DRAGEN v4.1.7
Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Bio-IT Platform v4.1.7.

Changes are relative to DRAGEN™ v4.1.5. If you are upgrading from a major version prior to DRAGEN™ v4.1, 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 software package includes downloadable installers for Phase 3 and Phase 4 on-site servers:
- DRAGEN™ SW for x86 Centos 7 - dragen-arch2-4.1.7-9.el7.x86_64.run
- DRAGEN™ SW for x86 Oracle 8 - dragen-arch2-4.1.7-9.el8.x86_64.run

The following configurations containing DRAGEN™ 4.1.7 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

Below is a summary of the changes included in v4.1.7. This is a minor update to DRAGEN™ v4.1:

- Important bug fixes across features. All bug fixes released with DRAGEN™ v4.0.5 have also been included in this 4.1.7 release.
- Robustness and usability improvements for NovaSeq-X on-instrument analysis.
- Enabling some additional callers in the WGS workflow.

Updates and Fixes

NovaSeq-X on-instrument

- The changes and/or fixes listed in the sections below apply to server, cloud, and on-instrument workflows.
- Various improvements have been made to address on-instrument system robustness.
- Improvements to the Sample Sheet Validator, to avoid blocking runs on unexpected user settings:
  - Allow any unknown key,value in the [BCLConvert_Settings] section of the sample sheet.
  - Allow any unknown column header in the [Data] section of the sample sheet.
  - Allow non-numeric values such as "na" for numeric fields in the sample sheet.
- Fixed several issues relating to mixing single and dual index samples using per-sample-settings in BCL, as detailed below.
- Enable SMN, GBA, CYP2B6 callers in the WGS Germline workflow when all callers are enabled.

BCL Conversion

- Make the combined index collision checking default to enabled for all lanes. Implement a new `IndependentIndexCollisionCheck` option to replace `CombinedIndexCollisionCheck`
  - This important change reverts a strict check on dual index collisions added to BCL based on customer feedback. With this change, the default behavior matches bcl2fastq2 and adds an option to change the behavior.

DRAGEN™ version	Index collision check behavior
3.9.x	Relaxed by default. No option to change. Matches bcl2fastq2
3.10.x and 4.0.x	Strict by default. No option to change.
4.1.5	Strict by default. New option `CombinedIndexCollisionCheck` introduced to optionally relax the strictness
4.1.7 and 4.2.x	Relaxed by default. Remove `CombinedIndexCollisionCheck` option, add new `IndependentIndexCollisionCheck` option to allow optional strict checking. Default matches bcl2fastq2

- Fix for index sequences missing from fastq headers when using `--no-sample-sheet` setting.
- Fix for BCL behavior being different than bcl2fastq2 with respect to "`Sample_Name`" and "`Sample_Project`". In the special case of "`Sample_Name`" == "`Sample_ID`", bcl2fastq2 does not create a "`Sample_ID`" subdirectory. This change makes bcl-convert behavior the same.
- Fix for false barcode collision reports when one sample's index is entirely trimmed out and another sample's index exists.
- Fix for BCL not aborting when single-index datasets have barcode collisions.
- Fix for incorrect yieldQ30/qscoresum stats when there is UMI in the first part of a read and TrimUMI is enabled (true by default).
- Fix for a false error when using global BarcodeMismatchesIndex2 and a sample does not use index BarcodeMismatchesIndex1, when the sample sheet contains both single & dual-index samples.

- Fix for BCL failing with a "`vector::reserve`" message for mixed index strategies.
- Fix for BCL outputting many duplicate error messages for missing CBCL files.
- Fixes related to the Per-Sample Settings feature introduced with the NovaSeq-X instrument:
    - Fix for extra metrics being output when running with Per-Sample Settings on a NextSeq550 dataset.
    - Fix a validation bug where Per-Sample Settings incorrectly flags errors when any read (genomic or index) is fully masked in one or more samples, but not in all samples. For example, an inconsistently fully-masked genomic read can cause a spurious error message indicating that AdapterRead{1,2} must be specified or not specified for a sample. An inconsistently fully-masked index read can cause a spurious error message indicating that BarcodeMismatchesIndex{1,2} must be specified or not specified. This error can be wrong and prevents conversion from continuing and exits with an error code.
    - Fix for mixed single & dual-index samples with combinatorial inputs in a lane via Per-Sample Settings, to cause some collisions to go undetected.
    - Fix for mixed single & dual-index samples in a lane using Per-Sample Settings not working properly, sometimes resulting in missing output for the single-index sample.
    - Fix for too-long-index-reads error "`No more than 27 total bases can be used as index bases`" when using fewer than 27 consecutive bases.
    - Fix for Top Unknown Barcodes output listed cycles being based upon the first sample listed, and not necessarily including all bases being used for indexing, when Per-Sample Settings are used.
    - Fix for Per-Sample Settings not isolating lanes when determining index cycles.
    - Fix for Per-Sample vs Global Settings in BCL producing different FASTQs and Demux Metrics when variety of reads are fully masked.

