DRAGEN v4.2.7 Software Release Notes

Introduction

These release notes detail the changes to software components for the Illumina® DRAGEN™ Secondary Analysis v4.2.7 Patch Release.

Changes are relative to DRAGEN™ v4.2.4. If you are upgrading from a version prior to DRAGEN™ v4.2, please review the release notes for v4.2.4 for a comprehensive overview of the new features and changes available in v4.2.

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

The software package includes downloadable installers for Phase 3 and Phase 4 on-site servers:
- DRAGEN™ SW for x86 Oracle 8 - dragen-4.2.7-9.el8.x86_64.run
- DRAGEN™ SW for x86 Centos 7 - dragen-4.2.7-9.el7.x86_64.run

The following configurations containing DRAGEN™ 4.2.7 are also available on request:
- Centos 7 Amazon Machine Images (AMI) for f1 instances, available in 12 regions
- Centos 7 Microsoft Azure Image (VM) available in West US 2
- Centos 7 and Oracle 8 RPM packages for use with Amazon Web Services (AWS) f1 instances, for customer generated AMIs or customer generated docker images
- DRAGEN™ Kernel drivers for el7 and el8, for use with customer generated AMIs or QuickStart
- Pre-built docker images with Centos 7 and Oracle 8 for on-site, AWS usage
- Pre-built docker image with Centos 7 for Microsoft Azure cloud usage

Deprecated platforms:
- Support for DRAGEN™ Server v1 FPGA cards have been deprecated since DRAGEN™ v3.10
- Support for Ubuntu has been deprecated since DRAGEN™ v3.9
- Support for x86 CentOS 6 has been deprecated since DRAGEN™ v3.8

Contents

Overview

DRAGEN™ v4.2.7 is a minor patch release that offers bug fixes on v4.2 as listed below.

For full extensive details on each feature of pipeline, please consult the v4.2.4 release notes and latest Illumina DRAGEN™ Secondary Analysis User Guide available on the support website at https://support.illumina.com/downloads/illumina-dragen-bio-it-platform-user-guide.html

Resource Files

DRAGEN™ 4.2 releases require updates to key resource files to function correctly and achieve the optimum performance. All resource files are available for download at the Illumina DRAGEN™ Product Files support site here: https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/product_files.html

The following resource files are updated for DRAGEN v4.2 (v4.2.4 and later):

Resource	Description	File name(s)
Hash Tables v9	Pre-built v9 multigenome hash tables for hg38, hg19, hs37d5, CHM13. The hash table builds include DNA, RNA, CNV, HLA tables.	hg38-alt_masked.cnv.graph.hla.rna-9-r3.0-1.tar.gz hg19-alt_masked.cnv.graph.hla.rna-9-r3.0-1.tar.gz hs37d5-cnv.graph.hla.rna-9-r3.0-1.tar.gz chm13_v2-cnv.graph.hla.rna-9-r3.0-1.tar.gz
SNV Systematic Noise Baseline collection v1.1.0	A collection of noise baseline BED files for hg19, hs37d5, hg38 and for WGS and WES respectively	systematic-noise-baseline-collection-1.1.0.tar The tar archive contains the following files: snv_wes_nextera_hg19_max_v1.1_systematic_noise.bed.gz snv_wes_nextera_hg19_mean_v1.1_systematic_noise.bed.gz snv_wes_nextera_hg38_max_v1.1_systematic_noise.bed.gz snv_wes_nextera_hg38_mean_v1.1_systematic_noise.bed.gz snv_wes_nextera_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wes_nextera_hs37d5_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg19_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg19_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg38_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg38_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hs37d5_mean_v1.1_systematic_noise.bed.gz snv_wgs_hg19_max_v1.1_systematic_noise.bed.gz snv_wgs_hg19_mean_v1.1_systematic_noise.bed.gz snv_wgs_hg38_max_v1.1_systematic_noise.bed.gz snv_wgs_hg38_mean_v1.1_systematic_noise.bed.gz snv_wgs_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wgs_hs37d5_mean_v1.1_systematic_noise.bed.gz
SV Systematic Noise Baseline collection v2.0.0	A collection of noise baseline BEDPE files for hg19, hs37d5, hg38 for WGS	sv-systematic-noise-baseline-collection-2.0.0.tar The tar archive contains the following files: WGS_v2.0.0_hg19_sv_systematic_noise.bedpe.gz WGS_v2.0.0_hg38_sv_systematic_noise.bedpe.gz WGS_v2.0.0_hs37d5_sv_systematic_noise.bedpe.gz
Custom Multigenome Reference Builder resources v1.1.0	Fasta, graph BED, mask BED files for hg38, hg19, hs37d5, CHM13, needed for custom multigenome	hg38-custom-reference-genome-1.1.0.tar.gz hg19-custom-reference-genome-1.1.0.tar.gz hs37d5-custom-reference-genome-1.1.0.tar.gz chm13_v2-custom-reference-genome-1.1.0.tar.gz

