TruSight Oncology 500 Local App

User Guide

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

Table of Contents

Document # 1000000067616 v00

iv

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

Introduction

The Illumina® TruSight™ Oncology 500 Local App analyzes sequencing reads of DNA libraries that are isolated from FFPE tissue and prepared using the TruSight Oncology 500 assay. The software produces a variant call file (*.vcf) for small variants. Reports include mutations per megabase scores for tumor mutational burden (TMB) and the percentage of unstable sites for microsatellite instability (MSI) status.

The application can start the secondary analysis workflow from an Illumina sequencing run folder containing base call files (BCL) or from FASTQ files staged in a FASTQ folder.

This document provides an overview of the TruSight Oncology 500 Local App workflow, computing requirements, installation, analysis software, analysis methods, and analysis outputs. Use of the app requires basic knowledge of the Linux operating system and the Docker or Singularity software container platform.

Settings for Sequencing

For optimal coverage, sequence eight DNA libraries per sequencing run with the following required settings.

▶ 2 x 101 bp

▶ Paired end

▶ Dual index

▶ For low-plex sequencing runs, use three or more libraries containing one of the following combinations of UPxx index primer mix sets to provide sufficient diversity. For example, the first library is assigned UP01, the second library UP02, and the third library UP03.

  ▶ Index set #1–UP01, UP02, UP03

  ▶ Index set #2–UP04, UP05, UP06

  ▶ Index set #3–UP07, UP08, UP09

  ▶ Index set #4–UP10, UP11, UP12

Computing Requirements

The following are the minimum operating requirements for TruSight Oncology 500 Local App analysis.

Compatibility with Illumina Systems

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

▶ NextSeq 500

▶ NextSeq 550

▶ NextSeq 550Dx (in RUO mode)

Software

▶ md5sum

▶ Linux CentOS 7.3 operating system, or later.

▶ Container engine Docker CE v17.03* or Singularity v2.4.2, or later.

▶ HTTP client such as Curl or Wget.

* If you have Network File System (NFS) mounted volumes for analysis output, use Singularity or contact Illumina Technical Support.

Hardware

▶ Minimum:
  ▶ 64 GB RAM
  ▶ 64-bit CPU with 12 physical cores, 24 threads (recommend SandyBridge or equivalent).
  ▶ 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 82 GB.
  ▶ The app download file is 32 GB, and is a single *.gzip compressed tar archive.
  ▶ The resource bundle requires 47 GB of space after decompression.
▶ A sequencing run with secondary analysis requires a minimum of 300 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.

Additional Resources

Visit the TruSight Oncology 500 Local App page on the Illumina support website for documentation, software downloads, training resources, and information about compatible Illumina products.

Document # 1000000067616 v00

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

2

For information on installing the product, see the Downloads section of the TruSight Oncology 500 Local App support page.

The following documentation is available for download from the Illumina website.

Resource	Description
TruSight Oncology 500 Local App Customer Release Notes (document # 1000000072787)	Provides information on software enhancements, limitations, and considerations for the current release.
TruSight Oncology 500 Local App SampleSheet.csv (document # 1000000072379)	Setup template for running the analysis.

Sample Sheet Requirements

A TruSight Oncology 500 Local App 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 instrument. See the example TruSight Oncology 500 Local App sample sheet (SampleSheet.csv) on the Illumina support site.

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

2   In the Header section, enter 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**.

3   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–Enter **101**.

  ▶   Number of sequencing cycles for Read 2–Enter **101**.

4   In the Settings section, enter the following parameters.

  ▶   Adapter–Adapter sequence for Read 1. Enter **AGATCGGAAGAGCACACGTCTGAACTCCAGTCA**.

  ▶   AdapterRead2–Adapter sequence for Read 2. Enter **AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT**.

  ▶   Read1UMILength–Enter **7**.

  ▶   Read2UMILength–Enter **7**.

  ▶   Read1StartFromCycle–Enter **9**.

  ▶   Read2StartFromCycle–Enter **9**.

5   In the Data section, enter the following parameters for each sample in plate wells.

Document # 1000000067616 v00

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

3

- ▶ Sample_ID—Enter a sample ID with the following characteristics. The Sample_ID becomes a part of output file names. Sample IDs are not case sensitive.
  - ▶ Unique for the run.
  - ▶ < 40 characters with no spaces.
  - ▶ Alphanumeric characters, underscores, and/or dashes only.
  - ▶ If you use an underscore or dash, enter an alphanumeric character before and after the underscore or dash. Example: Sample1-T5B1_022515.
  - ▶ Cannot be called "all", "none", "unknown", "undetermined", "stats", or "reports".
- ▶ **[Optional]** Sample_Name—Must match the Sample ID or be blank.
- ▶ **[Optional]** Sample_Plate—Name of the sample plate.
- ▶ **[Optional]** Sample_Well—Name of the sample well.
- ▶ 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—Adapter sequence for Index 1.
  - ▶ Index2—Adapter sequence for Index 2.
  - ▶ **[Optional]** I7_Index_ID
  - ▶ **[Optional]** I5_Index_ID
- ▶ **[Optional]** Description—Enter a sample description.