Germline Small Variant Caller

- Fixes related to Machine Learning (ML):
    - Fix an issue where the computation of PL from GP and PRI is missing, for hethom calls where ML prediction does not match the VC call.
    - Fix some accuracy discrepancies between runs in VCF vs gVCF output mode when ML is enabled.
    - Fix handling of `PL` and `GP` in `0/0` calls, which lead to an accuracy regression on Joint Calling.
- Fix an issue where some variants are not emitted, when evidence BAM is enabled.
- Fix an issue where all reads are disqualified in regions with ForceGT only events.
- Pedigree Joint Calling: Improve denovo SNV INDEL performance.

Somatic Small Variant Caller

- Fix a memory leak during on-sequencer enrichment somatic runs leading to potential out-of-memory errors.
- Fix an out-of-memory error when evidence BAM is enabled on high depth samples.
- Refactor TMB and Germline Filtering, to reduce peak memory usage and resolve out-of-memory issues.
- Fixed an issue where the MNV length overflows a variable, leading to a corrupted TAG and a downstream Germline Filtering that asserts.

Structural Variant Caller

- Fix a segmentation fault in Tumor Only mode due to long assembly size causing a 32bit integer overflow.

CNV Caller

- Remove unwanted assert during input file checking for Panel of Normals.

Targeted Callers

- CYP2D6: Fix for on-instrument Germline workflow exceeding memory threshold.
- GBA: Fix for GBA regression for LB-01223

RNA

- Fix for assert when RNA + down sampling is enabled and the input files are empty. Allow DRAGEN™ to handle empty input without crash.

Single-Cell

- Fix for a missing column for Feature/Peak ID in scRNA/scATAC output, causing compatibility issues for downstream tools.

Mapper and Aligner

- Fix for incorrect CIGAR string produced by mapper, leading to crash in Variant Caller. The issue was only present when using specific mapper settings for PE overhang trimming.

Gvcf Genotyper

- Fix a memory leak during input VCF reading.
- Fix for unnormalized variants in msVCF output of Lettuce samples, when `--gg-discard-ac-zero=true`.
- When writing to allele counts and frequencies to the output msVCF file in some circumstances non-ref values were not correctly processed.
    - This occurred when the global ref allele is different from the batch ref allele. The non-ref allele is represented by the symbolic base sequence 'X' which does not change under right renormalization of the base sequences when the ref allele is lengthened to match the global equivalent. As such, no-ref must have been treated separately. Fix for this issue.

Other Bug Fixes

- Fix for incorrect HLA genotyping output format when minor allele has insufficient support.
- Fix crash in down sampler when HLA is enabled.
- Fix overflow of 16-bit number when aggregating insert stats numerator value across many read groups.
- Fix for failed uninstallation of DRAGEN™ versions 3.0 to 3.3.
- Fix for license server challenge error on Microsoft Azure cloud, due to rare race condition.
- An invalid check for 10 required columns for the `--qc-cross-cont-vcf` file header leads to an exception. Fixed the check to require 8 columns. Also improved error handling for invalid file inputs, with clearer messages.
- Fix for watchdog not stopping a hanging process on the cloud.

Known Issues

Known issues of the DRAGEN™ v4.1.7 release

Comp	Issue ID	Summary	Resolution / Workaround
BCL	DRAGEN -26566	When sample sheet has same sampleID in the same lane multiple times, but with different output files (e.g. R1 fully masked out in one entry, but not in another), the validator fails to detect this case and does not error out. Subsequent on-instrument secondary analysis fails in fastqc generation	No workaround, except to change sample sheet to make output files match. Handling of this case planned for future version.
BCL	DRAGEN -26220	When using mixed indexing strategies, the index hopping counts .csv metrics for Undetermined reads may differ slightly between bcl-convert and NovaSeq-X on-instrument	No workaround. Fix planned for future version
BCL	DRAGEN -25363	BCL omits lines with zero reads in Demux tile stats and Quality tile stats .csv metrics	No workaround. Fix planned for future version
BCL	DRAGEN -23388	BCL will crash when "--no-sample-sheet true" & 0 indexes supplied	No workaround. Fix planned for future version
BCL	DRAGEN -22480	Customers with high CPU core count systems have reduced BCL performance due to a thread limit, since v3.10	No workaround. Fix planned for future version
BCL	DRAGEN -20663	BCL does not abort when Combined Index Collision Check is enabled on a dual index run with one index removed	Uncaught user input error. Operation proceeds normally.
BCL	DRAGEN -19157	Filenames for interleaved FASTQs that are Ora compressed, are not the same as the original file names. For original filenames ending in "R1_001.fastq","R2_001.fastq" the decompressed file names are "R_1.fastq","R_2.fastq", dropping the identifier "001". This could potentially lead to duplicate file name conflicts	No workaround. Fix planned for future version
BCL	DRAGEN -19103	BCL crashes in Robust mode when *.filter file is missing for single lane dataset	No workaround. Fix planned for future version
BCL	DRAGEN -18920	bcl-convert outputs different PF cluster YieldQ30 and QualityScoreSum stats in the legacy stats file ConversionStats.xml as compared to bcl2fastq2.	No workaround. Fix planned for future version
BCL	DRAGEN -13771	A crash during bcl error checks can lead to hang, due to timing race condition	No workaround. Fix planned for future version

