DRAGEN v3.9.3 Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Bio-IT Platform v3.9.3.

Changes are relative to DRAGEN™ v3.8.4. If you are upgrading from a version prior to DRAGEN™ v3.8.4, please review the release notes for a list of features and bug fixes introduced in subsequent versions.

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

The 3.9.3 software package includes:
- DRAGEN™ SW Intel Centos 7 - dragen-3.9.3-4.el7.x86_64.run

The following configurations are also available on request:
- Amazon Machine Image (AMI)
- Microsoft Azure Image (VM)
- RPM packages for Centos 7 for Amazon Web Services (AWS)

Deprecated platforms:
- Support for IBM PPC has been deprecated since DRAGEN™ v3.7
- Support for Intel CentOS 6 has been deprecated since DRAGEN™ v3.8
- Support for Ubuntu has been deprecated since DRAGEN™ v3.9

Contents

Overview....3
New Callers and Major Features....3
Improvements and Feature Additions....7
Interface Changes....13
Issues Resolved....13
Known Issues....15
SW Installation Procedure....17

Overview

Below is a summary of the changes included in v3.9.3. DRAGEN™ v3.9 offers new callers, as well as speed and accuracy gains and new feature introductions across most callers. For full extensive details, please consult the latest Illumina DRAGEN™ Bio-IT Platform User Guide available on the support website at https://support.illumina.com/downloads/illumina-dragen-bio-it-platform-user-guide.html

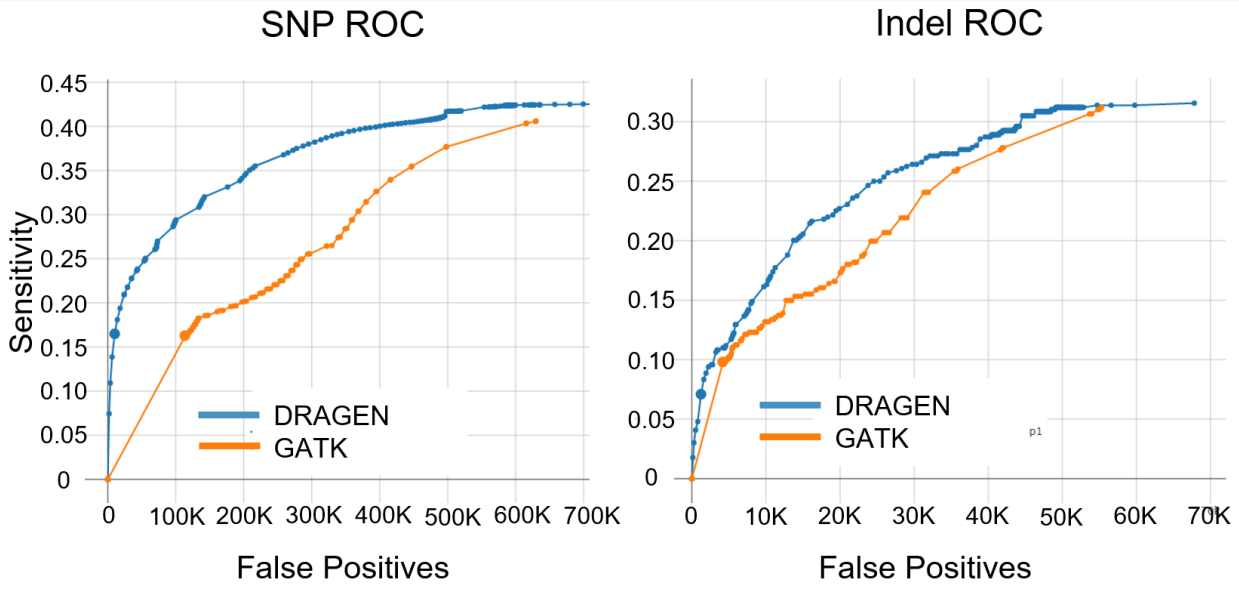
New Callers and Major Features

DRAGEN on Microsoft Azure

- DRAGEN is now available on Microsoft Azure. The NP-series of FPGA backed VMs are powered by Xilinx U250 FPGAs to enable hardware accelerated DRAGEN processing. NP10s, NP20s, and NP40s can all be used to run DRAGEN. Initial availability regions are West US2, East US, West Europe (Amsterdam), Southeast Asia (Singapore)
- The run time for processing NA12878 at 50x coverage is 40 min for germline small variants, 66 min for when enabling all callers, starting from FASTQs
- Contact your Illumina Sales Representative to learn about getting a license for use on Azure, and access to the DRAGEN Image Gallery

RNA

- v3.9 Improves on existing features and now includes **Variant Calling** and **Forced Genotyping**
- New De novo RNA VC calls variants directly without needed matched DNA samples. It leverages DRAGEN RNA Mapper and Somatic Variant Caller algorithms to accurately call variants from RNA-seq reads. It leverages FPGA acceleration to execute 19x faster than GATK



RNA VC ROCs showing sensitivity vs false positive count on a HCC1187 WTS sample. Germline calls from matched DNA sample used as truth set. Both DRAGEN and GATK use RNA-mapped DRAGEN BAMs as input

