DRAGEN v4.4.11 Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Secondary Analysis Software v4.4.11.

Changes are relative to DRAGEN™ v4.4.4. If you are upgrading from a version prior to DRAGEN™ v4.4, 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-4.4.11-12.multi.el8.x86_64.run

The following configurations containing DRAGEN™ 4.4.11 are also available on request:

- AlmaLinux 8 Amazon Machine Images (AMIs) for F2 instances, available in 12 regions
- el8 compatible RPM packages for use with Amazon Web Services (AWS) F instances, for customer generated AMIs or customer generated docker images
- DRAGEN™ Kernel drivers for el8, for use with customer generated AMIs or QuickStart

End-of-Life announcements:

- AWS F1 instances have been phased out of service and made obsolete on December 20, 2025
- F2-compatible DRAGEN versions of BSSH and ICA apps have been made available for a subset of current versions.
- Prior DRAGEN versions for use on-prem and BYOL have been obsoleted. Upgrade pathways and EOL information are summarized on the DRAGEN™ help pages here: https://help.dragen.illumina.com/reference/eol-transition

Deprecated platforms:

- AWS F1 instance types have been deprecated on December 20, 2025

Contents

DRAGEN v4.4.11 Updates 4
Overview 4
Changes to support NovaSeqX on instrument analysis 4
DRAGEN Bug Fixes 4
DRAGEN v4.4.7 Updates 4
Overview 5
DRAGEN Bug Fixes 5
BCL Convert Changes 5
DRAGEN v4.4.6 Updates 8
Overview 8
Amplicon and Pillar Panel Support 8
5-base Methylation Support 9
New 16S Metagenomics Pipeline 10
SNV VC 11
CNV, CNV Cytogenetics and CNV+SV 11
Explify Analysis Pipeline 13
RNA Gene Fusion 13
Illumina PIPseq Single Cell RNA (scRNA) 14
MRD (Minimal Residual Disease) Pipeline 15
MSI (Microsatellite Instability) 15
BCL 15
Compression 15
Other Bioinformatics fixes 16
Other Platform fixes 17
Resource Files 19
Reference Genome Recommendations 21
Multi-Version Installer for on-premises servers 23
Known Issues 25
SW Installation Procedure 30
Release History 30

DRAGEN v4.4.11 Updates

Overview

DRAGEN™ v4.4.11 is minor patch release that contain bug fixes for customer issues and minor changes to support NovaX on-instrument deployment. This version is available for on-premises dragen server installation, AWS AMI BYOL and NovaSeqX on-instrument and cloud applications.

Changes to support NovaSeqX on instrument analysis

Component	Description
On Instrument Scheduling Software	- Reverts a demultiplexing parameter which causes the Top_Unknown_Barcodes.csv file to be large. - Enabling high watermark recording in all modes - Changing the polling time for high watermark recording to every 2 seconds - Changing logged names of tasks (used with high watermark) - Change non-BCLConvert dependency graphs to handle specific failure cases (does not affect working case) - Contains change to not overwrite early metric CSV files and IndexMetricsOut.bin file during final merge.

DRAGEN Bug Fixes

Bug fixes generally applicable to dragen core.

Component	Description
Map/align	Fix for a rare intermittent issue where the BAM is missing one read mate. Observed on dragen v4.3 and v4.4
CNV	Disable germline WES ASCN mode. WES ASCN analysis (with target BED and PON) is not supported on v4.4 but the dragen options validation did not block it. This change now blocks this mode.
CNV/SV	Fix error/crash during SV/CNV integration step on certain samples. SV/CNV merge aborted due to a negative-length CNV. Added a condition preventing negative-length CNVs
CNV	Fix some metrics discrepancies in the deletion (DEL) and duplication (DUP) counts when the ASCN caller is enabled. Do not count predefined segments in CNV metrics
CNV	Adding CNV WES ASCN soft failure to generate an empty output, when input data generates 0 counts for most of the regions. Fixes a hang or segfault on invalid data.

DRAGEN v4.4.7 Updates

Overview

DRAGEN™ v4.4.7 is minor patch release that contain bug fixes for customer issues and introduces a new functionality for BCL convert to auto detect indexes and samples.

DRAGEN Bug Fixes

Component	Description
CNV	Allow Enrichment Germline CNV runs with PON to succeed when there are zero input reads, instead of asserting. With zero input reads, print a warning that GC bias correction can't be done, and write an empty VCF file.
MRJD	Fix for incorrect/corrupted VC metrics when SNV caller and MRJD are enabled simultaneously.
Indel Realigner	Fix a crash in Indel Realigner due to candidate.Length() <= consensus.size(). Extend the consensus sequences to match candidate length.
SV	Fixed the scenario when the requested tag contains a different type of payload than expected. e.g. XN:i vs XN:Z, due to read alignments from bwa-mem. Fixes error "std::exception::what: Can't parse unsigned long from string:"
SNV Somatic	Kernel segmenter (somatic contamination) dragen run logs were incorrect. Fixed to allow easier debugging.
Licensing	Remove the redundant "since <date>" from dragen_lic output. The updated output is easier to read and interpret.

BCL Convert Changes

DRAGEN™ BCL v4.4.7 introduces a new functionality that enables auto detection of indexes (new samples) and generates a new sample sheet containing those detected samples.

