

TSO500 Local Run Manager Software Release Notes

V2.2.0

For TruSight Oncology 500 Assay

January 19, 2021

Introduction

These Release Notes detail the key features and known limitations to software components for the v2.2.0 Local Run Manager module. This software is intended for use with the TruSight Oncology 500 Assay.

NEW FEATURES:

- Update to complex variant calling to include a new EGFR variant (EGFR:c.2309_2310AC>CCAGCGTGGAT).

DEFECT REPAIRS:

- BCL-Convert version updated to v3.5.7 to prevent analyses from crashing due to a rare tile drop out issue.
- The "Splice Support Reads" and the "Reference Reads Transcript" columns in the Combined Variant Output files have been corrected (previously the values for these were reversed).
- Read-stitching component (Gemini) updated to prevent an analysis from crashing if an indel was detected at the end of a chromosome.
- Read collapsing component (ReCo) updated to correct a metric that is used during read collapsing.
- Gene name for gene amplifications corrected from "MYCL1" to "MYCL".
- Manta RNA fusion calling step has been corrected to address a rare race condition which can cause analysis to fail.
- 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).
- Updated Read Stitching and Indel Realignment Component (Gemini) to create reproducible BAM file and VCF results.

KNOWN ISSUES:

- Rare instances of low multiplexed runs (3 DNA samples in a run) cause delayed results when performing read collapsing.
- The NextSeq instrument software memory consumption may prevent the TSO500 Local Run Manager module from starting. An error message in the module logs beginning "Failed to start the virtual machine..." indicates this issue (located in the directory "{Run_ID} \Analysis_N \YYYYMMDD_HHMMSS \Module_Logs"). After encountering the issue, please perform the following steps.
 - Use the "Exit to Windows" command to shut down the instrument software.
 - Restart the NextSeq Control Software.
 - Re-queue the run.

- In rare circumstances, the Run QC section of the Metrics Output file may fail to populate. Please review the RunQCMetrics.json file to troubleshoot the run.
- The descriptions for the “AD” and “DP” fields in the Splice VCF header are reversed.

PRODUCT LIMITATIONS:

- Performance not verified using read lengths other than 2 x 101.
- 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 variation in read start and end positions in an enrichment assay is sufficient to make incorrect variant allele frequency in output variants a low-probability situation.
- Germline estimation using high tumor purity (>70%) can impact estimation, due to somatic and germline variants appearing with similar variant allele frequency.
- 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
- Poor quality wild type reads may align as chimeric and be miscalled during RNA analysis
- When filtering fusion variant candidates, strand consideration is not used in annotation, which allows contiguous sequences to be intragenic, when they are on different strands
- Manta will not call fusions where both breakpoints map to the same gene transcript: FIP1L1-PDGFR (when both breakpoints overlap Ensembl transcript ENST00000507166) and GOPC-ROS1 (when both breakpoints overlap Ensembl transcript ENST00000467125).
- Manta may not always call fusions in the following situations, which have been exclusively observed in synthetic commercial controls:
 - Multiple fusion breakpoints from a single fusion gene pair with breakpoints within approximately 150 base pairs of each other (observed with ETV6-ABL1 and ETV6-NTRK3 fusions).
 - Multiple fusions from two different gene pairs with breakpoints within approximately 150 base pairs of each other (observed with IRF2BP2-NTRK1, TFG-NTRK1, SQSTM1-NTRK1 fusions).
 - Breakpoint(s) are located in region(s) with high homology (observed with fusions with breakpoints on SEPT14 exon 10).
- Software may include small variants outside, but near, manifest regions in samples where small variant candidates partially overlapping manifest boundaries are evaluated.
- Variant reporting is limited by a manifest file and a block list file. The manifest file excludes regions where the probe set does not effectively capture targets, and the block list file excludes specific positions from variant calling. TSO 500 probes target at least 97% of the CDS of 474 genes. Please contact your local Illumina representative for more information if needed.
- 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	CACTTTTCCAAAAGCACCTGATCCT AGTACCTTCCCAAATCTAAATTTT CTCTTGAAACTCCCATTGAGATC ATATTCATATTCGTTTCATC	FLT3	NM_004119.2	COSV54069050
chr13:28608124	C	CTTCCCAAATCTACTGTTGCGTTC ATCACTTTTCCAAAAGCACCTGATC CTAGTACC	FLT3	NM_004119.2	COSV54044227
chr13:28608129	C	CAAATCAAAAGCACCTGATCCTA GTACCTTCCC	FLT3	NM_004119.2	COSV54054381

chr13:28608129	C	CAAACCTCTAAATTTTCTCTTGAAAA CTCCCATTATCCTAGTACCTTCCC	FLT3	NM_004119.2	COSV54043729
chr13:28608129	C	CAAACCTCTAAATTTTCTCTTGAAAA CTCCCATTTTCCAAAAGCACCTGAT CCTAGTACCTTCCC	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] TSO 500 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	1045746	Trey Howard	Initial Release
01	CN 1127621	Francis Peters	Added disclaiming language for gene coverage.