	DRAGEN RNA VC	GATK RNA VC

Map/Align	1.5 min	N/A
Variant Calling	9 min	172 min

- RNA Forced Genotyping can confirm expression of DNA Somatic variants in RNA samples. Variants are genotyped directly from the VCF.
- Command line options allow users to set the input type (FASTQ/BAM/CRAM), include an (optional) splice annotation file, and include an (optional) ForceGT VCF.
    - Use the following options to enable either VC and/or ForceGT: `--enable-rna=true --enable-variant-caller=true --vc-forcegt-vcf=<forcegt_vcf_file>`
    - Usage Note: *"**tumor"** inputs must be used when enabling RNA VC, regardless of whether the input sample is from germline or somatic origin. This is required to ensure that the correct algorithm is run. The somatic algorithm can be used to produce germline or somatic calls. Gene Fusion Detection and Quantification can still be enabled alongside RNA VC*
- v3.9 RNA now reports gene fusions in **target repetitive regions**
    - Clinically relevant fusions, such as CIC—DUX4 and EGFR—SEPTIN14, with one or both breakpoints in repetitive regions are challenging to detect.
    - v3.9 implements a targeted alignment approach to report fusions in repeats of interest.
        - By default, target genes DUX4, CIC and SEPTIN14 are automatically applied for hg19 and GRCh38
        - Alternatively, the user may provide a list of gene names or a BED file with target intervals or disable the default gene list.
    - Example usage: `--rna-repeat-genes <GENE_NAME_FILE>`, OR `--rna-repeat-intervals <REPEAT_BED_FILE>`, where `GENE_NAME_FILE` may be a file with list of target gene names or gene IDs, one per line.

Homologous Recombination Deficiency (HRD) Biomarker

- DRAGEN HRD Scoring computes a genomic instability score from WGS T/N and T-Only somatic CNV calling, achieving high concordance with widely used microarray approaches
- Genomic Instability Scores (HRD Scores) are the summed scores of Loss Of Heterozygosity (LOH), Telomeric Allelic Imbalance (TAI), and Large-scale State Transitions (LST)
- The command line option: `--enable-hrd true` is used in conjunction with `–cnv-use-somatic-vc-baf true` for T/N or --cnv-population-b-allele-vcf <SNP_VCF>. Additional HRD options are available for LOH/TAI/LST configurations. Please refer to the DRAGEN User Guide for more information

Survival Motor Neuron (SMN) Caller

- V3.9 introduces a new caller for Spinal Muscular Atrophy (SMA) detection and carrier status from WGS data. SMN caller is capable of accurately identifying the SMN1 and SMN2 copy numbers from WGS data by analyzing read depth and 8 informative sites of sequence difference between SMN1/SMN2. The SMA and carrier status is reported along with the SMN1 and SMN2 copy numbers
- The SMN caller requires WGS data with at least 30x coverage. Supported references are GRCh37, hg19, hg38. The copy numbers are highly concordant with orthogonal methods such as digital PCR or MLPA.
- The caller can be optionally enabled as part of the Germline pipeline using `--enable-smn true`
- The output is a *.smn.tsv file containing the following columns:

```
#Sample   isSMA   isCarrier   SMN1_CN   SMN2_CN
HG03583   False   True        1         1
```

Uniparental Disomy (UPD) Detection

- V3.9 introduces detection of UPD detection in small VC output. At a whole chromosome level, the small variant caller output can be leveraged to measure the ratio of heterozygous/homozygous SNV sites. A per-contig het/hom ratio is output by default in a `*.vc_hethom_metrics.csv` output when running the Germline SNV VC (`--enable-variant-caller true`). Sub-chromosomal UPD detection and further detection modes are planned for future releases.

Single Cell RNA

v3.9 Introduces three new scRNA features:

- Genotype-free demultiplexing, which assigns sample identities to cells in single-cell RNA-Seq experiments where multiple samples are pooled together in one library.
    - With this method, there is no need to input the samples' genotypes. A reference genotype file from a genetically similar cohort can be used instead. With Genotype-free sample demultiplexing, cells are clustered and genotype probabilities are computed for each cluster, then cells are probabilistically assigned to clusters as in genotype-based sample demultiplexing
    - Command line options: `--single-cell-demux-reference-vcf <reference.vcf>` and `--single-cell-demux-number-samples <N>`
- Feature counting (dell-surface/antibody reads)
    - Feature counting is used for measuring the expression of cell-surface proteins or antibodies along with the expression of genes from the same cell. DRAGEN's feature counting module is flexible since it supports various feature read configurations including the ones that have a dedicated feature Unique Molecular Identifier (UMI).
    - Command line options to specify the feature barcode reference file and read group: `--single-cell-feature-barcode-reference <antibodyReference.csv>` and `--single-cell-feature-barcode-groups grpAntibody`
- Support cell-hashing to identify sample identities
    - Cell-hashing is a new technique for the assignment of sample identities to cells in a single-cell RNA experiment. It is based on tagging a part of the reads from each cell with a specific oligo sequence. DRAGEN's cell-hashing module accepts cell-hashing reads as either separate FASTQ files or as being mixed in the same FASTQ file with transcriptomic reads
    - Command-line options allow the user to select cell-hashing reference file, specify cell-hashing UMI on read 2, and specify cell-hashing read group: `--single-cell-cell-hashing-reference <cellHashingReference.csv>`, `--single-cell-feature-barcode-groups grpCellHashing`, `--single-cell-feature-barcode-r2umi=0_11`

Structural Variant Caller

- New Liquid Tumor mode. Liquid Tumor Mode improves somatic SV recall for tumor/normal analyses where Tumor-in-Normal (TiN) contamination is present. When liquid tumor mode is enabled, SV recall can be substantially improved for tumor-normal sample pairs with TiN contamination. The following example illustrates the improvement, based on usage of germline samples with high quality truth sets (NA12878 and NA24385):

