DRAGEN v4.3.6 Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Secondary Analysis Software v4.3.6.

Changes are relative to DRAGEN™ v4.2.9. If you are upgrading from a version prior to DRAGEN™ v4.2, 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-4.3.6-11.multi.el8.x86_64.run
- DRAGEN™ SW for x86 CentOS 7 - dragen-4.3.6-11.multi.el7.x86_64.run

The following configurations containing DRAGEN™ 4.3.6 are also available on request:
- AlmaLinux 8 and CentOS 7 based Amazon Machine Images (AMIs) for f1 instances, available in 12 regions
- CentOS 7 Microsoft Azure Image (VM) available in West US 2 for BYOL
- el8 and el7 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 and el7, for use with customer generated AMIs or QuickStart

DRAGEN™ v4.3.6 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 ends on June 30, 2024. DRAGEN™ v4.3.6 is the final release with CentOS 7 installers. Future releases will support el8 and el9.
- 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 v4.3.6. DRAGEN™ v4.3 offers significant improvements in accuracy, added features for a more comprehensive solution, and efficiency improvements. 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

Accuracy

- Next generation multigenome (graph) improves SNV accuracy.
- Improvements in ML models reduce FP and FN.
- Population aware mapper yields accuracy gains.
- Customer multigenome can reduce bias in population studies.
- Mosaic variant calling for low allele frequencies.
- Improved RNA Gene Fusion accuracy.
- SV accuracy improved in repeat regions.
- CNV accuracy improved in segmental duplication regions.
- New imputation reference panel improved accuracy from low-pass data.

Comprehensiveness

- New specialized callers for tandem repeats, DUX4.
- Variant calling in segmental duplication regions.
- RNA splice variant calling.
- Support for expanded annotations with PrimateAI-3D and Splice AI.

Efficiency and accessibility

- Improved performance for large cohort analysis.
- Improved ORA compression ratio for methylation and non-human data.

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

Updated Resource Files

DRAGEN™ v4.3 requires updates to key resource files to function correctly and achieve the optimum performance. Additional resource files are made available for v4.3. 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

The following resource files are updated:

Resource	Description	File name(s)
Hash Tables v10	Pre-built v10 multigenome and linear hash tables for hg38, hg19, hs37d5, CHM13.	Multigenome (graph): hg38-alt_masked.cnv.graph.hla.rna-10-r4.0-1.tar.gz hg19-alt_masked.cnv.graph.hla.rna-10-r4.0-1.tar.gz hs37d5-cnv.graph.hla.rna-10-r4.0-1.tar.gz chm13_v2-cnv.graph.hla.rna-10-r4.0-1.tar.gz Linear (non-graph): hg38-alt_masked.cnv.hla.methylated_combined.rna-10-r4.0-1.tar.gz hg19-alt_masked.cnv.hla.methylated_combined.rna-10-r4.0-1.tar.gz hs37d5-cnv.hla.methylated_combined.rna-10-r4.0-1.tar.gz chm13_v2-cnv.hla.methylated_combined.rna-10-r4.0-1.tar.gz
Multigenome Reference Builder Collection v4	HT mask BED, Graph BED, Graph exclusion BED, Graph msVCF and FASTA files for building hg38, hg19, hs37d5, chm13 references.	hg38-multigenome_reference_collection-v4.tar.gz hg19-multigenome_reference_collection-v4.tar.gz hs37d5-multigenome_reference_collection-v4.tar.gz chm13_v2-multigenome_reference_collection-v4.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 v2.0.0	A collection of Somatic noise baseline BEDPE files for WGS hg19, hs37d5, hg38. New file for Heme WGS hg38.	sv-systematic-noise-baseline-collection-3.0.0.tar The tar archive contains the following files: IDPF_WGS_hg38_v3.0.0_systematic_noise.sv.bedpe.gz WGS_hg19_v3.0.0_systematic_noise.sv.bedpe.gz WGS_hg38_v3.0.0_systematic_noise.sv.bedpe.gz WGS_hs37d5_v3.0.0_systematic_noise.sv.bedpe.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
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
Microsattelite Files v1.0.0	Microsattelite files for hg19, hs37d5, hg38 and for WGS and WES respectively	microsatellite-files-1.0.0.tar.gz The tar archive contains the following files: WGS_v1.0.0_hg38_microsatellites.list WGS_v1.0.0_hg19_microsatellites.list WGS_v1.0.0_hs37d5_microsatellites.list

		WES_v1.0.0_hg38_microsatellites.list WES_v1.0.0_hg19_microsatellites.list WES_v1.0.0_hs37d5_microsatellites.list
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: oradataV2.tar.gz Non-human and human methylated: oradata_arabidopsis_thaliana.tar.gz oradata_bos_taurus.tar.gz oradata_caenorhabditis_elegans.tar.gz oradata_danio_rerio.tar.gz oradata_gallus_gallus.tar.gz oradata_glycine_max.tar.gz oradata_homo_sapiens.tar.gz oradata_homo_sapiens_bisulfite.tar.gz oradata_mus_musculus.tar.gz oradata_oryza_sativa.tar.gz oradata_rattus_norvegicus.tar.gz oradata_sus_scrofa.tar.gz oradata_triticum_aestivum.tar.gz oradata_zea_mays.tar.gz Combined all species: oradata_all_species.tar.gz

NOTES:

- ML Model files are included in the installer by default since v4.2 and does not need to be downloaded.

Updated Installer

DRAGEN™ v4.3 has updated the installer for on-premises systems which results in a user interface break.

- Current state: one version on one server at a time
- Future state: multiple versions on each server
- Backward compatibility with previous versions coming soon. Select current versions will be available with new installers.

On-premises Installation

The software can be installed on an on-premises server by executing the `.run` installer for the desired version.

Installation procedure:

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

The `.run` file includes a script that administers un-installation of an existing software, integrity checking of the package and files, installation of the new `dragen` software version. The software is installed in part by use of the Linux RPM Package Manager (rpm). Several rpm packages comprise the installation of a single software version. The RPM packages also configure the system, such as raising user `ulimits`, and the `.run` script starts services needed for functionality, such as the Licensing daemon `dragen_licd`, and the hugepages daemon, `dragend_hp`.

NOTE: Root privileges are still required for the installation.

Single Version Installation

Up to DRAGEN™ v4.2, only one version of the software can be installed at a time. Executing the `.run` file will remove any existing installed version and (re)install the new version.

After installation, the application and associated files are available at `/opt/edico`.
The single version installer will add `/opt/edico` to the Linux $PATH, so that the user can just call `dragen` without specifying the full path.

New Multi-version Installation

Starting with DRAGEN™ v4.3 and later, multiple compatible versions of the software can be installed at a time. 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 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`

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/

Major Features and Updates

Reference Genome

- Hash Tables v10.
  - Hash tables must be updated to use v4.3. Existing hash tables built with v4.2 or older are not supported.
  - Improvements to the multigenome reference, expanded population, and changes to allow personalization, resulted in interface changes to the reference hash tables. The hash table interface is updated to version 10 (HTv10).
  - 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 be built in two ways:
    - On server using the hash table builder and the Multigenome Reference Builder Collection v4 input files, or
    - On BaseSpace using the DRAGEN Multigenome Reference Builder
  - See the User Guide for details on how to prepare your own reference genome.
  - See Table1 and Table 2 below for recommended usage.
- The multigenome reference is now extended from 32 to 128 samples.
  - Covers 26 different ancestries around the globe.
  - Capture more genetic diversity, reduce ancestry bias and improve SNV accuracy.
  - Reduce SNV SNP FP+FN by 35%, and Indel FP+FN by 12%, on average across population.
  - Improved multigenome hash table format, allows scaling to large population panels.
  - These reference updates are denoted as reference v4.

