DRAGEN TSO500 ctDNA Analysis Software

Release Notes

V1.1.0

For TruSight Oncology 500 ctDNA Assay

November 13, 2020

Introduction

These Release Notes detail the key features and known limitations to software components for the DRAGEN TSO500 ctDNA v1.1.0 Analysis Software.

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

- Software Version: 1.1.0

Docker Image ID: e758a789b0b3

The software installer includes:

- dragen_tso500_ctdna_1.1.0.tar – a tar file of the TSO500 ctDNA docker image.
- uninstall_DRAGEN_TSO500_ctDNA-1.1.0.sh – a script for uninstalling TSO500 ctDNA.
- resources/ - a directory containing all resources files necessary for TSO500 ctDNA.
- dragen-3.6.4-2.el7.x86_64.run – the DRAGEN installer.
- *.rpm – Mulitple RPM files used to install docker and its dependencies.
- install_DRAGEN_TSO500_ctDNA-1.1.0.run – The script used to install TSO500 ctDNA based on the contents listed here.

New Features:

- Update to DRAGEN software v3.6.
- Support for new NovaSeq 6000 v1.5 Reagent Kits.
- Bioinformatics improvements including:
    - Incorporation of Maximum Somatic Allele Fraction (MSAF) output (estimate of tumor fraction)
    - Improvement of MSI algorithm - increase sensitivity and specificity
    - Incorporate Small Variant filtering relative to fragment ends (improved variant calling) - improved assay specificity
- The addition of a small variant VCF file (in addition to the genome VCF file previously produced) in the Results folder.
- Support for new Sample Sheet v2 template.

Defect Repairs:

- Illumina Annotation Engine updated to address over 99% of known situations where the software may report incorrect C-Dot and P-Dot notation values for DNA variants on affected RefSeq transcripts (see Product Limitations below).
- An issue is fixed where the SampleAnalysisResults step enforced a hard limit on read depth, and intermediate and downstream files were not produced.

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, this may cause a run failure.
- Performance not verified using reads other than 2 x 151, paired end, dual index.
- The StitchedRealigned step produces the same output BAMs when using compute nodes with the same CPU configuration but may produce slightly different output BAMs on nodes with different CPU configuration.

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, 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
- 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.

[2] Released 27 August 2020.

Release History

Version	ER#	Author	Description of Change
00	DIR Workflow	Andrea Hatlen	Initial Release