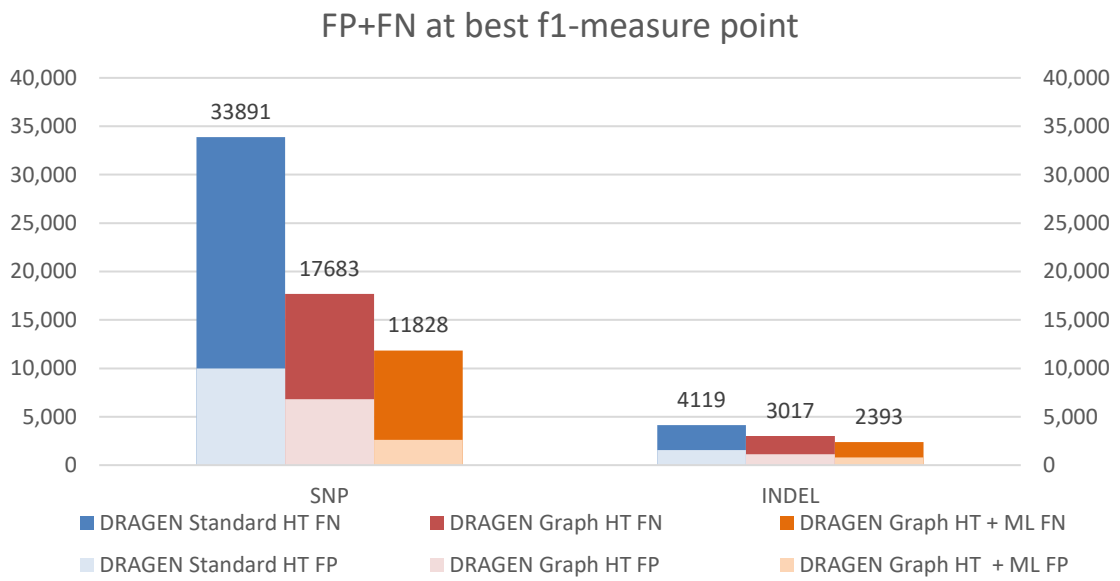
Method	TiN Contamination	Type	Recall	Precision	F-score
Default	0%	Deletions	0.527	0.862	0.654
		Insertions	0.446	0.834	0.58
Default	15%	Deletions	0.07	0.702	0.127

		Insertions	0.075	0.792	0.137
Liquid Tumor Mode	15%	Deletions	0.416	0.905	0.57
		Insertions	0.285	0.867	0.429

- Liquid tumor mode must be enabled by the user. The default tolerated TiN contamination level is 15% but can be adjusted for more heavily contaminated samples via command line option: `--sv-enable-liquid-tumor-mode=true`, `--sv-tin-contam-tolerance 0.25`

Small Variant Caller Machine Learning (ML) [alpha release]

- v3.9 Introduces an alpha release of VC integrated machine learning to refine small variants and quality scores. Powerful ML techniques were applied to produce a model that can optionally be used during germline small variant calling to update QUAL, GT, GQ fields in the VCF.
  - The ML refinement typically filters out approximately half of SNP FPs, with smaller gains on INDELs. FN counts are reduced by 10% or more.
  - Minimal extra processing is done (~5minutes added to run time on-site)
  - Applicable to WGS and WES
  - The original scores and genotypes are output in the VCF under DQUAL, DGT, DGQ fields
- Below results were achieved with the PrecisionFDA HG002 sample with high confidence truth set (training data excludes this sample):

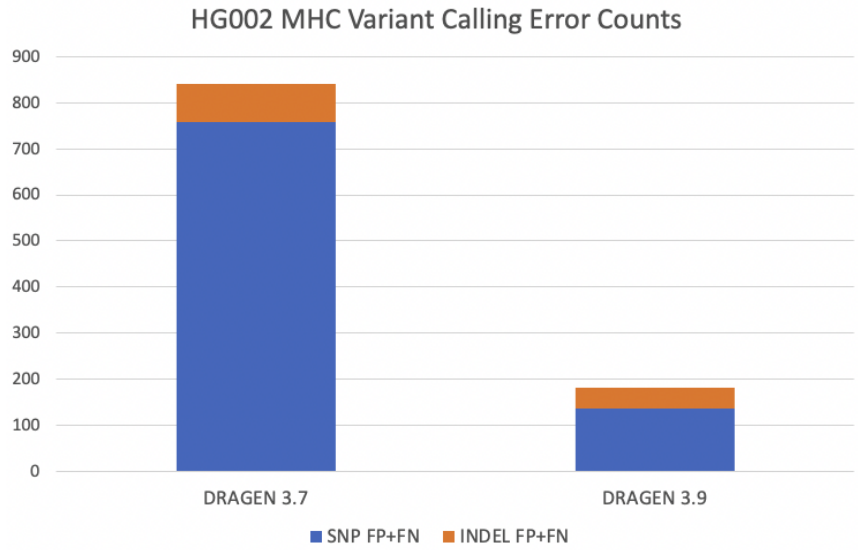


- Command line options to enable ML refinement:
  - `--vc-ml-dir </path/to/package/directory> --vc-ml-enable-recalibration true`
  - The hg38 package is available for download at the Illumina DRAGEN Support Page
  - Note: Only hg38 supported in the alpha release. Hard-filtering is applied at QUAL 3. Multiallelic variants ate not recalibrated.