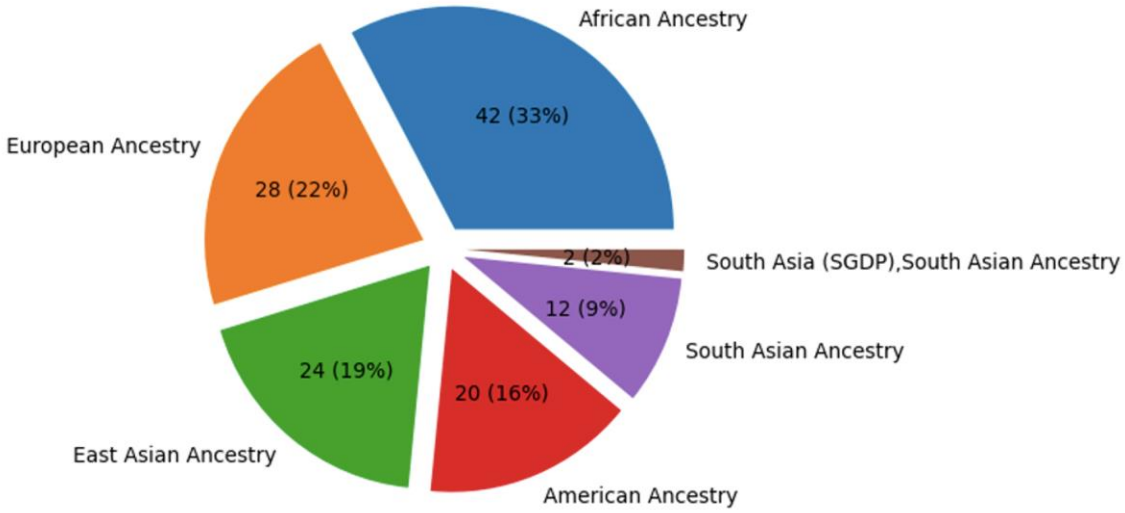


Table 1 v4.3 Reference Support and Recommended Use for Human Data

Human		hg19	hs37d5	hg38	chm13	Recommended Reference Type
Germline	SNV	Yes	Yes	Yes	Yes	Graph
	CNV	Yes	Yes	Yes	Yes*	Graph
	SV	Yes	Yes	Yes	Yes*	Graph
	Expansion Hunter	Yes	Yes	Yes	No	Graph
	Targeted Callers	Yes	Yes	Yes	No	Graph

	RNA	Yes	Yes	Yes	Yes*	Non-Graph
	De Novo	Yes	Yes	Yes	Yes*	Graph
	Joint Genotyping	Yes	Yes	Yes	Yes*	Graph
	Biomarkers (HLA)	Yes	Yes	Yes	Yes*	Graph
	Gvcf Genotyper	Yes	Yes	Yes	Yes*	Graph
Somatic	SNV	Yes	Yes	Yes	Yes*	Non-Graph
	UMI SNV	Yes	Yes	Yes	Yes*	Non-Graph
	CNV	Yes	Yes	Yes	Yes*	Non-Graph
	SV	Yes	Yes	Yes	Yes*	Non-Graph
Methylation	Methylation	Yes	Yes	Yes	No	Non-Graph
Annotation	Nirvana	Yes	Yes	Yes	No	n/a

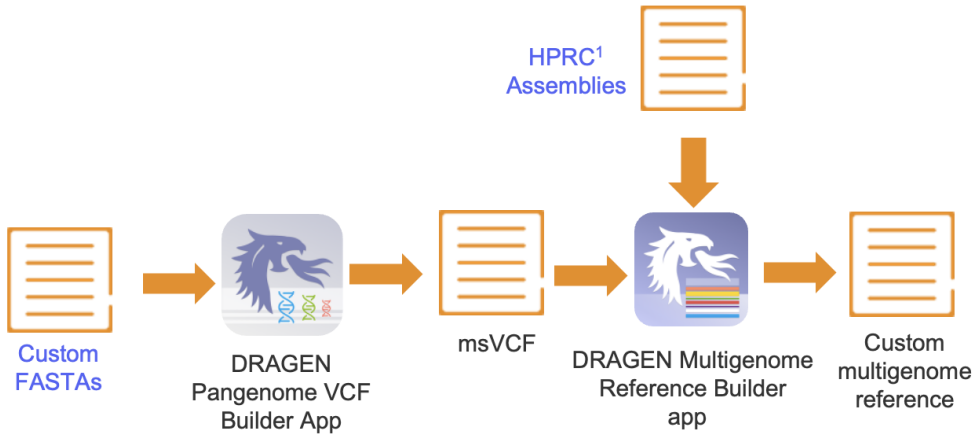
(*) DRAGEN™ supports the component execution; however, the component's accuracy has not been established.

Table 2 v4.3 Reference Support and Recommended Use for Non-Human Data

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

Build Custom Multigenome References

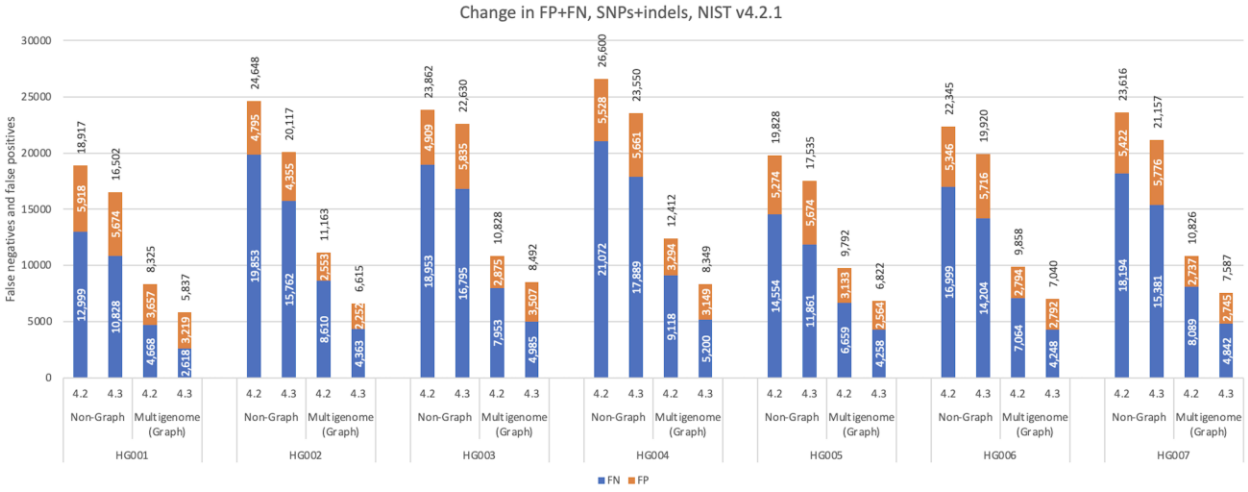
- DRAGEN™ v4.3 provides tools on BaseSpace to build custom multigenome (Graph) references tailored to your population.
  - Users can build high-quality multigenome references from haplotype-resolved assemblies.
  - Reduces ethnicity bias and allows customization for population-specific studies.
  - Support assemblies from the HPRC (Human Pangenome Reference Consortium) as input.
  - Two BaseSpace applications are available that allows for graph customization:
    - **`DRAGEN Pangenome VCF Builder`**
      - Generate phased msVCFs from custom assemblies (FASTA).
    - **`DRAGEN Multigenome Reference Builder`**
      - Select prebuilt datasets from HPRC.
      - Add custom datasets into the reference.
    - See https://basespace.illumina.com/apps/



1.Liao W.-W. et al. A draft pangenome reference. Nature 617, 312–324 (2023).
Blue color text indicates customer inputs, custom haplotype-resolved assemblies FASTA, UI input for HPRC

Germline Small Variant Caller

- v4.3 offers a significant accuracy improvement compared to prior versions:
  - Reduce SNV FP+FN by 35% on average cross population.
  - Reduce indel FP+FN by 12.1% on average cross population.
- Small variant calling accuracy improvements in v4.3 are achieved through:
  - ML model updates
  - Enhanced multigenome reference v4
  - Population aware multigenome mapper

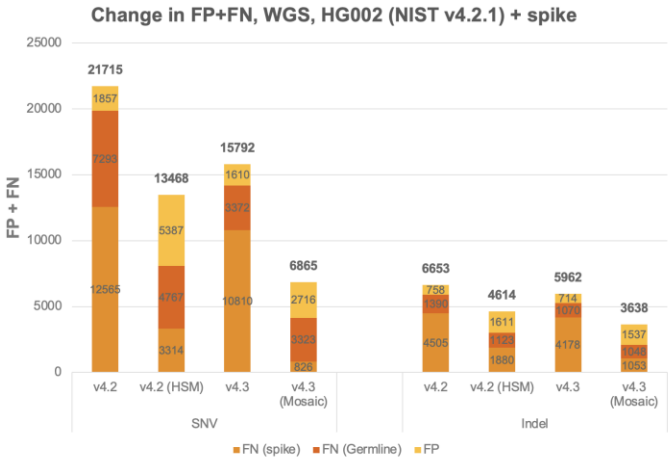


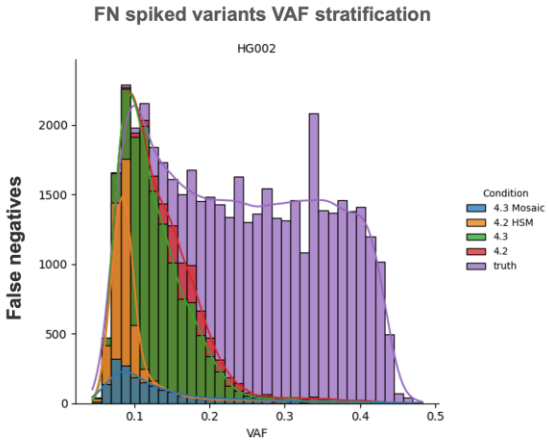
- Mosaic Variant Caller
  - Mosaicism is a postzygotic mutation that leads to some cells in the body having different DNA than others. Mosaic variants occur at low allele frequencies, making them hard to detect. To study the effects of mosaicism on biology and disease, mosaic variants need to be mapped at high sensitivity and high depth and is compute intensive.
  - v4.3 introduces an integrated mosaic caller.