CNV caller	DRAGEN -25042	Incorrect ploidy estimation on sample with large deletion, does not call the deletion	No workaround. Fix planned for future version
Alignment	DRAGEN -23757	Insert size estimates can be significantly inaccurate for a sample when there are sequencing dropouts (no reads or coverage) over the first tiles of a flow cell.	No workaround. Improvement planned for future version
Duplicate Marking	DRAGEN -23711	Very large samples can fail with the default dupmark-version=hash due to a system limitation. The system crash with "Assertion `pos < m_num_bits' failed.	Run with "dupmark-version=sort"
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 '.' haploid. However, according to the VCF 4.2 specification missing genotype should be coded with '.' for each missing allele i.e './.' for a missing diploid genotype.	No workaround. Very rare occurrence and low impact. Fix planned for future version
Gvcf Genotyper	DRAGEN -26325	Gvcf Genotyper truncates the names of contigs to the first colon. This leads to incorrect outputs for those contigs. Some references contain such HLA* contigs.	No workaround. Fix planned for future version
Gvcf Genotyper	DRAGEN -21922	Some incorrect LPL and LAA values in msVCF	No workaround. Fix planned for future version
Infra, SNV VC	DRAGEN -21518	Regression in run times on Microsoft Azure cloud nodes for Germline SNV	No workaround. Fix planned for future version
HW GRAPH	DRAGEN -18402	A very rare error in hardware graph has been seen, leading to assertion.	Re-run the sample
Hash Table	DRAGEN -26399	Hash table decompression error on some fasta files	Write the hash table uncompressed
Imputation	DRAGEN -22549	Imputation end to end pipeline adds only the first chromosome name to VCF the header, leading to problems with downstream tools	Re-header the VCF using bcftools
Infra	DRAGEN -19988	A crash on Microsoft Azure cloud can leave the system in a bad state that requires intervention and prevents subsequent jobs form succeeding. "ERROR: xclRegRW: can't map CU: 0"	Known issue for which a solution is not available
Joint Genotyping	DRAGEN -21909	Accuracy on denovo WGS joint genotyping changed, due to an ML qual adjustment made to improve NovaSeq-X indel performance	Planned FP/FN accuracy tradeoff for improved performance on NovaSeq-X data

Joint Genotyping	DRAGEN -19844	Joint genotyping is up to 30% slower compared to v4.0	No workaround. Fix planned for future version
On-instrument Analysis	DRAGEN -26321	Very rare instances have been encountered where the NovaSeq-X hardware gets into a bad state after a crash recovery, leading to PCIe errors.	The only remedy is to power cycle the CE so that the FPGAs can reload.
On-instrument Analysis	DRAGEN -25465	On-instrument NovaSeq-X runs can exceed a memory budget for BCL conversion and fail, when processing long reads such as 2x300	No workaround. Fix planned for future version
Ora Compress	DRAGEN -19279	File names are not preserved exactly as they were, for the interleaved decompression mode.	No workaround. Fix planned for future version
RNA Gene Fusion	DRAGEN -15168	Missed fusion in 1st exon of gene TLC1--TRBC2	No workaround. Fix planned for future version
scATAC	DRAGEN -23486	scATAC with combinatorial barcode position results in empty results	No workaround. Fix planned for future version
SNV VC	DRAGEN -25905	A single short target BED entry towards the end of a chromosome can cause a hang, for high depth samples.	Workaround is to either have more BED regions throughout the chromosome or increase bin memory
SNV VC	DRAGEN -23630	An invalid alignment used to build the graph genome, leads to an incorrect allele frequency. Only one such instance has been found.	No workaround. Fix planned for future version
SNV VC	DRAGEN -22841	In rare cases, MNVs are wrong when the merging distance is greater than graph TLEN	No workaround. Fix planned for future version
SNV VC	DRAGEN -17705	When output VCFs are not compressed, the md5sums are not available.	No workaround.
SV Caller	DRAGEN -18913	Any regions overlapping the hotspot BED files for DRAGEN-SV will be called, even with minimal support. This introduces of 1 FP across our FLT3-ITD suites	No workaround.
UMI	DRAGEN -23614	Some UMI samples with ultra-high sequencing depths, can run into out-of-memory condition on on-site systems with 256GB RAM.	No workaround.

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package
- The archive integrity can be checked using: `./<DRAGEN 4.1.7 .run file> --check`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.1.7 .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

Release History

Revision	Release Reference	Originator	Description of Change
00	1087776	Cobus De Beer	Initial release