DRAGEN TSO 500 Analysis Software Release Notes

V2.5.2

For TruSight Oncology 500, TruSight Oncology 500 HRD, and TruSight Oncology 500 High-Throughput

November 9, 2023

Introduction

These Release Notes detail the key changes to software components for the DRAGEN TSO 500 v2.5.2 Analysis Software.

This software is intended for use with the TruSight Oncology 500, TruSight Oncology 500 High-Throughput, and TruSight Oncology 500 HRD assays.

- Software Version: 2.5.2
- Docker Image ID: f9c61738b43b
- DRAGEN version: 3.10.16

The software installer script, install_DRAGEN_TSO500-2.5.2.run, includes the following:

- dragen_tso500_2.5.2.tar – a tar file of the DRAGEN TSO500 docker image.
- uninstall_DRAGEN_TSO500-2.5.2.sh – a script for uninstalling DRAGEN TSO500.
- check_DRAGEN_TSO500-2.5.2.sh – a script for self-testing DRAGEN TSO500.
- build-hashtable_DRAGEN_TSO500-2.5.2.sh – a script for building the hash table.
- Docker 20.10.15 and its dependencies
- install.sh – a script used to install DRAGEN TSO500 and Docker based on the contents listed.
- resources/ – a directory containing all resources files necessary for DRAGEN TSO 500 Analysis Software.
- dragen-3.10.16-8.el7.x86_64.run – the DRAGEN installer for servers running CentOS 7.
- dragen-3.10.16-8.el8.x86_64.run – the DRAGEN installer for servers running Oracle Linux 8.

NEW FEATURES (COMPARED TO DRAGEN TSO 500 ANALYSIS SOFTWARE V2.1.0 AND V2.1.1):

BaseSpace Sequence Hub (BaseSpace) functionality

- BaseSpace Run Planning tool now supports sample sheet v2 generation for DRAGEN TSO 500 Analysis Software (v2.1.1+) including when run on the Local DRAGEN server. Generated sample sheets can be uploaded to NextSeq 550/550Dx and NovaSeq 6000/6000Dx to configure run and provide input for running the DRAGEN TSO 500 Analysis Software.

Bioinformatics pipeline

- Manifest was updated to expand variant calling to +/-10 bp into introns (previously – 2 bp). This change satisfied several pending requests:

  - Position 17-41258470-T-C for a Belgium founder mutation in BRCA1 NM_007294.3: c.212+3A>G is now included
  - Intron-exon junctions positions for exons 13 and 7 of BRCA1 and BRCA2, correspondingly, are now included
- CNV calling for amplifications was expanded from 59 genes to 514 genes.
- CNV calling for deletions in 514 genes was added. "Low validation" flag assigned to this variant type in intermediate files is now removed.
- MSI JSON file containing MSI results was added to the Results folder.
- Several bioinformatics features were added with *beta* status. Beta features have not been verified by Illumina due to limited access to samples *or* lack of an appropriate orthogonal method to perform testing, and the use of *in silico* testing alone is not sufficient for verification purposes. Please refer to the Beta feature section below. Beta features include (all available with TSO 500 HRD kit):
  - Tumor fraction
  - Ploidy
  - Absolute copy numbers
  - Gene-level loss of heterozygosity (LOH) events
- Metric PCT_PF_UQ_READS (%) was removed as DRAGEN software performs alignment and read collapsing in one step and no longer calculates percent of reads that are unique. A similar metric UMI_FAMILY_SIZE can be used for assessing number of unique reads.
- DRAGEN sex prediction algorithm was added to the TSO 500 pipeline. The sex is predicted based on the read count information in the sex chromosomes and the autosomal chromosomes.

DEFECT REPAIRS (COMPARED TO DRAGEN TSO 500 ANALYSIS SOFTWARE V2.1.0 AND V2.1.1):

- All SNVs, insertions and deletions that are part of MNVs are now reported both individually and as merged variants in final VCF and combined output.
  - Fusion calls with mDNA partners are now filtered out from filtered fusion output.
  - Fusion calls with incorrect breakpoints due to indel nearby, now filtering is performed on reads to report correct breakpoints.
  - A fix was implemented for RNA fusion calling detection by updating maximum overhang from 8bp to 16bp to include more supporting reads in the fusion scoring. In addition, the alternative to reference reads ratio now includes a maximum threshold during fusion scoring.
  - Heuristic fixes were also introduced to address RNA fusion false positives, including soft clipped reads in intronic regions or with MAPQ 0 are not included as supporting read evidence, do not accumulate discordant reads as evidence if either left or right breakpoint evidence is outside annotated exon (or in intronic regions), and the MAPQ threshold is increased for a fusion in an "unenriched gene" (genes that are not covered by probe panel).