	hash table building from own population VCFs	
Imputation Reference Panels v1.2 and v2.0	Genetic maps and reference panels for hg38	`irp-hg38-1.2.1.tar` `irp-hg38-2.0.0.tar`

NOTES:

- ML Model files for DRAGEN™ v4.2 are now included in the installer by default and does not need to be downloaded.
- Multigenome references can now be built with the hash table builder. Pre-built hash tables are provided for reference.
- When upgrading from v4.1 or older, hash tables must be re-built to use DRAGEN™ 4.2. Existing hash tables built with v4.1 or older are not supported.

Changes

Changes made, and/or issues that are fixed in the DRAGEN™ v4.2.7 release:

Comp	ID	Description
Amplicon SNV VC	DRAGEN-26438	Amplicon VC parameters have been updated to make the pipeline more robust for new panels, instead of only Ampliseq. Read position filter is disabled, low fraction reads filter is set, and proper pair flag is set after amplicon soft clipping
BCL	DRAGEN-26567	Fix for BCL abort with "number of decompression threads must be at least one"
BCL	DRAGEN-26020	Fix BCL abort without an error message, when --bcl-only-lane is set to a lane not included in SampleSheet or RunInfo.xml
BCL	DRAGEN-24476	Improve BCL compute performance for very high sample counts (150K)
BCL	DRAGEN-28210	Fix for masking of R1 resulting in missing entries in fastq_list.csv file
BCL	SET-7028 DRAGEN-21364	Fix for BCL allocating insufficient memory for larger index2 reads. When index2 (i5) has more than 18 cycles, there was under allocation of buffer space leading to crash. Note that this change increases support to 27 bases per index, rather than 27 bases total.
BCL	DRAGEN-26294	Fix for bcl-convert not outputting all fastqs when CreateFastqForIndexReads is on, and some samples have fully masked indices
Check Fingerprint	DRAGEN-26754	Add support for observed_sample vcf file as input, to allow comparison of observed_sample vcf vs expected_sample vcf without need for dragen aligner to generate the observed_sample.
Cloud	DRAGEN-26711	Fix issue where FPGA shell version check leads to false crash on AWS.
Cloud	DRAGEN-25774	Allow space character(s) in URL path for input file streaming.
Cloud	DRAGEN-26199	Fix a bug with occasional failure reading Azure identity document.
Cloud	DRAGEN-28102	Fix for hang when input streaming large BAM files over S3 and http.
CNV	SET-5997, DRAGEN-26376	Do not output panel_of_normals filter meta-information line in VCF header when PON is not enabled.
CNV	DRAGEN-27562	Enable filtering of duplicate alignments in b-allele counter
CNV	DRAGEN-23062	Fix to CNV contig skip logic, so that running from BAM or FASTQ produce the same counts.
CNV	DRAGEN-28053	Fix to make option "cnv-enable-gcbias-correction=false" really disable GC-correction in Somatic CNV (WGS)
CNV	DRAGEN-28705	Fix CNV stats overflowing 32bit variable for very high coverage samples.
CNV	DRAGEN-27924	Fix hang during segmentation