- New optional BCLConvert_Settings option "AutoDetectDemuxMode" to enable auto detection of indexes (new samples), and for the generation of a new sample sheet.
  - Three operational modes are available
  - {"None", "OverwriteSamples", "CorrectAndExtendSamples", "ExtendSamples"}
  - The default is "None" (feature disabled)
- ExtendSamples mode
  - Enabled by setting AutoDetectDemuxMode=ExtendSamples, but it also executes after "OverwriteSamples" or "CorrectAndExtendSamples" modes are completed.
  - How it works:
    - BCL analyzes the output of the Top Unknown Barcodes obtained with the original SampleSheet.
    - It processes barcodes sorted in descending order by the number of reads and adds new samples one by one that satisfy these criteria:
      - None of the indexes contain N.
      - None of the indexes are a homopolymer (allowing for 1 base error), e.g., AAAAAAAA or AAAAAAAG are rejected.

        - None of indexes contains > 70% of G bases, e.g. AAGGGGGG (75%) is rejected, AAAGGGGGGG (70%) is allowed.
        - The sample has >= AutoDetectMinNumReads reads.
        - The index of the new sample does not collide with any of the existing samples.
        - The number of the newly added samples including the current one is <= AutoDetectMaxCandidates.
      - After new samples are generated, a new samplesheet is generated.
      - When global OverrideCycles setting is used, autodetect takes it into account and generates indexes that adhere to the global setting.
      - With per-sample OverrideCycles, when all original samples in a lane use the same setting, new samples in that lane will also use the same OverrideCycles setting.
- OverwriteSamples mode
  - Enabled by setting AutoDetectDemuxMode=OverwriteSamples
  - Requirements
    - Index or Index2 can only contain the letter 'A'.
    - BCLConvert_Data columns may only contain
      - Lane
      - Sample_ID
      - Index
      - Index2
    - BCLConvert_Settings may only contain
      - SoftwareVersion
      - FastqCompressionFormat
      - AutoDetectDemuxMode
      - AutoDetectMinNumReads
      - AutoDetectMaxCandidates
      - AutoDetectDownsampleNumClusters
      - AutoDetectCorrectionThreshold
      - AutoDetectAcceptanceThreshold
      - AutoDetectAdditionalRejectsReported
      - GenerateFastqcMetrics
      - GenerateFastqcCoverageStats
      - FastqcDownsampling
      - KeepFastq
  - How it works:
    - BCL removes all samples from BCLConvert_Data section (hence overwrite mode)
    - Then BCL continues with the behavior as described above for ExtendSamples.
- CorrectAndExtendSamples mode
  - Enabled by setting AutoDetectDemuxMode=CorrectAndExtendSamples
  - How it works:
    - BCL executes an algorithm to detect and apply sample sheet corrections.
    - BCL checks if the following errors are present, by iterative testing of whether the Undetermined exceeds AutoDetectCorrectionThreshold
      - reverse-complemented Index2
      - reversed OverrideCycles Index2
      - both
    - Reversed OverrideCycles Index2 correction is applied only when it changes the OverrideCycles string and there are no UMI cycles in its Index2 part.
    - All corrections to Index 2 or OverrideCycles are applied to either all samples or none
    - When reverse-complemented Index2 correction is accepted, the system reverse-complements Index2 values

        - When reversed OverrideCycles Index2 correction is accepted, the system changes OverrideCycles Index2 either for all samples or in global setting, depending on which variant is used in the input sample sheet
        - When corrections are done, BCL continues with the behavior as described above for ExtendSamples.
- New sample sheet generated
    - A new sample sheet is placed in the final output reports directory named Reports/SampleSheet.autodetect.csv.
    - The new samplesheet is identical to the original samplesheet with these changes:
        - New samples are added to the list of samples with BCLConvert_Data section.
        - With OverwriteSamples, all original samples are deleted.
        - The AutoDetect* options are removed from the BCLConvert_Settings section:
- Settings
    - The following optional settings are available. It is recommended to use the defaults.

Setting	Description	Value	Default
AutoDetectDemuxMode	Enable an autodetect mode	String {"None", "OverwriteSamples", "CorrectAndExtendSamples", "ExtendSamples"}	"None"
AutoDetectMinNumReads	Minimum number of reads for a sample to be added in the ExtendSamples mode	Positive integer or string "Auto"	"Auto"
AutoDetectMaxCandidates	Maximum number of samples added (in the extend mode) per lane	Positive integer	1000
AutoDetectDownsampleNumClusters	BCL does not process all tiles to determine indexes, because the runtime would be prohibitive. Instead, a down sampled demux is performed. Down sampled demux uses the same algorithm as full demux, but only analyzes tiles until a specific number of clusters are processed	Positive integer	48,000,000
AutoDetectCorrectionThreshold	Specifies a minimium ratio of initial undetermined reads for corrections to be attempted	Positive float	0.95
AutoDetectAcceptanceThreshold	Minimium ratio of undetermined reads for a correction to be accepted	Positive float	0.2

What did not change

- **References and Resource files**
  No updates to the Reference Genomes or Resource files.
  No changes to the recommended usage of Pangenome or Linear references.

v4.4.4 through v4.4.7 use the same files.

- **Bioinformatics output**
  DRAGEN bug fixes only
  No accuracy changes to any of the bioinformatics components
  No output file or format changes
  Updating to this patch from v4.4.6 will not cause any batch effects

DRAGEN v4.4.6 Updates

Overview

DRAGEN™ v4.4.6 was a minor release that contained bug fixes, improvements and new features across most of the callers and pipelines.