- In BRCA LR annotation for segments that are equally close to the baseline in the LR calculator, which annotates gain or loss of an exon-level CNV, has been updated to output the calls as "GAIN" and flag in the filter column of the vcf as "undetermined".
- Fixed an error with small variant calling step from processing NTC samples or samples with 0 reads within the manifest.
- Fixed an error caused by processing an NTC non-HRD sample with "Sample Feature" in the sample sheet set to HRD.
- Fixed errors with MNV calling when phased variant merging distance is greater than 10bp. Ensure haplotype from sufficiently long k-mers are used for merging phased variants.
- Fixed an instance where DRAGEN small variant calling did not correctly handle overlapping mates and mistakenly detected strand bias and subsequently filtered the variant. Strand bias was measured because the strongest support for the variant consistently came from one strand. This should however not be measured as strand bias since both mates were in agreement.
- Fixed an instance where DRAGEN small variant calling did not detect an insertion due to a too high graph pruning fraction. The problem was solved by updating the parameter to allow more sensitive haplotype generation.
- Fixed an error causing CNV pipeline assertion failure depending on the CNV panel of normals file name.
- Illumina Annotation Engine 3.2.6 (aka Nirvana) includes the following bug fix:
  - A fixed RefSeq version (105.20220307) was incorporated that fixed canonical transcript assignments for some prominent genes and variants. For example, for variant BRAF NP_004324.2. V600E, the canonical transcript and HGVS notations are now fixed in the CombinedVariantOutput file (was presented as BRAF NP_001361187.1. V640E in the previous version).
- Fixed several errors related to the sample sheet validator in BaseSpace Run Planning tool.
- Fixed an error of analysis failure when the input folder name has only numeric characters.
- Fixed an error of SARJ file displaying NA instead of Run ID in some cases.

- Fixed an error in GIS estimation where the algorithm was not properly filtering SNPs that had more than two variants detected in the sample (affecting on average less than one out of the over 20,000 SNPs used for GIS estimation). This caused a rare event (<0.5%) of a GIS being reported with a difference greater than 5 for a sample.

KNOWN ISSUES:

- Moving or modifying files during the analysis may cause the analysis to fail or provide incorrect results.
- Using control-c during a running analysis may cause an FPGA error. To recover from an FPGA error, shut down and restart the server.
- The sample sheet should not have blank rows between samples in the [Data] section. Blank rows in the [Data] section may cause a run failure.
- The sample sheet should not have blank rows after samples in the [Data] section. The workflow will fail if blank rows are present after the [Data] section.

- An erroneous warning is displayed when running the software - "Access to undefined parameter `report_folder`"

PRODUCT LIMITATIONS:

- Sample sheets generated for auto-launch are not compatible and cannot be reused without changes for DRAGEN TSO 500 Analysis Software v2.1.1+ on a Local DRAGEN server, and vice versa.
- This software version is only compatible with DRAGEN version 3.10.16.
- Performance not verified using reads other than 2 x 101.
- The values in the Run Metrics section will be listed as 'NA' if the analysis was started from FASTQs or if the analysis was started from BCLs but the InterOp files are missing or corrupted.
- The TSO 500 RNA workflow is unstranded. Fusions or splice variants could involve antisense transcripts instead of the reported genes.
- Poor quality wild type reads may align as chimeric and be miscalled during RNA analysis.
- TMB number may be inflated in samples with >5% supplementary (chimeric) alignments due to the larger number of false positive indels.
- Germline estimation that is used for TMB calculation uses the latest publicly available population data and is estimated to be representative of targeted population. The impact of rare germline mutations is expected to be limited for the TMB estimation.
- Germline estimation is difficult when tumor purity is > 85% causing expected variant allele frequency for somatic and germline variants to converge.
- Some regions are known to be difficult to sequence. One example region is the TERT promoter region. Although sequencing can occur at the TERT promoter region, this location might result in low coverage due to the GC rich content of the sequenced region. Another example region is the PMS2 gene which has high homology to pseudogenes and reads may not align properly. In general, the TSO 500 panel is designed to target unique regions, and the software accounts for background noise during small variant calling for each genomic position. This design is meant to prevent false positive calls. Analytical performance of the assay is evaluated panel-wide rather than for each gene or exon. However, due to these challenges certain regions covered in the product manifest are excluded from analysis due to high background noise. All excluded variants are identified in the VCF using a flag. This block list includes the following genes: HLA-A, HLA-B, HLA-C, KMT2B, KMT2C, KMT2D, chrY and positions with VAF > 1% occurred in six or more of the 60 baseline samples. The block list of excluded sites can be obtained on request from your local Illumina representative. In addition, HLA-A, HLA-B, HLA-C, KMT2B, KMT2C, KMT2D, HIST2H3A, HIST2H3C, and TERT (only covers promoter region) have insufficient probe coverage for CNV reporting.
- Lower sensitivity and specificity may be seen in CNV amplifications and deletions with less than 20 probes and higher noise profiles. Contact your local Illumina representative for more details.
- The Illumina Annotation Engine (aka Nirvana) may report incorrect HGVS c. and HGVS p. notation for small variants occurring in RefSeq transcripts that exhibit transcript sequences

differing from the genomic reference (i.e., RNA-edits). Currently the HGVS c. error rate is 0.00527% and the HGVS p. error rate is 0.00737%.