Install TruSight Oncology 500 Local App

Installation of the TruSight Oncology 500 Local App 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, refer see https://docs.docker.com/engine/installation/linux/linux-postinstall/. To load the docker image, you might require sudo/escalated privileges.

Singularity Considerations

Your local admin with root privileges sets up the Singular server. Singularity allows non-privileged users to run singularity containers in regular user role. After configuration, a regular user can check Singularity configuration and image info. For more information on Singularity editions, see https://www.sylabs.io/singularity/.

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 your customer-specific link.

```
curl -o ${DOWNLOAD_PATH}/TruSight_Oncology_500_${VERSION}.tgz
  "https://${APPLICATION_DOWNLOAD_URL}"
```

3. Extract the application bundle.
   After extraction, the bundle contains the following files.
   - ▶ md5sum.txt
   - ▶ TruSight_Oncology_500.sh
   - ▶ TruSight_Oncology_500_Docker.tgz
   - ▶ TruSight_Oncology_500_Resources.tgz
   - ▶ TruSight_Oncology_500.img

4. Extract the resources file with the following command.

```
cd ${APP_PATH}/${VERSION}
tar -xzvf TruSight_Oncology_500_Resources.tgz [-C ${RESOURCE_FOLDER}]
```

5. Verify the files by calculating the md5 hash value with the following command.

```
md5sum TruSight_Oncology_500_Docker.tgz
md5sum TruSight_Oncology_500.img
md5sum TruSight_Oncology_500_Resources.tgz
md5sum TruSight_Oncology_500.sh
```

6. Check that the calculated md5sum matches the value within the md5sum.txt file. If the checksum is mismatched, contact IlluminaTechnical Support.

```
{CheckSumValue} ./TruSight_Oncology_500_Docker.tgz
{CheckSumValue} ./TruSight_Oncology_500_Resources.tgz
{CheckSumValue} ./TruSight_Oncology_500.sh
{CheckSumValue} ./TruSight_Oncology_500.img
```

7. **[Docker]** Load the Docker image using the following command.

```
gunzip ./TruSight_Oncology_500_Docker.tgz
cat ./TruSight_Oncology_500_Docker.tar | sudo docker load
```

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

9. **[Singularity]** Make sure that the Singularity image file is contained in the applications bundle.

```
which singularity
singularity inspect TruSight_Oncology_500.img
```

10. Display the help for the Bash script with the following command.

```
./TruSight_Oncology_500.sh --help
```

Requirements to Run TruSight Oncology 500 Local App

Start the TruSight Oncology 500 Local App 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 TruSight Oncology 500 Local App 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.

▶ Full path to the analysis folder <ANALYSIS_OUTPUT_PATH> location. Make sure that the folder is empty.

▶ Select one of the following, *not both*.
  ▶ Full path to the sequencing system run folder location.
  ▶ Full path to an existing FASTQ folder location.

Additional requirements for running the software depend on whether you start from BCL files in the run folder or FASTQ files.

Starting From BCL Files

The following inputs are required for running the TruSight Oncology 500 Local App 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 TruSight Oncology 500 Local App 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.

FASTQ File Requirements

When starting from FASTQ files, follow the recommendations for FASTQ file organization, sample sheet considerations, and bcl2fastq 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#_Read#_Lane#.fastq.gz.

The TruSight Oncology 500 Local App 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 15.

Sample Sheet Considerations for FASTQ

FASTQ file names require a unique Sample ID. When running bcl2fastq to generate FASTQ files, use the template SampleSheet.csv so that FASTQ file names contain the sample ID (from the Sample_ID column). Only use FASTQ files from samples that contain UMIs.

bcl2fastq Compatibility and Arguments

Generate FASTQ files with bcl2fastq version 2.20 or later using only bcl2fastq arguments and commands.