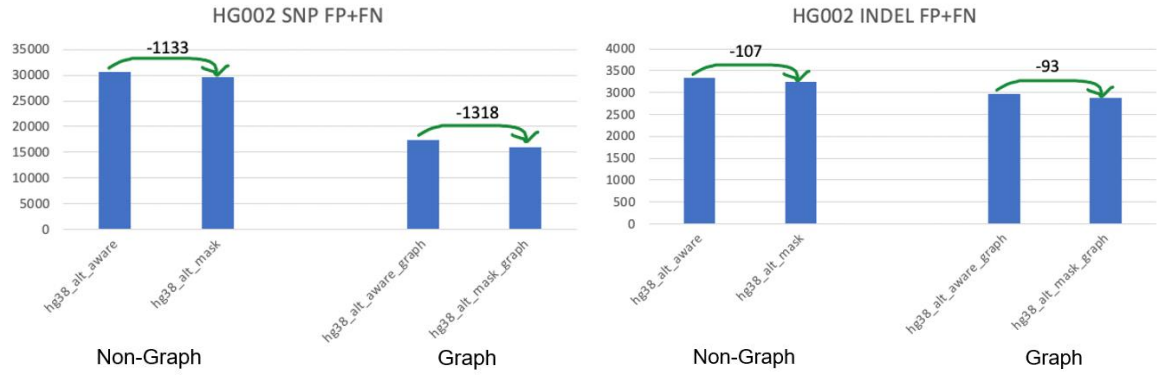
Improvements and Feature Additions

Small Variant Calling Updates

- Graph Improvements for the Major Histocompatibility Complex (**MHC**)
  - DRAGEN v3.7 introduced graph-based read mapping, which reduced read mapping errors and improved SNV/indel calling accuracy in the MHC, for hg19 and hg38 references.
  - DRAGEN v3.9 further improves MHC VC accuracy as follows:
    - More alternative common Alt-contigs have been added across entire MHC region for better read mapping accuracy
    - New 'SD' BAM tag (per-read SNV density) enables improved read filtering for better VC specificity
    - Selective masking of hg19/hg38 native MHC Alt-contigs further boosts VC sensitivity
  - Up to 80% accuracy improvement in MHC region

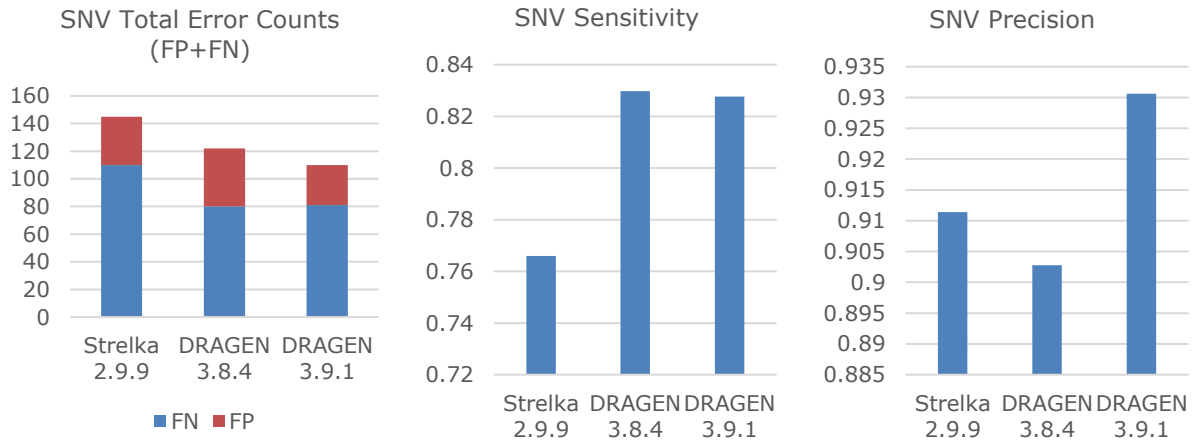


- Native **alt-contig masking** to improve mapping and VC accuracy
  - For genomes with ALT-contigs, DRAGEN has leveraged lift-over of alignments from ALT contigs to corresponding positions on chr1-22,XY to boost accuracy. For ALT-contigs native to hg19/hg38, we now mask ALT-contig regions corresponding to chr1-22,XY regions. Remaining unmasked ALT regions serve as decoys. Handling of DRAGEN graph ALT-contigs added to hg19/hg38 is still done via ALT alignment lift-over.



  - Usage recommendations:
    - **ALT-masking** is now recommended as the default native ALT-contig handling method in v3.9
    - Alt-aware Hash Tables can be re-built using v3.9. By default, hash table builder now uses an internal BED file to mask with N's all portions of hg38 ALT-contigs that are similar to chr1-22,XY.
    - Native ALT-contig masking improves read mapping and variant calling accuracy for both graph and non-graph genomes
    - ALT-masked Graph Hash Tables is available for download on the Illumina DRAGEN Support Page

- Improved precision in **Mitochondrial calling** by using Strelka2 model
  - Mitochondrial calling has been updated from Mutect2-like to Strelka2-like genotyping algorithm, consistent with other DRAGEN somatic pipelines. The improved calling can target allele frequencies down to 1% when coverage permits.
  - The accuracy improvements are benchmarked as shown below, using new pedigree-based truth sets for Platinum Genomes. Results shown are aggregated over 10 different samples (NA12880-NA12888 and NA12893)



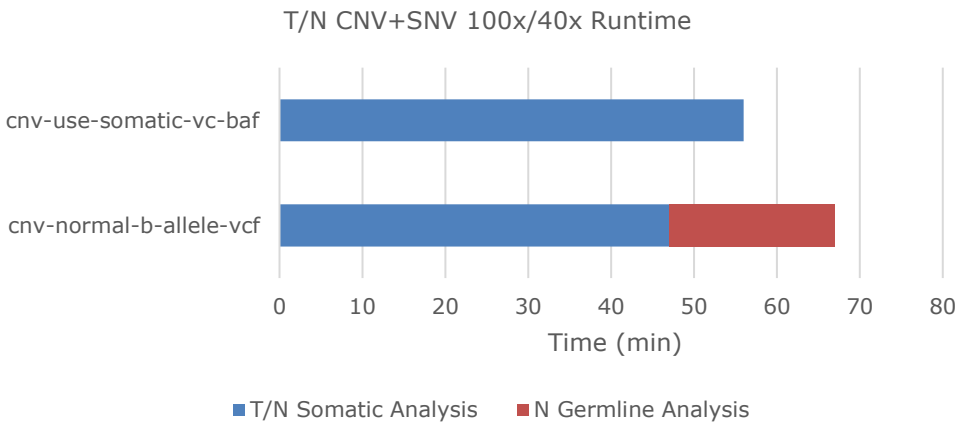
  - **NOTE Format changes:**
    - Call confidence is now reported as SQ (somatic quality, expressed as Phred-scale posterior probability) instead of LOD
    - Multi-allelic calls are now possible (more than one vcf record per position)
  - **NOTE Impact due to format changes:**
    - Combine GVCF has been deprecated in DRAGEN v3.9, and older versions do not support combination of DRAGEN v3.9 GVCFs with these Mitochondrial format change. Instead, GvcfGenotyper is recommended for aggregation of non-pedigree multi-sample GVCF or VCF -or- Joint Genotyping directly from GVCF for small sample sets.
- Multi-Nucleotide Variant (**MNV**) detection
  - DRAGEN supports output of phased variant records in the germline VCF and gVCF file. When two or more variants are phased together, the phasing information is encoded in a sample-level annotation, FORMAT/PS. Phased variant records that belong to the same phasing set can be combined into a single VCF record
  - You can specify the maximum distance between phased variants to be combined using the `--vc-combine-phased-variants-distance` command-line option. The default value is 0, which disables the option. When enabled, the option combines all phased variants in the phasing set that are within the provided distance value. Alternately a BED file could be

