# TruSight Oncology 500 v2.0 Local App

## User Guide

ILLUMINA PROPRIETARY

For Research Use Only. Not for use in diagnostic procedures.

## Introduction

The Illumina® TruSight™ Oncology 500 Local App generates sequencing libraries for DNA and RNA from formalin-fixed, paraffin-embedded (FFPE) tissue samples. The software supports run setup, sequencing, and analysis for the prepared DNA and RNA libraries. DNA library analysis outputs include tumor mutational burden, variant call files for small and complex variants, microsatellite instability, and gene amplifications. RNA library analysis outputs include fusions and splice variant call files. The TruSight Oncology 500 module supports 3–8 DNA libraries, 3–16 RNA libraries, and select combinations of DNA and RNA libraries per run.

## Installation Requirements

The following subsections are the minimum operating requirements for Local Analysis Software analysis.

## Output Compatibility

The software is compatible with sequencing outputs from the following sequencing systems.

▶ NextSeq 500

▶ NextSeq 550

▶ NextSeq 550Dx (in RUO mode)

## Software

▶ md5sum

▶ Linux CentOS 7.4 operating system, or later

▶ Container engine Docker CE v17.03 or Singularity v2.6.1 or later
  ▶ If you have Network File System (NFS) mounted volumes for analysis output, use Singularity or contact Illumina Technical Support

▶ HTTP client such as Curl or Wget

## Hardware

▶ Minimum:
  ▶ 64 GB RAM
  ▶ 64-bit CPU with 12 physical cores, 24 threads
  ▶ Dual Intel Xeon 64-bit CPU; E5-2620 v4 (released Q1 2012) or newer. Processor must support the AVX instruction set.
  ▶ 1 TB HDD free space
▶ Recommended:
  ▶ 128 GB RAM, 2400 MHz DDR4 RDIMM ECC
  ▶ 64-bit CPU with 24 physical cores, 24 threads
  ▶ 1 TB HDD free space
  ▶ Enterprise-grade solid state drive (SSD)

## Storage

▶ Total disk space for app installation is 157 GB.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

3

- ▶ The app download file is 68 GB, and is a single *.gzip compressed tar archive. The zipped file can be deleted after decompression.
- ▶ The app download file requires 89 GB of space after decompression.
- ▶ A NextSeq 500 or 550 sequencing run with secondary analysis requires a minimum of 400 GB of free space.
  - ▶ A sequencing run can produce 40-100 GB of output.
  - ▶ Secondary analysis of a sequencing run can produce 100-200 GB of output.
- ▶ Network File System (NFS) is recommended for mounted volumes. Use of Common Internet File System (CIFS) can result in pipeline failures during the alignment step.

## Additional Resources

The Local Analysis Software support pages on the Illumina support website provide additional system resources. These resources include software, training, compatible products, documentation, and product files, including a sample sheet template. Always check support pages for the latest versions.

| Resource | Description |
| --- | --- |
| *Local Analysis Software SampleSheet.csv (document # 1000000105888)* | Setup template for running the analysis. |
| *NextSeq System Denature and Dilute Libraries Guide (document # 15048776)* | Guidelines on the number of libraries and possible DNA/RNA combinations per sequencing run. |

## Install Local Analysis Software

Installation of the Local Analysis Software requires root or superuser privileges and that Docker or Singularity is running.

## Docker Considerations

Specify a user ID at Docker run time to limit user privileges associated with the account. For information on Docker permission requirements and alternatives to running as root, see the Post-Installation steps for Linux page of the Docker Docs website. You may require sudo/escalated privileges to load the docker image.

## Singularity Considerations

Singularity requires an admin user with root privileges for set up, but allows non-admin users to run singularity containers. After configuration, a non-admin user can check Singularity configuration and image info.

## Installation Instructions

1  Download the application bundle from the link you receive from Illumina in an email.
   The link expires after 72 hours.

2  Run the following command, substituting the link provided to you by Illumina.
```
curl -o ${DOWNLOAD_PATH}/trusight-oncology-500-ruo-${VERSION}.zip
   "https://${APPLICATION_DOWNLOAD_URL}"
```

3  Extract the application bundle.
   After extraction, the bundle contains the following files:
   - ▶ ./TSO500_RUO_LocalApp/md5sum.txt

- ▶ ./TSO500_RUO_LocalApp/trusight-oncology-500-ruo.img
- ▶ ./TSO500_RUO_LocalApp/trusight-oncology-500-ruo-dockerimage-ruo-{VERSION}.tar
- ▶ ./TSO500_RUO_LocalApp/TruSight_Oncology_500_RUO.sh
- ▶ ./TSO500_RUO_LocalApp/resources (folder)

4    Verify the files by calculating the md5 hash value with the following commands.
```
md5sum trusight-oncology-500-ruo.img
md5sum trusight-oncology-500-ruo-dockerimage-ruo-{VERSION}.tar
md5sum TruSight_Oncology_500_RUO.sh
```

5    **[Docker]** Load the Docker image using the following command.
```
| sudo docker load
```

6    **[Docker]** Make sure that the Docker image has the following values:
- ▶ IMAGE—trusight-oncology-500-ruo
- ▶ TAG—${VERSION}
- ▶ IMAGE ID—{Corresponding to VERSION}

7    **[Singularity]** Make sure that the Singularity image file is contained in the applications bundle.
```
which singularity
singularity inspect trusight-oncology-500-ruo.img
```

8    Display the help for the Bash script with the following command.
```
./TruSight_Oncology_500_RUO.sh --help
```

## Running Local Analysis Software

Start the Local Analysis Software software with the Bash script called TruSight_Oncology_500.sh in the application bundle. The Bash script can be executed on the command line and can run the software with Docker or Singularity. For arguments, see *Shell Script Command Line Arguments* on page 9. You can start from BCL files or from the FASTQ folder. The following requirements apply for both methods.

- ▶ Full path to the Local Analysis Software resources folder location <RESOURCE_FOLDER>. The resources folder contains the manifest file, the hg19 reference genome, sample sheet template, and other analysis resources. Modification of the folder causes analysis errors.
- ▶ The Local Analysis Software automatically creates an analysis folder with the path <ANALYSIS_ OUTPUT_PATH> location. If the folder is not empty, Local Analysis Software will produce an error.
- ▶ Select one of the following locations:
  - ▶ Full path to the sequencing system run folder location.
  - ▶ Full path to an existing FASTQ folder location.

## Create a Sample Sheet

A Local Analysis Software sample sheet is required for each analysis. A sample sheet is a comma-separated values file (*.csv) that contains information to set up and analyze a sequencing run.

The sample sheet includes a list of samples and their index sequences. Delete index IDs you do not require. The maximum number of samples per analysis run is eight.

Provide the sample sheet during the run setup steps on the sequencing system. See the example sample sheet (SampleSheet.csv) on the TruSight Oncology 500 Product Files page of the Illumina support site.

1    Open theLocal Analysis Software example sample sheet (SampleSheet.csv) on the Illumina support site.

2   Save the sample sheet with the name SampleSheet.csv in the sequencing run folder.

3   In the Header section, enter all of the following parameters:
    ▶ IEMFileVersion–Illumina Experiment Manager software version.
    ▶ Investigator Name–User name.
    ▶ Experiment Name–Experiment name.
    ▶ Date–Date in DD-MM-YY format.
    ▶ Workflow–Enter **GenerateFASTQ**.

4   In the Reads section, enter a number that is one more than the number of cycles for analysis.
    ▶ Number of sequencing cycles for Read 1–101.
    ▶ Number of sequencing cycles for Read 2–101.

5   In the Settings section, enter the following parameters:
    ▶ AdapterRead1–Adapter sequence for Read 1–
      AGATCGGAAGAGCACACGTCTGAACTCCAGTCA.
    ▶ AdapterRead2–Adapter sequence for Read 2–
      AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT.
    ▶ U7N1Y93;I8;I8;U7N1Y93–Specifies the sequencing and indexing cycles that should be
      usedwhen processing the data.