- Small VC reports mosaic variants with >20% allele frequency, by default.
- Optionally, low-AF Mosaic Detection Mode can be enabled with custom defined thresholds.
- Enabled for both WGS and Enrichment analyses.
- Usage:
  - `--vc-enable-mosaic-detection` Enable/disable mosaic detection. The default is true with mosaic AF filter threshold set to `0.2`.
  - `--vc-mosaic-af-filter-threshold` Set the allele frequency threshold for the application of the `MosaicLowAF` filter to mosaic calls. All `MOSAIC` tagged variants with `AF` smaller than the `AF` threshold are filtered with the `MosaicLowAF` filter. The default mosaic `AF` filter threshold is set to `0.2` when the germline variant caller is enabled. The AF default threshold is set to `0.0` when the mosaic detection mode is enabled with `--vc-enable-mosaic-detection=true`
  - `--vc-mosaic-qual-filter-threshold` Set the `QUAL` threshold for the application of the `MosaicHardQUAL` filter to mosaic calls. All `MOSAIC` tagged variants with `QUAL` smaller than the threshold `QUAL` are filtered with the `MosaicHardQUAL` filter. The default mosaic `QUAL` filter threshold is set to `3.0`.
  - `--vc-mosaic-target-bed` Optional target BED file to restrict the output of `MOSAIC` tagged variant calls only in the specified regions
- The mosaic variant caller replaces the High Sensitivity Mode (HSM) introduced in v4.2. See the table below for a comparison of settings and functionality.

Version	Mode	Command line options	QUAL threshold	MAPQ0 included	Mosaic Detection	Mosaic AF filter threshold
4.2	Default Small Variant Caller	`--enable-variant-caller=true`	3	No	No	N/A
4.2	High Sensitivity Mode Enabled	`--enable-variant-caller=true --vc-enable-high-sensitivity-mode=true`	0.4	Yes	Yes (Alpha)	N/A
4.3	Default Small Variant Caller	`--enable-variant-caller=true`	3	Yes	Yes (Full)	20%
4.3	Mosaic Detection Mode	`--enable-variant-caller=true --vc-enable-mosaic-detection=true`	0.4	Yes	Yes (Full)	0%

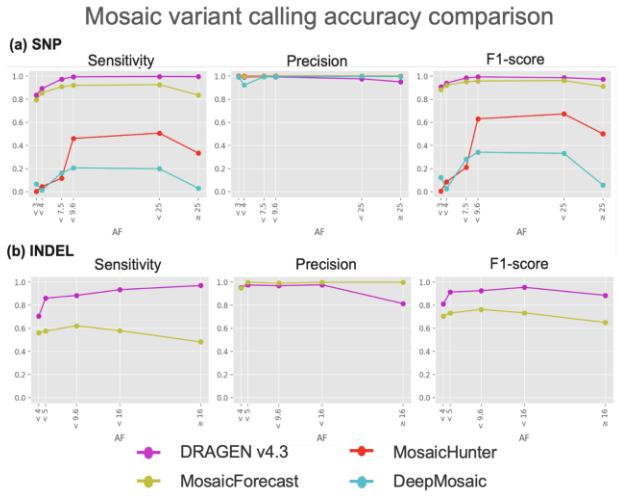
- Higher sensitivity for variant calls at low AF, with low added FPs.





* Tested with DRAGEN v4.2 on NIST HG002 dataset, with 3,936,174 variants including ~50k spiked variants

- DRAGEN™ is more accurate than other mosaic callers with faster turnaround time.

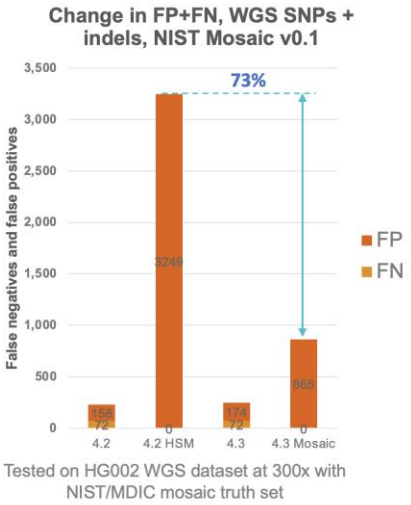


Mosaic variant calling runtime comparison

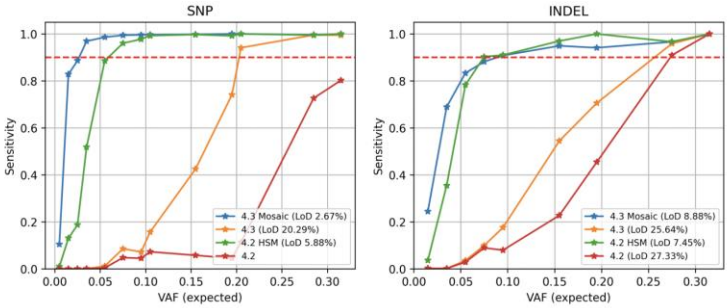
	Runtime (hr)	Hardware
DRAGEN v4.3	0.3	DRAGEN Server v4
DeepMosaic	5.5	Multi-core CPU + NVIDIA A100 GPU
MosaicForecast	12.8	Multi-core CPU

* DRAGEN runtime includes map/align. Other runtimes were for mosaic variant calling only.

- Greatly improved precision for high-depth samples with fast turnaround time



Tested on HG002 WGS dataset at 300x with NIST/MDIC mosaic truth set



Limit of Detection at 2.67% AF for SNPs and 8.8% AF for Indels on 1100x WES admixtures*.

	Runtime
300x WGS	6 hours
1100x WES	20 minutes

*Ha, YJ., Kang, S., Kim, J. *et al.* Comprehensive benchmarking and guidelines of mosaic variant calling strategies. ***Nat Methods* 20**, 2058–2067 (2023).
Ha, YJ., Oh, M.J., Kim, J. et al. Establishment of reference standards for multifaceted mosaic variant analysis. ***Sci Data* 9**, *35 (2022).*

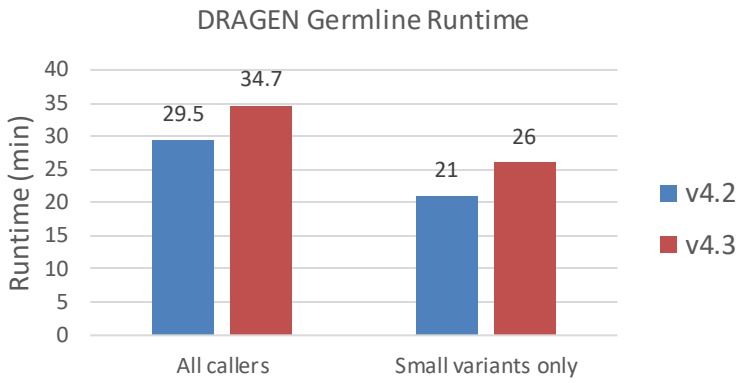
- Personalized Reference (Beta)
  - The feature builds a 2-haplotype personalized reference to impute variants, which is used as priors in the Variant Caller, and creates a new personalized ML model.
  - Reduces FP+FN by ~ 20% for SNPs, ~ 7% for INDELs.
  - Supports both WGS and WES analyses.
  - Usage:
    - `--enable-personalization=true` (default to false)
  - Note: Enabling this feature causes increased runtime for WGS and WES pipelines. Run time with on-site server on 30x WGS approx. 45 minutes (small VC only)

Mapper/Aligner

- New population aware multigenome mapper to implement the advancements in the multigenome reference.
- New graph alignment tag `ga:Z`
  - `--generate-ga-tags` (true by default for DNA, false for RNA) when applicable. This tag is used to describe the best alt contig alignment which improved the score of a primary-contig alignment at its liftover position. It can also be used to describe read alignments to alt contigs for which there is no liftover and the primary alignment is unmapped.
- Increase the sampling size for mapper insert-length calculations from 100K to 2M read pairs, to decrease sensitivity to variability and local drop out on flow cells. It improves insert-length concordance between BAM vs FASTQ input.

Run Time Updates

- Approx. 15-20% runtime increase for on-premises systems.
- Accuracy improvements and more comprehensive outputs result in an increased compute complexity across map/align, machine learning, and SNV variant calling, leading to increased run times for this release.
- The average runtime for HG001-HG007 (NIST v4.2.1) is shown below, compared to v4.2.