provided to specify the phase combination distances over regions, using `--vc-combine-phased-variants-distance-bed <phasing.bed>`.

- **Somatic gVCF** output
  - In contrast to a regular VCF file, a gVCF file covers all positions in the genome, listing both variants and positions determined to be homozygous reference. An artificial <NON_REF> tag is used to represent any (potentially unobserved) non-reference allele. DRAGEN assesses the probability that a position is homozygous reference with respect to a given limit of detection (minimum allele frequency). This new gVCF output can be used to gain confidence in the absence of a variant.
  - By default, the gVCF output is block compressed. Base pair resolution is available at the cost of file size. To enable GVCF output in somatic mode use `--vc-emit-ref-confidence GVCF` (default is VCF), and DRAGEN will output a somatic `*.hard-filtered.gvcf` file.

CNV Caller Updates

- Improved workflow for Somatic CNV in a single execution with SNV
  - In Tumor/Normal analysis, somatic CNV previously required the matched normal to first go through the germline SNV caller, necessitating an additional run. Now, when running concurrently with the somatic SNV caller in Tumor/Normal mode, the matched normal germline heterozygous sites can be leveraged directly by the somatic CNV caller in the same DRAGEN execution. Additionally, somatic SV can also be added in the same execution, simplifying workflows for your somatic analysis
  - Identical purity/ploidy models are selected when compared to legacy mode of operation using a matched normal germline SNV VCF
  - Since the germline analysis of the matched normal sample is no longer required for CNV, the run time of the workflow can be reduced as shown below using HCC1187 Cell Line T/N 100x/40x:



  - Use `--cnv-use-somatic-vc-baf true` instead of `--cnv-normal-b-allele-vcf` to enable this new workflow
- Somatic CNV for Enrichment Applications
  - v3.9 now supports somatic CNV for whole exome and targeted panels. The solution leverages panel of normals to estimate systematic biases due to library prep. Estimation is done on a per-region basis to determine dynamic DUL/DEL thresholds for genotyping. The somatic calling and scoring model for tumor inputs was improved. CNV events are reported as relative copy ratios (fold change) with respect to the panel of normals
  - Can be combined with `cnv-segmentation-bed` for gene level analysis

- No additional command line options are needed, and this is supported with any tumor input mode (FASTQ, BAM, CRAM)

Structural Variant Caller Updates

- Improved somatic Internal Tandem Duplication (ITD) detection
  - V3.9 is even more sensitive to somatic ITDs
  - Somatic ITDs 50 bases and larger can be detected by the SV caller as insertions and tandem duplications. Assembly and scoring have improved for all such ITDs. In tumor/normal mode, ITD sensitivity is further boosted in `hotspot` regions, set by default around the exons of FLT3, ARHGEF7, and KMT2A. Hotspot regions can be disabled or set to a custom region file. For tumor/normal analysis of known FLT3-ITD SVs in TCGA, DRAGEN v3.9 now calls 8 out of 10 SVs, compared to 3 in DRAGEN v3.8
  - Usage:
    - High ITD sensitivity enabled by default. For typical WGS samples, it is no longer required to set `--sv-use-overlap-pair-evidence true`
    - To set custom somatic ITD hotspot regions: `--sv-somatic-ins-tandup-hotspot-regions-bed <BED FILE>`
    - To disable somatic ITD hotspot regions: `--sv-enable-somatic-ins-tandup-hotspot-regions false`
- Higher forced genotyping sensitivity
  - More sensitive genotyping of common SV insertions
  - DRAGEN SV supports genotyping of Insertions, Deletions, Tandem Duplications and Breakends. SVs can be genotyped as a standalone operation or integrated with SV discovery. In v 3.9, insertion scoring improvements yield a 2.4% improvement in insertion recall when genotyping the LRGT SV set on HG002. These scoring improvements also apply to novel SV discovery and somatic ITD calling

Methylation Updates

- HT Building
  - Improved how hash tables are built:
    - Use `--ht-methylated true` to generate a hash table used for multi-pass mode during alignment.
    - Use `--ht-methylated-combined true` to generate a hash table used for single-pass mode during alignment.
    - To build a methylation hash table for both use in single-pass and multi-pass mode, enable both options
    - Single-pass mode now can fully support alt-aware references
  - **NOTE:** Single-pass mode is recommended for Methylation
- Strand-aware dedup
  - More accurate alignment with strand-aware dedup
  - Dedup for methylation has been improved. Methylation dedup now correctly distinguish the top strands reads (ie. original-top or complementary-to-original-top reads) from bottom strands reads (ie. original-bottom or complementary-to-original-bottom reads).
  - This leads to better alignment and increased accuracy in simulated tests
- Smaller reports, improved metrics
  - The output cytosine report can be an extremely large file, which typically contains detailed information about the called methylation status of the cytosines from the input reads. We have enabled compression of the cytosine report with an optional `--methylation-compress-cx-report=true` (default: false) In a typical 30X WGBS run, the output cytosine report is ~30GB, with the compression this is reduced 5X to ~6GB, without much run time impact

