DRAGEN v4.4.4 Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Secondary Analysis Software v4.4.4.

Changes are relative to DRAGEN™ v4.3.16. If you are upgrading from a version prior to DRAGEN™ v4.3, please review the release notes for a list of features and bug fixes introduced in subsequent versions.

DRAGEN™ Installers, Resource Files, and Release Notes are available here:
https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html

DRAGEN™ User Guide is now available here:
https://help.dragen.illumina.com

The software package includes downloadable installers for Phase 3 and Phase 4 on-site servers:
- DRAGEN™ SW for x86 Oracle 8 - dragen-4.4.4-11.multi.el8.x86_64.run

The following configurations containing DRAGEN™ 4.4.4 are also available on request:
- AlmaLinux 8 Amazon Machine Images (AMIs) for f1 instances, available in 12 regions
- AlmaLinux 8 Microsoft Azure Image (VM) available in West US 2 for BYOL
- el8 compatible RPM packages for use with Amazon Web Services (AWS) f1 instances, for customer generated AMIs or customer generated docker images
- DRAGEN™ Kernel drivers for el8, for use with customer generated AMIs or QuickStart

DRAGEN™ v4.4.4 is also made available on:
- Illumina BaseSpace and ICA platforms
- AWS and Azure Marketplaces
  - On AWS see "DRAGEN Complete Suite"
  - On Azure see "DRAGEN Public VM Image - PAYG"

Deprecated platforms:
- Support for CentOS 7 ended on June 30, 2024. DRAGEN™ v4.3 is the final release with CentOS 7 installers.
- Support for DRAGEN™ Server v1 FPGA cards have been deprecated since DRAGEN™ v3.10
- Support for Ubuntu has been deprecated since DRAGEN™ v3.9
- Support for CentOS 6 has been deprecated since DRAGEN™ v3.8

Contents

Overview

Below is a summary of the changes included in v4.4.4. DRAGEN™ v4.4 offers continuous innovation powering a comprehensive genome with accuracy, speed and ease. For full extensive details on each feature of pipeline, please consult the latest Illumina DRAGEN™ Software User Guide available at https://help.dragen.illumina.com

Highlights

- **Easy-to-use oncology apps**
  Pre-configured push-button analysis for Heme WGS, Solid WGS T/N and a new pipeline for MRD
- **Enhanced Multiomics**
  Support for new assays –Single Cell 3'RNA Prep, Protein Prep, 5-Base (Methylation)
- **Streamlined WGS Germline**
  All targeted callers in standard WGS DRAGEN workflow
- **Redefining SV Accuracy**
  30% boost to SV calling driven by multigenome mapper with pangenome reference
- **Personalized Pangenome**
  Official release with significant accuracy gains and improved run time to WGS germline analysis
- **AWS F2 Instances**
  DRAGEN v4.4 supports AWS F2 instances for faster cloud analysis times

Please review the section on Known Issues and limitations of the release.

Updated Resource Files

DRAGEN™ v4.4 requires updates to key resource files to function correctly and achieve the optimum performance. Additional resource files are made available for v4.4. All resource files are available for download at the Illumina DRAGEN™ Product Files support site here: https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/product_files.html

Resource	Description	File name(s)
Hash Tables v11	Pre-built v11 pangenome and linear hash tables for hg38, hg19, hs37d5, chm13_v2.	Pangenome: hg38-alt_masked.cnv.graph.hla.methyl_cg.rna-11-r5.0-1.tar.gz hg19-alt_masked.cnv.graph.hla.methyl_cg.rna-11-r5.0-1.tar.gz hs37d5-cnv.graph.hla.methyl_cg.rna-11-r5.0-1.tar.gz hs37d5_chr-cnv.graph.hla.methyl_cg.rna-11-r5.0-1.tar.gz chm13_v2-cnv.graph.hla.methyl_cg.rna-11-r5.0-1.tar.gz Linear: hg38-alt_masked.cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz hg19-alt_masked.cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz hs37d5-cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz hs37d5_chr-cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz chm13_v2-cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz hg38-mm39-alt_masked.cnv.hla.methyl_cg.methylated_combined.rna-11-r5.0-1.tar.gz
Pangenome Reference Builder Collection v5	HT mask BED, Graph BED, Graph exclusion BED, Graph msVCF and FASTA files for building hg38, hg19, hs37d5, chm13_v2 references.	hg38-pangenome-reference-collection-v5-1.tar.gz hg19-pangenome-reference-collection-v5-1.tar.gz hs37d5-pangenome-reference-collection-v5-1.tar.gz chm13_v2-pangenome-reference-collection-v5-1.tar.gz
SNV Systematic Noise Baseline collection v2.0.0	A collection of Somatic noise baseline BED files for hg19, hs37d5, hg38 and for WGS and WES respectively. New files for Heme and FFPE WGS for hg38.	systematic-noise-baseline-collection-2.0.0.tar The tar archive contains the following files: IDPF_WGS_hg38_v.2.0.0_systematic_noise.snv.bed.gz FFPE_WGS_hg38_v2.0.0_systematic_noise.snv.bed.gz WGS_hg38_v2.0.0_systematic_noise.snv.bed.gz WGS_hg19_v2.0.0_systematic_noise.snv.bed.gz WGS_hs37d5_v2.0.0_systematic_noise.snv.bed.gz WES_hg38_v2.0.0_systematic_noise.snv.bed.gz WES_hg19_v2.0.0_systematic_noise.snv.bed.gz WES_hs37d5_v2.0.0_systematic_noise.snv.bed.gz
SV Systematic Noise Baseline collection v3.1.0	A collection of Somatic noise baseline BEDPE files for WGS hg19, hs37d5, hg38.	sv-systematic-noise-baseline-collection-3.1.0-1.tar.gz The tar archive contains the following files: WGS_FF_Heme_hg19_v3.1.0_systematic_noise.sv.bedpe.gz WGS_FF_Heme_hg38_v3.1.0_systematic_noise.sv.bedpe.gz WGS_FF_Heme_hs37d5_chr_v3.1.0_systematic_noise.sv.bedpe.gz WGS_hg19_v3.1.0_systematic_noise.sv.bedpe.gz WGS_hg38_v3.1.0_systematic_noise.sv.bedpe.gz WGS_hs37d5_v3.1.0_systematic_noise.sv.bedpe.gz
Targeted Caller Systematic Noise Baseline collection v1.0.0	A collection of systematic noise baseline json files for hg38, hg19 and hs37d5 for use with WES analysis.	tc-systematic-noise-baseline-collection-v1.0.0-1.tar.gz The tar archive contains the following files: hg19_v1.0.0_systematic_noise.targeted.json.gz hg38_v1.0.0_systematic_noise.targeted.json.gz hs37d5_v1.0.0_systematic_noise.targeted.json.gz

CNV Population SNP VCF v1.0.0	Population SNP VCF for Somatic TO CNV for hg38, hg19, hs37d5 and chm13	Files from the GATK resource bundle uploaded for convenience: hg38_1000G_phase1.snps.high_confidence.vcf.gz hg19_1000G_phase1.snps.high_confidence.vcf.gz hs37d5_1000G_phase1.snps.high_confidence.vcf.gz chm13_1000G_phase1.snps.high_confidence.vcf.gz
CNV panel of normals (PON) v4.4	Collection of pre-constructed CNV PON files for WES	CNV_PON-Twist_ILMN_Exome_FFPE_2_5_Panel-DRAGEN_v4.4_v1-1.tar.gz CNV_PON-Twist_ILMN_Exome_Mito_2_5_Panel-DRAGEN_v4.4_v1-1.tar.gz CNV_PON-Twist_ILMN_Exome_2_5_Panel-DRAGEN_v4.4_v1-1.tar.gz
SNV Exclusion BED collection v1.0.0	Somatic SNV ALU region exclusion BED files for hg38, hg19, hs37d5	bed-file-collection-1.0.0.tar.gz The tar archive contains the following files: v1.0.0_hg38_Alu_regions.bed.gz v1.0.0_hg19_Alu_regions.bed.gz v1.0.0_hs37d5_Alu_regions.bed.gz
Microsatellite Files v1.1.0	Microsatellite files and panels of normals for hg19, hs37d5, hg38 and for WGS and WES respectively	microsatellite-files-1.1.0=1.tar.gz The tar archive contains the following files: WGS_v1.1.0_hg38_microsatellites.list WGS_v1.1.0_hg19_microsatellites.list WGS_v1.1.0_hs37d5_microsatellites.list WGS_FFPE_NovaSeq_6K_hg19_MSI_baselines_v1.1.0/ WGS_FFPE_NovaSeq_6K_hg38_MSI_baselines_v1.1.0/ WGS_FFPE_NovaSeq_6K_hs37d5_MSI_baselines_v1.1.0/ WES_v1.0.0_hg38_microsatellites.list WES_v1.0.0_hg19_microsatellites.list WES_v1.0.0_hs37d5_microsatellites.list WES_FFPE_hg19_MSI_baselines_v1.1.0/ WES_FFPE_hg38_MSI_baselines_v1.1.0/ WES_FFPE_hs37d5_MSI_baselines_v1.1.0/
Imputation Reference Panel v2.1 and Genetic Map v2.0	Genetic map and reference panel for hg38	genetic_maps-hg38-2.0.tar irp-hg38-2.1.2.0.tar
ORA compression references	Compression references for human, methylated and non-human	Human: oradata_homo_sapiens_V1.tar.gz (optimized for DRAGEN v3.10+) lenadata.tar.gz Human bisulfite: oradata_homo_sapiens_bisulfite_V1.tar.gz Non-human: oradata_arabidopsis_thaliana_V1.tar.gz oradata_bos_taurus_V1.tar.gz oradata_caenorhabditis_elegans_V1.tar.gz oradata_carina_moschata_V1.tar.gz oradata_danio_rerio_V1.tar.gz oradata_gallus_gallus_V1.tar.gz oradata_glycine_max_V1.tar.gz oradata_homo_sapiens_V1.tar.gz oradata_homo_sapiens_bisulfite_V1.tar.gz oradata_mus_musculus_V1.tar.gz oradata_oryza_sativa_V1.tar.gz oradata_rattus_norvegicus_V1.tar.gz oradata_sus_scrofa_V1.tar.gz oradata_triticum_aestivum_V1.tar.gz oradata_zea_mays_V1.tar.gz Combined all non-human: oradata_all_species_V2.tar.gz
RNA gene annotation files v1.0	GTF gene annotations from GENCODEGenes	gene-annotation-files-collection-v1.0-1.tar.gz The tar archive contains the following files: hg38-mm39/gencode.hg38_v44.mm39_vM30.annotation.gtf.gz hg19/gencode.v19.annotation.gtf hs37d5/gencode_nochr.v19.annotation.gtf hs37d5_chr/gencode_hs37d5_chr.v19.annotation.gtf

		hg38/gencode.v44.annotation.gtf.gz

Multi-Version Installer for on-premises servers

Starting with DRAGEN™ v4.3 and later, multiple compatible versions of the software can be installed at a time on the DRAGEN on-premises server. Executing the `.run` file will add the new version to the system.