CNV/SV	DRAGEN-26423	CNVSV Filter large SVCLAIM=J records
File IO	DRAGEN-27216	Fix file write error on CIFS/Samba storage
HLA	DRAGEN-26627	Fix to reload the HLA reference for back-to-back runs, to avoid corruption of the HLA reference
HLA	DRAGEN-27177	Fix anchored hash table address initialization for HLA, which lead to run-run differences
iGG	DRAGEN-27731	Replace problematic characters "-" ":" "*" with "_" in msVCF region name
iGG	DRAGEN-27846	Write . (missing) if there is no QUAL value for a site
iGG	DRAGEN-24248	Fix hang on RHEL8
Imputation	DRAGEN-27380	Skip imputation when there is only a single position in a chunk, and error when there are no chunks for a region
Imputation	DRAGEN-26340	Fix excessive logging, by limiting number of warnings "no genotype likelihood in the target dataset" is printed.
Imputation	DRAGEN-26091	Support for regions called "chrX_par1" or "chrX_nonpar" in chunk regions
Infra	DRAGEN-27496 DRAGEN-26321	Improve robustness of dragend daemon during aborts and resets
Infra	DRAGEN-27060	Migrate daemon logs for dragend to syslog, add to logrotate list
Infra	DRAGEN-15910	Add dragen_licd.log to logrotate list
Infra	DRAGEN-26903	Replace IntervalTree with cgranges to lower cpu usage in some callers
Infra	DRAGEN-28627	Fix segfault on FASTQ list with empty string
Infra	DRAGEN-28711 DRAGEN-26184	Fix "Received Alarm clock signal, aborting" crash, when input steaming from https
Licensing	DRAGEN-26961	Fix TSO500 usage reporting on cloud
Licensing	DRAGEN-28145	Fix issue where dragen_licd daemon termination cause incomplete dongle transactions
Map/Align	DRAGEN-26836	Fix POS/PNEXT mismatch when PE-overlap trimming is enabled in the mapper
Map/Align, HLA	DRAGEN-27056	Fix to allow HLA to be enabled together with graph alignment (ga) tag
Map/Align, SW mode	DRAGEN-26592	Fix a hang in Software Mode (Beta) mapper during insert stats estimation
Methylation	DRAGEN-26954	Fix a crash in Methylation Multi-pass pipeline
MSI	DRAGEN-27583	Support gzipped MSI microsatellites input file

ORA	DRAGEN-27894	Generate ORA.md5sum file(s) when ORA is enabled with map-align
ORA	DRAGEN-28208	Allow ORA Compression jobs to run in parallel with a dragen FPGA job
Pheno HRD	DRAGEN-27240	Modify PhenoHRD b-allele file format to be compatible with visualizations
QC Metrics	DRAGEN-26806	Fix to print the missing "COVERAGE SUMMARY" metrics to stdout for Germline runs
RNA	DRAGEN-26411 DRAGEN-26977 DRAGEN-26825	Fix variation in output of RNA Gene Fusion on cloud vs. local
scRNA	DRAGEN-28210	Fix to scRNA with fastq_list.csv. Where fastq_list.csv has only Read2 files, the Read2 files get cleared leading to run error with no input data.
SNV VC	DRAGEN-18272	Fix rare cases where some tabix files produced by DRAGEN can't be opened with htslib
SNV VC	DRAGEN-27189	Fix for incorrect insert stats summary metrics with VC evidence BAM
SNV VC	DRAGEN-27500	Fix segfault when using force genotype, add checks for haploid FGT genotypes
SNV VC	DRAGEN-27582	Disallow enabling ML in somatic VC run
Somatic VC	DRAGEN-26915	Add tumor pruning option "vc-enable-tumor-pruning-fraction". Setting the option to true causes the pruning threshold to be calculated based on only the tumor depth instead of the tumor + normal depth for the VC assembly graph for somatic tumor/normal mode
Somatic VC	DRAGEN-21168 DRAGEN-23627	Significant run time improvement for Somatic T/N when BAM/CRAM output is enabled. Reduce run time by 1 hour for 80x-40x run.
SV	DRAGEN-26993	Add sorting criteria in VCF output using SR and PR support counts. This added criterion resolves the challenge posed by duplicate SV candidates and outputs the best SV in the final Tumor SV VCF based on the number of read support. This change does not affect the accuracies of benchmarking samples but will recover some FNs in some specific samples.
SV	DRAGEN-27511	Add processing rate during genome scan and edge processing to stdout, to help debugging of slow runs
SV	DRAGEN-27575	Revert two parameters that control graph complexity, to resolve excessive SV run times
SV	DRAGEN-27014	Fix an SVCLAIM=D CNV call with a MatchSv, to be SVCLAIM=DJ
TMB	DRAGEN-28990	Fix issue where DRAGEN does not error out when germline tagging is enabled but "--tmb-skip-db-filter=true"
Software Mode	DRAGEN-27826	Fix a timeout during Software Mode (Beta) license checking.