- BRCA1 and BRCA2 large rearrangements (exon-level CNVs) with two segments that diverge equidistant from baseline in opposite directions in highly rearranged genomes would occasionally report a "GAIN" due to variation in the calculated distance from baseline. These samples are expected to have high genomic instability and will be filtered as "undetermined".
- False negatives for BRCA1 and BRCA2 large rearrangements (exon-level CNVs) with a single or partial exon loss or gain and VAF lower than 61% are observed at higher rate than presented in product specification (sensitivity of 95% at VAF 50% or higher for fewer than 3 exons) due to the higher amount of noise associated with the smaller segment size. Pathogenic variants with single or partial exon CNVs are expected to have a prevalence of 0.17% in ovarian cancer samples (Jones et al., Genes Chromosomes Cancer.2023;62:589–596). The current implementation was designed to reduce false positives and has shown to have a high gene-level specificity (100%) with internal testing.
- Genomic Instability Score and BRCA large rearrangements (exon-level CNVs) have not been verified with input over 80ng of FFPE.
- GIS analysis has not been verified using libraries with UDP indexes.
- The estimates for tumor fraction and ploidy may be less reliable for samples with lower Genomic Instability Score as they will have fewer genome rearrangements.
- The contamination score threshold will fail approximately 1% of HRD samples due to the variant allele frequency (VAF) shifts of highly rearranged genomes and not true contamination of foreign human DNA. Visual investigation of VAFs across the genome can be performed to determine if a shift of VAFs is due to true contamination.

BETA FEATURES:

Several bioinformatics features were added with *beta* status. Beta features have not been verified by Illumina due to limited access to samples *or* lack of an appropriate orthogonal method to perform testing, and the use of *in silico* testing alone is not sufficient for verification purposes. Beta features are only available with the TSO 500 HRD kit and include:

- Tumor fraction
- Ploidy
- Absolute copy numbers
- Gene-level loss of heterozygosity (LOH) events

Tumor fraction (beta)
Tumor fraction is calculated as described in the User Guide, section "HRD Metrics Report" and leverages the Myriad Genetics algorithm. Tumor fraction is output in the Logs_Intermediates/Gis/SAMPLE/SAMPLE.gis.json and Combined Variant Output file. Disclaimers about the beta status are provided in Combined Variant Output file but not in other files.

Ploidy (beta)
Ploidy is calculated as described in the User Guide, section "HRD Metrics Report" and leverages the Myriad Genetics algorithm. Ploidy is output in the in the Logs_Intermediates/Gis/SAMPLE/SAMPLE.gis.json and Combined Variant Output file. Disclaimers about the beta status are provided in Combined Variant Output file but not in other files.

Absolute Copy Numbers (beta)
Absolute copy numbers are calculated by leveraging the Myriad Genetics algorithm. The algorithm segments the entire genome using the HRD panel and provides an A and B allele estimate for each segment. After the TSO 500 pipeline determines CNV calls (using the TSO 500 panel), the segment covering the gene is identified, and the A and B allele numbers of the segment overlapping the gene are reported. If the gene is within 300 kbases from the segment boundary, the estimate is unreliable and "-1" is output. Absolute copy numbers are output in the Logs_Intermediates/Gis/SAMPLE/SAMPLE.abcn_annotated.vcf, Logs_Intermediates/Gis/SAMPLE/SAMPLE.abcn_genes.tsv and Combined Variant Output file. Disclaimers about the beta status are provided in Combined Variant Output file but not in other files.

Gene-level loss of heterozygosity (beta)
Gene-level loss of heterozygosity is calculated based on the minor copy number reported in the abcn_annotated.vcf. If the minor copy number is 0 then the gene is assumed to have a loss of heterozygosity. Gene-level loss of heterozygosity is output in the Logs_Intermediates/Gis/SAMPLE/SAMPLE.abcn_genes.tsv and Combined Variant Output file. Disclaimers about the beta status are provided in Combined Variant Output file but not in other files.

Release History

Revision	Release Reference	Originator	Description of Change
00	CN 1095955	Svetlana Bureeva	Initial Release