Multi Region Joint Detection (MRJD) Caller

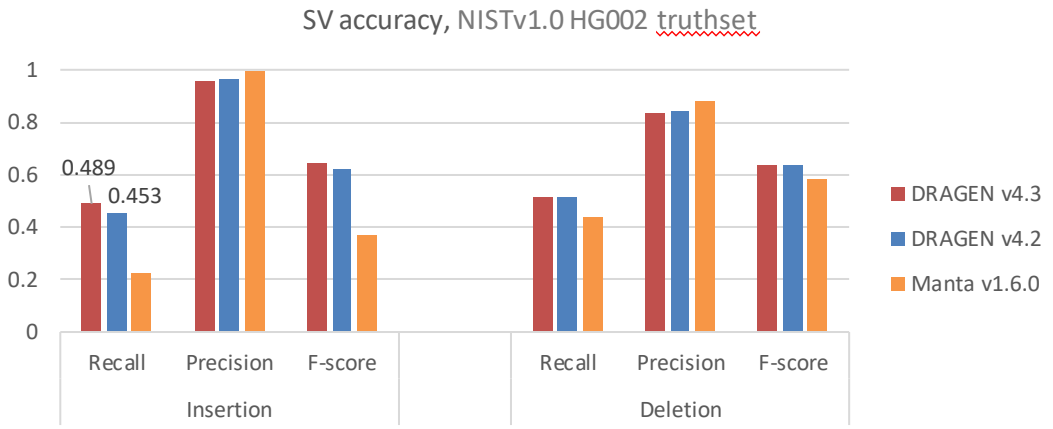
- Segmental duplication regions represent 5% of the genome and have poor mappability. MRJD implements a haplotype-based de novo small variant calling from collected reads potentially mapped to paralogous regions.
- Runs as standalone pipeline on DRAGEN™ server (not in integrated with Germline Small VC)
- MRJD is compatible with the hg38, hg19 and GRCh37 reference genomes.
- Offers a high sensitivity mode option.
- Genotyping is done for 6 medically relevant genes in those regions:

Gene	Application | Conditions
PMS2	Hereditary Cancer Screening | e.g. Lynch Syndrome for Colorectal/Endometrial Cancer
SMN1 (small variants)	Carrier Screening | Spinal Muscular Atrophy
STRC	Carrier Screening | Nonsyndromic hearing loss
NEB	Carrier Screening | Nemaline myopathy
TTN	Newborn Screening & Rare Diseases | Cardiomyopathy
IKBKG	Newborn Screening | Incontinentia pigmenti, hypohidrotic ectodermal dysplasia

- Usage:
  - `--enable-mrjd` Enable MRJD caller. Default is disabled.
  - `--mrjd-enable-high-sensitivity-mode` Enable MRJD high sensitivity mode. Default is disabled.
  - MRJD cannot run together with the Small Variant Caller in this version. The recommended usage is to run DNA Mapping and Small Variant Calling workflow first, and then run MRJD using the aligned BAM file generated from DNA Mapping workflow as input.
  - See the User Guide for a complete description of the methods and usage.

SV Caller

- Accuracy Improvements
  - Robust assembly methods in repeat regions improves insertion recall.
  - SV retains higher recall in Mobile Element Insertions (MEI) when compared to dedicated methods.
  - Additional assembler compute adds minimal runtime compared to v4.2
- Significant reduction in runtime with CRAM library updates
- Systematic noise BEDPE files for SV somatic mode can now be generated for the SV caller for a specific library prep, sequencing system and panels – following a similar process as SNV somatic noise building.
  - Run DRAGEN™ somatic tumor-only on normal samples with `--sv-detect-systematic-noise` set to true to generate VCF output per normal sample.
  - Build the BEDPE file using the VCFs and the `--sv-build-systematic-noise-vcfs-list`: List of input VCFs from previous step. Enter one VCF per line.



CNV Caller

- Detect copy number variations in segmental duplication regions.
  - DRAGEN™ v4.3 rescues ~1Mbp of CNV bins previously excluded from analysis.
  - Improves CNV detection across 43 clinically relevant genes.
  - Enabled by default for germline WGS analysis with the hg38 reference.

Gene	No. samples	Correct by DRAGEN™	Method & Citation
CYP2A6	20	20 (100%)	getRM Pratt et al. 2016
RHD/RHCE	40	38 + 1* (97.5%)	Molecular Inversion Probes (MIP) Nuttle et al. 2013
FCGR3A/B	40	39 (97.5%)	TaqMan qPCR Qi et al. 2016

- Enhanced CNV calling using WES data.
  - New filters reduce false positive rates:
    - Likelihood Ratio (`FORMAT: LR`) – Log10 likelihood ratio of ALT to REF
    - Dinucleotide Biases (`FORMAT: GC/CT/AC`) – Measure of dinucleotide biases that are outside of typical ranges will be filtered out.

    - Panel of Normals (PoN) updates
        - Improved PoN validation. Cross checking of critical options against the case sample under analysis to ensure matched parameters.
        - PoN Metrics. Additional panel of normals statistics calculated per target interval.
- Allele-specific copy number (ASCN) calling in somatic WES analysis.
    - ASCN analysis extends the utility of copy number alteration by enabling the detection of loss-of-heterozygosity (LOH)
    - New for tumor-only analysis. Also available for tumor-normal analysis.
    - Enabling purity estimation, reporting of LOH regions, and HRD Scoring
    - Requires panel-of-normals for exome analysis.

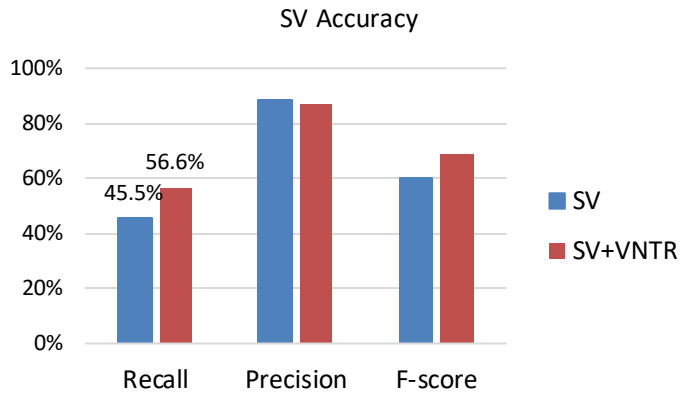
DUX4 Caller

- New DUX4 caller in somatic tumor only WGS analysis.
- DUX4 gene (DUX4-r) rearrangements are involved in a subtype of acute lymphoblastic leukemia (ALL). The caller identifies the events of potential structural rearrangements between DUX4 and other genes (including IGH).
- DUX4 rearrangement detection is enabled by machine learning.
- Supported for the hg38 reference. Run in parallel other analysis with minimal overhead.
- Demonstrated high sensitivity and specificity:
    - 100% Recall (52 cases of `IGH::DUX4` fusions and 1 case of `IGH::QSOX1::DUX4` fusion*)
    - 100% Precision (FP = 0)
- Usage:
    - `--enable-dux4-caller=true` Enable DUX4 caller. Disabled by default.
    - The output contains positive calls that represent translocation events across gene pairs. Example VCF output:

```
chr3    75667931        ExtendedDUX4:IGH:Bnd_W  T       T[chr14:105586938[
    .       .       SVTYPE=BND;CIPOS=.,.;EVENTTYPE=TRA;MATEID=ExtendedDUX4:I
GH:Bnd_X;TotalReadsNum=9061     GT:PR:SR:SRPB   1/1:1:0:110363.09
chr3    75667932        ExtendedDUX4:IGH:Bnd_V  G       ]chr14:105586937]G
    .       .       SVTYPE=BND;CIPOS=.,.;EVENTTYPE=TRA;MATEID=ExtendedDUX4:I
GH:Bnd_U;TotalReadsNum=9061     GT:PR:SR:SRPB   1/1:1:0:110363.09
```

Variable Number Tandem Repeat (VNTR) Caller

- New VNTR caller detects expansions and contractions in tandem repeat (TR) regions. For specified TR regions in the genome, the VNTR Caller estimates the size of the haplotypes in each region and provides variant calls, including the number of copies of the repeat for the sample in question.
- The VNTR Caller only considers TR regions included in a pre-specified VNTR catalog file.
- Supports hg38, hg19, GRCh37 references with curated VNTR catalog included.
- Automatically integrates calls into SV VCF, improving SV recall by > 10%.



