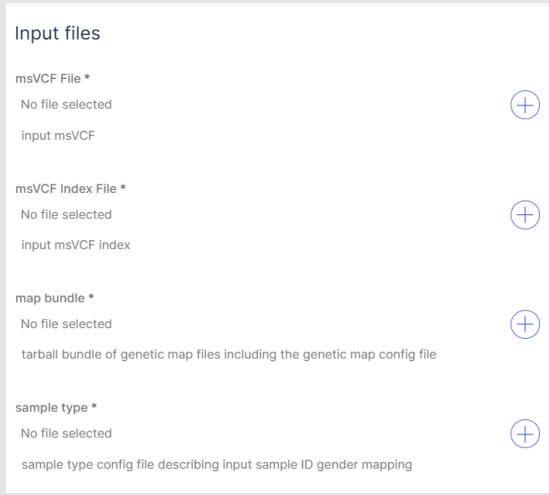
### Non-DRAGEN features or bug fixes

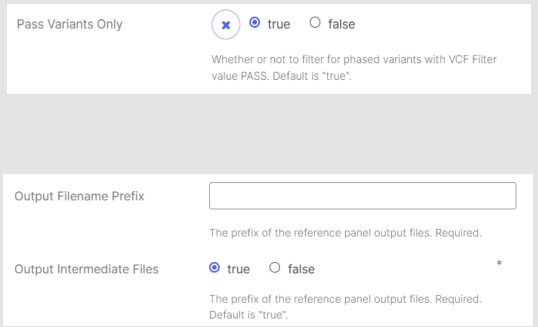
- Supports contig-specific joint-VCF input
- Reduced input requirements (no reference FASTA, no credentials file)
- Removed redundant normalization
- Output file basename control
- Intermediate file output control
- Properly names sub-contig output files (eg. chrX_par1)

**Availability**

ICA

### User interface changes



# DRAGEN Imputation Reference Panel Builder

ICA

## Key benefit of updates

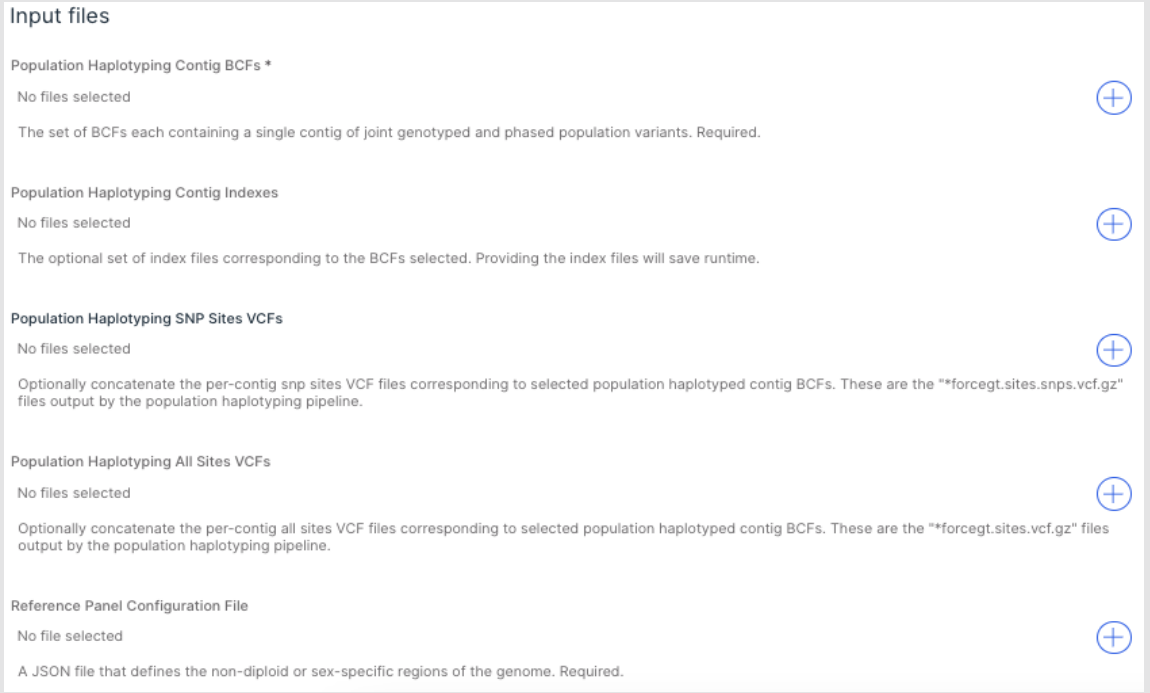
### New DRAGEN features

- Packs the per-contig output from the DRAGEN Population Haplotyping pipeline into a **custom imputation reference panel** TAR.GZ that can be used with DRAGEN Imputation pipeline or download.
- Supports packing for WGS, or selected contigs.
- Per-contig msBCF file are included in the TAR.GZ.
- Optionally create a snps/all site vcf file if the per-contig snp/all-variant site vcf files are provided.
- A config JSON file can be supplied to describe ploidy.

**Availability**

ICA

### User interface



# DRAGEN Baseline Builder

## Key benefit of updates

**New DRAGEN features**

- Upgrade to 4.4.4
- Support v11 reference hash tables (4.4 feature)
- Split support for CNV WGS and CNV Enrichment
- Always creates combined counts file in CNV modes

**Non-DRAGEN features**

- Support setting ICA storage size

**User interface changes**