After installation, the application files are available at `/opt/dragen/{version}` and FPGA files are located at `/opt/bitstream/{bitstream version}`.
The multi-version installer will NOT add `/opt/dragen/{version}` to the Linux $PATH, since multiple versions can be present at a given time. User should manage the desired paths to the specific version they want to run.

Notes on multi-version installation:

- Installers originally released for DRAGEN™ v4.2 and earlier are single version packages.
- Single version packages and multi-version packages cannot be mixed.
  - Installation of a prior single version package will remove all the multi-version packages.
  - Installation of a multi-version package will remove any installed single version package.
- After installing a multi-version package, see a list of installed versions at any time by running `/usr/bin/dragen_versions`
- To remove any multi-version package, call `yum remove` on its Path.
- A multi-version installer can be identified by the presence of `multi` in the file name, e.g. `dragen-4.3.6-11.multi.el8.x86_64.run`
- *Root privileges are required for the installation.*
- *Multi-version installers are only applicable to on-premises DRAGEN servers, not cloud.*

Example:

```
$ dragen_versions
The output format of this command may change. Use --json for machine readable output.

Dragen Version           Size (MB)  Install Date         Path
4.3.2                    1378.03    2024-03-10 18:26:17  /opt/dragen/4.3.2
4.4.3                    1381.41    2024-03-18 20:56:39  /opt/dragen/4.4.3
4.3.5                    1379.25    2024-03-11 15:20:24  /opt/dragen/4.3.5

Bitstream Version        Size (MB)  Install Date         Path
07.031.732 (0x18101306)  598.95     2024-03-10 18:26:03  /opt/bitstream/07.031.732
07.031.745 (0x18101306)  598.95     2024-03-18 20:56:18  /opt/bitstream/07.031.745

To remove a dragen version, call `yum remove` on its Path.
```

- Location of dragen and resource files

DRAGEN Version	on-premises server	cloud instance
v4.3 and later	/opt/dragen/{version}	/opt/edico/
v4.2 and earlier	/opt/edico/	/opt/edico/

- Availability of multi version installers for older releases

Multi-version installer capability has been backported to multiple older versions. Please reach out to Customer Support to inquire about access to an installer for your version.

Major Features and Updates

DRAGEN v4.4 offers new features and accuracy improvements for DRAGEN callers, new analysis pipelines and support for assays, and new tools.

Reference Genome

- Hash Tables v11.
  - The hash table interface is updated to format version 11 (HTv11). Hash tables must be updated to use v4.4. Existing hash tables built with v4.3 or older are not supported.
  - Added pre-built hash tables:
    - *hs37d5_chr*: hs37d5 with "chr" prefixed in the contig names, and the mito contig renamed to "chrM".
    - *hg38_mm39*: Unified human and mouse reference. Used for scRNA assay tests.
  - Pre-built hash tables for all supported human references are available at the Illumina DRAGEN™ Product Files support site and are recommended for use.
  - Multigenome references can also be built in two ways:
    - On server using the hash table builder and the Pangenome Reference Builder Collection v5 input files, or
    - On BaseSpace using the DRAGEN Pangenome Reference Builder app
  - See the User Guide for details on how to prepare your own reference genome.
  - See Table1 and Table 2 below for recommended usage.
- The pangenome reference has the following updates.
  - Include Structural Variants in msVCF to improve SV calling recall and accuracy.
  - Masking of the *chrUn_JTFH01000070v1_decoy* improves accuracy in MHC regions.
  - Fix to allow correct calling of FLG haplotype specific in African ancestry samples.
  - Remove trailing NNNN in alt contigs.
  - Filter haplotype-rich regions with more than 224 haplotypes in a 1Kbp window for runtime speedup for alignment.
  - Limit the number of pop-alt contigs with same position and CIGAR, to limit the total length of the concatenated reference to be < 2^32, which helps remove possibly redundant SV representations.
  - Other minor updates.
  - These reference updates are denoted as reference v5.
- Since v4.4, DRAGEN will error out if a linear reference is provided when running a component for which a pangenome reference is recommended as listed in the table below. If the user is sure that a linear reference is reference is desired, the error can be suppressed by setting `--validate-pangenome-reference=false`

Table 1 v4.4 Reference Support and Recommended Use for Human Data

Human		hg19	hs37d5	hg38	chm13	Recommended Reference Type
Germline	SNV	Yes	Yes	Yes	Yes	Pangenome
	CNV	Yes	Yes	Yes	Yes*	Pangenome
	SV	Yes	Yes	Yes	Yes*	Pangenome
	Expansion Hunter	Yes	Yes	Yes	No	Pangenome
	Targeted Callers	Yes	Yes	Yes	No	Pangenome
	RNA	Yes	Yes	Yes	Yes*	Linear
	De Novo	Yes	Yes	Yes	Yes*	Pangenome
	Joint Genotyping	Yes	Yes	Yes	Yes*	Pangenome
	Biomarkers (HLA)	Yes	Yes	Yes	Yes*	Pangenome

	Gvcf Genotyper	Yes	Yes	Yes	Yes*	Pangenome
Somatic	SNV	Yes	Yes	Yes	Yes*	Linear
	UMI SNV	Yes	Yes	Yes	Yes*	Linear
	CNV	Yes	Yes	Yes	Yes*	Linear
	SV	Yes	Yes	Yes	Yes*	Linear
Methylation	Methylation	Yes	Yes	Yes	No	Linear
Annotation	Nirvana	Yes	Yes	Yes	No	n/a

(*) DRAGEN™ supports the component execution; however, the component's accuracy has not been established.

Table 2 v4.4 Reference Support and Recommended Use for Non-Human Data

Non-Human		Supported	Recommended Reference Type
Germline	SNV	Yes	Linear
	CNV	No	n/a
	SV	Yes	Linear
	Expansion Hunter	No	n/a
	Targeted Callers	No	n/a
	RNA	Yes	Linear
	De Novo	Yes	Linear
	Joint Genotyping	Yes	Linear
	Biomarkers (HLA)	No	n/a
	Gvcf Genotyper	Yes	Linear
Somatic	SNV	No	n/a
	UMI SNV	No	n/a
	CNV	No	n/a
	SV	No	n/a
Methylation	Methylation	No	n/a
Annotation	Nirvana	Yes	n/a

Easy-to-use Oncology Apps

- The DRAGEN command line as a toolkit supports WGS and WES analysis of genetic alterations in Tumor DNA/RNA and supports biomarkers. This comprehensive collection of supported somatic analyses includes SNVs and Indels, CNVs, SVs, LOH, Fusions and Rearrangements, TMB, MSI, HRD for multiple types of input preparations.
- Using the DRAGEN command line as a somatic toolkit is good for experienced bioinformaticians and is modular by design. However, the recipes (command lines and parameters) and resource files for each oncology application is unique and needs to be carefully constructed.
- DRAGEN now offers streamlined application-specific oncology pipelines that makes it easy to use. The end-end workflow will enable all the relevant DRAGEN analyses, run the right command lines and parameters and use the appropriate validated resource files.
- These oncology pipelines are distributed as installable applications (.iapp) for the on-premises DRAGEN server.
- The applications (.iapp) and their resource files (.ires) are pre-packaged for downloading via an application downloader, available at the Software Downloads Page: https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/downloads.html
- DRAGEN v4.4.4 launches oncology pipelines for:
    - **Heme WGS**
        - BCL, FASTQ, or BAM/CRAM input

        - Produce DRAGEN alignment, SNPs and INDELs, Structural Variants, DUX4 Rearrangements, Internal Tandem Duplications, Copy Number Variants, Purity/Ploidy Aware, Loss of Heterozygosity.
    - **Tumor/Normal WGS**
        - FASTQ, or BAM/CRAM input
        - Produce DRAGEN alignment, UMI (optional), SNPs and INDELs, MNVs, Structural Variants, DUX4 Rearrangements, Somatic and Germline results, HLA, HRD, CNVs and CNAs, Biomarkers (incl. TMB and MSI), Star alleles, VNTR, QC coverage, ORA compression and simplified interpretation.
    - with more pipelines coming soon.
- The applications will install a DRAGEN Application Manager (DAM) CLI that can be used to manage and execute comprehensive applications available today or in the future.
- Please see the relevant Release Notes for each application and refer to the DRAGEN Software User Guide for up-to-date information about the pipelines.

