DRAGEN TSO500 ctDNA Analysis Software on ICA

Release Notes

V1.2

For TruSight Oncology 500 ctDNA Assay

August 29, 2022

Introduction

These Release Notes detail the key features and known limitations to software components for the DRAGEN TSO500 ctDNA v1.2 Analysis Software on ICA. This software is intended for use with the TruSight Oncology 500 ctDNA Assay. Below is a summary of the changes included in DRAGEN TSO 500 ctDNA v1.2 Software on ICA. For full extensive details, please consult the latest DRAGEN TSO 500 ctDNA v1.2 Software User Guide available on the support website.

NEW FEATURES:

1. You can run software on ICA v2.

DEFECT REPAIRS:

- Addressed annotation errors for CRLF2 and MAP3K14. The annotation component was updated to resolve errors.

KNOWN ISSUES:

- DRAGEN TSO500 ctDNA v1.1 does not detect expected fusion. Specifically, in regions with high levels of potential fusion calls it was noted that the filtering strategy resulted in the FKBP15-ALK ctDNA fusion being excluded despite the documented fusion filtering criteria being met. Tests showed that missed calls are rare events and that adjusting the filtering parameters allowed the fusion to be accurately called.
- Incorrectly Annotated Genes - Nirvana 3.2.3 & 3.2.5 does not properly account for All RNA-Edits
- Users have the ability to edit intermediate files. - If intermediate files are edited, analysis may fail and delay results.
- The sample sheet should not have blank rows between samples in the [Data] section, this may cause a run failure.
- The cloud workflow will fail if blank rows are present after the [Data] section in the sample sheet.
- The metrics output step module shows no error message when input file is missing.
- FastqGeneration issue: Missing .bcl files can cause FastqGeneration failure, but pipeline does not generate a MetricsOutput.tsv file with failed steps.
- There is incorrect date format in the SampleSheet. Selected output files do not have desired date format expected: yyyy-mm-dd.
- The FastQ validation step allows lane numbers that contain non-numeric characters.
- The analysis output folder name has format:

TSO500_ctDNA_Analysis_YYYYMMDD-HHMMSS, but it should be: TSO500_ctDNA_Analysis_YYYYMMDD_HHMMSS.

- When comparing DRAGEN TSO 500 ctDNA on-premise v1.2 and ICA v1.2 results, one may note very minor differences between these deployments in the following output files: MergedSmallVariants.genome.vcf, MergedSmallVariants.vcf, MergedSmallVariantsAnnotated.json.gz or other output files. This is due to insignificant rounding differences between the local and cloud DRAGEN mapping or UMI collapse logic. These differences do not affect the final calling of variants.
- There is incorrect time format in MergedAnnotation log file.
- There is an empty Sarj biomarkers section.
- The software does not notify the user when InterOp files for RunQC are missing or corrupted.
- Yente has merge duplicated information in the header. The header of the merged vcf from yente includes all the header lines from the gvcf concatenated with all the header lines from the PhasedVariants step. Since the PhasedVariants step uses the SmallVariantsFilter gvcf as input, the complex vcf already includes the header information from the gvcf. Thus, this information is duplicated in the merged vcf from the yente output.
- The Manta output has the following header issues: The BND svtype not listed in ALT section of header
- SARJeant logs incorrect version: SARJeant is logs version 1.0.0.0 for itself in the SARJeant-<timestamp>.log file.
- To manage your own storage, please purchase the Enterprise Annual Subscription (20038994) or the ICA Professional Annual Subscription (20044876). Managing your own storage is not dpossible with the Basic Subscription (20044874).
- To set up an S3 folder name, please provide a path to where the data should be stored. The "myfoldername" will need to be the same as the folder name entered during S3 bucket configuration which is <folderNameUponProjectCreation>. The sequencing run data will need to be replicated into the path entered during the setting /<bucketName>/<keyPrefix>/<folderNameUponProjectCreation>/ if the run data is not stored there.

PRODUCT LIMITATIONS:

- The sample sheet must be configured as described in the User Guide.
- 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 , as well as for specific regions of the RET gene. Component and phased variants would both be contained in the output. The pipeline specifically evaluates the EGFR gene regions for the list of well characterized complex variants listed in the table below.

Chromosome	Position	Reference Allele	Alternative Allele
chr7	55242482	CATCTCCGAAAGCCAACAAGGAAAT	C
chr7	55242466	GAATTAAGAGAAGCAACAT	G
chr7	55242465	GGAATTAAGAGAAG	AATTC
chr7	55242465	GGAATTAAGAGAAGCAAC	AAT
chr7	55242469	TTAAGAGAAGCAACATCTC	T
chr7	55242467	AATTAAGAGAAGCAACATC	A
chr7	55242469	TTAAGAGAAG	C
chr7	55242467	AATTAAGAGAAGCAACATC	T
chr7	55242465	GGAATTAAGA	G
chr7	55242467	AATTAAGAGAAGCAACATCTC	TCT
chr7	55242467	AATTAAGAGAAGCAAC	T
chr7	55242464	AGGAATTAAGAGAAGC	A
chr7	55242466	GAATTAAGAGAAGCAA	G
chr7	55242464	AGGAATTAAGAGA	A
chr7	55242469	TTAAGAGAAGCAA	T
chr7	55242465	GGAATTAAGAGAAGCAACATC	AAT
chr7	55242469	TTAAGAGAAGCAACATCT	CAA
chr7	55242463	AAGGAATTAAGAGAAG	A
chr7	55242468	ATTAAGAGAAGCAACATCT	A
chr7	55242462	CAAGGAATTAAGAGAA	C
chr7	55242465	GGAATTAAGAGAAGCAA	AATTC
chr7	55242469	TTAAGAGAAGCAA	C
chr7	55242467	AATTAAGAGAAGCAAC	A
chr7	55242469	TTAAGAGAAGCAACATCTCC	CA
chr7	55242468	ATTAAGAGAAG	GC
chr7	55242465	GGAATTAAGAGAAGCA	G
chr7	55242468	ATTAAGAGAAGCAAC	GCA
chr7	55242465	GGAATTAAGAGAAGCAACA	G
chr7	55249011	AC	CCAGCGTGGAT

In order to ensure variant calling specificity, limitations on the number of observed mismatches to the reference sequence have been implemented. Due to these limitations the following EGFR complex variant cannot be called:

Chromosome	Position	Reference Allele	Alternative Allele
chr7	55242467	AATTAAGAGAAGCAACA	TTGCT

- Limitations on the number of observed mismatches apply as well to the RET complex variants. The RET complex variants will be provided in the output if detected. However, these variants have not been analytically validated. The underlying nucleic acid changes that result in the relevant amino acid changes in RET are less well characterized such that a comprehensive evaluation of which nucleic acid changes would not be in the final output VCF could not be performed.
- 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
- 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 TSO 500 ctDNA 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.

Release History

Version	Workflow	Author	Description of Change
00	DIR Workflow	Andrea Hatlen	Initial Release
01	DIR Workflow	Andrea Hatlen	Added the cloud workflow Known Issue.
02	CN - 1055723	Mario Duff	Updated EGFR and RET complex variant product limitation.
03	CN 1073094	Ramon Sanchez	Updated for v1.2 release