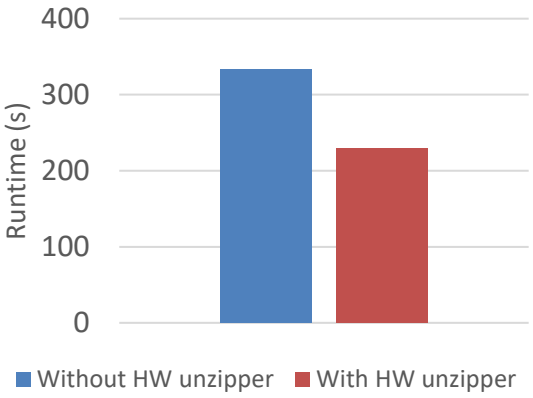
- Bugs in reported metrics have been fixed. Single-pass now correctly emits methyl metrics (eg. methylated C % in CpG), which is concordant with the multi-pass. Multi-pass now correctly emits mapping metrics (eg. aligned read rate), which has been removed from v3.7

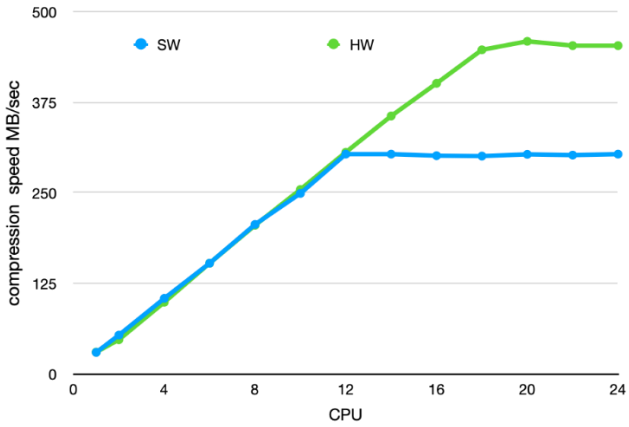
BCL Conversion Updates

- Choose tiles to include or exclude
  - Include only tiles matching a `--tiles` regular expression, or exclude tiles matching an `--exclude-tiles` expression, or both, with one flexible expression syntax for both options. This can be used to exclude problematic regions of a flow cell, produce small data subsets quickly for testing, parallelize tiles across multiple machines
    - Specify lane & tile numbers, with range for any digit
    - Specify multiple terms, separated by +
    - Example usage:
      - Enable first tile of each side & swath: `--tiles [0-9][0-9]01`
      - Exclude first tile of lane 2: `--exclude-tiles s_2_1101`
      - Enable lane 1 through 2: `--tiles s_[1-2]`
      - Enable 2nd tile of first lane & all of lanes 2-8: `--tiles s_1_1102+s_[2-8]`
- Improved metrics
  - Additional metrics provided:
    - Per-read quality stats in new `Quality_Stats.csv` file
    - Added derived stats to all metrics files (percentages)
    - Improved format of Adapter & Demultiplex stats files
    - Increased Top Unknown Barcodes output to 1000
  - All metrics provided in CSV format for easy machine & human parsing

Ora Compression Updates

- Paired-read compression
  - v3.9 implements optional interleaved compression of paired read files. The input can be a list of paired reads. Decompression of an interleaved paired read file outputs the two paired read files. Interleaved paired read file can be used as an input of the DRAGEN mapper when specified with option on command line
  - Example usage:
    - Use both `--ora-input <READ1.fastq.gz>` and `--ora-input <READ2.fastq.gz>` to compress paired reads into one interleaved `.ora` file
    - Use both `-1 <interleaved.fastq.ora>` and `--interleaved` to map/align one interleaved Ora file
- FPGA HW unzipping
  - v3.9 enables use of hardware accelerated unzippers when using the standalone FASTQ to Ora workflow. This offloading of the unzip allows more compute threads to be allocated to the Ora compression and improves the run time for standalone Ora compression. This is enabled by default and can be disabled with using `--ora-use-hw false`





- Input streaming from AWS S3 and Azure Bblob storage
  - This allows easier usage of input files located on AWS or Azure storage. Compression/decompression output is written locally
  - Example usage:
    - `--ora-input s3://path/file.fastq.gz`
- Direct BCL to FASTQ.ora workflow [alpha]
  - V3.9 implements an alpha release of direct from BCL to FASTQ.ora compressed output workflow.
  - To evaluate this feature, add the following command line to the DRAGEN BCL command line: `--ora-reference=/staging/lenadata --fastq-compression-format=dragen`

Gvcf Genotyper Updates

- Sample renaming
  - User can provide a path to a file providing a tab/comma-separated mapping of samples that should be renamed on the merged output.
  - Usage:
    - `--gg-sample-rename-mapfile <file>`, where the format of the file is: `Original,Target[,Filename]` OR `/regex pattern/,substitution`
- gnomAD compliant ChrM output format
  - New VCF output format on chrM. The format is gnomAD compliant with one ALT allele per VCF line at multi-allelic sites.
  - Example chrM output (showing only relevant fields)

```
CHR  POS REF    ALT             FILTER          GT
chrM 301 A      ACC,<NON_REF>   PASS            0/1 0/0
chrM 302 A      AC,<NON_REF>    base_quality    0/1 0/1
chrM 302 A      ACC,<NON_REF>   base_quality    0/1 0/0
```

- User can choose to split output msVCF into a custom set of regions
- Tabix index for output VCF is now generated by default