6   In the Data section, enter the required parameters for each sample in plate wells. Both required and
    optional parameters along with any relevant details for each parameter are listed in the table below.

| Sample Parameter | Details |
| --- | --- |
| Sample_ID | The Sample_ID becomes a part of output file names. Sample IDs are not case sensitive. Sample IDs must have the following characteristics:<br>• Unique for the run<br>• No spaces<br>• Alphanumeric characters, underscores, and/or dashes only<br>• If you use an underscore or dash, enter an alphanumeric character before and after the underscore or dash. Example: Sample1-T5B1_022515.<br>• Cannot be called "all", "default", "none", "unknown", "undetermined", "stats", or "reports" |
| Sample_Type | Enter a sample type of either DNA or RNA. |
| Sample_Name | **Optional**. Name of sample |
| Sample_Plate | **Optional** Name of the sample plate. |
| Sample_Well | **Optional** Name of the sample well. |
| Pair_ID | Optional unless paired sample output is desired or pair column is present in the sample sheet. *If pair column is present, all samples must have a pair ID.* Use to pair samples and named combined output files. Recommended format: <PairID>, <PairID>-DNA,<PairID>-RNA. |
| Index adapter sequences | Enter TruSight Oncology 500 compatible index adapter sequences for samples. Select unique index pairs for each sample. Example indexes are listed in SampleSheet.csv. |
| Index | Index 1 sequence. |
| Index2 | Index 2 sequence. |
| I7_Index_ID | **Optional** |
| I5_Index_ID | **Optional** |
| Description | **Optional** Description of the sample. |

7   In the Data section, for each sample plate well, enter the Sample_ID, , Sample_Plate (optional), Sample_Well (optional), TruSight Oncology 500 compatible index adapter sequences for samples, and a Description of the sample (optional).

8   Save the sample sheet with the name SampleSheet.csv in the sequencing run folder.

## Starting From BCL Files

The following inputs are required for running the Local Analysis Software using BCL files. The requirements apply to both Docker and Singularity.

The run folder contains data from the sequencing run. If you start with the run folder, make sure that the folder contains the following files.

| Folder/File | Description |
|---|---|
| Config folder | Configuration files. |
| Data folder | BCL files. |
| Images folder | Raw sequencing image files. |
| Interop folder | Interop metric files. |
| Logs folder | Sequencing system log files. |
| RTALogs folder | Real Time Analysis (RTA) log files. |
| RunInfo.xml file | Run information. |
| RunParameters.xml file | Run parameters. |
| SampleSheet.csv file | Sample information. If you want to use a sample sheet that is not in the run folder, provide the full path. |

## Starting From FASTQ Files

The following inputs are required for running the Local Analysis Software App using FASTQ files. The requirements apply to both Docker and Singularity.

▶   Full path to an existing FASTQ folder.

▶   FASTQ files conform to the folder structure in *FASTQ File Organization* on page 7.

▶   The sample sheet is in the FASTQ folder path, or set the path to the sample sheet with the `--sampleSheet` override command.

The Local Analysis Software uses Illumina BCL Convert for FASTQ Generation. Using files produced by Illumina bcl2fastq will not produce the same results and is discouraged.

For sequencing runs containing both RNA and DNA, BCL Convert must be run twice. Using the local docker app demultiplex only option to generate FASTQs is strongly recommended.

## FASTQ File Requirements

When starting from FASTQ files, follow the recommendations for FASTQ file organization, sample sheet considerations, and BCL Convert compatibility and arguments.

## FASTQ File Organization

Store FASTQ files in individual subfolders that correspond to a specific SAMPLE_ID. Keep file pairs together in the same folder.

```
${full_path_to_fastqfolder}/${Sample_ID}/${Sample_ID}_Sample#_Lane#_Read#_
001.fastq.gz
```

The Local Analysis Software requires separate FASTQ files per sample. Do not merge FASTQ files.

The NextSeq 550 generates two FASTQ files per flow cell lane, so that there are eight FASTQ files per sample.

### Sample1_S1_L001_R1_001.fastq.gz

- ▶ Sample1 represents the Sample ID
- ▶ The S in S1 means sample, and the 1 in S1 is based on the order of samples in the sample sheet, so that S1 is the first sample.
- ▶ L001 represents the flow cell lane number.
- ▶ The R in R1 means Read, so that R1 refers to Read 1.

For an example of the eight files that are output for one sample, see *Folders* on page 1.

## Sample Sheet Settings

Use BCL Convert to produce FASTQ files for TruSight Oncology 500. **Using bcl2fastq will not produce the same results and is discouraged.**

The sample sheet must have the correct settings for both RNA and DNA. Use the following sample sheet settings in BCL Convert to produce input for TruSight Oncology 500:

Table 1   BCL Convert Sample Sheet Settings

| Setting | Value |
| --- | --- |
| AdapterRead1 | AGATCGGAAGAGCACACGTCTGAACTCCAGTCA |
| AdapterRead2 | AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT |
| AdapterBehavior | trim |
| MinimumTrimmedReadLength | 35 |
| MaskShortReads | 35 |
| Setting | Value |
| OverrideCycles | U7N1Y93;I8;I8;U7N1Y93 |

Table 2   RNA-Specific Sample Sheet Settings

| Setting | Value |
| --- | --- |
| OverrideCycles | Y101;I8;I8;Y101 |

## Command Line Settings

Use the following command line settings in BCL Convert to produce input for TruSight Oncology 500:

```
--bcl-input-directory {input run folder path] --output-directory {output
folder path} --sample-sheet {path to sample sheet} --strict-mode true
```

## Run Local Analysis Software in Docker

You can run the workflow analysis from an Illumina sequencing run folder containing BCL files or from FASTQ files staged in a FASTQ folder.

Paths in the command line must be absolute paths. Avoid special characters such as &, *, #, and spaces.

1   Wait for any running Local Analysis Software container to complete before launching a new analysis. Run the following command to check running containers:

```
sudo docker ps
```

2  Select from one of the following five options:

▶ **Start from BCL files in the run folder.**
```
${APP_PATH/{Version}/TruSight_Oncology_500_RUO.sh --user =$UID --remove
   \
--analysisFolder /full/path/to/output/analysisFolder \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder
```

▶ **Start from BCL files using a separate sample sheet.**
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh --user =$UID --
   remove \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder \
--analysisFolder /full/path/to/output/analysisFolder \
--sampleSheet /full/path/to/SampleSheet.csv
```

▶ **Start Demultiplex only.**
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh --user $UID --remove
   \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--runFolder /full/path/to/runFolder --demultiplexOnly \
```

▶ **Start from FASTQ files.**
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh --user =$UID --
   remove \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--fastqFolder /full/path/to/fastqFolder
--analysisFolder /full/path/to/output/analysisFolder
```

> **NOTE**
> The FASTQ folder must have FASTQ files organized by Sample ID defined in the
> SampleSheet.csv. The SampleSheet.csv must be present in the FASTQ folder if not supplied
> using the –sampleSheet option.

▶ **Start from FASTQ files with only paired samples.**
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh --user =$UID --
   remove \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--fastqFolder /full/path/to/fastqFolder \
--samplePairIDs ="Pair_1,Pair_2"
--sampleSheet /full/path/to/SampleSheet.csv
```

## Shell Script Command Line Arguments

| Argument | Description |
| --- | --- |
| --resourcesFolder | Full path to ${RESOURCE_PATH}/resources created above. |
| --analysisFolder | Full path to analysis folder. This path contains the output from the run analysis. Make sure that the analysis folder is empty. |

| Argument | Description |
|---|---|
| --fastqFolder<br>--runFolder | Full path to an existing FASTQ folder or the sequencing system run folder. Provide either -runFolder, or -fastqFolder, but *not* both. |
| --sampleSheet | **[Optional]** Full path to SampleSheet.csv file. If you do not specify a path to the SampleSheet.csv, locate the file directly under the --fastqFolder or --runFolder path. |
| --engine | **[Optional]** The default is Docker. To specify Singularity, set this input to singularity. |
| --user | **[Optional]** For Docker only. Expects a $UID value. Software leverages this user ID to run subsequent processes. |
| --remove | **[Optional]** For Docker only. Removes the subsequent docker container after a run. Identical to passing -rm to a docker run command. |
| --version | **[Optional]** Displays the version of this software. |
| --help | **[Optional]** Displays a help screen with available options. |
| --samplePairIDs | **[Optional]** Defined in the sample sheet. Provide the comma-delimited Sample Pair IDs that should be processed on this node. Example: "Pair_1, Pair_2" |

## Run Local Analysis Software in Singularity

You can run the workflow analysis from an Illumina sequencing run folder containing BCL files or from FASTQ files staged in a FASTQ folder.