Highlights of v4.4.6

- Numerous bug fixes and improvements to address customer feedback from v4.4.4 and v4.3
- Support for 12 Pillar OncoReveal panels, with easy-to-use pre-packaged resources and settings to support streamlined analysis
- Introducing support for the Illumina 5-base Methylation prep
- Introducing a new 16S metagenomics pipeline
- Improvements to scRNA PipSeq and Perturb-Seq pipeline
- Improvements to the MRD pipeline
- 5 new onboard and cloud applications on MiSeq i100 using v4.4.6

What did not change in v4.4.6

- **References and Resource files**
  No updates to the Reference Genomes or Resource files.
  No changes to the recommended usage of Pangenome or Linear references.
  v4.4.4 and v4.4.6 use the same files.

Amplicon and Pillar Panel Support

v4.4.6 adds a major update to the Amplicon pipeline. DRAGEN adds built-in support for 12 Pillar oncoReveal panels with PONs and noise baselines files pre-packaged with DRAGEN, and simple single command line option for the panel selection. Adds an Amplicon application on-board the MiSeq i100 instrument with easy-to-use Pillar oncoReveal panel support and updates the Amplicon BSSH application. The following updates are included:

- Add built-in support for 12 Pillar oncoReveal panels.
  - Amplicon pipeline now just selects the panel type in the analysis, and all the dragen settings are auto configured.
  - Package all required resource files for SNV systematic noise files, SV systematic noise files, CNV PONs (combined counts), Amplicon BED files, with dragen installer.
  - Supports: BRCA CNV, Heme, Lymphoid, Core LBx, Essential LBx, MPN, Fusion LBx, Multi-Cancer with Fusion, Multi-Cancer with CNV, Myeloid, Nexus, Solid Tumor v2

    - See the DRAGEN User Guide for full details: https://help.dragen.illumina.com/product-guide/dragen-v4.4/dragen-amplicon-pipeline#dragen-amplicon-panel-specific-settings
- Accuracy improvements
    - Fix for an expected variant call that was not detected, by excluding amplicon for somatic soft clipping proximity events
    - Fix issue where Germline has downsampling bias, and reports much higher VAF that expected
    - Fix for FP somatic variants in FFPE leave-out sample. Use the read cigar for soft clip ratio calculation
    - Fix issue where deletion is called in amplicon where the deleted base is in the probe region of Pillar MPN sample. Generally, exclude deletion calls, if deleted base is in the probe region
    - Fix a CNV FP from a normal FFPE leave-out sample, by applying sum test on the small interval
    - Fix false positive on AmpliSeq Myeloid Panel. Add a ref median read position check for read position filter
    - Fix CNV FPs that were introduced with control bed feature. Use scaling factor to adjust qual for small regions
    - Fix Amplicon RNA fusion FNs, which were detected by PiVAT and not DRAGEN.
    - Fix for some truth FLT3-ITD variants reported only in hard-filtered.vcf.gz file but not sv.vcf sv for some Amplicon panels.
    - Fix anchor length calculation to improve fusion calling. This rescues detection of the EML4::ALK gene.
    - Fix for known fusion BRAF-KIAA1549 missing on Pillar OncoReveal Multi-Cancer with CNV and RNA Fusion.
- Fix a watchdog hang during variant calling on Pillar Essential LBx Essn32 sample
- Add "imbalance ratio" output to RNA Amplicon pipeline

5-base Methylation Support

Illumina DNA Methylation Prep and Illumina DNA Methylation Prep with Enrichment use a proprietary enzymatic reaction to convert methylated cytosines to thymidine in a single step, while retaining the identity of unmethylated cytosines. They generate sequencing-ready libraries from gDNA or cfDNA. These library prep methods enable the detection of both methylation and genomic variants with a single library prep.

See https://www.illumina.com/science/genomics-research/articles/5-base-solution.html and https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-5-base-methylation-flyer-m-gl-03401/illumina-5-base-methylation-flyer-m-gl-03401.pdf

v4.4.4 added initial support for the new Illumina 5-base (Methylation) prep, on limited pipelines for preview.

v4.4.6 now introduces 5-base prep support in germline / somatic / UMI-enrichment workflows and adds a new license requirement.

- The DRAGEN 5-base Methylation pipeline can be run from the BSSH/ICA Methylation apps, as well as from the DRAGEN command line.
- When running with 5-base prep inputs, simply enable the option "--methylation-conversion=illumina" to auto-configure the pipeline for the prep.
- 5-base Methylation analyses now use the same Reference Genome as the Germline and/or Somatic pipelines and components. Legacy Methylation solution for TruSeq DNA Methyl and TruSeq Methyl Capture library prep kits remain unchanged

- Please see the User Guide for more details, examples and command line options

The following changes are made to the 5-base pipeline in v4.4.6:

- Improve 5mC reporting accuracy by using the XM tag information
- Remove redundant "--controls-report" option
- Fix an issue where re-mapping of a methylation BAM generates incorrect CA control mapper tags
- Don't allow check-fingerprint with methylation
- Methylation VC
    - When methylation VC is enabled, set the default methylation report MAPQ threshold automatically to 20, to make the CX report more concordant with the M5mC gVCF.
    - Include known Methylated bases and overlapping reads when counting bases, which is used for 5MC reporting in GVCF homref blocks. This improves the concordance on beta values from ~60% to 100%.
    - Improve SNP FN and INDEL FP_FN for Methylation variant calling, by refining Smith Waterman scores in VC for 5-base analysis.
- Tumor/Normal analysis
    - Improve 5-base SNV FP/FN for Tumor/Normal analyses, by separating Tumor and Normal sample beta estimations
    - Split methylation metrics for Tumor/Normal runs. Add "TUMOR" and "NORMAL" labels to the Tumor/Normal methylation metrics output
    - Increase default binner memory budget for 5-base runs with Tumor/Normal input, to avoid hanging during the sort phase.
    - Fix bug in the Tumor/Normal M5mC reporting where normal reads were included and caused incorrect 5mC reporting
- UMI analysis
    - Fix UMI methylation-aware VC accuracy, when running from a bam/cram file without mapping.
    - Fix XM tags for duplex collapsed UMI reads

