DRAGEN v4.2.9 Software Release Notes

Introduction

These release notes detail the changes to software components for the Illumina® DRAGEN™ Secondary Analysis v4.2.9 Patch Release.

Changes are compared to DRAGEN™ v4.2.7. If you are upgrading from a version prior to DRAGEN™ v4.2, please review the release notes for v4.2.7 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-4.2.9-9.el8.x86_64.run
- DRAGEN™ SW for x86 Centos 7 - dragen-4.2.9-9.el7.x86_64.run

The following configurations containing DRAGEN™ 4.2.9 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 (Amazon Machine Image) or customer generated docker images
- DRAGEN™ Kernel drivers for el7 and el8, for use with customer generated AMIs or QuickStart

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 4
Resource Files 4
Changes 6
Known Issues 7
SW Installation Procedure 11
Release History 11

Overview

DRAGEN™ v4.2.9 is a minor patch release that offers bug fixes and improvements on v4.2 as listed below.

For full extensive details on each feature of pipeline, please consult the v4.2.4 and v4.2.7 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

When upgrading from prior versions, please note that DRAGEN™ 4.2 releases require updates to key resource files to function correctly and achieve the best 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 (Structural Variants) 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	Fasta, graph BED, mask BED files for hg38, hg19,	hg38-custom-reference-genome-1.1.0.tar.gz hg19-custom-reference-genome-1.1.0.tar.gz

Reference Builder resources v1.1.0	hs37d5, CHM13, needed for custom multigenome hash table building from own population VCFs	`hs37d5-custom-reference-genome-1.1.0.tar.gz` `chm13_v2-custom-reference-genome-1.1.0.tar.gz`
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 (Machine Learning) 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

The following issues are fixed since DRAGEN™ v4.2.7.

Issues that are fixed in the DRAGEN™ v4.2.9 release:

Comp	ID	Description
RH Caller	DRAGEN-25216 SET-8666	Fix crash in RH caller on ultra-low coverage samples
RNA Gene Fusion	DRAGEN-32116	Improve Gene Fusion accuracy, and fix accuracy regression from DRAGEN™ v3.10

RNA Gene Fusion accuracy had regressions on some sample types compared to v3.10. This has been fixed and improved.

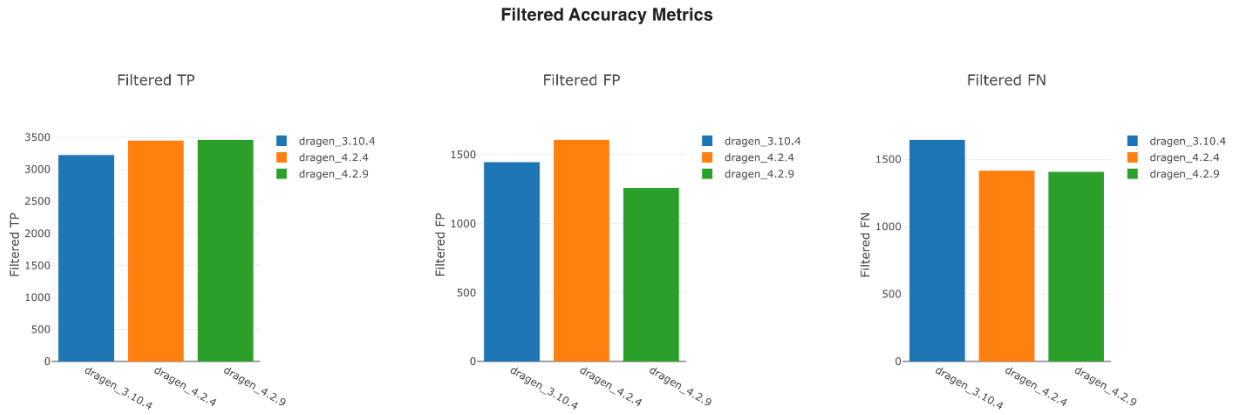

Figure 1: Post-filter accuracy of RNA Gene Fusions over 205 test samples.

Changes made, and/or issues that are fixed in the DRAGEN™ v4.2.7-6 release for NextSeq1k2k, which are also included in v4.2.9:

Comp	ID	Description
Instrument workflow	DRAGEN-30506	NextSeq1k2k on-instrument Enrichment and Amplicon runs did not pass in metrics needed for correct read length estimates for a later step. If running any sample that is not 150bp the VCF output would mismatch between instrument and cloud. Workflow command line change.
Instrument workflow	DRAGEN-30504 DRAGEN-30347	NextSeq1k2k on-instrument Somatic Enrichment and Somatic Amplicon runs enabled sorting twice, leading to longer run times. Workflow command line change.

		Also needed a DRAGEN code change to use defaults, to prevent crashing on samples that have too small number of reads to an output estimates file used in a later step.
scRNA	DRAGEN-30614	Single cell RNA crashes during fastq_list processing for use case where UMI are in read and only read2 files present.
scRNA	DRAGEN-30663	NextSeq1k2k on-instrument single cell RNA workflow does not clean up "scrna.BAM" files when output is set to "none"

Known Issues

Known issues of the DRAGEN™ v4.2.9 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.
Amplicon	DRAGEN-31298	Target coverage metrics are incorrect for tiling amplicons (overlapped designed amplicons)	No workaround.
Amplicon	DRAGEN-32594	Amplicon alignment on tiling amplicons (overlapped designed amplicons) have elevated FPs	No workaround.
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.
BCL	DRAGEN-30269	Some combinations of indexing schemas have a significant run time regression compared to DRAGEN v3.7. For example, Mixed: single 6bp, single 8bp and dual 8bp+8bp index type on 25B flowcell increased run time from 4h to 9.5h	The issue is due to excessive allocator usage. A workaround to improve the run time on-premises is to tcmalloc. tcmalloc must be installed, then set: " LD_PRELOAD=/usr/lib64/libtcmalloc.so.4 dragen <parameters...>"
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 (Allele Depth) values in the joint VCF are not correct. 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.
SV Caller	DRAGEN-25649	Small increase in FP and FN for fusion calls on TSO500 assay using DRAGEN v4.x	No workaround
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 4.2.9 .run file> --check`"
- Install the right file based on your Linux OS with the command: "`sudo sh <DRAGEN 4.2.9 .run file>`"

Release History

Revision	Release Reference	Originator	Description of Change
00	1105388	Cobus De Beer	Initial release