New MRD (Minimal Residual Disease) Pipeline

The DRAGEN MRD (Minimal Residual Disease) pipeline detects residual cancer cells in solid tumors, enabling the monitoring of treatment efficacy and disease progression. This pipeline utilizes a tumor-informed Whole Genome Sequencing (WGS) approach.

To detect trace ctDNA in plasma, the analysis targets sites and alleles identified as somatic variants in the patient's initial tumor (the tumor fingerprint). Due to the need for significantly higher sensitivity compared to standard ctDNA variant calling, a dedicated application is required to detect these rare molecules (down to tumor fractions as low as 10^-4).

The MRD Detect component provides ultra-sensitive detection of tumor ctDNA and generates multiple quality control (QC) metrics that can be used to assess the validity of the results.

The MRD pipeline does not include a pre-build workflow but rather defines the required computational steps, command line options, pre-built WGS/WES noise files for FFPE and FF.

- **Initial Diagnosis**
  At initial diagnosis, a solid tumor biopsy and a matched normal sample are collected. The DRAGEN small variant caller identifies specific genetic mutations (SNVs) unique to the patient's cancer from this matched sample pair. This set of unique markers constitutes the "tumor fingerprint." It is recommended to prepare libraries with greater than 80X average tumor coverage and greater than 30X average normal coverage. Tumor samples can be FFPE (Formalin-Fixed Paraffin-Embedded) or fresh frozen. Buffy Coat (BC) matched normal samples are recommended.

- **Follow-up Plasma Samples**
  After treatment (e.g., surgery, chemotherapy, stem cell transplant), follow-up plasma samples are collected at various time points to detect residual cancer cells. The tumor fingerprint from the initial diagnosis is used to target the variant sites where residual disease is assessed. Follow-up samples are also evaluated against QC thresholds to ensure sufficient quality. An inter-sample contamination detection step is included to identify potential sample contamination. It is recommended to sequence plasma samples at approximately 50X average WGS coverage.

See the User Guide for examples.

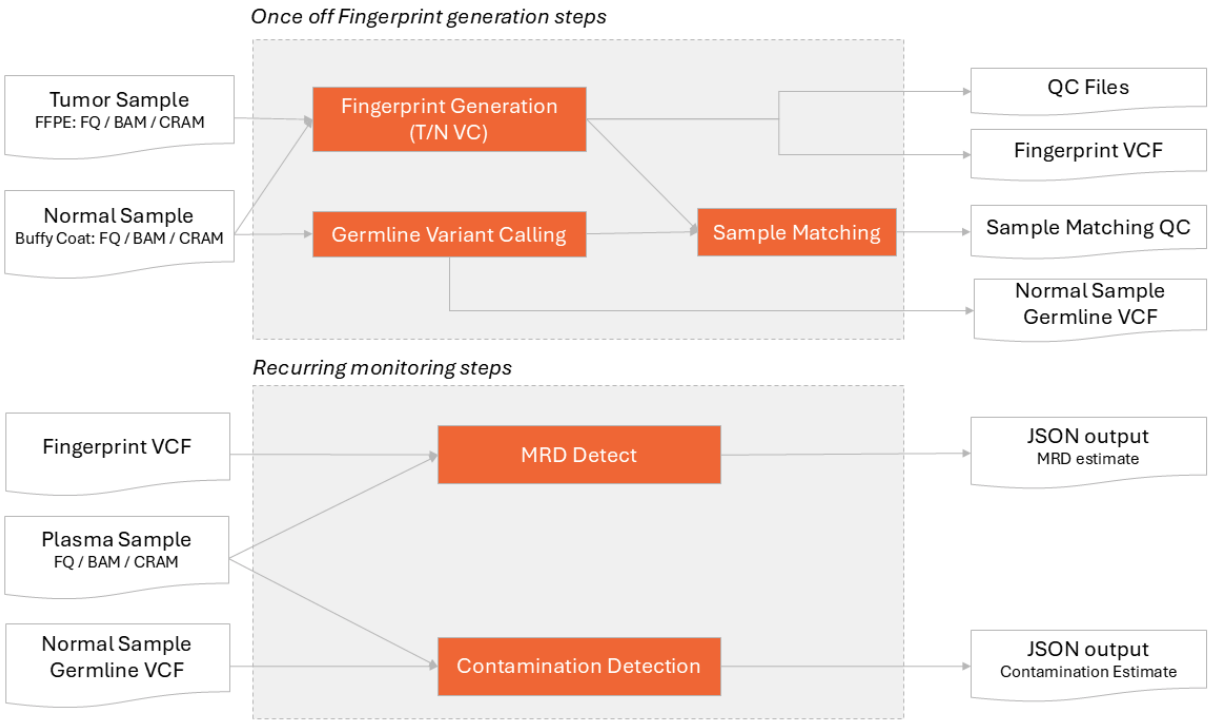


Figure 1. Main DRAGEN steps for an MRD pipeline

Table 3. 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% [95% CI: 99.1%-100%]	100% [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

