DRAGEN v3.4.5 Software Release Notes

August 26, 2019

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Bio-IT Platform since the package containing DRAGEN™ v3.3.11

If you are upgrading from a version prior to DRAGEN™ v3.3.11, please review the release notes for DRAGEN™ v3.3.11 for a list of features and bug fixes introduced in that version.

DRAGEN™ v3.3.11 Installers and Release Notes are available here:
https://support.illumina.com/downloads/illumina-dragen-v3-3-installers.html

The 3.4.5 software package includes:
- DRAGEN™ SW Intel Centos 6 - dragen-3.4.5.el6.x86_64
- DRAGEN™ SW Intel Centos 7 - dragen-3.4.5.el7.x86_64
- DRAGEN™ SW IBM PPC Centos 7 - dragen-3.4.5.el7.ppc64le.run

Contents

Highlights

- Small Variant Caller
  - Added Regions of Heterozygosity (ROH) output for human genomes.
  - Added callability metrics output for Germline runs with gVCF output.
  - Added B-Allele Frequency output for single SNP alt alleles present in the Germline and Somatic output VCFs. The output is a BigWig-compressed file.
  - Improved robustness of sample naming to avoid BAM/VCF sample name mismatches. Deprecated “vc-sample-name” and “vc-tumor-sample-name” options.
- Structural Variant Caller
  - Accuracy Improvements:
    - Substantially improved insertion recall through refinements to clipped read handling, contig assembly and breakpoint processing.
      - Improvements result in a germline insertion F-score improvement of 9% compared to DRAGEN 3.3.
    - Added depth-based filtering to Improve precision for large germline deletions and duplications.
      - For deletions 10kb and larger, precision improves by over 19% compared to DRAGEN 3.3.
  - Runtime improved by 15-30% via assembly fixes and a new read caching mechanism.
  - Improved the consistency of run times. Outlier samples that used to run in ~2 hr nowrun in 6 min without loss of accuracy.
  - Denovo scoring:
    - Added sex chromosome handling.
    - Added support for quads and quints.
    - Added DQ and DN fields as proband FORMAT fields to the VCF.
  - Added metrics output for SV calls.
- RNA
  - Gene Expression Quantification:
    - Over 2x speed improvement.
    - Added support for single-end libraries, and auto-detection of the library type.
    - Added support for Variational Bayes model fitting (new default).
  - Output compatible with Differential Expression tool DESeq2
  - Improved Gene fusion calling accuracy by integrating features from Manta RNA
    - 75% reduction in false positives while increasing sensitivity over 20%.
    - Speed improvements by about 1 minute per sample.
- DeNovo Variant Filtering:
  - A separate command-line execution that is run after small VC, CNV and SV joint callers’ outputs are available.
  - Improves specificity of the joint called small variants which overlaps with ploidy-changing structural and copy number variants in the pedigree.
- BCL:
  - Added support for UMI in indexed reads.
  - Added metrics outputs for demultiplexing, index hopping, top unknown barcodes .
- Graph Expansion Hunter:
  - Added ability to generate multi-sample repeat-genotyping VCFs.
  - Updated to Expansion Hunter v3.1.2.
- Mapper/Aligner metrics:
  - Additional mapping metrics to improve ability to do sample QC without need to run external tools:
    - Q30 Gbases excluding clipped and duplicated, % of Q30 bases on read1/2/both
    - % coverage at mapq for autosome

        - % mismatches on read1/2, Fragment length median
        - % reads aligned to different chromosomes, that are soft-clipped, containing indel
    - Coverage information, such as mean coverage, has been removed from the mapping_stats.csv and is now included in a standalone file: coverage_metrics.csv.
- Somatic Variant Calling:
    - Add support for CRAM tumor/normal input to the somatic variant caller. New command line option “--tumor-cram-input” added.
    - Output separate tumor/normal map/align stats when T/N mode is enabled.
- Other:
    - CRAM md5sum is now generated along with the CRAM alignment output.
    - BAM/CRAM header contigs are checked against the reference to avoid mismatches.
    - Support CNV de novo calling on chr Y.
    - Added VC metrics for counting Het.Hom insertions and Het deletions.
    - All variant callers (small VC, CNV, SV) now support denovo calling on quint pedigrees, using the same pedigree file.
        - 2 samples
        - 3 samples, with at most 1 proband
        - 4 samples, with at most 2 proband
        - 5 samples, with at most 3 proband

Non-backward compatible changes

DRAGEN v3.4.5 has a change to the handling of sample naming that may impact backwards compatibility with existing customer command line scripts.