- Usage:
  - VNTR caller requires whole-genome sequencing (WGS) data aligned to a human reference genome with at least 30x coverage.
  - Can run directly from FASTQ input with the mapper, or from pre-aligned BAM/CRAM input. Can enable the VNTR caller in parallel with any other germline variant callers as part of a WGS germline analysis workflow.
  - `--enable-vntr=true` Enable VNTR calling. Disabled by default.
  - The VNTR caller reports calls following the VCFv4.4 spec:
    - `chr1 2141862 vntr_00000216 A <CNV:TR>,<CNV:TR> 0.946 PASS SVTYPE=CNV;EVENTTYPE=VNTR,VNTR;SVLEN=634,634;END=2142496;REFRUC=15.1;RUS=CAGCCGAAGCGCCTCCTTTCACTCCAGGGTCCACACATCCAG;RUL=42;RB=508,549;CN=0.801,0.866;CNVTRLEN=-126,-85;RUC=12.1,13.1;RUCCHANGE=-3.0,-2.0 GT:SVFT:VQ:CN:TOTALRB:TOTALCNVTRLEN:TOTALRUC 1/2:PASS:0.946:1.667:1057:-211:25.2`
    - `REFRUC`: reference copy number
    - `RUC`: total copy number found per haplotype
    - `RB`: total length of each haplotype

De Novo STR Expansion Detection

- Short tandem repeats (STRs) are regions of the genome consisting of repetitions of short DNA segments called repeat units. STRs can expand to lengths beyond the normal range and cause mutations called repeat expansions. Repeat expansions are responsible for many diseases, including Fragile X syndrome, amyotrophic lateral sclerosis, and Huntington's disease.
- The ExpansionHunter de novo algorithm is implemented in DRAGEN™ for STR detection. The caller allows the discovery of expanded STR regions from paired end reads across a cohort of samples. It is designed to work with PCR-free samples of 100-200bp paired-end reads at >30X coverage.
- Usage:
  - Cohort analysis is done in two stages:
  - Compute per-sample profiles while mapping: `--enable-str-profiler=true` while separating each into cases and controls output folders.
  - Compare cohorts of profiled samples:

```
--enable-str-profiler=true
--str-profiler-analysis=<outlier|casecontrol>
--str-profiler-cases-directory=<directory_with_cases_profiles>
--str-profiler-controls-directory=<directory_with_controls_profiles>
```

  - The caller outputs motif and locus tables, where STR expansions can be identified.

Targeted Callers

- New unified VCF output for three targeted callers HBA, GBA and CYP21A2
  - Records can be `PASS` even with low QUAL and low GQ values.
  - These variants require further assay if placement in target gene is required.
  - VCF format consistent with MRJD region-ambiguous variants.
  - HGVS identifiers reported in `ALLELE_ID` INFO field.
  - Reported genotypes are consistent with any overlapping SV DELs detected by the caller.
  - Polyploid genotypes and associated quality scores for "joint" analysis of homologous sites are reported in separate VCF fields (e.g. `JGT`, `JGQ`, `JPL`).
- HBA structural variant in VCF output
  - Forced genotyping of all 5 supported structural variants:
    - `-α3.7` DEL and `ααα3.7` DUP
    - `-α4.2` DEL and `ααα4.2` DUP
    - `--SEA` DEL
  - `ALLELE_ID` INFO field used to label variant alleles.
  - `SVModelQual` FILTER applied when data does not fit any of the supported genotypes.
  - Small variant calling for these 3 genes are enabled by small VC by default, even if targeted callers are not explicitly enabled. Only for WGS.
- Small variant VCF output for recombinant haplotypes:
  - Enabled for GBA and CYP21A2.
  - Gene conversions reported as small variants in target gene rather than SV breakends.
  - Deletion-like recombinant haplotypes reported as small variants in target gene with overlapping deletion alleles in pseudogene rather than SV DEL.

Somatic VC

- The columnwise detection feature is now enabled by default, improving small variant recalls in repetitive regions.
  - Repetitive local sequence leads to cycles in graph assembly of reads and may result in missed variants.
  - Columnwise detection identifies variants directly from the BAM pileups to supplement the graph.
- Somatic hotspots are now allele-specific, reducing FP impact of the hotspots feature.
- Skip the multiallelic filtering and clustered event penalty for variants in somatic hotspots.
- New option `--vc-output-variant-read-position` to enable outputting the variant read position in the INFO field.
- Somatic WGS HEME Tumor support
  - SNV, SV and CNV callers support easy configurations for Heme tumor only analysis
  - HEME specific systematic noise files have been added for download
  - `--heme-cnv=true` Configures CNV settings for Liquid Tumors (e.g., AML/MLL).
  - `--heme-sv=true` Configures SV settings for Liquid Tumors (e.g., AML/MLL).
  - See the "DRAGEN Recipes" section in the User Guide for HEME Tumor Only analysis

RNA

- Significantly improved RNA fusion detection accuracy via extensively trained machine learning model
  - Uses an updated non-linear model with more features.

- Significantly enhanced truth set, using simulations and real data.
- Additional datasets across library preps (WTS, mRNA, Panels), and read lengths, tissue types, sample types, etc.
- Pure score-based filtering enables better recall/precision tradeoffs.
- Improved precision and recall across entire truth set.

Attribute	Prior Model (pre v4.3)	New Model (v4.3)
Model	Linear	Non-linear
Features Used	8	39
Truth Description	Gene-pair only	Genes + breakpts
Total Truth Set Size	223	>12,000
PASS/FAIL Filters	Score + 5 post-filters	Score only

RNA Splice Variant Caller

- New RNA Splice Variant Caller (Beta)
- Enables detection of alternative splicing or intra-genic fusions
- Targets well-known splice variant biomarker events:
  - EGFR VIII
  - MET exon14
  - AR-V
- General (non-targeted) option available with increased false positives
- Usage:
  - Required resource input files:
    - Anchored-RNA reference HT (for mapper). Available as part of pre-built references.
    - Gene Annotation file (i.e. GTF or GFF)
  - Optional recommended new resource input files:
    - `rna-splice-variant-knowns` Knowns SJ file ("allowed list")
    - `rna-splice-variant-normals` Normals SJ file ("blocked list")
    - `rna-splice-variant-regions` BED file for making calls
  - `--enable-rna-splice-variant=true` When added to an RNA Map/Align job, the software runs a Splice Variant caller by taking advantage of its fast and highly accurate splice-aware read mapper/aligner that aligns to the whole genome to identify novel alternative Splice Junction (SJ) candidates.

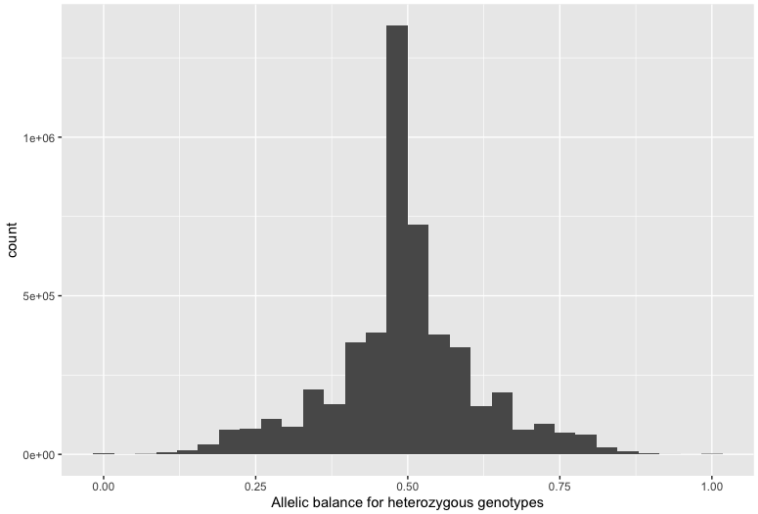
Gvcf Genotyper for Cohort Analysis

- Improved multi-sample VCF (msVCF) output.
  - New options to compress and customize multi-sample VCF (msVCF) output.
    - Sparse compression of msVCF output (*Michael Lin, 2020*), yields 20-30x size reduction for large cohorts.

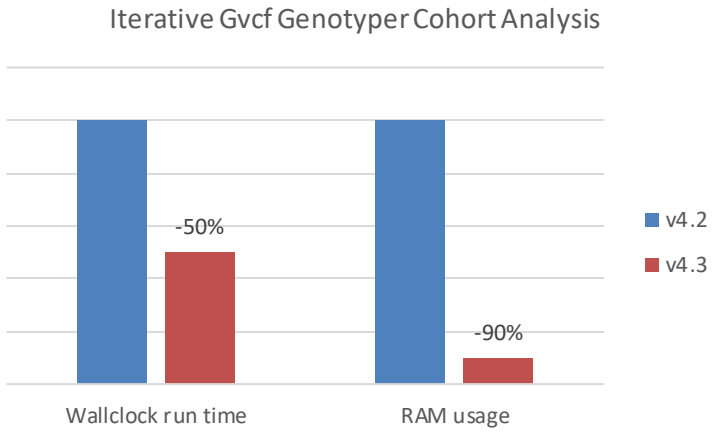
      ```
      --gg-output-type=spVCF
      --gg-squeeze-msvcf=true
      ```
    - Easy customization of msVCF output, easier ingestion into third-party tools
      - How to customize msVCF output:

        ```
        --gg-msvcf-info-fields=AC;AN;NS;NS_GT;NS_NOGT
        --gg-msvcf-format-fields=GT:LAD:LPL:LAA:QL
        ```
- New msVCF metrics enable filtering for highly accurate call sets.
  - Allelic balance can identify incorrect Genotype calls.
  - Genotypes with unusual allelic balance are candidates for filtering.