New 16S Metagenomics Pipeline

The DRAGEN 16S pipeline is a rapid, kmer-based informatics solution designed for microbial classification and community profiling from mixed flora and metagenomic sample types. The pipeline delivers powerful secondary analysis of Illumina 16S sequencing data, with steps for read QC (optional), taxonomic classification, result filtering (optional), and reporting.

v4.4.4 Introduced the new 16S Metagenomics pipeline via the BSSH application DRAGEN 16s Plus: https://help.idm.illumina.com/dragen-16s-plus/dragen-16s-plus

v4.4.6 adds 16S pipeline performance and accuracy improvements, expanded test coverage, and offers minor updates and bug fixes as summarized below. The Application is also offered on-board the MiSeq i100 instrument.

- See the v4.4 DRAGEN User Guide for full details: https://help.dragen.illumina.com/product-guide/dragen-v4.4/dragen-16s-pipeline
- v4.4.6 updates
    - 16S pipeline performance and accuracy improvements.
        - Improves accuracy by adding a second-pass k-mer classification when the first-pass classifies too high on the taxonomic tree.
        - Improves performance by removing the unnecessary sample composition step from the pipeline, creating an output file directly instead of deriving it from an

intermediate file, pre-computing taxonomic relationships, and removing some large vestigial data structures
    - Support for paired-end samples.
        - Better handling of the input fastq list CSV file.
        - Support paired-end read QC
    - Several minor updates to output AP JSON file
    - Change reporting threshold from abundance to read count. The parameter "16S-report-threshold" is now an integer that represents the read count threshold used to filter out detections at low levels.
    - Fix a bug with custom database building
- Fix a bug with incorrect taxonomic representations in the output files.

SNV VC

v4.4.6 contain various bug fixes, usability improvements, and run time improvements. Somatic improves MNV counting, and Germline adds a high sensitivity mosaic ML model

- Somatic
    - Fix an issue where INFO.MQ and INFO.FractionInformativeReads differ between GVCF and VCF records for a couple of variants. Handle GVCF and GVCF_NONE the same for somatic in convertEventsToAlleles
    - Improve run time of Somatic calling. Do not include MAPQ=0 reads in somatic callable regions
    - Make the option "vc-remove-all-soft-clips" also work for the somatic mode, to address problems with samples that contain a large fraction of soft clipped bases.
    - Fix a segfault when building a systematic noise file
    - Change handling on the counting of MNV calls in vc_metrics to ensure concordance with RTG vcfstats.
        - MNVs with equal length REF and ALT are counted to the MNPs
        - MNVs with unequal length REF and ALT are counted to the Indels category
        - Output MNPs as a separate category in the vc_metrics.csv file to match RTG vcfstats output
        - No longer skip MNV component calls in the vc_metrics counts. We are filtering them with the new mnv_component filter flag as of v4.4, so there is no longer an issue of double counting them in the VC metrics and we can treat them the same as other variants.
    - Fix invalid Nirvana VCF output, when "--variant-annotation-enable-vcf-output=true" is enabled in a somatic run with germline tagging
- Germline
    - Add new SNV high sensitivity mosaic ML model, that is used with "--vc-enable-high-sensitivity-mosaic-detection=true"
    - Fix crash in read realignment during evidence BAM generation, due to very short haps.
    - Fix invalid PL values in haploid regions, which leads to GvcfGenotyper failure with Extreme FORMAT/PL value
    - Fix germline tagging crash when CNV or SV is enabled but SNV is not enabled

CNV, CNV Cytogenetics and CNV+SV

v4.4.4 introduced Germline WGS Cytogenetics in the CNV caller, added in-run construction of CNV panel-of-normals (PoN) for Enrichment applications, added WGS panel-of-normals (PoN) support, and made Somatic model improvements. v4.4.6 makes improvements to these functions and fixes various issues