Somatic Small Variant Caller

- 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;FractionInfo
rmativeReads=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;FractionInformativeR
eads=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;FractionInfo
rmativeReads=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
```

- 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
    - User Interface changes for TMB/Tagging
        - Passing `--enable-variant-annotation true` will do full annotation on all DRAGEN (G)VCF outputs
        - Omitting `--enable-variant-annotation` will automatically only annotate the VCF file with reduced set of annotations needed for TMB/tagging
- Auto-detecting sample type and applying optimal noise method
    - Newer versions of the systematic noise include allele specific information along with two columns for noise frequency, one for the "mean" noise and one for the "max" noise.
    - During a VC run, DRAGEN will automatically detect the input sample type as either WGS or WES/panel and will apply the optimal noise values based on sample type and run context.
    - For WGS data the "max" noise is used by default; for WES/panel data or whenever UMI is enabled, the "mean" noise is used.
    - Usage:

        The following command line options are no longer required
        `--vc-systematic-noise-method max/mean` (option not used)
        `--build-sys-noise-method max/mean` (option deprecated)
- Deprecated pre-filtered VCF output file
    - DRAGEN SNV VC used to output both pre-filtered `<prefix>.vcf.gz` and post-filtered `<prefix>.hard-filtered.vcf.gz` files. The pre-filtered file is now removed for Germline and Somatic pipelines.
- Output STR repeat unit and repeat length information to the INFO field of VCF records
    - Matches INFO annotations output by mutect2
    - Feature is controlled with `--vc-enable-str-info-tag=true (default=true)`
    - Variants that lie within an STR will be tagged with the `STR` flag along with `RU={repeat_unit};RPA={repeat_len}`
- Somatic Allele Transition Noise Metrics File
    - Nucleotide (NTD) error bias estimation can be especially important for FFPE samples.
    - DRAGEN estimates the rate of all possible transitions and transversions in the sample and produces new a header-less CSV file named `*.allele-transition-noise-metrics.csv`.
- Germline Tagging Counts Metrics File

  - In tumor-only mode, DRAGEN includes both germline and somatic variants in the output and. When germline tagging is enabled, DRAGEN will add a `GermlineStatus` tag to the `INFO` field of the VCF.
  - It will also create a header-less CSV file named `*.vc_germline_tagging_metrics.csv` which indicates the overall count of germline and somatic variants
- Tumor-aware checkfingerprint
  - Tumor-only mode
  - Tumor-aware settings assume tumor samples with loss of heterozygosity
  - Usage:
    ```
    --checkfingerprint-enable-tumor-aware true
    --checkfingerprint-loss-of-het-rate [float] (default 0.5)
    ```
- Support for Tumor + Normal BAM/CRAM input to mapper
  - 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 sample types
    - All UMI types

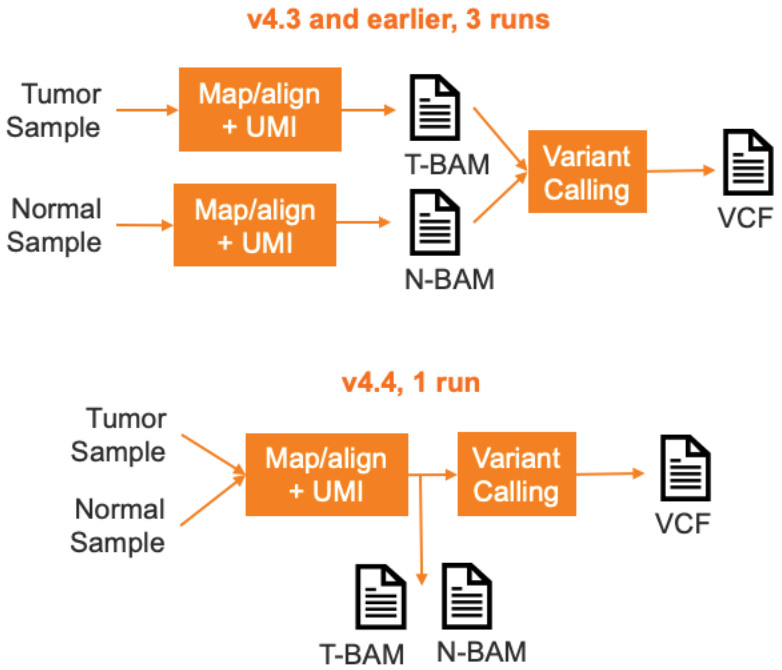


Figure 2. DRAGEN now supports T/N input with UMI in a single run

Illumina PIPseq Single Cell RNA

- DRAGEN now supports scRNA analysis from Illumina Single Cell 3' Prep
  - The DRAGEN PIPseq scRNA Pipeline is designed to process data obtained with the proprietary Illumina Single Cell 3' RNA Prep library based on PIPseq technology
  - The command line option `--scrna-enable-pipseq-mode=true` runs the scRNA pipeline with methods that are optimized to process the unique structure of the library.
- Features of the DRAGEN scRNA analysis
  - 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

  - 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)
  - Barcode and feature counting and correction with new IMIs-per-molecule (IPM) algorithm automatically corrects IMI counts
  - DRAGEN Reports supports highly accurate and visual reports for single cell QC
    - Trimming, FASTQC, mapping and QC metrics, Barcode Rank plot, UMAP plots
  - DRAGEN outputs are supported in Connected Multiomics for further data exploration
- PIPseq CRISPR Mode
  - DRAGEN also supports processing samples from Illumina's CRISPR Single Cell kits using PIPseq technology.
  - Setting `--scrna-enable-pipseq-crispr-mode true` activates this mode.
  - Activating PIPseq CRISPR mode automatically configures DRAGEN for processing feature reads containing the CRISPR guide RNA (gRNA) sequences.
- Requires a license called "PipSeq" and consumes quota from that license. The license is provided free of charge to DRAGEN customers.
- Please refer to the User Guide for command line examples and a full description of the pipeline.

Bulk RNA

- Improved accuracy of bulk RNA gene fusion detection.
  - Improved RNA mapper insert size model and long-range pairing leads to RNA caller improvements

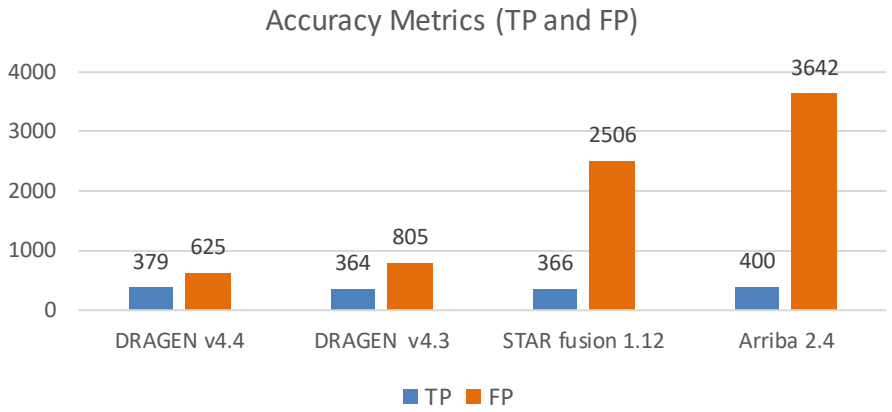


Figure 3. Improved accuracy of Bulk RNA and vs 3rd party solutions

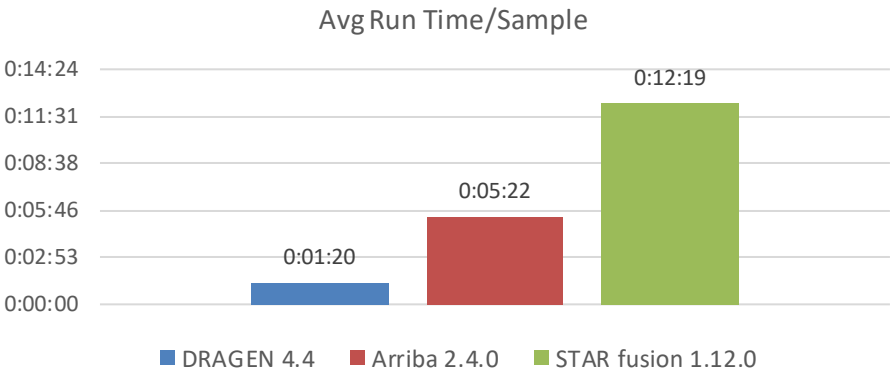


Figure 4. Run Time of DRAGEN vs 3rd party solutions

- 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
  - Updated output of `<prefix>.fusion_candidates.final`
    - `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

… 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	…

Updated New Fields

- RNA Quantification and Fusion calling can now run from BAM input without map/align.
    - Allows RNA Aligner benchmarking vs. STAR, Bowtie, etc. by comparing downstream callers with different aligner outputs

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
- v4.4 detects 3.2x more splices than v4.3
- Outputs:
    - Intragenic splice variants: *splice_variants.tsv*