Use the following arguments to generate FASTQs using bcl2fastq version 2.20.0.422.

```
#${OUTPUT_DIRECTORY_PATH}: Absolute path for output generated fastqs
#${RUN_FOLDER_PATH}: Absolute path to run folder containing BCLs
#${SAMPLE_SHEET_PATH}: Absolute path to sample sheet (Typically
   contained within run folder)
${BCL2FASTQ_PATH) --output-dir ${OUTPUT_DIRECTORY_PATH} --runfolderdir
${RUN_FOLDER_PATH} --sample-sheet ${SAMPLE_SHEET_PATH} --barcode-
   mismatches 1 --adapter-stringency 0.9 --mask-short-adapter-reads 35
   --minimum-trimmed-read-length 35
```

Run TruSight Oncology 500 Local App 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 TruSight Oncology 500 Local App container to complete before launching a new analysis. Run the following command to check running containers before starting TruSight Oncology 500 Local App.

```
sudo docker ps
```

2   Select from one of the following three options.
   ▶   Start from BCL files in the run folder.

```
${APP_PATH/{Version}/TruSight_Oncology_500.sh --user=$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.

```
${APP_PATH}/{Version}/TruSight_Oncology_500.sh --user=$UID --remove \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder \
--analysisFolder /full/path/to/output/analysisFolder \
--sampleSheet /full/path/to/SampleSheet.csv
```

   ▶   Start from FASTQ files.

```
${APP_PATH}/{Version}/TruSight_Oncology_500.sh --user=$UID --remove \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--analysisFolder /full/path/to/output/analysisFolder \
--fastqFolder /full/path/to/fastqFolder
```

Run TruSight Oncology 500 Local App 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 three options.
   ▶   Start from BCL files in the run folder.

```
./TruSight_Oncology_500.sh --engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder \
--analysisFolder /full/path/to/output/analysisFolder
```

   ▶   Start from BCL files using a separate sample sheet.

```
./TruSight_Oncology_500.sh --engine singularity \
--resourcesFolder ${RESOURCE_FOLDER}/resources \
--runFolder /full/path/to/runFolder \
```