- Robustness improvements for CNV pedigree analysis

- Recover rejected pairs in call merging
- Fix segfault during CNV segmentation. Skip segmentation of contigs with degenerate parameters (to avoid divide by zero)
- Fix VCF source in ASCN+SV to be DRAGEN_ASCN_SV
- Fix crash during ASCN smoothing on exomes, by removing obsolete merge pairs from rejected pairs
- Fix crash when using non-human reference, by disabling autosomal length checks
- Fixed an issue where "cnv-counts-method" midpoint or overlap was not supported from BAM input
- Add command line option validation for SLM "cnv-slm-eta" and "cnv-slm-omega" arguments to prevent incorrect inputs
- Fix a watchdog timeout in CNV B-allele modeling in WES or high depth data
- Produce header only VCF on NTC samples
- Fix: CNV BRCA LR module from BAM with map-align disabled, the output cnv.LR.VCF does not list the sample name in the header
- Fix CNV segfault during TargetIntervals parsing on very small contigs
- Fix for PON generation from target.counts.gz inputs (non GC-corrected)
- Add a CNV post-VCF filter for target regions of interest. New command line option "cnv-post-vcf-target-bed", to filter output VCF entries that do not overlap with provided target region. It impacts to 3 output files of .cnv.vcf.gz, .gff3, and .cnv_metrics.csv file
- Fix Bigwig to bedgraph conversion bug, caused by assigning off-by-one start coordinate to back-to-back intervals
- Improve error messages for PON, when input CASE and PON do not have identical target intervals
- Add bigwig output file ".depth.seg.bw" from the depth.seg file previously generated, for depth+BAF segmentation
- Add SVCLAIM=D to CNV calls, for contigs without any SV calls
- The CNV+SV workflow now takes max CNV QUAL instead of max germline QUAL
- Add INFO/SVCLAIM field to segmentation-bed records, which were missing in the cnv_sv.vcf.gz
- Fix for "chromArmBinCount" missing in cnv_sv VCF header, by removing cyto pre-defined segments from the cnv_sv.vcf.gz
- Add resolution-dependent homozygosity index for Cyto VCF. The original implementation of the homozygosity index reports a single index, based on the largest calls (by default ≥ 2Mb) that are found in the standard CNV VCF. This update adds additional indexes (one for each resolution) on the CYTO VCF. E.g "##HomozygosityIndex(25k)=0.001015"
- Use total autosomal length when computing homozygosity index
- Emit SVTYPE for backward compatibility
- Add support for integration of segmental duplication rescued intervals in germline WGS ASCN.
- New minimum autosomal median count (10) for germline ASCN
- Emit GAINLOHs as ALT=DUP (instead of ALT=LOH) if user requests it. A new boolean option "cnv-ascn-emit-gainloh-as-dup" is provided (default=false)
- Update calls with MOSAIC/HET annotation only above a certain length
- Allow user to output Cyto results only for selected resolutions. New option "cnv-cyto-keep-resolutions" in the set {25k,50k,500k,1M, 1M_depth} can be specified as comma separated list. E.g. "cnv-cyto-keep-resolutions=25k,1M" to select two sets
- ASCN - Filter short arm of acrocentric chromosomes (chr13,14,15,21,22) when not compatible with long arm
- Update the default value of "cnv-enable-kmer-binning" to true for germline WGS with PON normalization. Applies to PON generation
- Disallow germline ASCN CNV in 5-base methylation workflow

Structural Variant Caller

v4.4.6 brings various bug fixes, robustness updates, run time and memory usage improvements to the SV caller

- Fix a FLT3-ITD SV FN case on targeted UMI assay
- Fixed missing KMT2A-PTD involving exon 3-6 in clinical NGS setting.
- Ignore stats and do not crash, if a sample contains no reads. This adds support for control samples
- Fix for crash "SizeDistribution.cpp line 27 -- cumulativeCount <= totalCount" by using 64 bit counters in SV stats.
- Fix occasional segfault in "SVLocusSetMerger::merge"
- Fix for T/N failing in the SV caller when running on tumor sample with UMI inputs and "qc-coverage-ignore-overlaps=true" provided on the command line. Maintain mapper output order for alignments in overlap analyzer.
- Fix an out-of-memory crash during SV Systematic Noise building step. Reduce memory usage by 50% and improve run time by 35%
- Fix out-of-memory assertion in SV caller during refinement for SVCandidates, by limiting breakend region growth for extreme scenarios.
- Fixed OOM errors on very high coverage WGS samples
- Fixed high memory usage spike on some somatic samples, which may lead to out-of-memory errors
- Fix for issue where moderate to high depth SVs have SOMATIC_SCORE=0. This is caused by triggering the MAX_DEPTH filter during scoring and early aborting the scoring. Now downsampling the SR when above max depth, instead of aborting scoring.
- Enable Nirvana Annotation on the DUX4 vcf
- Improve run time of SV merging and dedup step
- Improve run time of SV caller on samples with excessively high PCR duplication ratios.
- Improve run time when running directly from BAM input.
- Add early exit when detecting erroneous soft clip pattern in phase0 sampling, to avoid very long run times on noisy data

Explify Analysis Pipeline

The pipeline is offered via the DRAGEN command line, and also via the Microbial Enrichment Plus application on BSSH https://help.idm.illumina.com/dragen-microbial-enrichment-plus/dragen-microbial-enrichment-plus.

v4.4.6 adds the DME+ application on-board the MiSeq i100 instrument, and makes the following updates:

- Run time improvement for the standalone k-mer classifier, through more efficient use of available CPUs
- Add viral variant calling and consensus thresholds. Three command line options are added for the pipeline: "--explify-viral-consensus-depth-threshold", "--explify-viral-vc-depth-threshold", and "--explify-viral-vc-af-threshold". The variant calling depth and AF thresholds are passed onto the viral variant calling module and used to determine whether to consider each variant. The consensus depth threshold is passed onto the reporting step of the pipeline (the final step), where it is used to hard mask viral consensus sequences.
- Update hash table params used with flu and SARS-CoV-2, to achieve better accuracy with some Influenza subtypes.
- The pipeline can now outputs post-quality fastqs with human reads removed. New command line option "--explify-post-qc-fastq-mode"

RNA Gene Fusion

v4.4.4 introduced a new ML model for Splice VC and made significant improvements in Gene Fusion calling. v4.4.6 further adds minor improvements, including improvements on RNA amplicon panels

- Splice variant calling
  - Fix for splice variant not called on NSD1 gene
  - Add NPM1 to repeat-intervals BED file for fusion
  - RNA amplicon and WTS updated to use same parameters
  - Fix for splice variant fusions not found in fusion_candidates.final
- Gene Fusion
  - Add NoAssembly for readthrough fusions
  - Improve checks for existence of enriched gene/region input files
  - Add FUSION_SEQ, FUSION_SEQ_LEN, FUSION_LEEWAY to VCF output
- Quantification
  - Add RNA quantification metrics to metrics.json output, so that quant.metrics.csv values are propagated over

Illumina PIPseq Single Cell RNA (scRNA)

v4.4.4 Introduced support for scRNA analysis from Illumina Single Cell 3' Prep. It also supported a mode for processing samples from Illumina's CRISPR Single Cell kits using PIPseq technology. v4.4.6 adds some new features, robustness improvements and bug fixes.