    - Intergenic splice variants: *splice_variant_fusions.tsv*
- New Features Reported:
    - Unique mapping read counts
    - 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)
- Gene fusion + splice variant caller
    - Genes passed to fusion caller
        - 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)
        - New 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 passed to fusion caller
        - 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`
    - Usage:
        - **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=<file>` containing `SLC45A3,ELK4` genes.

Germline Small Variant Caller

- Accuracy improvements in difficult-to-map regions

- Regions that have other homologous regions in the reference genome for the given read length, number of mismatches, and number of indels.
- Regions with segmental duplications
  - 5.58% reduction in SNPs FP+FN in difficult-to-map regions
  - 4.39% reduction in INDELs FP+FN in difficult-to-map regions

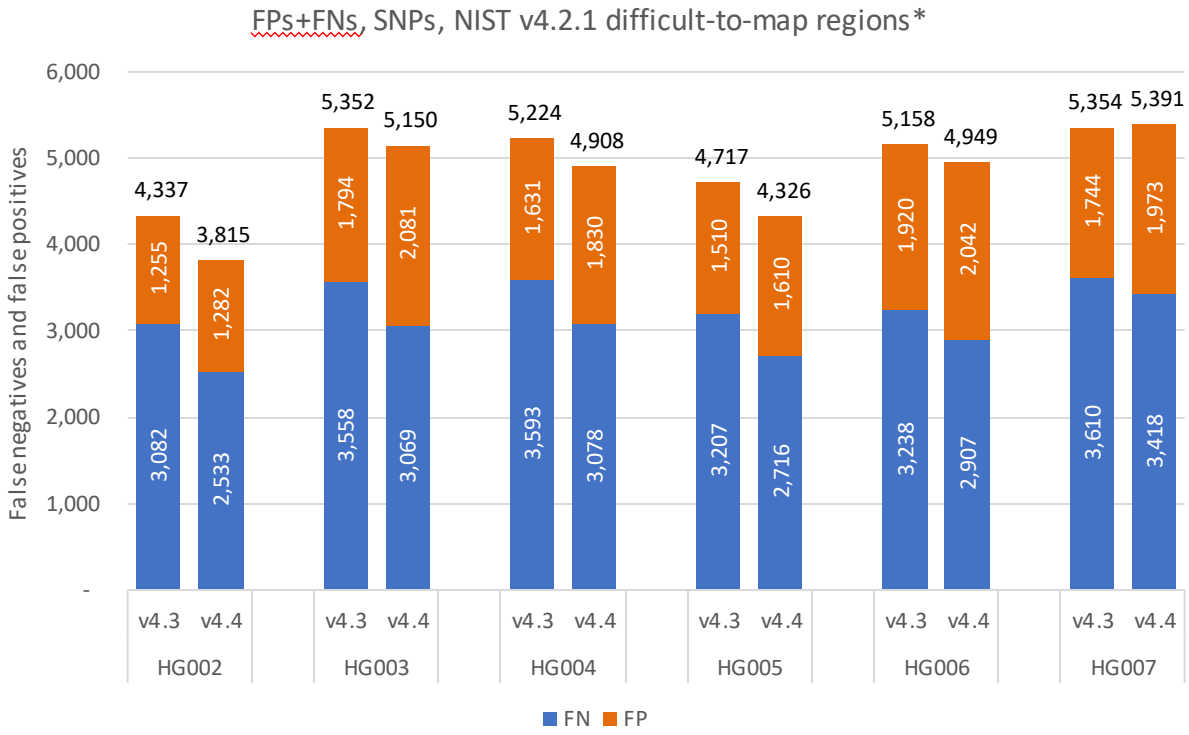


Figure 5. Reduction in SNP FP+FN in difficult-to-map regions

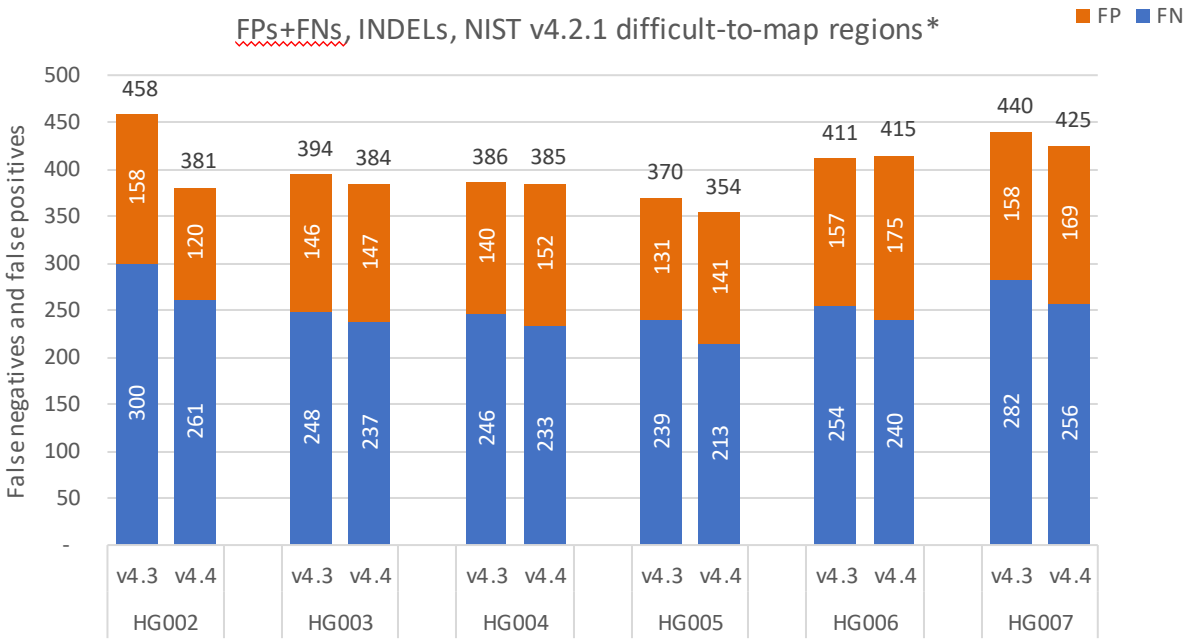


Figure 6. Reduction in INDEL FP+FN in difficult-to-map regions

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

- Official release of the Personalized Germline small variant caller
  - The feature builds a 2-haplotype personalized reference to impute variants, which is used as priors in the Variant Caller, and creates a new personalized ML model.
  - Reduces FP+FN by ~ 20% for SNPs, ~ 7% for INDELs.
  - Adds less than 4 minutes to default small variant calling runs on a DRAGEN P4 server
  - Supports both WGS and WES germline analyses.
  - Requires the pangenome v5 hash table.
  - Usage:
    - `--enable-personalization=true` (default to false)
- Personalized pangenome and ML Model in DRAGEN 4.4 results in a significant accuracy improvement
  - 17.76% reduction in FP+FN (SNPs+INDELs) using NIST T2TQ100-v1.1 truth (HG002)
  - 33.96% reduction in FP+FN (SNPs+INDELs) using NIST v4.2.1 truth (HG002)
  - 24.18% reduction in FP+FN (SNPs+INDELs) using NIST v4.2.1 difficult-to map regions
  - 18% reduction in FP+FN (SNPs+INDELs) using NIST Challenging Medically Relevant Genes (CMRG)

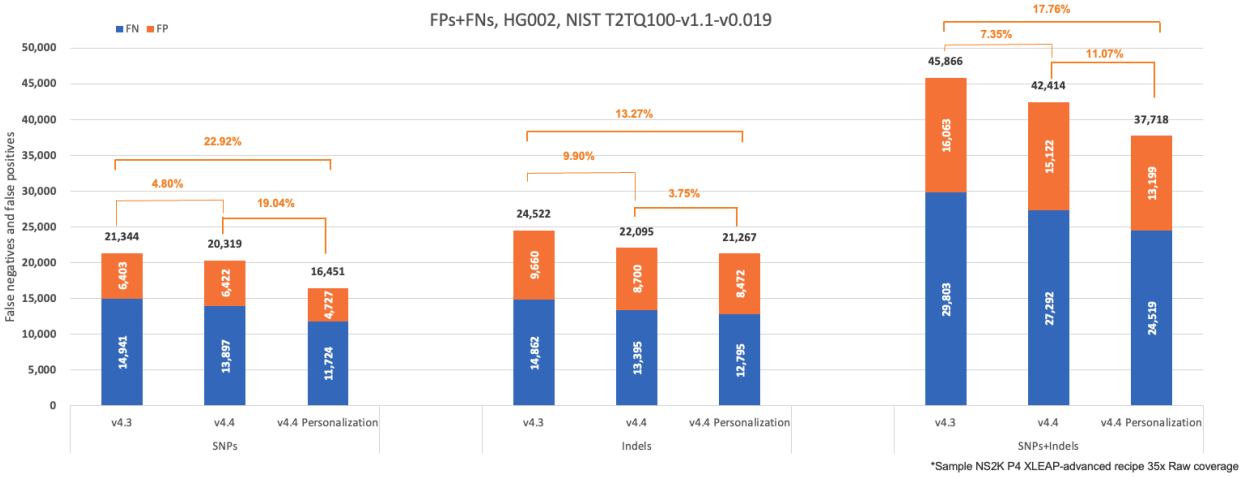


Figure 7. Germline Small Variant Caller Accuracy using NIST T2TQ100

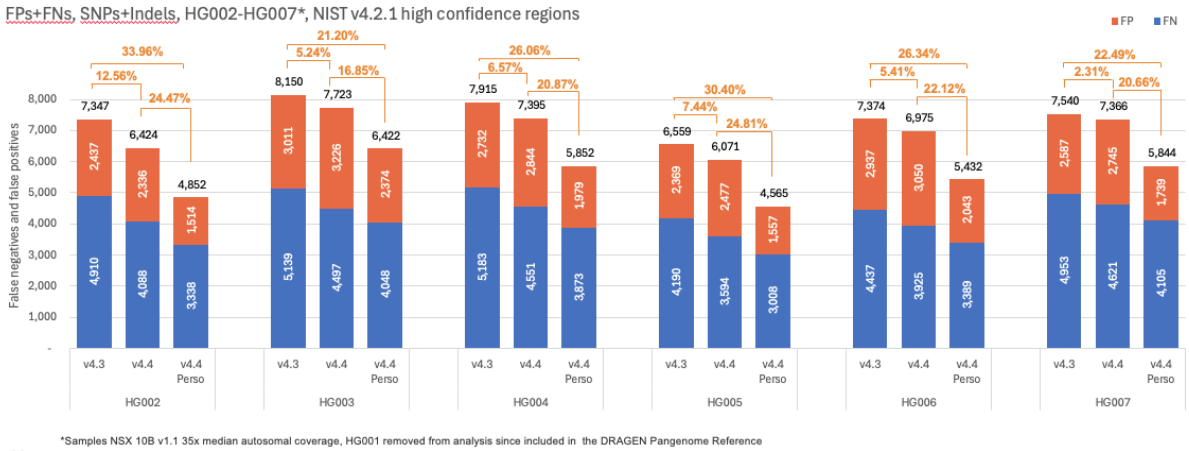
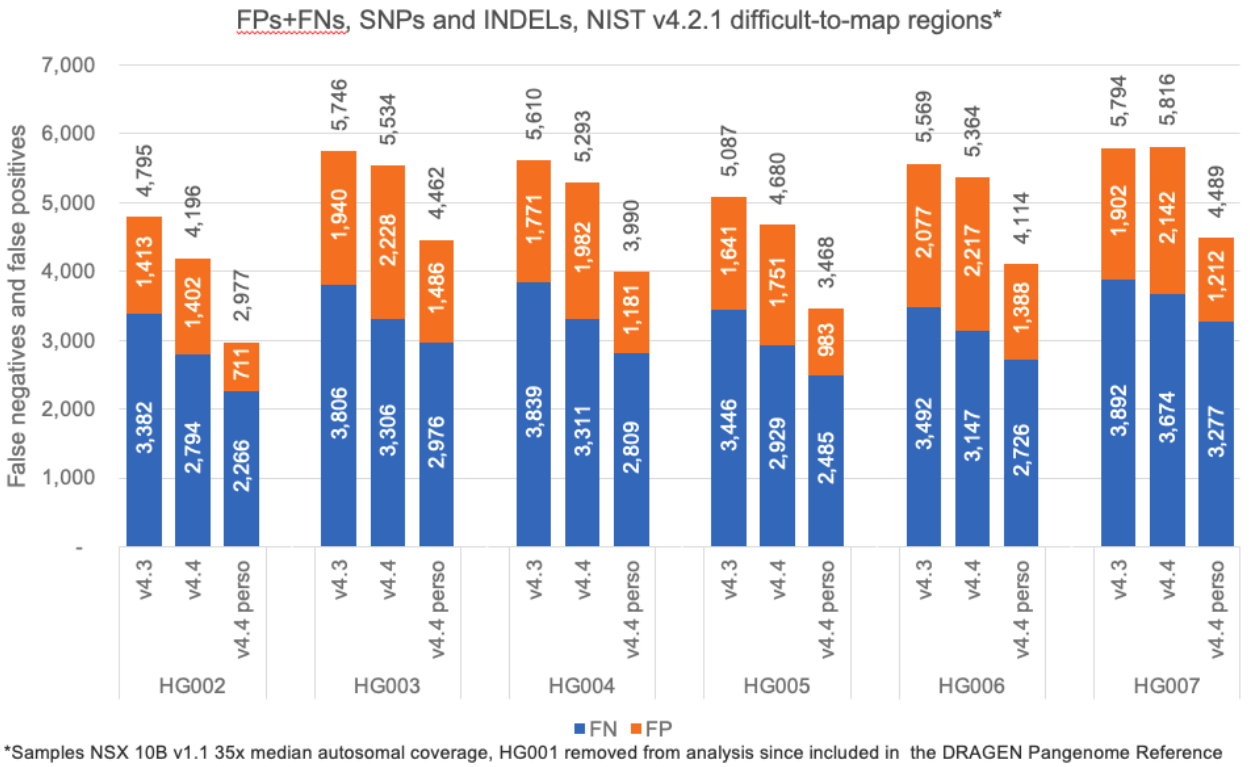


Figure 8. Germline Small Variant Caller Accuracy using NIST v4.2.1



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

Figure 9. Germline Small Variant Caller Accuracy on difficult-to-map-regions

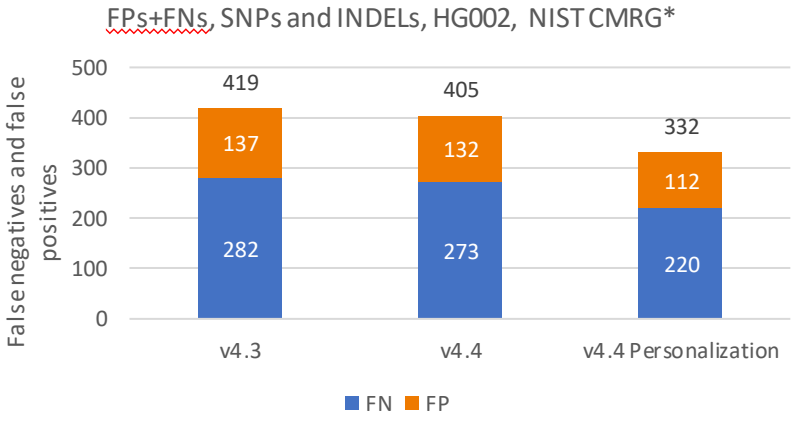


Figure 10. Germline Small Variant Caller Accuracy in CMRG

Germline Structural Variant Caller

- Major leap in structural variant calling accuracy
  - DRAGEN 4.4 improves on SV F-score by more than 30% (on HG002 with both NIST T2TQ100 and CMRG benchmarks)
  - Structural variant population haplotypes incorporated in DRAGEN's pangenome reference increase accuracy

- Significant recall gains in SV insertion detection including MEIs
- Updates in read alignment, breakpoint detection, population-guided assembly, and genotyping models enhance the accuracy of SV calls

Table 4. SV Accuracy

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

Table 5. SV MEIs

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

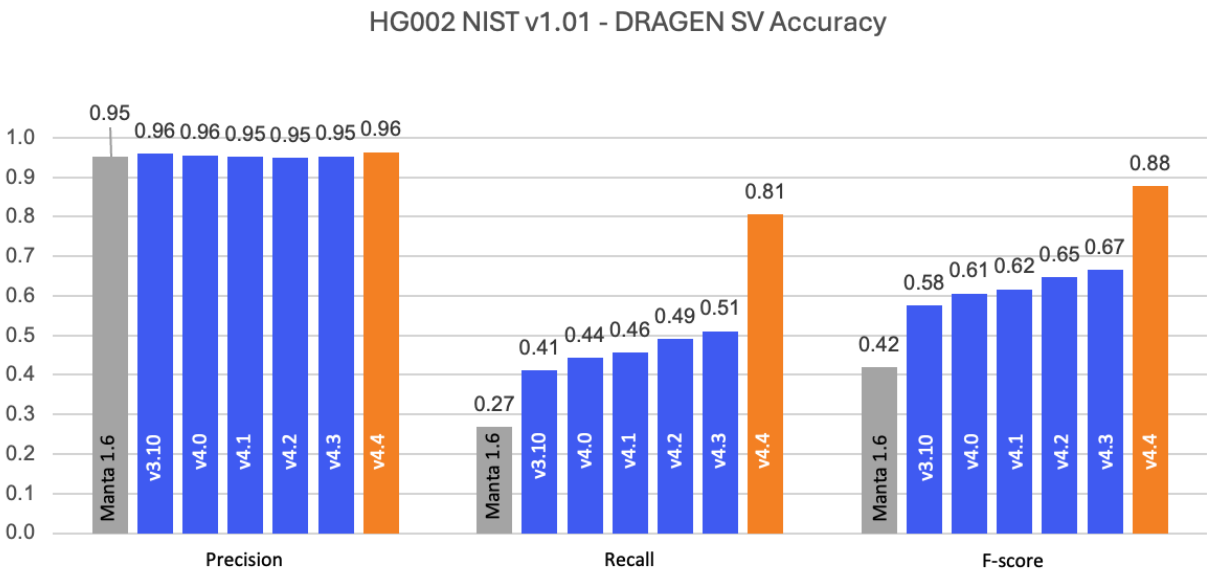


Figure 11. Accuracy Improvements of DRAGEN SV Calling

Germline Enrichment CNV

- BSSH and ICA Germline Enrichment apps now offer in-run construction of CNV panel-of-normals (PoN)
  - 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

Germline WGS Cytogenetics

- Germline CNV caller now supports B-allele frequency estimation, enabling 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
- > 97% agreement with results from karyotyping and chromosomal microarray
  - Evaluated on 98 samples with 146 CNVs ranging from 40kbp to whole-chromosome aneuploidies
- Newly introduced cytogenetics modality reports multiple views of the same sample
  - Fine resolution views for shorter alterations (≥ 25kb)
  - Coarse resolution views for larger alterations (≥ Mbs)
- See the User Guide for complete information about Cytogenetics Modality

Supported events
DUP ≥ 1Mb
DEL ≥ 1Mb
DUP [50kb, 1Mb)
DEL [25kb, 1Mb)
AOH ≥ 500kb
Mosaicism ≥ 20%

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

Figure 12. > 97% Concordance against karyotyping and CMA

Targeted Calling

- **New WES Targeted Calling support**
  - Support for HBA and SMN on enrichment data
  - 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
  - 95% benchmark concordance for SMN
  - How to run WES Targeted Calling from the 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
- **#1 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.
  Usage:
  ```
  dragen \
  -r {REF} \
  --bam-input {BAM} \
  --output-directory {OUTDIR} \
  --output-file-prefix {PREFIX} \
  --enable-map-align=false \
  --targeted-generate-exome-counts true
  ```
  Output file: `<output-file-prefix>.targeted.exome.counts.json.gz`
- **#2 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.
  Usage:
  ```
  dragen \
  --output-directory {OUTDIR} \
  --output-file-prefix {PREFIX} \
  --targeted-pon-counts-list run1.targeted.exome.counts.json.gz
  ```
  Output file: `<output-file-prefix>.targeted.pon.json.gz`
- **#3 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.
  Usage:
  ```
  dragen \
  -r {REF} \
  --bam-input {BAM} \
  --output-directory {OUTDIR} \
  --output-file-prefix {PREFIX} \
  --targeted-pon run2.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 Software Support Site page)
  Output files: `<output-file-prefix>.targeted.json`
  `<output-file-prefix>.targeted.vcf.gz`