- To enable output of allelic balance in the msVCF:
  `--gg-msvcf-info-fields=ABHom;ABHet;ABHetP`
- To filter msVCF based on maximum P-value of allelic balance:
  `--gg-hard-filter="ABfilter:all:ABHetP < 0.05"`



- Performance Improvements
  - The Iterative Gvcf Genotyper has been successfully used for processing multiple extra-large cohorts. Specifically, a cohort of 500k WGS samples has been processed with the software on the Illumina Connected Analytics (ICA) platform. The analysis resulted in 874K jobs, aggregated 1.4 billion variants, and completed in less than 3 months.
  - This effort led to significant run time, robustness and feature improvements.



- For further details, please refer to the User Guide, section "Joint Analysis" and "Population Mode".

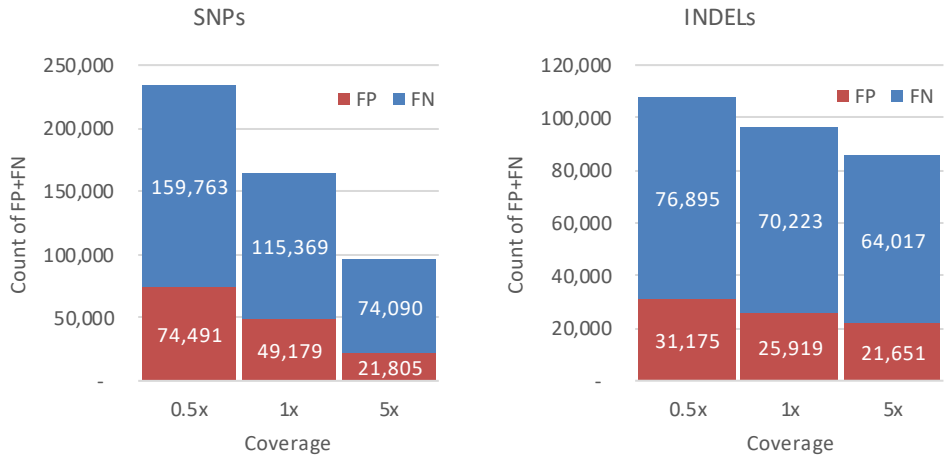
Imputation and Population Haplotyping

- Population Haplotyping pipeline on ICA
  - Builds haplotypes from a set of population samples.
  - Build a custom reference panel for imputation.

  - Multi-step pipeline integrated into an end-to-end workflow on BSSH and ICA, tuned for optimization of runtime and accuracy, and leverages the multi-node infrastructure of ICA.
- Imputation pipeline on BSSH and ICA
  - Infers variants of low pass sequencing data.
  - Supports up to 100 input FASTQ samples or 1000 input VCF samples.
  - New reference panel (IRPv2.1, human hg38) improves indel imputation accuracy.
    - ~14000 fewer indel FN on HG002 using IRPv2.1 vs IRPv2.0
  - New option to batch imputation samples for a quicker turnaround time.
  - Estimated cost for imputation pipeline from FASTQ (incl. FGT): 5 iCredits/sample.

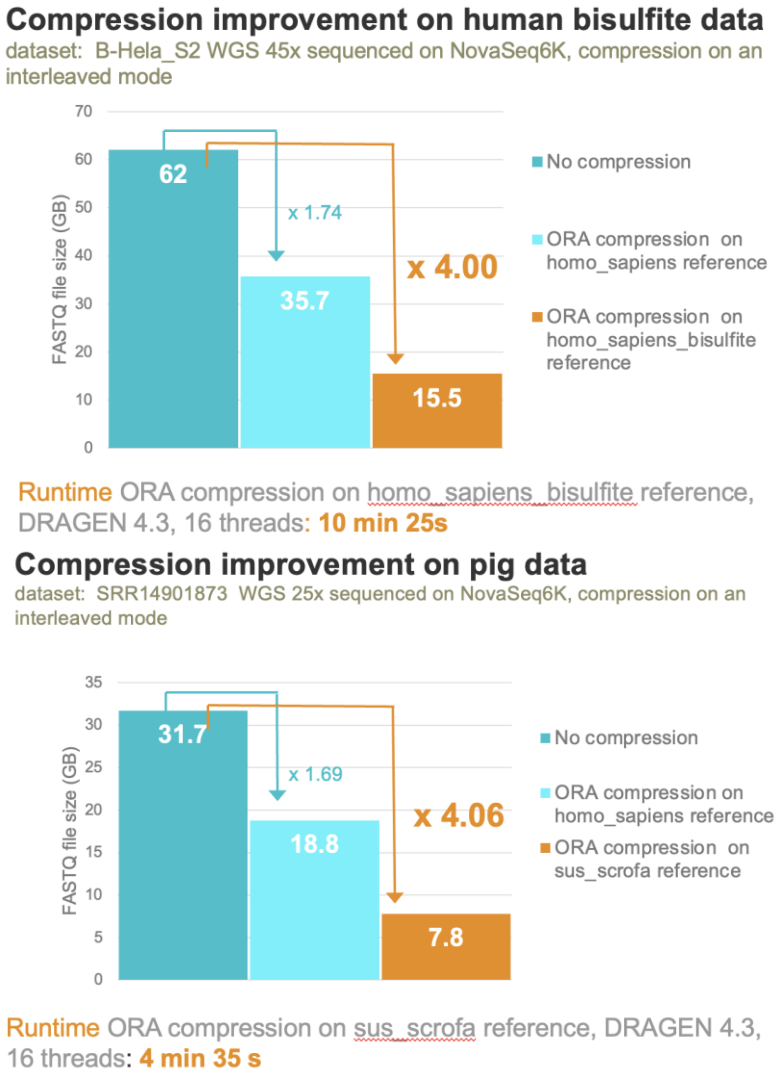
	IRPv2.1	**IRPv2.0**
Data source[1]	3,202 samples	3,202 samples
Multi-allelic SNP positions	Yes	Yes
INDELs	Yes – all INDELs	Yes - AF>3%
ChrX	Yes	Yes
Total Number of Variants	125,715,255	111,279,429

(1) Data from 1000 Genome Project, processed with Gvcf Genotyper



ORA Compression

- ORA Compression now supports more datatypes and species.
- Dataset specific compression reduces file size and storage cost by up to 75%.



- List of supported references and associated sting name to use with `--ora-compression-species <string>`. The most up to date list of supported reference can be found at the Illumina DRAGEN™ Product Files support site.

Model	Valid string value	Size
Human	Homo_sapiens	6.5 GB
Human methylated data	Homo_sapiens_bisulfite	11 GB
Pig	Sus_scrofa	5.0 GB
Chicken	Gallus_gallus	3.8 GB
Rice	Oryza_sativa	1.9 GB
Arabidopsis	Arabidopsis_thaliana	478 MB
Wheat	Triticum_aestivum	13 GB
Cattle	Bos_taurus	5.3 GB

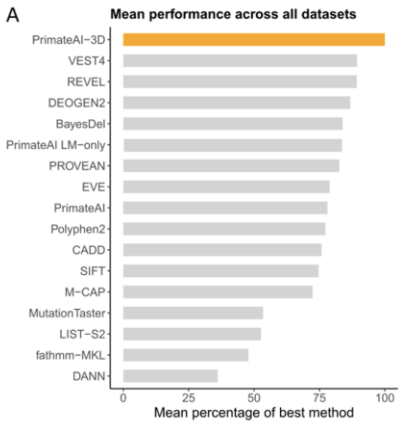
Soybean	Glycine_max	2.0 GB
Rat	Rattus_norvegicus	4.5 GB
Maize	Zea_mays	4.2 GB
Zebrafish	Danio_rerio	4.8 GB
Mouse	Mus_musculus	4.5 GB
Roundworm	Caenorhabditis_elegans	569 MB

- How to use the new references
  - Download the desired ORA reference `tar.gz` files, do not change any of the file names of the downloaded archives.
  - Move the file to the location you would like to contain the reference directory in, and then enter the following to extract the contents: `tar -xzvf <file>`
  - Set the `--ora-reference` command line option to the extracted `/oradata` folder path.
  - At compression, select which reference species to use with option `--ora-compression-species <species_scientific_name>` If unspecified, Homo sapiens reference will be used by default. The output file will always be valid.
  - At decompression, detection of the species used to compress the ORA file is automatic. The software will look for the appropriate species in the oradata folder pointed by `--ora-reference`.