1   Select from one of the following five options:
    ▶   Start from BCL files in the run folder.
```
${APP_PATH/${Version}/TruSight_Oncology_500_RUO.sh \
--engine singularity \
--analysisFolder /full/path/to/output/analysisFolder \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder
```
    ▶   Start from BCL files using a separate sample sheet.
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh \
--engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder \
--analysisFolder /full/path/to/output/analysisFolder \
--sampleSheet /full/path/to/SampleSheet.csv
```
    ▶   Demultiplex only.
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh \
--engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--runFolder /full/path/to/runFolder \
--demultiplexOnly
```
    ▶   Start from FASTQ files.
```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh \
--engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--fastqFolder /full/path/to/fastqFolder \
--sampleSheet /full/path/to/SampleSheet.csv
```
    ▶   Start from FASTQ files with paralleling samples.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

10

```
${APP_PATH}/{Version}/TruSight_Oncology_500_RUO.sh \
--engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--fastqFolder /full/path/to/fastqFolder \
--samplePairIDs "Pair_1,Pair_2"
```

## Analysis Methods

The Local Analysis Software workflow performs the following analysis steps, and then writes analysis output files to the folder specified.

► FASTQ Generation
► DNA Analysis Methods
  ► DNA Alignment and Realignment
  ► Read Collapsing
  ► Indel Realignment and Read Stitching
  ► Small Variant Calling
  ► Small Variant Filtering
  ► Copy Number Variant Calling
  ► Phased Variant Calling
  ► Variant Merging
  ► Annotation
  ► Tumor Mutational Burden
  ► Microsatellite Instability Status
  ► Contamination Detection
► RNA Analysis Methods
  ► Downsampling
  ► Read Trimming
  ► Alignment
  ► Duplicate Marking
  ► Fusion Calling
  ► RNA Fusion Filtering
  ► Splice Variant Calling
  ► Annotation
  ► Fusion Merging
► Quality Control
  ► Run QC
  ► DNA Sample QC
  ► RNA Sample QC

## FASTQ Generation

BCL files are demultiplexed and the software generates intermediate analysis files in the FASTQ format. FASTQ files use a text format to represent sequences. They contain reads for each sample and the associated quality scores. Any controls used for the run and clusters that did not pass the filter are excluded. Each FASTQ file contains reads for only one sample, and the FASTQ file name includes the name of that sample.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

11

The bcl-convert software uses the SampleSheet_dna.csv and SampleSheet_rna.csv generated by the analysis module for demultiplexing. The sample sheets can be found in the FastqGeneration step folder.

## DNA Analysis Methods

Figure 1  TruSight Tumor 170 Analysis Module DNA Workflow
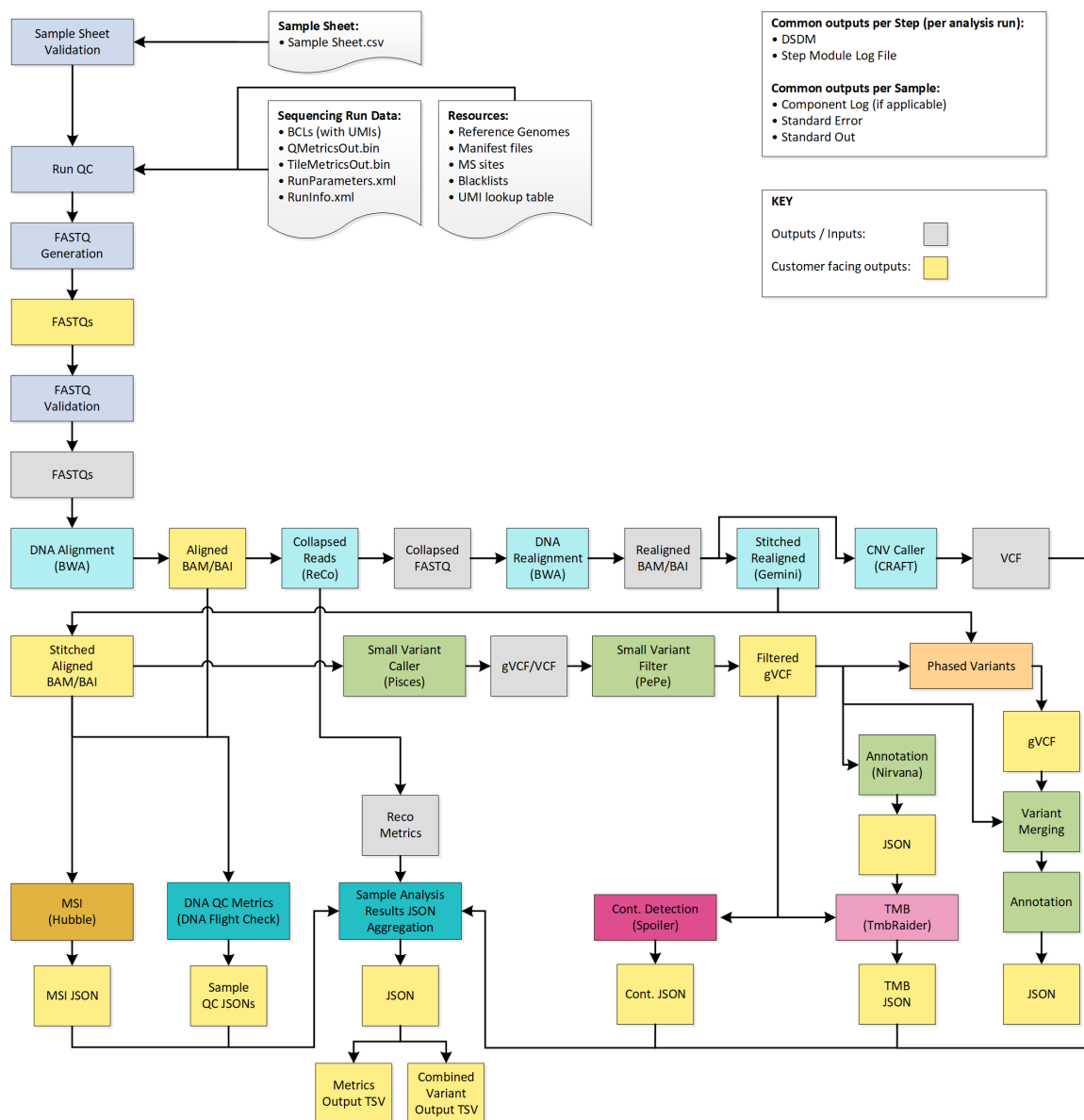


## DNA Alignment and Realignment

The alignment step uses the BWA-MEM alignment with the SAM Tools utility to align DNA sequences in FASTQ files to the hg19 genome. This alignment step generates BAM files (*.bam) and BAM index files (*.bam.bai), which are saved to the DnaAlignment folder. A BAM file is the compressed binary version of a SAM file that is used to represent aligned sequences.

The software performs a second alignment on FASTQ files after the Read Collapsing step completes. The second alignment enables the realignment of sample reads using only unique molecular identifier (UMI) collapsed reads.

For more information on BWA-MEM, see the Burrows-Wheeler Aligner website. For more information on SAM and BAM files, see the Sequence Alignment/Map specification page on GitHub.

## Read Collapsing

The read collapsing analysis step executes an algorithm that collapses sets of reads (known as families) with very similar genomic locations into representative sequences using UMI tags. This process allows for the accurate removal of duplicate reads without losing the signal of very low frequency sequence variations. Additionally, UMI collapsing further reduces FFPE deamination artifacts by utilizing duplex collapsing where information from complimentary strands are combined. The read collapsing step produces FASTQ files and associated metrics files in the CollapsedReads output folder. Read collapsing adds the following BAM tags:

▶ **RX/XU**–UMI.
▶ **XV**–Number of reads in the family.
▶ **XW**–Number of reads in the duplex-family, or 0 if not a duplex family.

## Indel Realignment and Read Stitching

The Gemini software performs local indel realignment, paired-read stitching, and read filtering to improve small variant calling results. A stitched read is a single read that has been combined from a pair of reads. Reads near detected indels are realigned to remove alignment artifacts. It takes in a single BAM file and the genome FASTA used to align it and outputs a corresponding single BAM file with stitched, pair-realigned reads. Read pairs with poor map quality or supplementary and secondary alignments from the input BAM are ignored.