- **MRJD**
  - 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
  - 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
- **STR motif composition**
  - Some STR loci have polymorphic motifs, meaning that the different repetitions of the motif have minor variations in their sequence. For example, an STR may contain some repetitions of AAAGG and some repetitions of AAGGG, all in the same haplotype.
  - 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
  - Example: (enabled by default for RFC1)

```
{
     "LocusId": "RFC1",
     "LocusStructure": "(AARRG)*",
     "ReferenceRegion": "4:39348424-39348479",
     "VariantType": "Repeat",
     "MotifAnalysis": [
       "AAAAG", "AAAGG",
       "AAGGG", "AAGAG",
       "AACGG", "ACGGG",
       "ACAGG", "AAAGGG"] # for HMM
}
```

  - Output: New `FORMAT` fields `MOTIFS`, `MF` (motif fractions) and `AMF` (motif fractions per allele)
- **Genotyping for 41 HLA and HLA-related genes**
  - Nucleotide-based alignments using DRAGEN map-align and expectation-maximization for HLA genotyping
    - HLA reference contains ImMunoGeneTics project (IMGT) allele sequences
    - Genotyping**:** at most 2 alleles returned per gene
      - Allele pair supported by max read count
      - Homozygosity check: posterior probability > 0.15
    - Requires map/align
  - 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
  - HLA Genotyping to Class 2 is now enabled by default
    - The option `--hla-enable-class-2` is now deprecated
- VCF output for PGx genes (CYP2D6, CYP2B6)

CNV Caller

- **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.

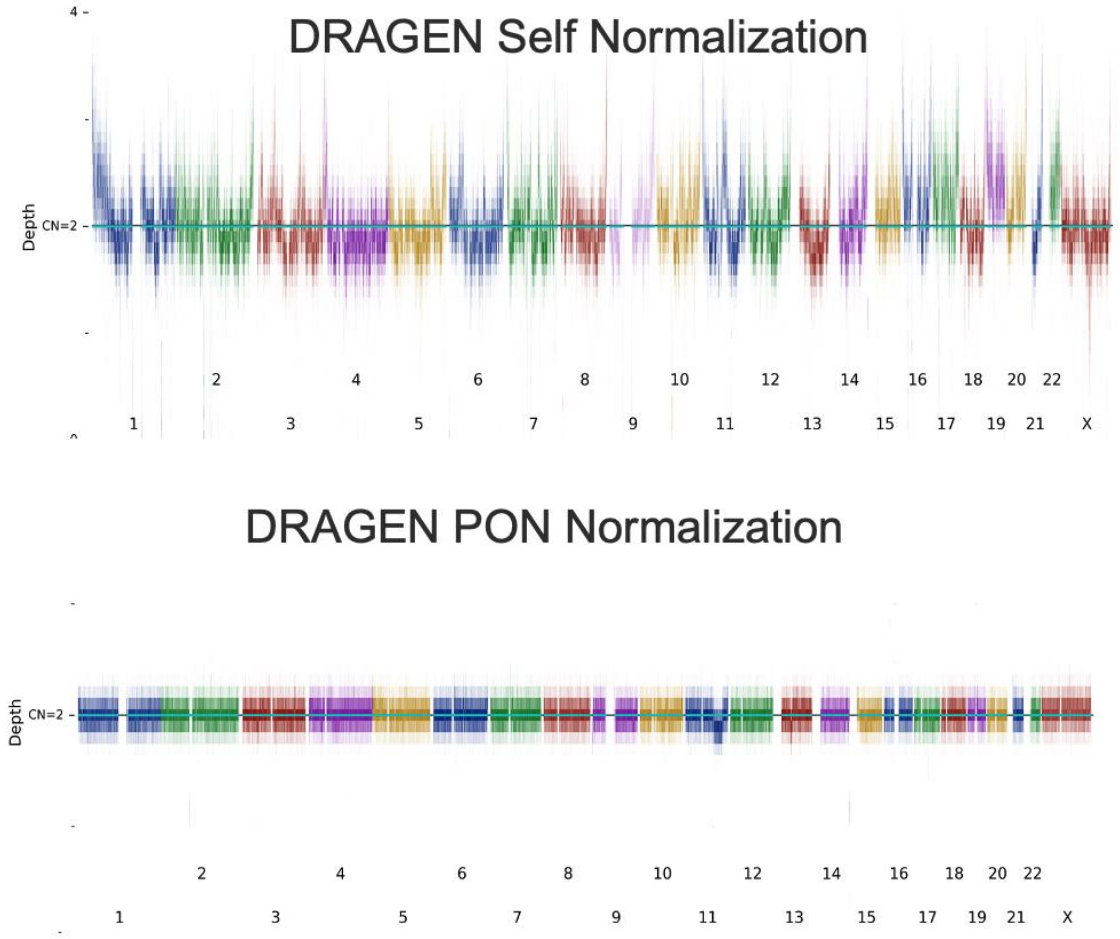


Figure 13. PoN Normalization vs Self Normalization

- **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.