CYP2D6 Caller Updates

- Improved CYP2D6 with variant phasing
  - Read-backed phasing is used to distinguish variants in CYP2D6 from CYP2D7 gene conversions. Variant sites are phased with nearby sites having a sequence difference between CYP2D6 and CYP2D7 that is fixed in the population

- CYP2D6 caller can now detect 3 additional star alleles (*27, *32, *139), for a total of 128 different star alleles defined by the Pharmacogene Variation (PharmVar) Consortium

Newly supported star allele	Number of samples in test corpus	1000 Genomes Project haplotype frequency
*27	6	0.12%
*32	4	0.08%
*139	1	0.02%

Nirvana Annotation Updates

- The Nirvana packaged with DRAGEN is updated to version 3.16.1. This version adds the following updates:
  - New somatic and germline data sources.
    - FusionCatcher gene fusions including 13 germline fusion sources, 22 somatic sources, and 3 oncogene sources
    - COSMIC cancer gene fusions including detailed information about associated tumors & tissues
  - New annotations for RNA gene fusion
    - Nirvana produces annotated gene fusion calls from Genomic SVs and RNA gene fusion caller output (new)
    - Gene fusions in v3.9 now include improved HGVS RNA notation, in-frame detection from precise SV calls, Gene fusions from non-coding transcripts

Platform Updates

- Ubuntu packages are removed
- v3.9 is the last version to support on-site servers with Phase1 FPGA cards

Interface Changes

Important interface changes to note for v3.9

- Mitochondrial calling updates result in variant format changes (see above)
- Combine GVCF has been deprecated in v3.9, and older versions do not support combination of v3.9 GVCFs with the Mitochondrial format change
- The single-pass methylation mode introduced in v3.8 is recommended for use (see known issues with multi-pass mode)

Issues Resolved

Issues found on DRAGEN™ v3.8.4 that are fixed in v3.9.3

Defect ID	Component	Description
DRAGEN-13647	BCL	Fix for crash when HW-accelerated decompression is enabled for DRAGEN BCL conversion

DRAGEN-13497	Dedup/UMI, SNV VC	Fix to make SNP error calibration give a warning instead of assert on small BED file
DRAGEN-13448	SV	Fix for run-run variation in SV metrics when running ITD samples
DRAGEN-13432	Methyl-Seq	Fixed assertion when --enable-methylation-calling=false is used in conjunction with multi-pass methylation and alt-aware hash tables.
DRAGEN-13383	SNV VC	Fixed issue where large samples with high coverage would hang during VC due to DRAM out of space
DRAGEN-13214	HLA	Fix issue where the HLA BED file being used by default is always grch38, and may not match the reference, when user does not specify BED file.
DRAGEN-13199	Methyl-Seq	Fix Seg Fault during methylation alignment when using alt-masked reference where the entire contig is N's
DRAGEN-13145	Methyl-Seq	Fix issue where single pass methylation may create tags for pairs with one mate ambiguously aligned, when one read has mapq0
DRAGEN-13103	BCL	Feature to allow uppercase TRUE for NoLaneSplitting due to Excel default formatting
DRAGEN-12904	BAM/CRAM	The maximum supported length for individual chromosomes is 512 Mbp (2^29 bases) for the bai/crai indexing format. Automatically disable BAI/CRAI output for unsupported references instead of silently generating invalid bai/crai
DRAGEN-12855	Amplicon	Fix issue where a hang during Amplicon run fills up the /var volume with the hang_diag file
DRAGEN-12834	DNA Alignment	Fix issue where Hard-trimmed reads <20bp does not get filtered
DRAGEN-12739	Metrics	Fix for wigToBigWig crash during metrics compression
DRAGEN-12640	Somatic	Fix crash in dragen somatic contamination, which happened due to a bug in the Eigen::BDCSVD library.
DRAGEN-12605	Force GT, Somatic	Fix crash when using somatic FGT on graph genome
DRAGEN-12534	GC Bias	DRAGEN germline GC Bias fails on E.Coli reference when mock-chromosome names are used, due to DRAGEN treating them as human. Add option to allow user to disable GC metrics.
DRAGEN-12531	Metrics	Update the Autosome Median Coverage calculation, to be the median of all depths over all autosomes, instead of a median of the autosomal contig mean coverages
DRAGEN-12525	Amplicon	Fix issue where unmapped reads were used in amplicon mode for coverage calculations
DRAGEN-12410	BCL	Fix for timeout in validate-sample-sheet mode with high sample count
DRAGEN-12408	CNV VC	Fix for invalid purity selected during VAF modeling, leading to crash during somatic CNV
DRAGEN-12298	BCL	Fix for TopUnknownBarcodes empty when # unknown barcodes is < 100
DRAGEN-12291	Dedup/UMI	Fix UMI watchDog hang issue when very large size fastq file used
DRAGEN-12124	BCL	Fix crash when using bcl-only-read=1 and UMIs in genomic read is 2
DRAGEN-11944	BCL	Fix for incorrect base masking when using TrimUMI,0
DRAGEN-11834	Somatic	Fix issue where DRAGEN somatic VC emits almost all variant calls filtered as 'multiallelic' when vc-sq-filter-threshold=0.0 and vc-sq-call-threshold=0.0

DRAGEN-11797	Somatic	Fix for incorrect GT applied to calls with no coverage
DRAGEN-10772	DNA Alignment	Fix for Wheat sample Out of Memory during map/align
DRAGEN-10432	CNV VC	Add support for CNV WGS single ended analysis