Gemini adds the following BAM tags:

▶ **XD**–For successfully stitched reads only. Directional support string indicating forward, reverse, and stitched positions.
▶ **XR**–For successfully stitched reads only. Pair orientation (FR or RF).

## Small Variant Calling

Pisces software performs somatic variant calling to identify variants at low frequency in DNA samples. Pisces calls small variants in the BAM files that are generated from the StitchedRealigned analysis step.

For each variant candidate, Pisces adds an US field under the Format column in the genome.vcf for the mutant support of the following read type counts: duplex-stitched, duplex-nonstitched, simplex-forward-stitched, simplex-forward-nonstitched, simplex-reverse-stitched, simplex-reverse-nonstitched, followed by total support of the same read type counts. The small variant calling genome.vcf at this step is only responsible to collect candidate and output corresponding read support information, and the final variant call is determined in following post-processing step.

The software component Psara is used to trim the gVCF based on the panel manifest. Variants are considered for inclusion if they overlap with the manifest or are contained within an overlapping indel. Small Variant Filtering determines the final variant call.

Document # 1000000095997 v00

13

For Research Use Only. Not for use in diagnostic procedures.

## Small Variant Filtering

The software component, Pepe, performs post-processing on the small variant calling VCFs with the intended result of background polishing and quality score adjustment. The software filters out variants when their error rates do not meet quality thresholds. This analysis step produces gVCF files and associated error rate files. The minimum read depth for reference calls is 100 and the limit of detection for VAF is 0.05 at that depth.

Pepe computes two quality scores for each variant candidate that dynamically adjusts for the following conditions:

▶ Background noise

▶ Nucleotide change

▶ Read support type

For each variant candidate, background noise at the same site is estimated from normal baseline samples of varying qualities. A p-value is calculated using the observed mutant depth, total depth, and background noise using binomial distribution and then converted to a variant quality score (AQ). The sample-specific error rate of each nucleotide change is estimated from different support categories in each sample by using all the positions with an allele frequency less than 5%. For each variant candidate, a likelihood ratio (LQ) score is computed by the corresponding error rate of the observed total and mutant read.

For variants with a Catalogue of Somatic Mutations in Cancer (COSMIC) count > 50, the LQ and AQ thresholds are 20 and the remaining sites are 60. For indel, at least one stitched mutant support is required. The net effect of the read collapsing and variant filtering reduce false positives in a typical FFPE DNA sample from ~1500 per Mb to < 5 per Mb.

## Copy Number Variant Calling

The CRAFT copy number variant caller performs amplification, reference, and deletion calling for target CNV genes within the assay. The CRAFT software component counts coverage of each target interval on the panel, performs normalization, calculates fold change values for each gene, and determines the CNV status for each CNV target gene. During normalization steps, coverage biases are corrected using potential variables such as sequencing depth, target size, PCR duplicates, probe efficiency, GC bias, and DNA type. A collection of normal FFPE and genomic DNA samples is used to correct some of these variables. For each target CNV gene, *in silico* data is trained to determine a gene specific threshold for amplification and deletion. The inputs are collapsed read in BAM format and the outputs are VCF files. Amplification are annotated as DUP in the VCF file. Deletions status (DEL) are provided for information only and always be marked as LowValidation in the VCF file.

## Phased Variant Calling

Scylla rapidly detects multiple nucleotide variants (MNVs) in a given sample. The workflow uses Scylla to detect specific, clinically relevant mutations in EGFR exon 19 that would otherwise be out of scope for the variant caller. Psara filters the small variant gVCF to a small region in exon 19 of EGFR. Candidate SNPs, MNVs, and indels from this subset of the gVCF are given to Scylla along with the BAM output from Gemini. Scylla uses the original BAM to determine which of these small variants should be phased together into longer MNVs.

For Research Use Only. Not for use in diagnostic procedures.

At a high level, Scylla identifies variants that are candidates for phasing in the input gVCF and arranges the variants into local neighborhoods. Scylla then mines the sample BAM file for any evidence that these small variants occur in the same clonal sub-populations with each other. This is done by clustering overlapping reads in the neighborhood into a minimal set of clusters, which contain the same variants.

Unlike Pisces, Scylla does not require that variants be on the same read to be phased. Once the phasing is complete, a new gVCF is generated.

## Variant Merging

The software merges the phased variants with the other small variants generated from small variant filtering step and produces a gVCF file. In this process, exact duplicates that match chromosome, position, reference allele, and alternative allele are removed. Only the following Epidermal Growth Favor Receptor (EGFR) variants are added if found from Phased Variant Calling. All other EGFR variants are filtered out in variant merging.

Table 3   EGFR Variants

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

| chr7 | 55242465 | GGAATTAAGAGAAGCAA | AATTC |
|------|----------|-------------------|-------|
| chr7 | 55242465 | GGAATTAAGAGAAGCAACATC | AAT |
| chr7 | 55242468 | ATTAAGAGAAGCAAC | GCA |

## Annotation

The Illumina Annotation Engine Nirvana performs annotation of small variants. The inputs are gVCF files and the outputs are annotated JSON files.

Each variant entry that is processed by Nirvana will be annotated with available information from databases such as dbSNP, gnomAD genome and exome, 1000 genomes, ClinVar, COSMIC, RefSeq, and Ensembl. Version information and general details can be retrieved from the header, which is the first line of the JSON file. Each annotated variant is included as a nested dictionary structure in separate lines following the header. Version information for each annotation database is shown in the following table.

| Database | Version |
|----------|---------|
| gnomAD | 2.1 |
| COSMIC | v84 |
| ClinVar | 2019-02-04 |
| dbSNP | v151 |
| 100 Genomes | Phase 3 v5a |
| RefSeq | VEP build 91 |
| Ensembl | VEP build 91 |

## Tumor Mutational Burden

The tumor mutational burden (TMB) analysis step generates TMB metrics from the annotated small variant JSON file and the gVCF file that is generated from the small variant filtering analysis step. The annotated JSON file is used to retrieve information regarding individual variants, such as allele counts in public databases and resulting consequences at a transcript level. The gVCF is used to evaluate the effective panel size denominator.

To remove germline variants from the TMB calculation, the software uses a combination of public database filtering and post-database filtering strategy that uses allele frequency information and variants in close proximity.

First, the component excludes any variant with an observed allele count ≥10 in any of the GnomAD exome, genome, and 1000 genomes database. To filter germline variants that are not observed in the database, the software identifies variants on the same chromosome with an allele frequency within a certain range. If a given variant is not filtered out based on occurrence in the databases, variants on the same chromosome with similar allele frequencies will be grouped, and if 5 or more similar variants are found to have been filtered, the variant of interest is removed from the TMB Calculation. Additionally, variants with an allele frequency ≥ 90% are removed from the TMB calculation as well. The TMB is calculated as follows:

TMB = Eligible Variants / Effective panel size

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

16

| Eligible Variants (numerator) | • Variants not removed by the filtering strategy.<br>• Variants in the coding region (RefSeq Cds)<br>• Variant Frequency >= 5%<br>• Coverage >= 50X<br>• SNVs and Indels (MNVs excluded)<br>• Nonsynonymous and synonymous variants<br>• Variants with COSMIC count >= 50 excluded |
|---|---|
| Effective Panel Size (denominator) | • Total coding region with coverage > 50X<br>• Excluding low confidence regions in which variants are not called |

Outputs are captured in a *_TMB_Trace.tsv file that contains information regarding variants used in the TMB calculation and a *.tmb.json file that contains the TMB score calculation and configuration details.

## Microsatellite Instability Status

The microsatellite instability status step determines microsatellite instability from the BAM file from the read stitching analysis step and generates an MSI metric file. The software assesses microsatellite sites for evidence of instability, relative to a set of baseline normal samples that are based on information entropy metrics. The percentage of unstable MSI sites to total assessed MSI sites is reported as a sample-level microsatellite score.

## Contamination Detection

The contamination analysis step detects contamination by foreign DNA in the VCF files that the small variant filter step generates. The software determines whether a sample has foreign DNA from the combination of contamination p-value (p-score) and contamination scores.