Iterative Gvcf Genotyper (iGG)

- iGG v4.4 now requires a license called "GvcfGenotyper" and consumes quota from that license
  - License check only on step1 (`--gvcfs-to-cohort-census`) or end-to-end (`--gvcfs-to-msvcf`)
- Improve iGG scalability for GWAS
  - Steps have been combined and made faster
- Support gVCF output of new DRAGEN germline and targeted caller FORMAT and INFO fields
  - Flexible gVCF metric import. 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
    - Usage:
        - 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
          (GT is unconditionally imported)
          ```

- iGG usability and robustness improvements
    - Improve error handling for customers with DRAGEN 3.7 gVCF input
    - Add option to discard mosaic variants `--gg-discard-mosaic`

BCL Convert

- Updates for greater flexibility and efficiency.
    - New SampleSheet 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
    - New output files
        - New `Demultiplex_Detailed_Stats.csv` file provides cycle & transition error details for mismatch tolerant demultiplexing (miscall metrics)
        - Extended columns to per-tile `Demultiplex_Tile_Metrics.csv` file, to include derived stats identical to those in the aggregate `Demultiplex_Stats.csv` file
    - Generate undetermined read stats when enabling `--bcl-only-matched-reads`
        - Even though Undetermined FASTQs are not produced, the stats will still be produced.
    - Memory usage optimization
        - New settings reduce memory usage for large sample batches

ORA Compression

- BCL to FASTQ.ORA up to 30% faster
    - Convert BCL into ORA skipping intermediate FASTQ.GZ files, and decreasing the run time
    - When using compression during BCL convert, add the command `--bcl-ora-direct true`
    - Limited to a max number of samples per lane of 40
    - Requires `--fastq-compression=dragen` (cannot be used with interleaved)
- BCL to FASTQ.ORA stores additional metadata 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

Nirvana Annotation

- DRAGEN output annotation updates
  - Added annotation for `.vntr.vcf` and `.targeted.vcf` output.
  - Removed VNTR annotation
  - Skip annotation of `hard-filtered.gvcf` output if there is also `hard-filtered.vcf` (no need to annotate both)
  - Ease-of-use updates
    - No longer required to specify `--variant-annotation-assembly` on the command line, the annotation assembly GRCh37/GRCh38 is auto-detected.
    - Specifying of annotation options for TMB and Germline Tagging purposes are no longer required. DRAGEN will automatically run an optimized Nirvana annotation for TMB/Tagging.
    - Behavior change
      - When `--enable-variant-annotation=true` is specified, all DRAGEN outputs will be annotated
- Nirvana is updated to v3.25.0
  - Interface change
    - Nirvana now has new binary for downloading annotation data and to validate data: `<INSTALL_PATH>/share/nirvana/DataManager`. Old *`Downloader`* binary is removed
    - To be able to download data successfully, a credential file with Illumina API key is required.
    - Please see the User Guide for a full description of the Illumina Connected Annotations
  - Updated gene models from RefSeq and Ensembl
  - Added Karyotype Annotation

AWS F2 for Cloud Analysis

- **New AWS EC2 f2 FPGA instance types.**
  - Second-generation FPGA-powered instance types have more CPUs more RAM larger FPGAs improved networking and EBS bandwidth and faster clock speeds, compared to AWS f1.
  - DRAGEN v4.4 has native support for acceleration on f2
  - All BSSH/ICA apps with v4.4 prioritizes execution on f2 instances
  - Improved turn-around times and capacity
  - Lower cost for BYOL customers
- **Significant reduction in turn-around time across all workflow types**

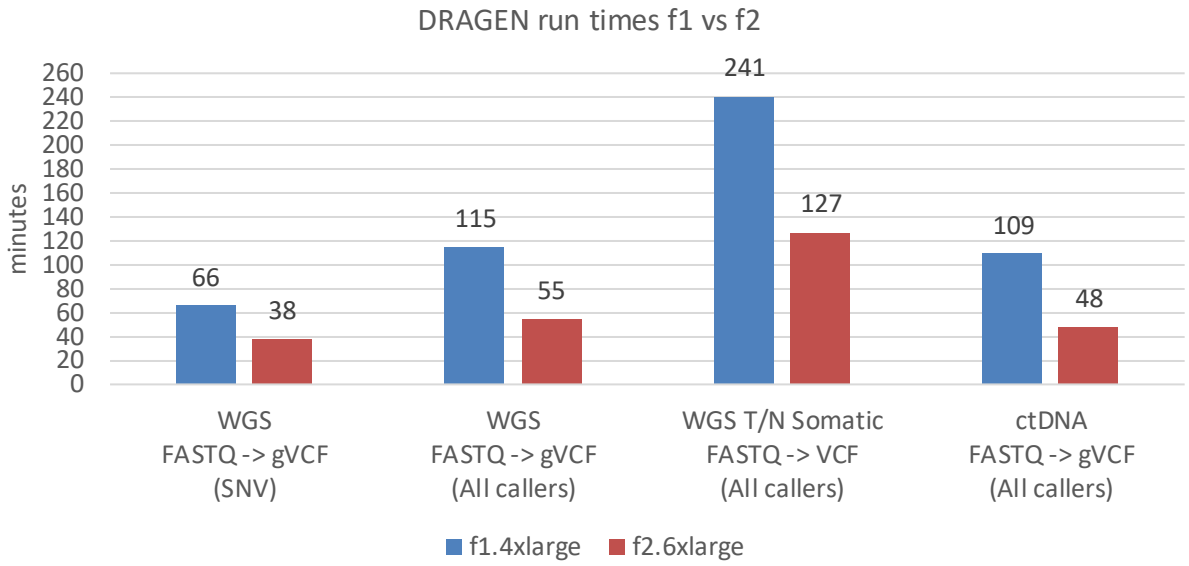


Figure 14. DRAGEN Run Time on AWS f2 vs f1

- AWS f2 support will be backported to selected prior DRAGEN versions. Please reach out to customer support for more information on versions, deployments and schedules.

Other Updates

- **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
  - Usage:
    - New option `--cram-version=3.1` (default is 3.0)
- **Cross Sample Contamination**
  - New option to enable contamination: `--qc-detect-contamination=true` (default false)
  - Uses GATK contamination method for both germline and somatic modes, and autodetects the population allele frequency vcf to use.
  - The recommended way to run contamination detection based on *GATK CalculateContamination*.
  - NOTE:
    - The legacy contamination module that is similar to VerifyBamID, is only supported in germline mode and no longer recommended. Enabled with `--qc-cross-cont-vcf=<vcf>`.
    - The newer contamination module that based on *GATK CalculateContamination* by specifying population allele frequency marker loci, has advanced support for somatic samples and is recommended. A VCF can be passed with `--qc-somatic-contam-vcf=<vcf>`, but the new option is recommended.
- **Unified JSON Metrics**
  - DRAGEN 4.4 generates a unified JSON metrics file `<output_prefix>.metrics.json` for every run
  - Unified JSON metrics can contain the following modules (when 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
    - Example

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

- **Fractional Downsampler for RNA Pipelines**
    - 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-deduped)
    - Reduce runtime and cost of analysis using high-depth samples
    - Any input supported by DRAGEN can be used
    - Usage:
        - 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%)
- **Operating System Support**
    - CentOS 7 support is deprecated
        - NOTE: On-premises builds, AWS AMIs and Azure VMs for el7 could be manually generated for v4.4 on request though customer support. 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 plans:
        - el9 is supported since v4.4 but not yet released
        - Illumina DRAGEN server OS image for Oracle 9 to be released in Q3/Q4 2025, and el9 rpms/installers to follow.

- **Licensing**
  - New License Server Domain – license.dragen.illumina.com
    - On-premises servers
      - HTTP @ lus.edicogenome.com is being replaced with HTTPS @ license.dragen.illumina.com
      - HTTPS domain addresses security concerns for connected servers
      - Only DRAGEN 4.4+ supports the new domain, previous versions of DRAGEN will need to continue to use HTTP @ lus.edicogenome.com
    - BYOL Cloud and Software-only Mode
      - 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
  - BYOL Cloud – Now 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 DRAGEN commands `--lic-server` or `--lic-credentials`.
    - Can export to a JSON format using the `-j` option.
    - Example

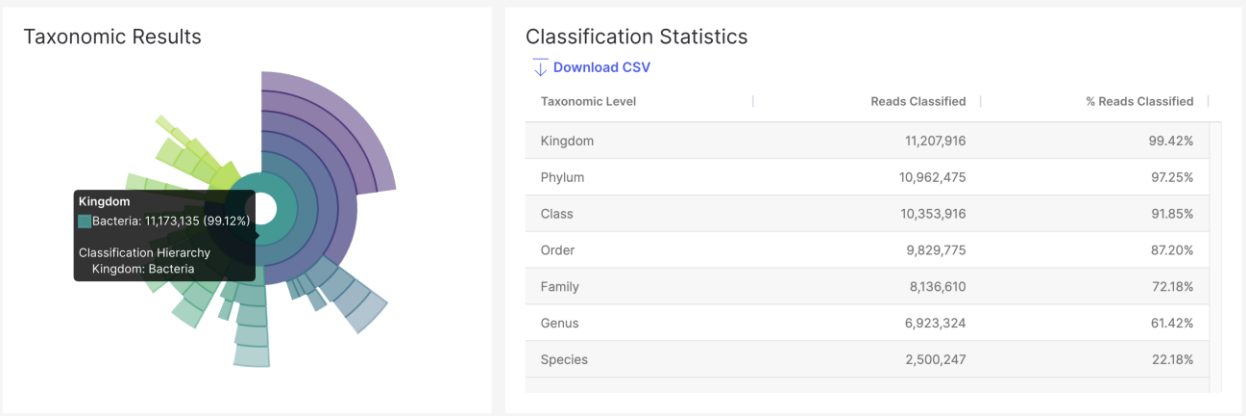
```
$ dragen_lic --lic-credentials my_credentials.cfg
User: <username>
DRAGEN Version: <version>
Time: <time>

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
```