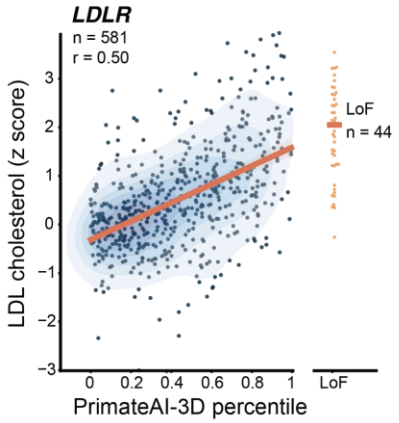
Connected Annotations (Nirvana)

- The latest version 3.23 of the annotation engine incorporates PrimateAI-3D as a key dataset.
- Benefits of annotation with the Primate AI-3D dataset:
  - Leveraging Primate AI-3D reduces variants of unknown significance (VUS) and predict pathogenicity of all protein coding variants.
  - Provides predicted impact on coding sequence and protein sequence following Human Genome Variation Society (HGVS) standards. Provides consequences relevant to each variant using Sequence Ontology standard nomenclature.
  - PrimateAI-3D reclassified >4 million human missense variants of previously unknown consequence as likely benign, resulting in a > 50-fold increase in the number of annotated missense variants compared to existing clinical databases.
  - The pathogenicity of the remaining 94% of variants were computed with deep learning, achieving state-of-the-art accuracy for diagnosing pathogenic variants in patients with genetic disease.
  - Validated to have the top classifier performance in six different benchmarks based on real-world rare and common disease patient cohorts.

PrimateAI-3D achieves top performance across six benchmarks for variant interpretation



PrimateAI-3D accurately predicts phenotypes in 450,000 individuals from the UK Biobank



- Connected Annotations is extremely accurate at an extraordinary speed: 17.29 minutes run time for 6.5 million DNA variants with 99.9983% average accuracy.

- **Important interface changes:**
  - DRAGEN™ v4.3 now includes Annotations v3.23 by default.
  - The Annotations `Downloader` and the `Nirvana` packaged with the installer supports free download and annotations using Primate AI-3D databases.
    - The downloader is present at `/opt/dragen/<version>/share/nirvana/`
  - Annotations v3.23 is not compatible with old Transcript databases.
  - The latest annotation databases must be downloaded to use v4.3 annotations.
  - The TMB caller requires annotation to be enabled and requires the latest annotation databases.

DRAGEN™ version(s)	Annotations version	AI annotations
4.3	3.23	spliceAI, primateAI3D
3.9, 3.10, 4.0, 4.1, 4.2	3.16.1	spliceAI, primateAI
3.8	3.14	spliceAI, primateAI
3.6, 3.7	3.9.0	spliceAI, primateAI
3.5	3.6.0	spliceAI, primateAI

Downsampling

- A new Fractional Downsampler method is available.
  - Fully integrated workflow, enable with any downstream secondary analysis
  - No additional run time
  - Subsampling is based on user-defined percentage of reads
  - Subsampling is applied to raw input reads pre-alignment, with no modification (no trimming, no filtering, pre-deduplicated)

- Reduce run time and cost of analysis on high-depth samples. Avoid the need to run separate downsampling tools.
- All input formats supported by the software can be used (FASTQ/BAM/CRAM)
- Usage:
  - `--enable-fractional-down-sampler=true` Enable the fractional downsampler.
  - To set percentages of number of reads to keep:
    `--down-sampler-normal-subsample=<fraction>`
    `--down-sampler-tumor-subsample=<fraction>`
    `<fraction>` is the approximate percentage of reads to keep expressed as fraction, e.g. 0.05 = 5%

Precision Metagenomics Pipeline

- Add support for additional panels: "VSPv2" and "Custom"
  - Supported by the option `--explify-test-panel-name`
- Databases for v4.3 have been updated
  - `RPIP-6.1.0`, `UPIP-8.1.0`, `VSPv2-1.2.0`
- New Kmer classifier database builder allows users to build their own DBs for the Kmer classifier.
  - The software can generate a custom indexed, searchable database of reference sequences to be used for Kmer classification
- Classification algorithm enhancements, threshold reporting flexibility, bacterial AMR improvements
- New pipeline options
  - `--explify-sensitivity-threshold` Set sensitivity threshold, applicable to VSPv2 panel
  - `--explify-custom-ref-fasta` Reference Fasta file required for custom reference DBs
  - `--explify-custom-ref-bed` Reference BED file required for custom reference DBs
- Run time and memory use improvements
- Please refer to the User Guide for extensive details.

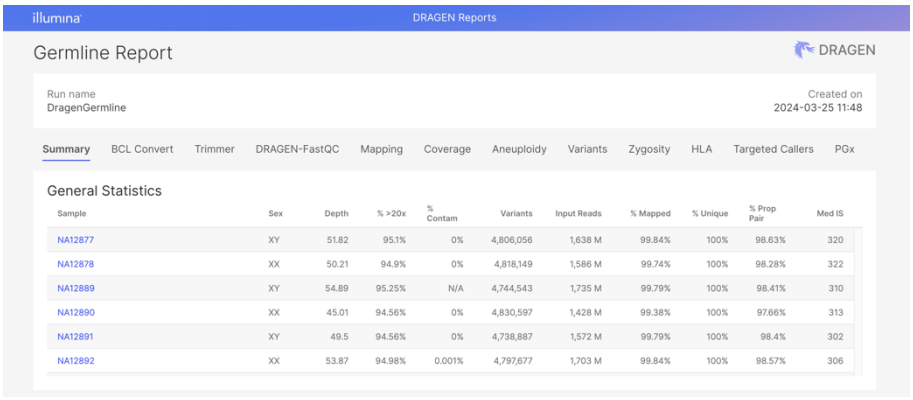
BCL Convert

- Support for `--bcl-enable-tile-metrics` command line option. `true` by default. When set to `false` the software will output only the header and no content for the following files:
  - `Demultiplex_Tile_Stats.csv`
  - `Quality_Tile_Stats.csv`
- Supporting longer index read length from 27 total bases to 27 bases per index.
- Include the following new columns in the `fastq_list.csv` file when present in the sample sheet:
  - `RGID`,`RGSM`,`RGPU`,`RGPL`,`RGLB`,`RGCN`,`RGPM`
  - A custom column of that starts with the substring "`RG`" and has 4 characters.
  - This will allow the tags to propagate to the BAM file which can aid with identification of experiments
- Support for order independent OverrideCycles format
  - Order Dependent: the order of the mask elements must match that of the RunInfo.xml (example: `U7N1Y143;I8;I8;U7N1Y143`)
  - Order Independent: order of the mask elements is flexible
    - Each read must be specified in the read description using "`R`" for genomic or "`I`" for index followed by the read number
    - The read description is ":" separated from the read mask (example: `R1:U7N1Y143;I1:I8;I2:I8;R2:U7N1Y143`)
  - Order independent and order dependent specifications cannot be mixed within the same sample sheet

- Library Rebalancing Stats report output according to the `LibraryInputVolume` setting in the sample sheet to support library and pooling QC on the iSeq 100 system
  - The `LibraryInputVolume` setting must be a real number
  - When the setting is specified, the `LibraryRebalancing_Stats.csv` metrics file will be output with the following columns:
    - Lane
    - SampleID
    - Index
    - Index2
    - # Reads
    - % Reads
    - Rebalancing Factor
    - Rebalancing Input Volume

DRAGEN™ Reports Tool

- DRAGEN™ Reports is a docker image that provides tools for generating rich, interactive and self-contained HTML reports from DRAGEN™ Secondary Analysis output files.
- These reports combine data from QC, trimming, mapping, variant and other modules to create a comprehensive summary of a multi-sample workflow, as well as more detailed reports for individual samples.
- Features:
  - Generate reports without re-running a whole workflow.
  - Mix and match samples from across workflows into one report.
  - Supports new callers in v4.3
  - Drop-in Replacement for FastQC / MultiQC.



- The DRAGEN™ Reports tool is provided as a Docker image that can be downloaded from the Illumina DRAGEN™ Product Files support site. Please refer to the User Guide for comprehensive usage instructions.