The contamination score is the sum of all the log likelihood scores across all positions. The p-score represents the significance that SNPs are distributed non-uniformly across the chromosomes. This could indicate a highly-rearranged genome and cause false positives for contaminaton.

In contaminated samples, there are SNPs that have variant allele frequency shifts from 0%, 50%, or 100%. The algorithm collects all the positions that overlap with common SNPs with variant allele frequencies of < 25% or > 75%. Then, the algorithm computes the likelihood that the positions are an error or a real mutation using the following qualifications:

▶ Estimates the error rate per sample.

▶ Mutation support.

▶ Total depth of each position selected.

# RNA Analysis Methods

Figure 2   TruSight Tumor 170 Analysis Module RNA Workflow
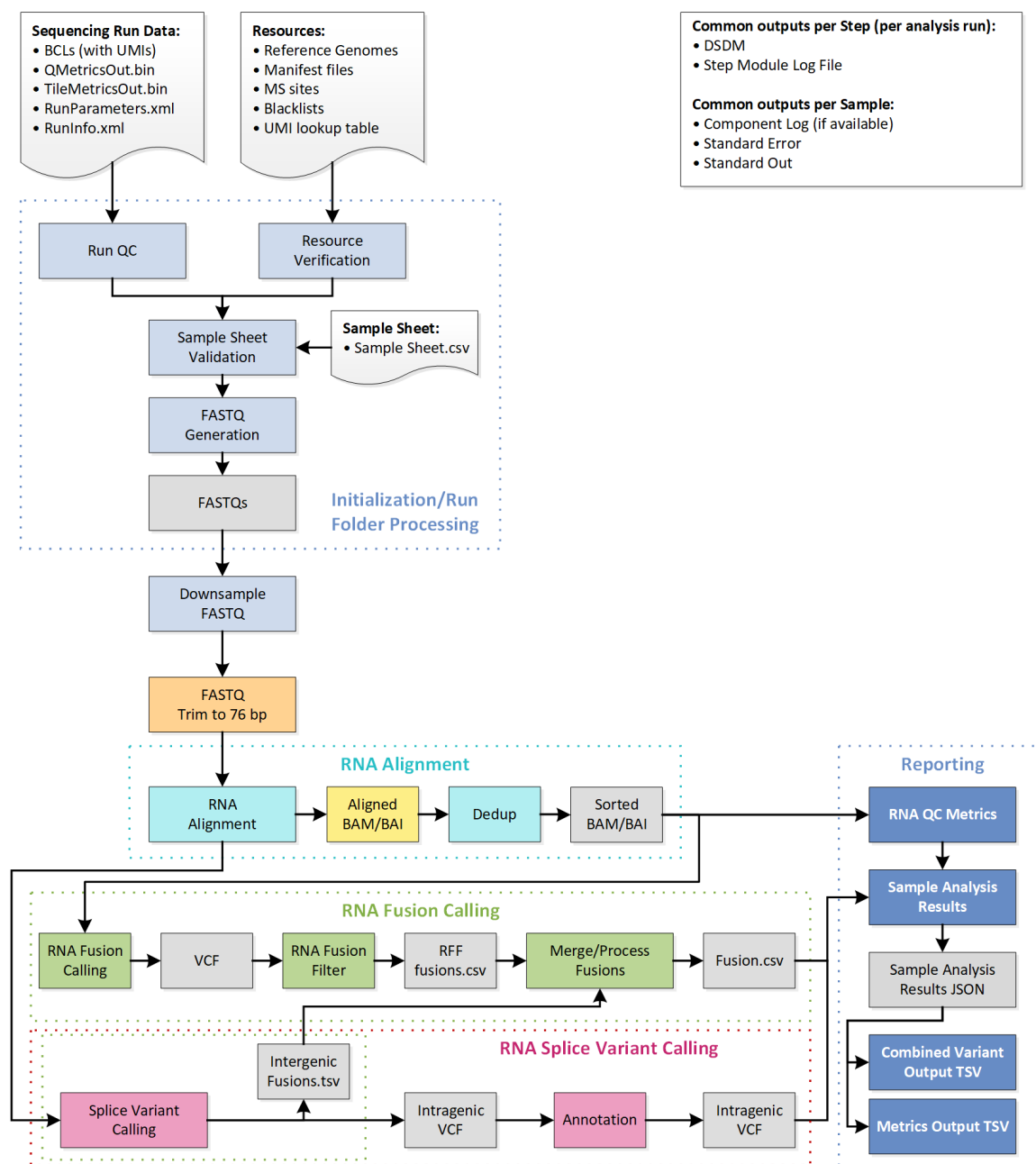


## Downsampling

Each sample is downsampled to 30 million RNA reads. This number represents the total number of single reads (ie, R1 + R2, from all lanes). When using the recommended sequencing configurations or plexing, the samples can have fewer reads than the downsampling limit. In these cases, the FASTQ files are left as-is.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

18

## Read Trimming

Reads are trimmed to 76 base pairs for further processing. From internal testing, fragment sizes in RNA FFPE samples hover around 100 bp, so the majority of reads at 76 bp are overlapping (see ). While STAR alignment does perform stitching to handle overlapping reads, internal testing using simulated data indicates that performance is improved with fewer overlapping reads.
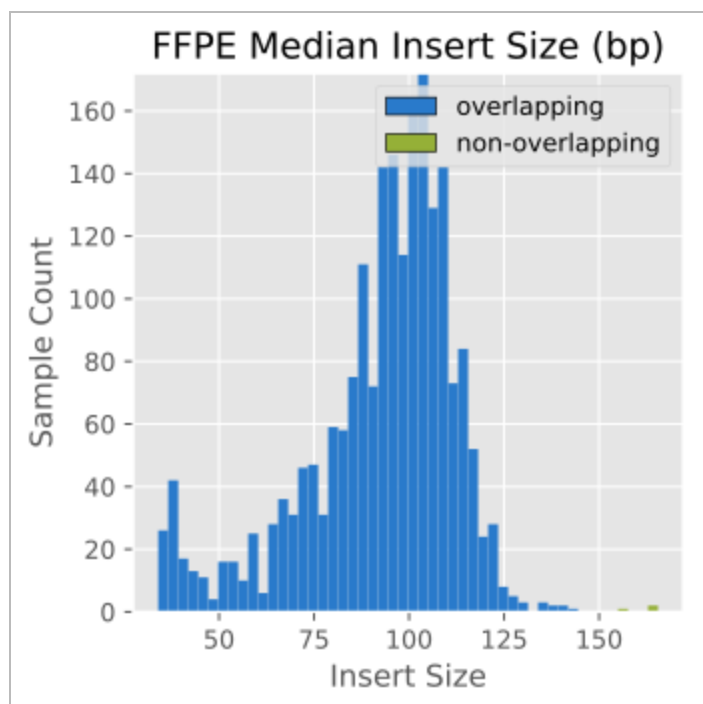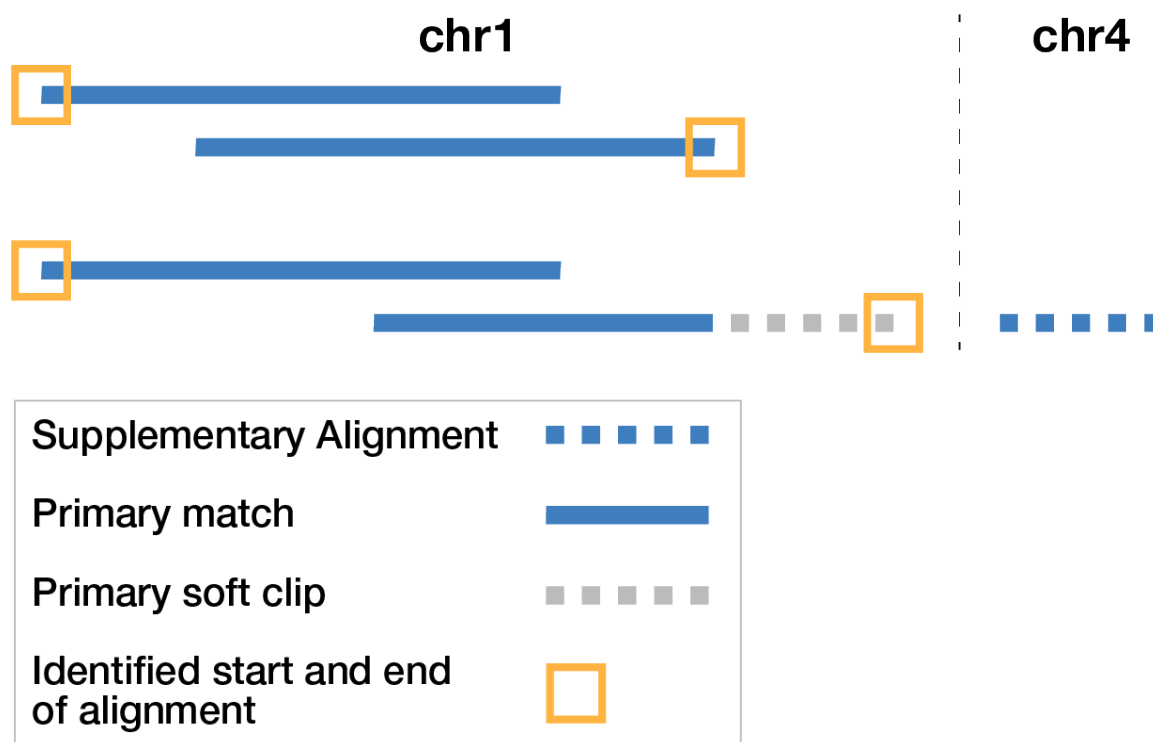