- Avoid Genome Mapping of CRISPR Reads. This fixes double counting of CRISPR counts and metrics reporting
- Add support for CRISPR guide RNA calling. There is a new option "--scrna-enable-crispr-guide-calling=true" to enable this feature
- Fix out-of-memory issues on AWS with CRISPR runs. Release free memory after deleting cell tables
- Add min alignment score option for PIPseq CRISPR mode to prevent false mapping of contaminated CRISPR reads with GEX barcodes
- Revamp and update scRNA metrics. Add percentages and additional metrics
- Print scRNA metrics to stdout
- Support feature sequences with mixed case
- Improve run time of single cell feature matching
- Improve run time and memory usage for all single cell runs
- Improve run time of single cell feature matching for guide RNA
- Improve run time of IPM calculations. The time correcting scRNA molecular identifier sequences is improved
- Fix a crash during target handling for FASTA based features
- Update PIPseq feature counting to use total counted feature reads and not the IMI corrected value
- Fix crash when analyzing with cell hashes in PIPSeq mode. Add support for cell hashing in PIPseq mode
- Clarify error message when feature reference file has wrong number of columns
- Allow existing single cell genotype-based and genotype-free demultiplexing to work in PIPseq mode by allowing PipSeqCountTable to be used as input to the demultiplexer. This does not implement genotype demuxing but allows the run to succeed.
- Fix feature read groups for BCL NoLaneSplitting case. Fixes issue where "--scrna-feature-barcode-groups" did not work as expected when "NoLaneSplitting=True"
- Update settings to improve demux doublet detection
- Allow barcode phasing independent of PIPseq mode
- Fix fractional downsampling for UMI FASTQs
- Output barcodeGeneCounts.csv file independent of PIPseq

MRD (Minimal Residual Disease) Pipeline

v4.4.4 introduced a new MRD pipeline to detect residual cancer cells in solid tumors. v4.4.6 adds accuracy improvements discovered from further expanded test and validation on larger sample sets

- Improve specificity on plasma samples (false positive rate)
- Improve Indel noise model to properly discriminate error rates for insertions and deletions.
- Update Indel QC metric and "detect" decision
- Reduce inflated false positive detection on certain samples
- Add noise level QC to summary
- Adjust mrd score-threshold default

MSI (Microsatellite Instability)

v4.4.6 aims to improve usability of the MSI module, and provides bug fixes

- Fix an issue where MSI fails to process records with read-stitching=true when starting from a bam file.
- Fix issue where MSI misses a newline character when splitting combined .dist files
- Turns on pair-by-name by default
- Turns on read-stitching by default to avoid double counting
- MSI runs in tumor-only mode if ref-normal is provided
- Made the following options configurable on command line: "msi-record-processors", "msi-min-input-ref-normal-files", "msi-min-assessed-sites", "msi-min-covered-normals", "msi-pvalue-threshold"

BCL

BCL convert fixes some regressions introduced in earlier versions, related to very long run times with very high sample counts and network file systems. Introduces flexible no-lane-splitting to output concatenated FASTQ files over a subset of lanes.

- Flexible no-lane-splitting: output concatenated fastq.gz files for subset of samples across lanes
- Support for output of newer RTA versions that support RQI file format (two-level PF selection)
- Fixes issue where some sample sheet inputs had undetected conflicts at startup, causing possible error during conversion
- Top Unknown Barcode tracking now prunes noisy inputs, preventing out-of-control memory usage in rare conditions
- Fixes LibraryRebalanceStats listed first index2 entry across all entries if no Lane column in sample sheet
- Performance regression for no-index flow cells
- Improved performance for lanes with very high sample count (10K+)
- Fixes crash on bcl.bgzf inputs with "no lane splitting" enabled and bcl-validate-sample-sheet-only=true

Compression

v4.4.4 updated the CRAM module. Specifically, it transitioned the solution to use a recent htslib that offers CRAM v3.1 support. v4.4.6 fixes issues with non-human samples due to this update and adds some run time improvements. Bug fixes for ORA compression issues are also included

- CRAM
  - Disable indexing for unsorted CRAM, for analyses such as Phix (references that include chromosomes with length greater than 512 Mbps (2^29 bases)), to avoid htslib failure.
  - Fix an intermittent CRAM exception caused by closing a pipe descriptor two times
  - Enable AVX2 for CRAM htslib. Speeds up CRAM output by 9% on average
  - Fix htslib CRAM hangs, when attempting to populate reference segments from the internet. The htslib mis-detects that a reference contig needs to be loaded. Now always use the reference provided to dragen in all conditions.
  - Fix extreme CRAM run times on samples like PhiX, by explicitly instructing htslib to use multi-seq (unsorted) mode.
- ORA
  - Fix fatal error when --ora-get-metadata is used on empty fastq.ora file
  - Fix a watchdog timeout with fastq.ora input
  - Fix random "Data corruption detected" issue
  - Fix BCL to ORA issue , where one of the mate file is missing very last read of the file