Software Mode	DRAGEN-26584	Fix intermittent crash during Software Mode (Beta) license checking.

Known Issues

Known issues of the DRAGEN™ v4.2.7 release

Comp	ID	Summary	Resolution/Workaround
Amplicon	DRAGEN-29540, SET-7374	In Amplicon analysis, a validated deletion at the end of the read was not present in alignment	No workaround. A fix has been made to future version.
Amplicon	DRAGEN-29572	In Amplicon analysis, there is an FN at the edge of amplicon targets if "vc-remove-all-soft-clips" is false.	Enable "vc-remove-all-soft-clips" to remove reads supporting the event.
BCL	DRAGEN-29926, SET-6881	BCL Convert crash with "ptr != dmindex.map.end" when checking per-sample barcode mismatches on valid barcode combinations.	Workaround to set BarcodeMismatchesIndex1,2 to 0. A fix has been made to future version.
CNV VC	DRAGEN-28695	Up to v4.0, if we cannot detect a purity or ploidy (DEGENERATE_DIPLOID) then we would FAIL all records in our VCF. In v4.2, we changed this behavior for WGS, but not for WES.	No workaround. A fix has been made to future version.
CNV VC	DRAGEN-29533	New since v4.2 a CNV standalone run (BAM->CNV only) will consume license bases, per design decision.	Enable CNV with FASTQ->VCF analysis to avoid double charge. The behavior will be reverted in future version due to customer request.
Compr	DRAGEN-24400	Runs on Azure occasionally crash with "corrupted size" message after streaming of ORA compression/decompression finishes.	No workaround
Compr	DRAGEN-26451	CRAM decompress & map/align with different references, can falsely run into an alt contig error check and crash, when hash table is used for cram decompression.	Alt contigs are erroneously counted on both references and can exceed a threshold. Use FASTA for CRAM decompression instead of hast table. A fix has been made to future version.
Compr	DRAGEN-28133	Running Ora with output files present and without the --force option has inconsistent exit codes	No workaround. Exit code for Ora compression sometimes 1 sometimes 11.
Down sampling	DRAGEN-19385	Down sampling from BAM input has a chromosome coverage bias. This is not the case when the input is FASTQ. The average coverage is the same. This impacts accuracy when using BAM input and down sampling.	No workaround. Future version has new down sampling implementation