## Alignment

The STAR Aligner aligns RNA reads to the human hg19 genome containing unplaced contigs (i.e., chrUn_gl regions) and uses GENCODEv19 to identify splice sites. STAR also includes stitching logic to handle overlapping reads.

## Duplicate Marking

RNA does not contain UMIs, so duplicate marking is performed using an internally developed tool based on the Picard duplicate marking algorithm. The start and end coordinates of alignments (adjusted for soft clipping) are used to determine whether fragments are overlapping or not. Fusion and splice variant calling only use deduped fragments to score variants. Only primary alignments are considered, supplementary and secondary alignments are not. The alignment with the highest read score is chosen as the unique fragment.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

19

Figure 3   Picard Alignment Based Duplicate Marking.



## Fusion Calling

Manta is a fusion caller which discovers, assembles, and scores large-scale SVs. Manta only considers candidate fusions with at least 3 unique supporting reads, one of which must be a split read (a single read crossing the fusion breakpoint). The inputs are BAM files and the outputs are VCF files.

## RNA Fusion Filtering

The RNAFusionFilter scores fusions and applies filters based on contig alignment to the genome and other features. It also determines which gene is on the 3' end and which gene is on the 5' end based on gene annotations and aligning the contig to the genome. The inputs are BAM files and VCF files and the outputs are .csv format.

Table 4   Scored Features

| Score Component | Fusion Feature | Scored Range | Coefficient |
|---|---|---|---|
| Split reads | SplitAlt | 0-10 | .15 |
| Paired reads | PairedAlt | 0-5 | .15 |
| Alt allele fraction | (splitAlt + PairAlt) / (SplitRef + PairRef) | 0-.01 | .1 |
| Fusion contig align length (bp) | ContigAlign1 ContigAlign2 | 12-76 | .4 |
| Breakend homology (bp) | BreakpointHomology | 2-20 | -.2 |

| | | | |
|---|---|---|---|
| Fusion contig align length (bp) (For fusions on different chromosome: dist-DistMax) | | 100000 - 2000000 | .1 |
| Coverage after breakend (bp) | min(CoverageGene1, CoverageGene2) | 0-500 | .1 |

## Splice Variant Calling

Splice variant calling is performed using internally developed software. The inputs are BAM files and SJ.out.tab files from STAR. Junctions from SJ.out.tab are filtered first using splice annotations from GENCODEv19, and then further filtered using a baseline from a cohort of non-tumor FFPE samples of varying tissue types. Splice junctions appearing on a whitelist are not filtered. The whitelist contains ARv7, MET exon 14 skipping, and EGFRvIII. The outputs are VCF files, which are the final output, and TSV file containing intergenic variants, which are used in fusion merging. Splice variants are scored from 0-10 as shown in the table below.

Table 5   Scored Features in Splice Variant Caller

| Score Component | Splice Feature | Scored Range | Coefficient |
|---|---|---|---|
| Split reads | split_unique_reads_alt | 0-10 | 1 |

## Annotation

The Illumina Annotation Engine performs annotation of splice variants. The inputs and outputs are VCF files.

## Fusion Merging

Fusions identified through the fusion calling and RNA fusion filtering are combined with the intergenic calls made during splice variant calling. Additionally, each precise fusion event from the RNA fusion filter is recalculated for read count support. The inputs are intergenic splice variant TSV files, fusion filter CSV files, and BAM files. The outputs are CSV files.

## Quality Control

The TruSight Oncology 500 includes several quality control analyses.

## Run QC

The Run Metrics report provides suggested values to determine if run quality results are within an acceptable range using InterOp files from the sequencing run folder. For Read 1 and Read 2, the report provides the average percentage of bases ≥ Q30, which is a quality score (Q-score) measurement. A Q-score is a prediction of the probability of an incorrect base call.

## DNA Sample QC

The inputs for DNA Sample QC are DNA alignment, read collapsed BAM, indel realignment, read stitching BAM, and CRAFT normalized BinCount.tsv files. The metrics and guideline thresholds can be found in the MetricsOutput.tsv file.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

21

| Metric | Description | Recommended Guideline Quality Threshold | Variant Class |
|---|---|---|---|
| CONTAMINATION_ SCORE and CONTAMINATION_ P_VALUE | The contamination score from based on VAF distribution of SNPs. The contamination p-value is used to assess highly rearranged genomes and only needed when contamination score is above USL. A p-score less than 0.05 suggest that the sample has likely large-scale rearrangements that could lead to high contamination scores without actual sample contamination. | Contamination Score ≤ 3106' OR Contamination Score > 3106 and Contamination p-value ≤ 0.049 | All |
| MEDIAN_EXON_ COVERAGE | Median exon fragment coverage across all exon bases. | ≥ 150 | Small variant Tmb |
| PCT_EXON_50X | Percent exon bases with 50X fragment coverage. | ≥ 90.0 | Small variant Tmb |
| MEDIAN_INSERT_ SIZE | The median fragment length in the sample. | ≥ 70 | Small variant Tmb |
| COVERAGE_MAD | Median Absolute Deviation. Represents the median normalized deviation across all regions used for CNV calling. | ≤ 0.210 | CNV |
| MEDIAN_BIN_ COUNT_CNV_ TARGET | The median raw bin count per CNV target. | ≥ 1.0 | CNV |
| USABLE_MSI_ SITES | The number of MSI sites usable for MSI calling. | ≥ 40 | MSI |

## RNA Sample QC

The inputs for RNA Sample QC are RNA alignment. Metrics and guideline thresholds can be found in the MetricsOutput.tsv file.

| Metric | Description | Recommend Guideline Quality Threshold | Variant Class |
|---|---|---|---|
| MEDIAN_ CV_ GENE_ 500X | The median CV for all genes with median coverage > 500x. Genes with median coverage > 500x are likely to be highly expressed. Higher CV median > 500x indicates an issue with library preparation (poor sample input and/or probes pull down issue). | ≤ 93 | Fusion Splice |
| MEDIAN_ INSERT_ SIZE | The median fragment length in the sample. | ≥ 80 | Fusion Splice |
| TOTAL_ ON_ TARGET_ READS | The total number of reads that map to the target regions. | ≥ 9000000 | Fusion Splice |

## Analysis Output

When the analysis run completes, the Local Analysis Software generates an analysis output folder in a user-specified location with a user-specified folder name.

For Research Use Only. Not for use in diagnostic procedures.

Within the analysis folder, each analysis step generates a subfolder within the Logs_Intermediates folder. The software writes the analysis final outputs to the Results folder within a Pair ID or Sample ID subfolder.

To view analysis output, complete the following steps.