Other Bioinformatics fixes

- Insert Stats
  - Fix issue where the full run stats double counts the initial insert stats
- Aligner
  - If user specified '--Aligner.match-score=0' on the command line, then match-n-score was set to zero under-the-hood. Fix the MatchNScore
  - Fix an unexpected 5-10x runtime hit when "--generate-md-tags" is enabled on some references. Optimize the MD generation
  - Fix segfault for nonexistent refdir during options validation
- Downsampling
  - Fix segfault when downsamling is used with UMI input
  - Fix "down-sampler-reads" 4G limit due to use of unit32
- Hash Table builder
  - Fix for scRNA hash table builder not working for custom gene
- Iterative Gvcf Genotyper
  - Silence warnings about missing variant INFO fields in gVCF homref records, to reduce size of log file for large batches
  - Fix crash on targeted calls with ploidy > 2
- Joint Genotyping (Pedigree calling)
  - Add support for pre-signed URLs as input for VCF streaming into joint caller
- JSON metrics
  - Fix JSON metric for "hetHomRatio", which did not match the CSV file.
- Nirvana annotation
  - Update Nirvana to 3.26.0
  - Support VCF output format for Nirvana running from DRAGEN. Controlled via new optional command line option "--variant-annotation-enable-vcf-output" (default=false)
- OncoVirus
  - Improve run time from WGS BAM files
- Paralog Caller
  - Updated error message when targeted callers are enabled but the data is WES. Disable the caller, do not fail the run
  - Fix a crash when running exome data with PON. Fixed a kmeans clustering issue
  - Allow calling CYP21A2 c.955C>T in isolation on a single haplotype.
  - Avoid crash "The Likelihood must be a positive number.-nan" for CYP21A2 total CN 1 (noisy data)
- Personalization

    - Allow enabling of HLA caller with personalization
- Repeat Genotyping
    - Stop using offtarget regions for polyalanine repeats in default STR catalogs. The VariantType will go from RareRepeat back to Repeat.
    - polyalanine loci affected by the change: ARX1, ARX2, FOXL2, HOXA13_1, HOXA13_2, HOXA13_3, RUNX2, SOX3, TBX1, ZIC2, ZIC3
- Spatial Pipeline
    - Spatial support for coordinate-based output
    - Multi-mapped read counting
    - Fix for segfault on low MAPQ reads
    - Apply barcode corrections on omitted reads (fix invalid barcodes in BAM)
- Star Allele Caller
    - Fix for star allele caller failing to make a call on MT-RNR1 when the GT of a variant is 1/1
- UMI
    - Enable fastq list input for (T+UMI)/N workflow. Fixes a bug with processing multiple lanes for the UMI read group
- Variant Merger
    - Improve run time of variant merging through vectorization.
    - Fix for "ERROR: Found an invalid ref allele" in variant merger. Get the ref as ACTG only in variant merger to account for SV VCF reporting ref alleles containing "N"
- VNTR
    - Fix a crash on zero normal reads in VNTR, when the pipeline is run with NTC (water)

Other Platform fixes

- Multi-version concurrency fix
    - Multi-version dragen has a latent concurrency issue where a dragen execution on one version can create an error or hang in another version, by changing the size of memory buffers used to send data to the FPGA.
        - Two versions can run concurrently in parallel, when the dragen license daemon (dragen_licd) and dragen_lic application would run or be called from one version, while secondary analysis is run from a different version.
        - This can happen more frequently with Fleet licensing, or when the user polls the license via dragen_lic asynchronously with analysis jobs
    - The issue potentially exist only when multi-version packages for v4.0 and earlier are installed at the same as v4.2 and later. (such as TSO500 v3.10 and dragen v4.4 simultaneous installs)
    - The fix is being rolled out to all existing versions.
- Licensing
    - Skip non-primary alignments in license base counting, when running from BAM input to VC without map/align. License bases will now match Total bases in MAPPING/ALIGNING SUMMARY
    - Fix MRJD from BAM not consuming any license quota
    - Fix issue where "dragen_lic -X" does not report all installed licenses
    - Fix a crash with "dragen_lic -b" or "dragen_lic -j" when a license has expired.
    - Fix "already installed" error when new license types are re-installed.
    - Security enhancement: License credential and URL must be provided by "--lic-credentials=<file>". The "--lic-server" option explicitly not allowed for some modes to prevent credentials leaks in terminal logs.
    - Support new license types:
        - "5base": required when running any Illumina 5-base methylation pipeline
        - "Spatial": required when running any single cell spatial pipeline
    - Allow "dragen_lic -i auto" to install new license types: "PipSeq", "GvcfGenotyper", "5base".