- Motivation:
    - With the addition of multiple callers, DRAGEN produces output VCF files for each caller separately. The way that sample names were handled across callers was not consistent, and a specific use of the command line options often lead to sample name mismatches between small VC and the other callers (CNV, SV, EH).
    - The problem is exacerbated with de novo calling, which requires a re-header of some of the output VCF files.
    - Specifically, if a user did not provide “—RGSM” when processing FASTQs, the sample name in the BAM would be hardcoded to SM0, while the sample name in the small VC output would be set to “—vc-sample-name”. SM0 would then also be used as sample name for CNV and SV callers, thereby creating a mismatch.
- Changes made:
    - Deprecated the use of “--vc-sample-name”.
    - Enforce strict checking for presence of “--RGSM” when processing from FASTQ. If "--vc-sample-name" is provided, then a check is done to make sure that it matches RGSM. This check is added to ensure that existing user scripts are updated correctly.
    - When enabling both map/align and small VC, the small VC and other callers’ output VCFs will have the same sample name as the BAM/CRAM, which is “--RGSM”.
    - When processing from BAM/CRAM, the variant callers’ output VCFs will match the BAM/CRAM sample name.
    - Removed the ability to produce a SNV VCF with a different sample name than what is in the BAM header. The user needs to re-header BAM/CRAM files if a different sample name is desired.
- Usage examples and behavior:

Prior to DRAGEN v3.4.x

Use case	RGSM	vc-sample name	BAM	VCF	Sample name outcome
FASTQ -> map/align + VC	provided	provided	RGSM	vc-sample-name	Output BAM and VCF match, only if RGSM==vc-sample-name
FASTQ -> map/align + VC	Not provided	provided	“SM0”	vc-sample-name	Output BAM and VCF mismatch
BAM -> map/align + VC	Not provided	provided	Input BAM	vc-sample-name	Output BAM and VCF could mismatch depending on input BAM
BAM -> VC	Not provided	provided	n/a	vc-sample-name	VCF could mismatch input BAM sample name

DRAGEN v3.4.x and later

Use case	RGSM	vc-sample name	BAM	VCF	Sample name outcome
FASTQ -> map/align + VC	provided	provided	RGSM	RGSM	Output BAM and VCF match. RGSM==vc-sample-name enforced
FASTQ -> map/align + VC	provided	n/a	RGSM	RGSM	Output BAM and VCF match.
BAM -> map/align + VC	Not provided	n/a	Input BAM	Input BAM	Output BAM and VCF matches input BAM sample name.
BAM -> VC	Not provided	n/a	Input BAM	Input BAM	Output BAM and VCF matches input BAM sample name.

Known Issues

- Intermittent abnormal exit during map/align phase. Frequency of occurrence has been noted as < 0.1% of run executions, in combination with a specific dataset. Observed only on Phase2 server with U200 FPGA board, not observed on AWS or Phase1 servers. This issue will not lead to incorrect results. If encountered, the system will abort with either of the following messages: “hwal_eagleaxi_board.cpp line 3437 – false – Timeout waiting on write ACK” or “stats_init_remapper.cpp line 98 – false – Unexpectedly high latency in map/align engines; could not flush stats”.

SW Installation Procedure

- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 3.4.5 .run file>`
- Cold boot (hard reset or power cycle) is required after installation. The updated FPGA shell image needs to load from flash, this is only achieved with cold boot.
- Installing prior releases after 3.4.5 was installed:
    - Installing a prior release, DRAGEN 3.3.7 or older, will require the following two 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

md5checksum:

```
6ab76245c142fe60250143112b324a5e dragen-3.4.5.el6.x86_64.run

3f5d04179c17a3a44f56a90b2eab972d dragen-3.4.5.el7.x86_64.run

ec477d313cde7ba1d71ad5bf52141746 dragen-3.4.5.el7.ppc64le.run
```