

DRAGEN TSO 500 Analysis Software on ICA

Release Notes

V2.1.2

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

Introduction

These Release Notes detail the key features and known limitations of software components for the DRAGEN TSO500 v2.1.2 Analysis Software on Illumina Connected Analytics (ICA). For full details, please consult the DRAGEN TSO 500 v2.1.2 on ICA Analysis Software User Guide available on the support website.

This software is intended for use with the TruSight Oncology 500 Assay and the TruSight Oncology 500 HRD Assay.

- Software Version: 2.1.2
- DRAGEN version: 3.10.9

NEW FEATURES (COMPARED TO DRAGEN TSO 500 ANALYSIS SOFTWARE V1.1.1 ON ICA):

ICA functionality

- Sequencer integration including data streaming to ICA has been expanded and now supports NextSeq 550/550Dx, NovaSeq 6000/6000Dx and DRAGEN TSO500 v2.1.2 Analysis Software on ICA
- Auto-launch functionality now supports automated start of the DRAGEN TSO500 v2.1.2 Analysis Software on ICA after completion of sequencing and data streaming utilizing settings provided in the sample sheet v2. The sample sheet can be generated manually or using BaseSpace Run Planning tool.
- DRAGEN TSO500 v2.1.2 Analysis Software bundle on ICA has been updated with a specialized BCL Convert module (BCL Convert v3_10_9 for TSO500) improving experience for FASTQ -> VCF analysis
- Improvements to the ICA user interface: removed single tile option due to limited use, removed requirement for user to specify locations of Resource folder and hashtable, optimized fields for analysis input
- Improvements to the content of error messages and logs
- Contextual help improvements when launching new analysis

BaseSpace Sequence Hub (BaseSpace) functionality

- BaseSpace Run Planning tool now supports sample sheet v2 generation for DRAGEN TSO500 v2.1.2 Analysis Software on ICA. Generated sample sheets can be uploaded to NextSeq 550/550Dx and NovaSeq 6000/6000Dx to configure run, provide input for data streaming to ICA, auto-launch and running of the DRAGEN TSO500 v2.1.2 Analysis Software on ICA.
- Analysis run folder and other analysis outputs (analysis error messages and logs) are now output in BaseSpace in addition to ICA

- DRAGEN TSO500 Analysis Software users can now monitor analysis progress via BaseSpace interface in addition to monitoring sequencing status

Bioinformatics pipeline

- The pipeline now includes DRAGEN variant callers for SNV, MNV, CNV, TMB, MSI, and gene fusion variant calling.
- Performance improvements:
 - Improvements in contamination QC and CNV QC to better handle samples with highly rearranged genomes
 - Improvements in MSI detection to better handle low-quality samples and decrease the rate of sample QC failure
 - Improvements in small variant specificity
 - Improvements in RNA fusion sensitivity
 - Capability of calling insertions/deletions > 25 bp and complex variants >3 bp panel-wide
 - Improved sensitivity of EGFR complex variants
- BRCA 1/2 large rearrangements (exon-level CNVs) called at 43% or higher VAF for 3 or more exons, 50% or higher VAF for fewer than 3 exons.

Genomic instability scoring powered by HRD technology from Myriad Genetics (not available in Japan; must be run with the TSO 500 HRD kit (MN 20076480).

DEFECT REPAIRS (COMPARED TO DRAGEN TSO 500 ANALYSIS SOFTWARE V2.1.1 ON ICA):

- The issue of GIS results being absent in the Combined Variant Output file when the software is run with the auto-launch functionality has been resolved

DEFECT REPAIRS (COMPARED TO DRAGEN TSO 500 ANALYSIS SOFTWARE V1.1.1 ON ICA):

- Illumina Annotation Engine 3.2.6 (aka Nirvana) includes the following enhancements and bug fixes:
 - Added genes and transcripts from the *NCBI Homo sapiens Updated Annotation Release 105.20201022* to provide the latest RefSeq content for GRCh37
 - Reduced the HGVS c. error rate by 54% and HGVS p. error rate by 20%. Currently the HGVS c. error rate is 0.00527% and the HGVS p. error rate is 0.00737%.
 - Fixed issues related to incorrect CDS coordinates in some edge cases
 - Improved detection of frameshifts when variants partially overlap the coding sequence

- Fusion caller can now call fusions when breakpoint(s) are located in region(s) with high homology.
- The Metrics Output TSV file has been updated so that 'steps_not_executed' does not include HRD related steps when the analysis does not include HRD samples.
- The issue where if a certain step of the RNA sub-workflow (FASTQ Downsample, RNA QC, Splicegirl, or Fusion Merge) failed for at least one sample, the step was listed in the "Steps Not Completed" section of the Metrics Output file for all samples in the run was resolved.
- The issue causing the TSO 500 analysis pipeline to crash when Illumina Annotation Engine 3.2.5 (used for TMB calculation) encountered rare annotation issues related to using Ensembl transcripts for a limited set of variants was resolved.
- The issue where an extra line at the end of the sample sheet caused Nextflow failures has been resolved.
- Warnings in the nextflow run logs about undefined parameter 'report_folder' have been fixed.
- Warnings in the nextflow run logs for 'unknown recognition error type: groovyjarjarantlr4.v4.runtime.LexerNoViableAltException' have been fixed.
- The error log for the SampleSheetValidation module was corrected to reflect the module name to be consistent with other error log files.
- The pipeline was updated to reduce the time to failure in case of sample sheet errors. This was accomplished by running validation ahead of all other steps and allowing the step to be executed on a more widely available node. The current 'time to failure' now corresponds directly to the size of the input run or FASTQ folder, as this must first be copied into a scratch location to support SampleSheetValidation.

KNOWN ISSUES:

- Moving or modifying files during the analysis may cause the analysis to fail or provide incorrect results.
- 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.
- The TSO 500 RNA workflow is unstranded. Fusions or splice variants could involve antisense transcripts instead of the reported genes.
- ICA Run Time depends on ICA instance availability, it will be affected by region and traffic.

PRODUCT LIMITATIONS:

- The sample sheet must be configured as described in the User Guide or by using BaseSpace Run Planning tool.
- Sample sheets generated for auto-launch are not compatible and cannot be reused without changes for DRAGEN TSO500 Analysis Software v2.1.1 on a Local DRAGEN server, and vice versa.
- This software version is only compatible with DRAGEN version 3.10.9.
- Performance not verified using reads other than 2 x 101.
- GIS analysis has not been verified using libraries with UDP indexes.
- 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.
- Unmapped long insertions are not likely to occur on shorter indels because there is sufficient reference-matching sequence in the reads.
- Incorrect calculation of variant allele frequency can occur in variants near the start and end of genomic reads, but there is a low probability of incorrect variant allele frequency in called variants due to sufficient variation in read start and end positions.
- Germline estimation uses 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.
- Germline estimation is difficult when tumor purity is > 85% causing expected variant allele frequency for somatic and germline variants to converge.
- Poor quality wild type reads may align as chimeric and be miscalled during RNA analysis.
- BRCA 1/2 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".
- 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.
- 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%.