1    Navigate to the directory that contains the analysis output folder.

2    Open the folder, and then select the files that you want to view.

## Combined Variant Output

The combined variant output file contains the variants and biomarkers in a single file that is based on a Paired Sample (if using PairID) with the file name *_CombinedVariantOutput.tsv. The following variant types and biomarkers are present in the output:

▶    Small Variants (including EGFR complex variants)

▶    Gene Amplifications

▶    TMB

▶    MSI

▶    Fusions

▶    Splice Variants

The combined variant output file also contains Analysis Details and Sequencing Run Details sections. The content of each is listed in the following table.

| Analysis Details | Sequencing Run Details |
|---|---|
| • Pair ID<br>• DNA Sample ID (if DNA is run)<br>• RNA Sample ID (if RNA is run)<br>• Output Date<br>• Output Time<br>• Module Version<br>• Pipeline Version (Docker Image Version #) | • Run Name<br>• Run Date<br>• DNA Sample Index ID (if DNA is run)<br>• RNA Sample Index ID (if RNA is run)<br>• Instrument ID<br>• Instrument Control Software Version<br>• Instrument Type<br>• RTA Version<br>• Reagent Cartridge Lot Number |

## Variant Filtering Rules

▶    **Small Variants**—All variants with the FILTER field marked as PASS in the merged genome VCF are present in the combined variant output.
  ▶    Gene information is only present for variants belonging to canonical transcripts that are within the Gene Whitelist-Small Variants.
  ▶    Transcript information is only present for variants belonging to canonical transcripts that are within the Gene Whitelist-Small Variants.

▶    **Copy Number Variants**—All copy number variants in the copy number variant VCF are only present if they meet the following conditions:
  ▶    FILTER field marked as PASS.
  ▶    ALT field is <DUP>.

▶    **Combined Variants**—All fusion variants that meet the following criteria:
  ▶    Passing Variant Call (KeepFusion field is true).
  ▶    Contains at least one gene on the fusion whitelist.

▶    **Biomarkers TMB/MSI**—Always present when DNA sample is processed.

▶ Fusion Variants—All fusion variants in the fusions csv are present if they meet the following condition:
   ▶ At least one gene partner is on the Gene Whitelist Fusion Section.
▶ **Splice Variants**—Passing splice variants that are contained on genes EGFR, MET, and AR.

## Metrics Output

The MetricsOutput.tsv file contains the following quality control metrics for all of the samples: QC metrics for small variant calling, TMB, MSI, CNV, and Fusion, along with RunQc analysis status and contamination. This TSV file also includes expanded DNA library QC metrics per sample, based on total reads, collapsed reads, chimeric reads, and on-target reads.

## DNA Output

## Merged Small Variant gVCF

File name: {SAMPLE_ID}_MergedSmallVariants.genome.vcf

The merged variant genomeVCF combines the small variant genome VCF (output of Variant Filtering step) and clinically relevant variants in EGFR exon 19 from Phased Variant calling. This contains information on all candidate small variants evaluated. The variant status is determined by the FILTER column in the genome VCF as follows:

| ALT | Filter | Note |
| --- | --- | --- |
| . | PASS | WT |
| ., A, C, G and etc. | LowDP | No call (DP < 100X, insufficient depth to confidently detect variants with VF >= 5%) |
| A, C, G and etc. | PASS | PASS variants |
| A, C, G and etc. | LowSupport | Filtered variant candidate:<br>• Fail AQ or LQ<br>• 0 stitched support for indel or variant in homopolymer context |
| A, C, G and etc. | Blacklist | Position with high background noise. Not available for variant detection |
| A, C, G and etc. | LowVarSupport | Filtered variant candidate with mutant support <1. |

## Merged Small Variant Annotated JSON

File name: {SAMPLE_ID}_MergedSmallVariantsAnnotated.json.gz

The merged small variants annotated file provides variant annotation information for all non-reference positions from the Merged genome VCF including Pass and non-Pass variants.

## TMB Trace

File name: {Sample_ID}_TMB_Trace.tsv

The TMB trace file provides comprehensive information on how the TMB value is calculated for a given sample. All passing small variants from the small variant filtering step are included in this file. To calculate the numerator of the TmbPerMb value in the TMB JSON, set the TSV file filter to use the IncludedInTMBNumerator with a value of True.

Document # 1000000095997 v00

For Research Use Only. Not for use in diagnostic procedures.

24

The TMB trace file is not intended to be used for variant inspections. The filtering statuses are exclusively set for TMB calculation purposes, and setting a filter does not translate into the classification of a variant as somatic or germline.

| Column | Description |
|---|---|
| Chromosome | Chromosome |
| Position | Position of variant |
| RefCall | Reference base |
| AltCall | Alternate base |
| VAF | Variant allele frequency |
| Depth | Coverage of position |
| CytoBand | Cytoband of variant |
| GeneName | Name of gene if applicable. A semicolon delimited list is used for multiple genes. |
| VariantType | Type of the variant: SNV, insertion, deletion, MNV |
| CosmicIDs | Cosmic IDs, if multiple concatenated by ";" |
| MaxCosmicCount | Maximum Cosmic study count |
| AlleleCountsGnomadExome | Variant allele count in gnomAD exome database |
| AlleleCountsGnomadGenome | Variant allele count in gnomAD exome database |
| AlleleCounts1000Genomes | Variant allele count in 1000 genomes database |
| MaxDatabaseAlleleCounts | Maximum variant allele count over the three databases mentioned above |
| GermlineFilterDatabase | TRUE if variant was filtered by the database filter |
| GermlineFilterPRoxi | TRUE if variant was filtered by the proxi filter |
| CodingVariant | TRUE if variant is in the coding region |
| Nonsynonymous | TRUE if variant has any transcript annotations with nonsynonymous consequences |
| IncludedinTMBNumberator | TRUE if variant is used in the TMB calculation |

# Copy Number VCF

File name: {Sample_ID}_CopyNumberVariants.vcf

This file contains CNV calls for DNA libraries of the amplification genes targeted by TruSight Oncology 500. The CNV call indicates fold change results for each gene classified as reference, deletion, or amplification.

The value in the QUAL column of the VCF is a Phred transformation of the p-value where $Q=-10 \times \log10$ (p-value). The p-value is derived from the t-test between the fold change of the gene against rest of the genome. Higher Q-scores indicate higher confidence in the CNV call.

In the VCF notation, <DUP> indicates the detected fold change (FC) is greater than a predefined amplification cutoff. <DEL> indicates the detected fold change (FC) is less than a predefined deletion cutoff for that gene. This cutoff can vary from gene to gene.

<DEL> calls have only been validated with *in silico* data sets. As a result, all <DEL> calls have LowValidation filter in the VCF.

Each copy number variant is reported in the form of fold change on normalized read depth in a testing sample relative to the normalized read depth in diploid genomes. Given tumor purity, you can infer the ploidy of a gene in the sample from the reported fold change.

Given tumor purity X%, for a reported fold change Y, the copy number n can be calculated by:

$$n=[(200 \times Y)-2 \times (100-X)]/X$$

For example,a tumor purity at 30% and a MET with fold change of 2.2x indicates that 10 copies of MET DNA are observed.

## RNA Output

### Splice Variant VCF

File name: {Sample_ID}_SpliceVariants.vcf

Contains all candidate splice variants targeted by the Analysis panel identified by the RNA analysis pipeline. The following filters can be applied for each variant call:

| Filter Name | Description |
|---|---|
| LowQ | Splice Variant Score is < the Passing Quality Score threshold value of 1. |
| PASS | Splice Variant Score is ≥ to the Passing Quality Score threshold value of 1. |

See the headers in the output for more information about each column.

### Splice Variant Annotated JSON

Each splice variant is annotated using the Illumina Annotation Engine and the following information is captured in the JSON if available:

- ▶ HGNC Gene
- ▶ Transcript
- ▶ Exons
- ▶ Introns
- ▶ Canonical
- ▶ Consequence

### All Fusions CSV

File name: {Sample_ID}_AllFusions.csv

This file contains all candidate fusions identified by the RNA analysis pipeline. Candidate fusions from the splice variant caller are listed in this output with relevant supporting information but are not considered high confidence. Two key output columns in the file describe the candidate fusions: Filter and KeepFusion.

The following table describes the output found in the Filter columns. The output is either a confidence filter or information only as indicated. If none of these filters are triggered, the Filter column displays PASS.