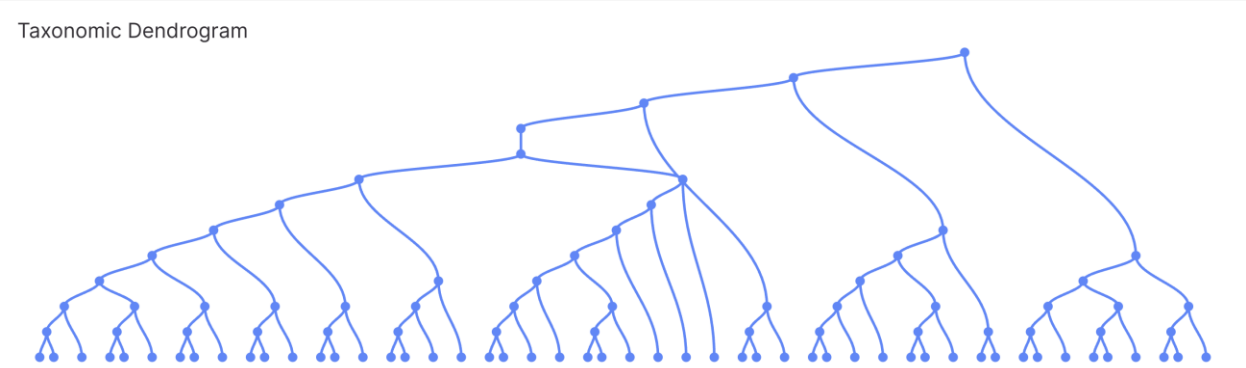
    - Alternatively, you can retrieve license information using our License Server endpoint specified below without the use of DRAGEN. License information is returned in a JSON format
    - GET request to https://license.dragen.illumina.com/api/v2/query_quota. Your user credentials must be provided as a Basic Authorization header. An example of this using the curl tool is shown below.

```
curl https://license.dragen.illumina.com/api/v2/query_quota --user
<username>:<password>
```

  - Fleet Licensing
    - Not DRAGEN version specific, mentioned here for information.
    - DRAGEN will be launching a fleet-based licensing in 2025.
    - One single license purchase and quota can be shared by multiple servers. This simplifies sales and management of multiple DRAGEN servers.
    - Requires the on-premises servers to be connected.

- **DRAGEN Reports updates**
  - Updated DRAGEN Reports .rpms available for download.
  - 16S Report
    - 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





Known Issues

Known issues of the DRAGEN™ v4.4.4 release

Component	Summary	Resolution/Workaround
Amplicon	Germline amplicon analysis can report a VAF much higher than expected	None. A fix is planned for next release.
BCL	If a directory is specified as input to "--sample-sheet", BCL Convert will hang at the beginning of a run while trying to copy that path as a file to <outdir>/Reports/SampleSheet.csv	Specify the sample sheet file.
BCL	BCL does not detect when LibraryInputVolume setting is blank	Blank/empty value is the same as not providing the setting, which is the same as the default setting of being disabled.
BCL	BCL conversion appends FASTQ files when using "—force". FASTQ output may get concatenated if user uses the same output directory twice for BCL.	Do not run BCL conversion multiple times using the same output folder
BCL	Sample sheet validation can pass, but demultiplexing fails	Barcode collision detection works slightly differently for ss validation and real runs. Some errors are not caught at ss validation step.
CNV VC	The cnv-exclude-bed option is not honored in segmental duplication results	None.
CNV VC	There might be two alterations co-occurring, AOH/LOH and DEL. Our current output format supports reporting a single alteration in such case and reports the strongest alteration between the two. In 4.4, for the same region we can now report both the DEPTH+BAF and the DEPTH-only call in the new Cytogenetics modality.	None.
CNV VC	segmentMean filter thresholds are not printed in the header when not specified in input	None. If required, add the filter command line option.
Compression	Fatal error when --ora-get-metadata is used on empty fastq.ora file	None. A fix is planned for next release.
DNA Alignment	Specifying '--Aligner.match-score=0' will set match-n-score to zero under-the-hood which leads to leads to invalid alignments	None. Do not use the setting. Informational only
DNA Alignment, RNA Alignment	FASTQ header parsing does not support tabs	None. Edit the FASTQ files from tools that inject tabs.

Downsampling	Exome downsampling is not giving right coverage when coverage downsampling with no genomic region is specified.	Use --down-sampler-genome-size that matches the size of the target BED region, or use the fractional downsampler
Downsampling	Option --down-sampler-reads has a limit of 4,294,967,295 due to the use of uint32.	Use the fractional downsampler for large files
Germline	Small SNP accuracy regression may be seen on HG001 truth samples. HG001 samples were removed from the ML training to remove overfit bias, therefore some regressions on this sample is expected.	None. Informational only
Imputation	Intermittent hang has been encountered for low pass sequencing samples using ForceGT (imputation).	Re-run the sample
Infrastructure	If an AWS node is configured to "IMDSv2 Required", S3 input file streaming does not work.	Typical configuration is "IMDSv2 Optional", in which case S3 input streaming works.
Infrastructure	Input streaming from s3 bucket is not working with IAM role, when the instance is in eu-west-2 (other regions ok).	None. Bug in AWS SDK.
Infrastructure	Azure direct to BLOB UL streaming has intermittent crashes.	None. The crashes happen in the Azure storage sdk libraries. Do not use UL direct to BLOB streaming on Azure.
Infrastructure	BAMs from 3rd party tools may produce RG SAM tag for a record that does not have a matching RG ID. For T/N analysis, tumor reads with an invalid RG SAM tags get misallocated to matched normal.	None. Dragen BAMs do not have this issue.
ML	Joint Genotyping with ML and personalization enabled have PL values that are inconsistent with GT.	None. Do not use personalization for joint calling
ML	Extra FNs have been observed for spiked-in Mosaic variants on 35x spiked-in samples.	Does not affect high-depth (e.g 300x) spiked-in samples which constitute the main use case for the mosaic caller.
Ora	ORA compression or decompression without --force option, and with output file already there, returns inconsistent error codes.	The system correctly stops before overwriting already existing output file.
Ora, SNV Germline	DRAGEN watchdog time outs have been observed when starting from ORA input and running high coverage samples.	None. A fix is planned for next release.
Paralog Caller	CYP2D6/CYP2B6 phenotype annotation doesn't handle >=3 copies of star allele	None.

Personalization	Dragen uses more system memory when personalization is enabled	None.
scRNA	When the option single-cell-threshold=inflection is set, there is a chance that the number of cells is overcounted because the second knee is chosen rather than the first knee in the cell calling knee-plot.	None. A fix is planned for next release.
scRNA	Intermittent hangs have been observed when processing T100 samples	None.
scRNA	Single cell RNA reads are reported as "R1" in the "mapping_metrics.csv" file even though the gene expression cDNA is part of R2.	None.
scRNA	Feature barcode read groups fails when no lane splitting is set to true in BCL.	None.
scRNA	HI and NH BAM tags are zeroed out in single cell	None.
scRNA	Incorrect results when using -1 and -2 (fastq-file1 and fastq-file2)	Use the recommended methods for scRNA input as per the user guide.
scRNA	Occasional run-to-run variation in mapping_metrics_csv have been observed, leading to small differences in SJ.out.tab, and unfiltered.SJ.out.tab files. The run-run variation is on a single read only.	None. A fix is planned for next release.
scRNA	Fractional downsampling does not work for single-cell RNA-seq	None. A fix is planned for next release.
scRNA	Feature counting is using IMI-corrected reads. Seen when enabling feature counting mode via scrna-feature-barcode-reference and scrna-feature-barcode-read-groups, but without enabling scrna-enable-pipseq-crispr-mode.	None. A fix is planned for next release.
SNV Germline	Lower QUAL scores have been observed for rare variants since v4.3	None. A fix is planned for next release.
SNV Somatic, UMI	Tumor+UMI/Normal from BAM/CRAM input crashes with setting --tumor-normal-has-umi=tumor.	None. A fix is planned for next release.
SNV VC	Crash observed in read realignment due to very short haps, when using --vc-output-evidence-bam true --vc-evidence-bam-force-output true	None. A fix is planned for next release.
Somatic	Somatic small VC has higher memory usage on v4.4 relative to v4.3 due to adding STR annotations to INFO field of VCF records and enabling MNV detection by default.	None.

SV	Very high coverage somatic samples that have excessive structural variants detected, may take a long time to run on cloud platforms such as ICA, due to limited CPU capability.	Regions with F2 instances mitigate the excessive execution time.
SV	T/N analysis may fal in the SV caller when running on tumor sample with UMI inputs	None. A fix is planned for next release.
TMB	TMB WES tumor-only accuracy is not as reliable as T/N WES TMB	None.
UMI	(T+UMI)/N fails when fastq list input is used with multiple lanes.	None. Start the run from FASTQ files
iGG	Sometimes negative PL values are produced in haploid regions, and is interpreted by htslib as missing value, leading to failure of Step 1 n iGG	None. A fix is planned for next release.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package.
- The archive integrity can be checked using: `./<DRAGEN 4.4.4 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.4.4 .run file>`

Release History

Revision	Release Reference	Originator	Description of Change
00	1121521	Cobus De Beer	Initial release