Known Issues

Known issues of the DRAGEN™ v3.9.3 release

Defect ID	Component	Issue Type	Description	Remedy / Workaround
DRAGEN-13896	CNV	Bug	Panel of Normals filtering is applied incorrectly during Somatic CNV analysis	Affects the output of the panel-of-normals feature, which is no longer recommended. Users should use self-normalization for Somatic CNV. The feature will be deprecated in future releases
DRAGEN-13886	SNV VC	Bug	GVCF differences between BAM (no mapping) and FASTQ input to enrichment germline pipeline runs	The issue mainly affects the VCF annotations of the NON_REF calls only. Hence overall low impact.
DRAGEN-13837	Gvcf Genotyper	Usability	We have seen incompatibility of Mito (chrM) output from GvcfGenotyper with Hail	Affects ChrM multi-sample VCF output. Old versions of Hail throw an error complaining about single "." and not ".,.". Recent version of Hail is able to process, but needs to set parameter: {array_elements_required=False}
DRAGEN-13799	Metrics JSON	Usability	Metric values for floating point values are slightly different between CSV files and metrics.json files due to rounding	Rounding differences in the least significant digits. Minor or no impact
DRAGEN-13797 DRAGEN-13796 DRAGEN-13900	Metrics JSON	Usability	Some new CSV metrics in 3.9 are not available in the metrics.json file, and some CSV metrics logged in metrics.json file that are not in datasettype	No impact on existing users of JSON metrics
DRAGEN-13771	Infrastructure	Stability	A DRAGEN job that runs in "software only" mode, and encounters a hang, will not be killed by the watchdog process.	User can safely kill the job manually, since it is not using hardware.
DRAGEN-13752	Somatic ForceGT	Usability	Application of the SOM tag is inconsistent in rare cases, when using Forced GT during somatic runs	To be addressed in subsequent release

DRAGEN-13717	RNA VC, SNV VC	Stability	A very rare HW graph error (seen once every 6-9 months in routine VC testing). VC encounters "ERROR: Invalid node flags"	If seen in field, recommendation is to re-run sample as it is expected to pass. There is a trap in place to catch the HW error, and there would be no incorrect result produced
DRAGEN-13621	Infrastructure	Stability	Low occurrence intermittent issue encountered when an Azure instance is brought up. The PR bitstream switch fails	If the PR failure issue happens, the instance needs to be killed and new instance started.
DRAGEN-13619	CNV	N/A	An accuracy test for one FFPE T/N sample fails our requirement by small margin, due to HET being enabled in the test.	FFPE samples should not run with HET enabled. In case customer has FFPE samples to be processed through CNV, then we recommend cnv-enable-het-calling=false for the processing
DRAGEN-13616	Methylation	Bug	Run-to-run variation is seen for methyl-seq mapping metrics, when using alt-aware references	Existing known issue with mult-pass methylation and alt-ware hash tables. Not a regression from prior releases. Alt-aware references are not recommended for Methyl-seq. The issue is also not present with the new single-pass methylation mode or new alt-masked references
DRAGEN-13493	Infrastructure	N/A	Reference loading step sometimes takes up to twice as long	Intermittent and rare. Adds 30 seconds variance to overall run time
DRAGEN-13476	Somatic	N/A	Somatic SNV VC runtime is slower on 3.9 relative to 3.8	There is an expected 3-6% slowdown on most somatic runs, due to algorithmic updates.
DRAGEN-13459	SNV VC	Bug	Differing variant caller results b/w germline joint workflow from FASTQ and BAM	The issue mainly affects the VCF annotations of the NON_REF calls only. And it is not a regression from prior releases.
DRAGEN-13343	Cloud Licensing	Stability	Intermittent license challenge start failure message with "HTTP error 400: Bad Request" seen on the cloud.	Occasionally, retries are encountered when communicating with license server. No timeout or failure has been observed in testing.
DRAGEN-13339	Map/Align	Stability	A watchdog hang has been encountered during DRAGEN M/A, due to "WARNING: There are more valid reads in file R1 than file R2", when that was not the case	Issue has only been seen once ever and not reproducible. If encountered, user can re-run the job.

DRAGEN-13211	Joint Genotyping	N/A	A decreased concordance in SNP and INDEL between single sample and the sample pulled from a joint called trio, on pre-filtered VCF	The decreased concordance on pre-filtered joint called vs single sample is due to more low-quality calls made in 3.9 vs 3.8 in pre-filtered VCF. They are being filtered in the hard-filtered output.
DRAGEN-13208	RNA	Accuracy	RNA gene fusion has 1 more FP count in 3.9 for one of the test samples	Expected small accuracy change due to bug fix. Still passes accuracy requirements
DRAGEN-12743	Infrastructure	Bug	We have seen a case where the Watchdog does not trigger during a Dragen VC hang, leading to indefinite hang	Very low occurrence and not reproducible. If encountered, User can manually kill the job.
DRAGEN-12601	Methylation	Bug	Methylation mapper intermittently produces non-deterministic global alignments when using multi-pass mode.	The use of single pass methylation is recommended.
DRAGEN-11904	Somatic ForceGT	Bug	Somatic Forced GT outputs a B-allele file with some FGT positions in the B-allele file is not consistent with the VCF.	B-allele frequencies for variants of interest could be determined from the VCF. This is expected to be low impact and will be fixed for future releases.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package
- The archive integrity can be checked using: `./<DRAGEN 3.9.3 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 3.9.3 .run file>`
- Please follow the installer instructions. Server power cycle may be required after installation, depending on the currently installed version. If an updated FPGA shell image needs to load from flash, this is only achieved with power cycle.
    - A power cycle is required when upgrading from v3.3.7 or older
    - A power cycle is required when downgrading to v3.3.7 or older
    - A power cycle is not required when upgrading from a release after v3.3.7
- Procedure to downgrade to v3.3.7 or older:
    - Requires the following three steps. The prior .mcs file needs to be flashed manually:
        - Install the prior release: `sudo sh <DRAGEN 3.3.7 .run file>`
        - `program_flash /opt/edico/bitstream/07*/*.mcs`
        - Power cycle