DRAGEN TSO500 Analysis Software ICA Release Notes

V1.1.1

For TruSight Oncology 500 Assay

August 29, 2022

Introduction

These Release Notes detail the key features and known limitations to software components for the DRAGEN TSO500 v1.1.1 Analysis Software on Illumina Connected Analytics (ICA) platform.

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

- Software Version: 1.1.1

NEW FEATURES:

- Ability to run DRAGEN pipelines for TSO 500 analyses on ICA
- GUI enabled to run TSO500 analysis on ICA

KNOWN ISSUES:

- The sample sheet should not have blank rows between samples in the [Data] section, this may cause a run failure.
- Performance not verified using reads other than 2 x 101.
- The TSO500 RNA workflow is unstranded. Fusions or splice variants could involve antisense transcripts instead of the reported genes.
- Fusion caller may not always call fusions where breakpoint(s) are located in region(s) with high homology.
- The cloud workflow will fail if blank rows are present after the [Data] section.
- The Run ID field in the sampleInformation section for the SARJ files for some local DRAGEN analyses is displayed as "NA".
- The user interface in ICA allows the user to start an analysis without specifying the RUN folder path or FASTQ folder path. This action results in a failed analysis.
- An analysis run may be marked as 'Succeeded' although an analyzed sample may have failed. This may occur when a sample may have failed analysis, but the pipeline was executed successfully.

PRODUCT LIMITATIONS:

- The sample sheet must be configured as described in the User Guide.
- NovaSeq S4 analysis time for 192 samples may require more than 12 hours to complete.
- 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. Product claims only indels up to 25 base pairs.
- Complex variants are specifically output only for a specific region of the EGFR gene, component and phased variants would both be contained in the output.
- 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 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
- The Illumina Annotation Engine (aka Nirvana) may report incorrect protein (P-Dot) and transcript (C-Dot) changes in HGVS nomenclature for small variants located on a RefSeq transcript where an RNA-edit has occurred. Most known variants on these transcripts are unaffected. A list of affected Canonical RefSeq transcripts and Cosmic Variants from those transcripts can be found below. A full explanation of this product limitation can be found in PQN2020-1090. [1]

Affected Canonical RefSeq Transcripts

Transcript ID	Gene Symbol
NM_002467.4	MYC
NM_003224.5	ARFRP1
NM_004119.2	FLT3
NM_006904.6	PRKDC
NM_198291.2	SRC
NM_021960.4	MCL1
NM_001025366.2	VEGFA

Affected Cosmic Variants from Canonical RefSeq Transcripts

The list of affected variants is based on an analysis of COSMIC database version 92 variants located along the Canonical RefSeq Transcripts listed above [2]. New variants are regularly submitted to COSMIC, and this list of affected variants may change over time.

Chr:Position	REF*	ALT**	Gene Symbol	Transcript ID	COSMIC_ID
chr1:150548890	A	ATCTA	MCL1	NM_021960.4	COSV57189597
chr6:43738444	C	T	VEGFA	NM_001025366.2	COSV104569261
chr8:48805817	G	GG	PRKDC	NM_006904.6	COSV58041377
chr8:128748839	GC	G	MYC	NM_002467.4	COSV104388447
chr8:128748840	C	A	MYC	NM_002467.4	COSV104388806
chr8:128748840	C	G	MYC	NM_002467.4	COSV104388204
chr8:128748841	T	C	MYC	NM_002467.4	COSV104388663
chr13:28608094	C	CACTTTTCCAAAAGCACCTGATCCTAGTACCTTCCCAAACTCTAAATTTTCTCTTGGAAACTCCCATTTGAGATCATATTCATATTCGTTCATC	FLT3	NM_004119.2	COSV54069050
chr13:28608124	C	CTTCCCAAACTCTACTGTTGCGTTCATCACTTTTCCAAAAGCACCTGATCCTAGTACC	FLT3	NM_004119.2	COSV54044227
chr13:28608129	C	CAAACTCAAAAGCACCTGATCCTAGTACCTTCCC	FLT3	NM_004119.2	COSV54054381
chr13:28608129	C	CAAACTCTAAATTTTCTCTTGGAAACTCCCATTATCCTAGTACCTTCCC	FLT3	NM_004119.2	COSV54043729
chr13:28608129	C	CAAACTCTAAATTTTCTCTTGGAAACTCCCATTTTCCAAAAGCACCTGATCCTAGTACCTTCCC	FLT3	NM_004119.2	COSV54075746
chr20:36030939	G	GTGGCC	SRC	NM_198291.2	COSV99050886
chr20:62331336	C	CC	ARFRP1	NM_003224.5	COSV53926174

*Reference base(s)

**Alternate base(s)

[1] DRAGEN TSO500 uses the Canonical RefSeq transcript when annotating variants passed into the Combined Variant Output file. The Illumina Annotation Engine selects canonical transcripts based on the following rules:

- Order all overlapping transcripts by coding sequence length.
- Pick the longest transcript that has an associated Locus Reference Genome (LRG) sequence.

- If no LRGs exist for the set of transcripts, pick the longest transcript that is coding.
- If there is a tie, pick the transcript with the smaller accession id number.

[2] Released 29 August 2022.

Release History

Version	CN#	Author	Description of Change
00	1073924	Darryl Leon	Initial Release