Gene Fusion	DRAGEN-29181	RNA filter info and candidate output has a minor run-run variation on Azure Cloud.	No workaround. Does not affect the accuracy. Not present on AWS or Local.
Gene Fusion	DRAGEN-29582	The option "rna-gf-enable-read-share-cleanup" is missing cleaning up some paired reads.	No workaround
GVCF Genotyper	DRAGEN-21091	When a site is missing in the input gVCF file for a sample and the site is output to the msVCF file, the genotype is coded as missing using '.', i.e. haploid	No workaround
GVCF Genotyper	DRAGEN-26325	GG does contig name truncation on HLA* alt contigs to the first colon. This could lead to incorrect outputs for those contigs	It is a long-standing issue we are highlighting. No workaround. Fix planned for future version
Gvcf Genotyper	DRAGEN-26768	iGG fails with input gVCFs generated by pre-3.3 DRAGEN	Since iGG v4.2, we will fail with GVCF inputs from pre-3.3 DRAGEN.
Gvcf Genotyper	DRAGEN-26929	Unnormalized indel variants on Gvcf Genotyper msVCF output	There are some additional FN indels in the msVCF that are not in the input gVCF, due to unnormalized indel variants for indels of certain type.
Hash Table Builder	DRAGEN-26399	Hash table decompression error on some FASTA input files.	Use option to write the hash table uncompressed during build. The uncompressed hash table is valid.
Inputs	DRAGEN-26218	Map/align errors if r2 FASTQ file contain more reads than r1 (not expected), but runs ok if r1 FASTQ contains more reads than r2 (expected)	Trim the FASTQ pair to contain the same number of reads.
Joint Genotyping	DRAGEN-24805	Higher number of denovo SNP calls observed in some trios since v4.2	Not a bug, this is expected. For information only.
Multi genome reference	DRAGEN-26308	WGS runtime increased with multigenome vs legacy genome	5% longer run time. For information only.
Paralog Caller	DRAGEN-25971	GBA reports a single recombinant haplotype with RecNciI+RecNciI instead of two recombinant haplotypes with RecNciI each for NA20273	Not a bug in the caller, but output formatting may be misleading. Format change planned for future version
Population Haplotyping	DRAGEN-25979	Non-deterministic output. Different output VCF PREFIX.ph_phase_common.vcf.gz every time it is run.	This is a feature of shapeit5 tool integrated in DRAGEN. For information only
RNA Quant	DRAGEN-24824	RNA quant - SJ.saturation.txt has minor differences with different num-threads value	No workaround. Negligible impact.

SNV Germline	DRAGEN-23801	Joint Calling in Mito is not giving proper VAF's, when one or more samples have a variant, but other samples have a homref call at the same position.	For some alleles, the AD values in the joint VCF are not accurate. Looking at the corresponding single sample gVCF can resolve the inconsistency.
SNV Germline	DRAGEN-26359	Small regression in INDEL sensitivity in v4.2 compared to v4.0	Single sample SNV has major accuracy improvements due to graph and reference and machine learning updates. In some cases, such as INDEL, there are very minor changes where precision may improve at cost of recall, and vice versa. For information only.
SNV Germline, MNV	DRAGEN-25661	Germline MNV - In some cases, germline variants within the phased range distance will not be combined into a single MNV event.	No workaround
SNV Somatic	DRAGEN-22241	Some T/N and T/O samples have >5% runtime regression relative to v4.0	No workaround
SNV Somatic	DRAGEN-24060	Small regression in INDEL FP across most T/O WGS and WES test datasets	Single sample SNV has major accuracy improvements in v4.2 due to graph mapper and reference updates. In some cases, such as INDEL, there are very minor changes where precision may improve at cost of recall, and vice versa. For information only.
SNV Somatic	DRAGEN-29115	Different number of variants in postfilter vs prefilter VCF when both MNV and Germline Tagging is enabled, due to IGNORE flag	For information only
SNV Somatic	DRAGEN-29580	Padding an interval does not always produce the same results as running with larger interval, for somatic VC.	Specific to Somatic VC. Use larger regions instead of padding. A fix has been made to future version.
SNV Somatic, MNV	DRAGEN-25710	Somatic SNV T/O MNV failing to merge two MNV calls, in the the edge case where we have a deletion upstream of another co-phased variant with an out-of-phase snp in between them that is covered by the REF allele of the upstream deletion.	No workaround
SNV VC	DRAGEN-25905	Hang observed on high depth samples, when target BED is used to run the SNV caller over regions which are close to the end of a chromosome.	Have more BED regions throughout the chromosome or increase bin memory.
SNV VC	DRAGEN-25933	VCF GQ values may not match VCF specification	In most positions, the probability that the position is a variant is very close to 1 and the impact is negligible. In corner cases where p(0/0) is not negligible, we have the wrong value in the GQ field. For information only.

Software Mode	DRAGEN-28949	Software Mode (Beta) encounters a license max timeout 99 error on some systems.	No workaround. A fix has been made to future version.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package.
- The archive integrity can be checked using: `./<DRAGEN 4.2.7 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.2.7 .run file>`

Release History

Revision	Release Reference	Originator	Description of Change
00	1096732	Cobus De Beer	Initial release