    - Augment the RunInfo with more run metadata.
- Cloud
    - AWS F2: Improve the run time of the HW graph on AWS F2 instances. Improvement seen with high coverage / somatic WGS sample types
    - Fix for cloud crash "Reset while connections are still opened". Affected runs with HLA enabled. Do not disconnect from the board at the end of HLA processing.
    - Fix an issue when on-prem dragen packages are used on cloud, dragen incorrectly tries to fork a dragend process. This cannot succeed and dragen is unusable. This fix allows users to use the same dragen docker image for all platforms.
- Platform
    - Only grab a board lock when we are using the hardware and a board is specifically requested. This fixes cases such as running ORA compression job in parallel with dragen analysis job, when the board is only specified due to license requirements, but no actual board resources are used. This use case was broken in v4.4.4
    - Fix a crash on Azure el9, where a cache line could not be read from within a confidential VM image.
    - Enable sort spill compression by default, to allow some samples such as scRNA, that currently require a PH4 server to be able run on a PH3 server, and run on cloud instances with smaller attached disk.
    - Make dragen on-prem more robust to CTRL-C cancelling of jobs, to avoid putting server in a bad state
    - Fix an orphaned PID collision. When dragen is run in docker container, then crashes, a second invocation can have a PID collision if host pid was not mapped in.
    - Fix an issue encountered in SLURM environment, where dragen is SIGKILL'ed by SLURM during crash recovery. Remove some steps from crash recovery which avoids the dragen board ending up in a non-recoverable state where a power cycle is necessary.

Resource Files

NOTE: v4.4.4, v4.4.6 and v4.4.7 use the same resource files. This information is included for convenience.

NOTE: Illumina 5-base Methylation analyses use the same Hash Tables as Germline or Somatic pipelines

DRAGEN™ v4.4 released updates to key resource files required for functionality and optimum performance. Additional resource files were 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`

Reference Genome Recommendations

NOTE: Illumina 5-base Methylation analyses now use the same Reference Genome as the Germline and/or Somatic pipelines and components. Legacy Methylation solution for TruSeq DNA Methyl and TruSeq Methyl Capture library prep kits remain unchanged

NOTE: No other changes to the recommendations.

- 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	5-base Germline	Yes	Yes	Yes	No	Pangenome
	5-base Somatic	Yes	Yes	Yes	No	Linear
	TruSeq DNA Methyl TruSeq Methyl Capture	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

Multi-Version Installer for on-premises servers

NOTE: No changes are made to the installer. This information is included for convenience.

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.

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	Fixed in v4.4.6.
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	Fixed in v4.4.6.
DNA Alignment	Specifying '--Aligner.match-score=0' will set match-n-score to zero under-the-hood which leads to leads to invalid alignments	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
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	Intermittent hangs have been observed when processing T100 samples	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
scRNA	Feature barcode read groups fails when no lane splitting is set to true in BCL.	Fixed in v4.4.6.
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	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
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.
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	Fixed in v4.4.6.
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.	Mitigated by F2 instances. Improvements in v4.4.6.

SV	T/N analysis may fal in the SV caller when running on tumor sample with UMI inputs	Fixed in v4.4.6.
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.	Fixed in v4.4.6.
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	Fixed in v4.4.6.

New issues discovered on v4.4.6 release

Component/s	Summary	Resolution/Workaround
Infrastructure	Option "combine-samples-by-name" no longer working since dragen v4.3	None.
BCL	BCL->Ora one-step mode does not error out when NoLaneSplitting is enabled and the input sample sheet does not contain lanes.	None. NoLaneSplitting not supported for ORA output
BCL/ORA	Fastq ORA compression error when bcl-only-matched-reads=true are used.	Use the "--force" option
BCL	BCL run time is very long on proteomics prep samples.	Do not use this version v4.4.6 for BCL on proteomics prep samples.
MSI	When "--amplicon-enable-cfdna-core=true" and collect-evidence is used, dragen doesn't set the microsatellite file for collect-evidence. Resulting in empty MSI output	Provide the --msi-microsatellite-file /path/to/pillar_core_microsatellites-sites-v1.list
SNV VC	VC Segfault has been observed in performBanding	None.

New issues discovered on v4.4.7 release

Component/s	Summary	Resolution/Workaround
SV	SV caller MaxDepth filter "sv-max-depth-factor" applies to normal reads only in T/N mode but somatic reads in T-only mode	None. To be fixed in next major release.
QC Metrics	GC bias reports for WES germline does not restrict assessment to the target BED regions	None. To be fixed in next major release.

SNV Somatic	SNV caller misses a deletion when phasing two MNV components. It is a rare edge case that only impacts MNVs composed of a deletion that deletes the anchor base of a downstream indel.	None. To be fixed in next major release.
RNA Splice Variants	RNA Splice variant caller has excessive run time on some samples	Run the analysis with downsampled fragments: "--enable-down-sampler true --down-sampler-fragments 50000000--enable-watchdog false"
MRD	Large plasma samples (500x) have seen watchdog timeout failures during MRD Detect	Disable watchdog timeout and monitor the run

New issues discovered on v4.4.11 release

Component/s	Summary	Resolution/Workaround
RNA Alignment	Run-run chimeric alignment difference when using graph reference. ~1/150M chimeric reads map differently in alignment, only when using graph reference on RNA.	Linear reference is still recommended on v4.4, the issue is not present with linear reference.
HLA	Dragen mode: Somatic T/N + solid WGS + UMI + HLA can encounter an assertion in the HLA component.	No workaround.
Gene Fusion	RNA runs with Splice VC can hang on excessively large samples.	Use `--down-sampler-fragments` to set a reasonable fragment size.
SNV Somatic	Segfault or out-of-memory error in Somatic Enrichment T/N analysis with `full_res` depth metrics and `--enable-metrics-compression=true`	Disable the metric compression or re-run the sample.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package.
- The archive integrity can be checked using: `./<DRAGEN 4.4.11 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.4.11 .run file>`

Release History

Revision	Release Reference	Originator	Description of Change
00	1134696	Cobus De Beer	Initial release