Other Updates

- Input multiple BAM/CRAM files for top-up cases
  - New options `--bam-list` and `--cram-list`
  - Provide text file with one column and header "`BamFile`" or "`CramFile`" and each file path on a new line.
  - The intended use is top-up. Note that some callers (i.e. DRAGEN™ small variant calling) are unable to process a bam-list or cram-list that is composed of input files containing multiple samples. In the case where identical read group IDs appear across multiple files and they need to be treated as distinct read groups, use the option `--prepend-filename-to-rgid=true` to distinguish between read groups.
- Software for on-premises server has some changes to support the new multi-version paradigm.
  - These changes are fully transparent to users and shared only for information as it may relate to sysadmin activities.
  - Communication with the FPGA card is still handled by a daemon process called `dragend`
  - `dragend` no longer runs as a Linux system service and is not registered and started automatically when the software is installed.
  - The `dragen` application now launches `dragend` as a process when it starts and ends the process when it ends.
  - `$ dragen -V` stdout only returns the client version of the `dragen` application.
  - `dragen_reset` for cloud is unchanged.
  - `dragen_reset` on-server is still automatic, and no longer need to be called since v4.2. For backwards compatibility with existing user scripts, tt can be called and will return 0 (without error).
- Cloud license credentials via input file
  - Customers using the BYOL cloud deployment can replace the "`--lic-server`" command line with a "`--lic-credentials=<file>`" option. The file helps avoids credentials leaks from command line console logs.
  - The file is a plain text file audited by the customer. The format is the same as other config files: `key = value` separated by new line. The following key names must be used: `credentials1`, `credentials2`
- Support for AWS IMDSv2 has been added. Does not yet support S3 input streaming with IMDSv2 Only.

Known Issues

Known issues of the DRAGEN™ v4.3.6 release

Component	Summary	Resolution/Workaround
Amplicon	Amplicon read target coverage report can be incorrect for tiling amplicons in some cases. Target assignment and accuracy is not impacted.	None.
Amplicon	FN is observed at the edge of amplicon targets due to "vc-remove-all-soft-clips=false" in v4.3. The occurrence is rare and only affect the variants at the edge of targets.	None.
Amplicon	Missing FLT3 insertion variant in SV calling when we have overlapping SV events. Occurrence is rare	None.
BCL	Some combinations of indexing schemas have a significant run time regression compared to DRAGEN v3.7	A fix is not available, but a workaround for on-premises can be done of this specific indexing schema.
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	BCL streaming over network sometimes hangs or takes hours, compared to minutes on local disk, when a large number of files are opened and streamed.	Recent investigations has shown that network file systems may hit performance issue when streaming from large number of file handles, likely inducing severe caching issues on the network filer. A BCL setting exists that can reduce the number of simultaneous over-the-network accesses which has shown to alleviate the problem, with run time impact. The best remedy is to avoid letting the network file system get into a bad caching state, by using run modes with fewer open file handles, or run from local disk.
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 Convert does not validate when "Logs" or "Reports" is provided for a Sample_Project, and the software will be unable to create the subdirectories if these strings are provided	None.
BCL	BCL Convert does not support the "--first-tile-only" option being specified for SP flow cells, but the new "--tiles" option can be used as a substitute	Use the new "—tiles" option
BCL	Does not error when no tile list exists in the RunInfo.xml file and "--tile" or "--exclude-tiles" is specified in the command line	Ensure that tiles are not specified when none exist.
BCL	When an index collision exists in a lane that has been excluded via the "ExcludeTiles" setting, the software will still error as it is a sample sheet validation error	None.

BCL	BCL Convert has differences in legacy stats compared to bcl2fastq2 (all of which are justified)	None. This is per design, please consult the BCL documentation for expected outputs
CNV VC	CNV VAF loci related to wrong purity/ploidy estimate when normal sample has problematic regions.	Rare occurrence. A possible workaround is to disable VAF-aware mode in CNV when needed.
Compression	CRAM decompress and map/align using different references, can falsely run into an alt contig error check, when two conditions are true: hash table used for cram decompression, and both hash tables combined contain alt contigs that exceed a threshold. The software counts contigs from both refs instead of only the map/align ref.	Supply the FASTA file to decompress the CRAM
DNA Alignment, Compression	DRAGEN "BAM to CRAM, back to BAM" is not lossless. There is a CIGAR mismatch for full soft clipped reads. The CRAM format does not store CIGAR and other fields for unmapped reads. This is the implementation per standard, and same in all CRAM tools. In v4.3 with graph ref and ga tags, fully clipped reads will be unmapped. But due to ga tags, the CIGAR is stored in the BAM record. When compressed with CRAM, and back to BAM, those CIGAR fields are lost, and the decompressed BAM mismatch with original BAM.	None.
Downsampling	Target coverage downsampler doesn't hit the correct coverages.	For v4.3 a new fractional downsampler is available to handle this use case better.
Downsampling	Crash during coverage downsampling.	For v4.3 a new fractional downsampler is available to handle this use case better.
GVCF Genotyper	Sample rename feature for iGG does not work.	None.
Hash Table Builder	Hang when building custom graph Hash Table with input msVCF containing mega-base long indels. In such case Hash Table builder runtime can be extremely long or hang completely.	Pre-filter the input msVCF to remove million-bases long indels.
HLA	There is a minor accuracy regression on HLA due to a miscall in the DRB1 gene after changing the reference from v3 graph to v4 graph, on hg38 reference.	None.
HW GRAPH	A rare segfault has been observed on cloud runs due to HW GRAPH error.	The rate of occurrence is < 1/1000 and not repeatable. 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.
QC Metrics	Overlapping mates' calculation does not correctly handle the situation where a supplementary alignment overlaps with a primary alignment. The new mapper HW changes exposed this bug when the alignment of some reads changed. This should be an extremely rare bug limited to situations where a supplementary alignment overlaps the primary read, and where the primary read start and ends within the range of the supplementary.	None. Rare and impact on coverage metrics is very limited

RNA Quantification	Minor differences in RNA quant SJ.saturation.txt with different num-threads value.	Use a consistent number of threads across runs if desired
SNV Germline	There is a risk of overwriting files for the denovo filtering component, when "dn-output-vcf" is not used but "output-dir" is used and set to the path of the input file "dn-input-vcf". It is not doing a check	Do not use "output-dir" for the denovo filtering feature. Follow the use guide instructions for specifying input and output paths
SNV Germline	Missing some expected mosaic variants in certain samples.	None. The mosaic caller is new. Continuous improvement of this new feature will happen in future releases.
SNV Somatic	Accuracy (FP) regression on some FFPE/FF Normal/Normal datasets in somatic due to columnwise detection, as result of large difference in noise profiles between the FFPE tumor and FF normal sample.	There is no major FP regression in any other TN datasets, and accuracy improved across vast majority of samples. For these extreme cases, columnwise detection could be disabled if needed
SNV Somatic	A MNV FP has been identified on the TSO500 assay.	None.
SNV Somatic	Somatic SNV T/O MNV failing to merge two MNV calls in the edge case where we have a deletion upstream of another co-phased variant with and an out-of-phase SNP in between them that is covered by the REF allele of the upstream deletion. In this scenario, we will end up excluding haplotypes based on the haplotype max klen values being less than the distance between 2 variants in the group that are not actually in phase with one another.	None.
SNV Somatic	Somatic T/O WES Indel Sensitivity regression seen in some samples, due to a single new FN introduced by germline tagging with new Nirvana resources including gnomAD v4.0.	None.
SNV Somatic	Minor SNP FP regressions observed on some datasets due to hotspots VCF file updates. The nett effect of the updated hotspots is positive over the majority of test samples.	The FP in question are variants that were newly added to the hotspot file or variants that were originally present but represented in a non-normalized fashion.
SNV Somatic	Bam output runtime increases by 7% for Tumor Only mode	None.
SNV Somatic	3 new SNP FP introduced in one test sample due to updates to the somatic hotspots file to include a few thousand additional heme-specific hotspots.	Future updates to the hotspots file may resolve the FPs. Accuracy changes are expected in 4.3. While most of the accuracy metrics improve significantly across the majority of datasets, there are isolated cases of minor regressions on some.
SNV Somatic	Multiple FGT tags attached to 1 forcegt call, in the scenario where a variant is present multiple times in the input vc-forcegt-vcf file, once as a single variant and again as part of multiallelic records. In this scenario, DRAGEN will output a forced call with the INFO.FGT tag applied twice.	None.
SNV Somatic	FP regression observed on a NormalNormal HCC1395BL PCRFree on 4.3 compared to 4.2	None. Accuracy changes are expected in 4.3. While most of the accuracy metrics improve significantly across the

		majority of datasets, there are isolated cases of minor regressions on some.
SNV Somatic	Most somatic SNV datasets have >10% increase in total runtime relative to previous release v4.2.	New features, including updated mapper, columnwise detection and allele-specific systematic noise and hotspots adds extra compute complexity and are expected to increase runtime.
SV	AWS S3 uplink streaming is not functional for many use cases of DRAGEN.	Recommended to use downlink S3 streaming only
TMB	Tumor only TMB is not as reliable as T/N TMB.	None.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package.
- The archive integrity can be checked using: `./<DRAGEN 4.3.6 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.3.6 .run file>`

Release History

Revision	Release Reference	Originator	Description of Change
00	1108491	Cobus De Beer	Initial release