Table 6   Filter Column Output

| Filter | Description |
|---|---|
| Imprecise | **(Confidence filter)** A low-resolution candidate, not an assembled fusion call. |
| RepeatOverlap | **(Confidence filter)** The fusion is tagged as overlapping with a repeat region. Only used as a confidence filter for nonuniquely mapping fusion candidates, otherwise information only. |

| | |
|---|---|
| WeakBreakend | **(Confidence filter)** The read/alignment evidence on one side of the fusion is weak. Usually this filter indicates that the reads only overlap the fusion by a few base pairs. Alternatively, it can indicate too much homology (no unique sequence). |
| Homology | **(Information only**) The fusion contig is a substring of another fusion contig. |
| DuplicateContig | **(Confidence filter)** The two contigs of the fusion are the same contig. |
| ContigIntragenic | **(Confidence filter)** The realignment of half-contigs produces alignments that map to the same gene on both sides (or within 1 kb if unannotated). |
| LowQ | **(Confidence filter)** Fusion supporting reads (unique) < 5 (+ 1 for every 10 million reads over 16 million reads). |
| LowDupReads | **(Information only)** Fusion supporting reads (duplicate) < 5. |
| NonExonic | **(Information only)** Fusion breakpoint does not fall within an exon. |
| LocalContigAlign | **(Information only)** Contig realignment found a nonfusion alignment for this contig. |
| LowFusionRatio | **(Information only)** Few strong evidence reads compared to wild type reads. |
| NoReferenceReads | **(Information only)** No reads on either side of the presumed breakpoint are marked as reference (structurally normal) reads. |

The KeepFusion column of the output has a value of True when the RNAFusionFilter score is ≥ 0.45,none of the confidence filters are triggered, and fusions called by the splice variant caller have a score of 1.

See the headers in the output for more information about each column.

Table 7  Fusion Columns

| Fusion Object Field | Source |
|---|---|
| Caller | **(Either RNAFusionFilter or SpliceGirl)** The algorithm used to identify the fusion. |
| Gena A | The gene associated with the A side of the fusion. A semicolon delimited list is used for multiple genes. |
| Gene B | The gene associated with the B side of the fusion. A semicolon delimited list is used for multiple genes. |
| Gene A Breakpoint | The chromosome and offset of the Gene A side of the fusion. |
| Gene B Breakpoint | The chromosome and offset of the Gene B side of the fusion. |
| Score | The quality of fusion as determined by the respective caller. Results from different callers are not equivalent. |
| Filter | The filter associated with the fusion as determined by the respective caller. Results from different callers are not equivalent. |
| Precise/Imprecise | **(RNAFusionFilter Only)** Whether the algorithm could identify the precise fusion coordinates. Coordinates listed for imprecise fusions are based on strongest statistical evidence. |
| Intragenic Call | **(SpliceGirl Only)** List any genes associated with a splice overlapping the fusion. A semicolon delimited list will be used for multiple genes. |
| Ref A Split | Gene A uniquely mapping reads spanning the junction. Does not support fusion. Duplicate reads included. |
| Ref A Pair | Gene A uniquely mapping reads paired across junction. Does not support fusion. Duplicate reads included. |
| Ref B Split | Gene B uniquely mapping reads spanning the junction. Does not support fusion. Duplicate reads included. |

| Ref B Pair | Gene B uniquely mapping reads paired across junction. Does not support fusion. Duplicate reads included. |
|---|---|
| Alt Split | Uniquely mapping reads split by the junction. Supports fusion. Duplicate reads included. |
| Alt Pair | Uniquely mapping reads paired across junction. Supports fusion. Duplicate reads included. |
| CandidateAlt | **(RFF Imprecise Reads Only)**The number of reads and pairs that potentially supports this candidate before refinement and scoring. |
| Contig | Sequence of fusion. Can be used to determine fusion directionality. (RFF Only) |
| ContigAlign1 | **(RFF Only)** Length of Gene A in contig. |
| CntigAlign2 | **(RFF Only)** Length of Gene B in contig. |
| KeepFusion | The determination whether the fusion should be kept or dropped from the list of fusions. |
| Ref A Dedup | Gene A uniquely mapping reads paired across or split by the junction. Does not support fusion. Duplicate reads are not included. |
| Ref B Dedup | Gene B uniquely mapping reads paired across or split by the junction. Does not support fusion. Duplicate reads are not included. |
| Alt Split Dedup | Uniquely mapping reads split by the junction. Supports fusion. Duplicate reads are not included |
| Alt Pair Dedup | Uniquely mapping reads paired across junction. Supports fusion. Duplicate reads are not included. |
| Fusion Directionality Known | Whether Fusion Directionality is known and indicated by gene order. |

When using Microsoft Excel to view this report, genes that are convertible to dates (such as MARCH1) automatically convert to dd-mm format (1-Mar) by Excel. Fusion whitelist genes are listed below:

- ▶ ABL1
- ▶ AKT3
- ▶ ALK
- ▶ AR
- ▶ AXL
- ▶ BCL2
- ▶ BRAF
- ▶ BRCA1
- ▶ BRCA2
- ▶ CDK4
- ▶ CSF1R
- ▶ EGFR
- ▶ EML4
- ▶ ERBB2
- ▶ ERG
- ▶ ESR1
- ▶ ETS1
- ▶ ETV1

- ▶ ETV4
- ▶ ETV5
- ▶ EWSR1
- ▶ FGFR1
- ▶ FGFR2
- ▶ FGFR3
- ▶ FGFR4
- ▶ FLI1
- ▶ FLT1
- ▶ FLT3
- ▶ JAK2
- ▶ KDR
- ▶ KIF5B
- ▶ KIT
- ▶ KMT2A
- ▶ MET
- ▶ MLLT3
- ▶ MSH2
- ▶ MYC
- ▶ NOTCH1
- ▶ NOTCH2
- ▶ NOTCH3
- ▶ NRG1
- ▶ NTRK1
- ▶ NTRK2
- ▶ NTRK3
- ▶ PAX3
- ▶ PAX7
- ▶ PDGFRA
- ▶ PDGFRB
- ▶ PIK3CA
- ▶ PPARG
- ▶ RAF1
- ▶ RET
- ▶ ROS1
- ▶ RPS6KB1
- ▶ TMPRSS2
- ▶ TP53

## Uninstall Local Analysis Software

The procedure to uninstall the TruSight Oncology 500 Local App removes the following assets:

▶ TruSight Oncology 500 Local App

▶ Containers (if you performed an analysis)

▶ Singularity image

▶ Docker image

▶ Bash Script

1 Check for active runs. When a run is active, at least one Docker container contains the image **trusight-oncology-500**.
```
$ sudo docker ps
$ CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
```

2 If there are active runs, stop run execution.
```
# using the container id, found in the above command:
$ sudo docker stop <container_id>
```

3 Remove all Docker containers with the **trusight-oncology-500** image.
```
# similar to the above, identify the container id:
$ sudo docker ps -a|grep trusight-oncology-500
# then remove each container:
$ sudo docker rm $(sudo docker ps -a|grep trusight-oncology-500| \ awk
    '{print $1}')
```

4 Remove Docker from the image and previous runs.
```
$ sudo docker rmi -f trusight-oncology-500:${VERSION}
```

5 Remove downloaded assets.
```
$ rm ${APP_PATH}md5sum.txt
$ rm -rf ${APP_PATH}/resources
$ rm ${APP_PATH}/{trusight-oncology-500-ruo-dockerimage-ruo-
    2.0.0.70.tar}
$ rm ${APP_PATH}/{trusight-oncology-500-ruo.img}
$ rm ${APP_PATH}/{Trusight_Oncology_500_RUO.sh}
```

## Troubleshooting

| Failure Type | Actions |
|---|---|
| Software | Open the log file ./{analysisFolder}/Logs_Intermediates/TruSight-Oncology-500-pipeline-<timestamp>.log file. The log file displays all commands run by the workflow and the exit code for each analysis step. If a step fails, review standard output and standard error log files in the folder ./{analysisFolder}/Logs_Intermediates/. |
| Samples | Open the final sample biomarker report log file ./{analysisFolder}/Results/SampleID/CombinedVariantOutput.tsvt. If a sample fails an analysis step, the step name appears in the [SAMPLE STATUS] section of the report in the Failed Steps field. Review the log files for the step in ./{analysisFolder}/Logs_Intermediates/{FailedStep}/. |

illumına®