```
    --analysisFolder /full/path/to/output/analysisFolder \
    --sampleSheet /full/path/to/SampleSheet.csv
```
▶ Start from FASTQ files.
```
    ./TruSight_Oncology_500.sh --engine singularity \
    --resourcesFolder ${RESOURCE_FOLDER}/resources \
    --analysisFolder /full/path/to/output/analysisFolder \
    --fastqFolder /full/path/to/fastqFolder
```

Shell Script Command Line Arguments

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

Analysis Methods

The TruSight Oncology 500 Local App workflow performs the following analysis steps and then writes analysis output files to the folder specified. For information on software components and file outputs for each analysis step, see *Software and Output Information for Analysis Steps* on page 14.

- ▶ Run QC
- ▶ FASTQ Generation
- ▶ DNA Alignment
- ▶ Read Collapsing
- ▶ Read Stitching
- ▶ Indel Realignment
- ▶ Small Variant Calling
- ▶ Small Variant Filtering
- ▶ Variant Annotation
- ▶ Complex Variant Calling
- ▶ Tumor Mutational Burden
- ▶ Microsatellite Status
- ▶ Contamination Detection
- ▶ Sample Quality Control
- ▶ Resource Validation

Run QC

The software performs the RunQC analysis step when you run the TruSight Oncology 500 Local App from BCL files, but not when run from FASTQ.

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 a wrong base call.

FASTQ File Generation

The FASTQ file generation step applies only when you run the TruSight Oncology 500 Local App from BCL files.

bcl2fastq software demultiplexes BCL files from the sequencing instrument to intermediate analysis files in the FASTQ format.

FASTQ files use a text format to represent sequences. They are the primary input for alignment. The files contain sequencing reads for each sample and their quality scores, excluding reads identified as inline controls or clusters not passing filter.

FASTQ file names include the sample ID. The software generates eight FASTQ files per sample. For more information, see *FASTQ File Requirements* on page 7.

Alignment

The alignment step uses the BWA-MEM alignment algorithm with the SAMtools 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). 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 http://bio-bwa.sourceforge.net/. For more info on SAM and BAM files, see https://samtools.github.io/hts-specs/SAMv1.pdf.

Read Collapsing

During the read collapsing step, an algorithm uses sets of reads with similar genomic locations (known as families) and UMI tags to consolidate information into representative sequences. This process allows for the accurate removal of duplicate reads without losing the signal of low frequency sequence variations.

Because TMB estimation requires extremely high specificity in variant calling which is difficult to achieve in FFPE samples, the software uses UMIs to reduce sequencing noise. Reads from complimentary strands are also collapsed through a duplex collapsing step, which greatly reduces FFPE deamination artifacts.

The software collapses BAM files from the alignment step and generates FASTQ files and an associated metrics file. Read collapsing adds the following BAM tags.

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

Read Stitching

A stitched read is a single read that has been bioinformatically joined from a pair of reads. Stitcher software takes paired reads from a BAM, stitches them, and writes out a new BAM file and BAM index file. In addition, the software filters stitched reads, removing reads with > 4 mismatches in a single read.

Indel Realignment

Reads ending near detected indels are realigned to remove alignment artifacts. The realignment identifies likely indels observed across all reads and realigns reads against those targets. This step is necessary because reads with indels at the end are sometimes aligned incorrectly during the first alignment. The realigned BAM and BAI files are saved to the alignment folder.

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 reading stitching analysis step. Psara software then filters the VCF files that Pisces variant calling generates to detect variants that are

Document # 1000000067616 v00
For Research Use Only. Not for use in diagnostic procedures.

11

at intron/extron boundaries. The software removes variants that do not overlap regions of interest.

Small Variant Filtering

The software performs post-processing on the small variant calling VCF files, 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 results in gVCF files and associated error rate files.

Variant Annotation

The Illumina Annotation Engine annotates VCF files. Files are first annotated after the variant filter analysis step, and then again after the complex variant calling step.

Complex Variant Calling

The complex variant calling analysis step uses BWA-MEM and SAMtools to align any previously generated FASTQ files to a complex variant reference sequence data set. The data set represents 29 complex indels of the EGFR gene. The analysis step uses EGFR variant reference sequences for the following.

▶ Creates alternative alignments and performs duplicate-marking on the resulting alignments.

▶ Compares the alternative alignments to variant calls to determine whether an EGFR complex variant is present.

The software sorts and indexes the data set, flagging duplicate reads in BAM files and generating BAM files with alignments and indexes. The software then calls complex variants in the BAM files and generates a variant call file.

Tumor Mutational Burden

The tumor mutational burden analysis step generates TMB metrics from the VCF files generated from the small variant filtering analysis step. The software generates results in a *.tsv file and contains the TMB score calculation and which variants are considered somatic or germline mutations.

In addition to the earlier read collapsing analysis step, the software reduces FFPE false positives in the TMB step. The software uses a filter method based on likelihood ratios that dynamically adjusts for the following conditions.

▶ Observed nucleotide change type.

▶ Strand.

▶ Error rate.

The net effect of the read collapsing and TMB analysis steps reduces false positives in a typical FFPE sample from ˜1500 per Mb to < 5 per Mb.

A combination of database and post-database filtering strategies leverage copy number variant (CNV) and allele frequency information to remove germline variants. For more information, refer to the white paper *Analysis of TMB and MSI Status with TruSight Oncology 500 (document # 1170-2018-009)* on the Illumina website.

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 proportion 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 and contamination scores.

The contamination score is the sum of all the log likelihood scores across all positions. The contamination p value (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.

Sample QC

The Sample Quality Control step generate a single metric file per run per sample from stitched reads, collapsed reads, and indel realigned reads.

Resource Validation

Although not an analysis step, the TruSight Oncology 500 Local App confirms the existence and format of required analysis inputs in the resource folder. The resource folder includes the following files.

▶ Manifest

▶ hg19 reference genome

▶ Complex variant reference sequences.

▶ Baseline definitions for variant calling and filtering.

▶ Databases that analysis steps require.

▶ Sample QC threshold guidelines.

▶ SampleSheet.csv

Document # 1000000067616 v00

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

13

Analysis Output

When the analysis run completes, the TruSight Oncology 500 Local App generates an analysis output folder in the user specified location with the folder name TruSightOncology500_Analysis_{Date}_{Time}.

Within the analysis folder, each analysis step generates a subfolder within the Logs_Intermediates folder. The sample report step writes the final sample report to the Results folder.

To view analysis output, complete the following steps.

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

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

Software and Output Information for Analysis Steps

System Name / Output Folder Names	File Output Type	Content	Software Component
RunQC	RunQCMetrics.json	Run quality results using InterOp files.	Not applicable
FastqGeneration	*.fastq.gz -two per sample	Original FASTQ files from bcl2fastq.	bcl2fastq
DnaAlignment DnaRealignment	*.bam *.bam.bai	DnaAlignment–Aligned BAM and BAI files from BWA for each sample. DnaRealignment–Realigned BAM and BAI files from BWA for each sample.	BWA-MEM SAMtools
CollapsedReads	*.fastq.gz -two per sample *_metrics.json	FASTQ files from ReCo for each sample.	ReCo
StitchedReads	*.stitched.bam *stitched.bam.bai	Stitched BAM and BAI files from Stitcher for each sample.	Stitcher SAMtools
IndelRealignment	*.bam *.bam.bai	BAM and BAI files from Hygea for each sample.	Hygea
VariantCaller	*.filtered.genome.vcf *.genome.vcf	Genome VCF files from Pisces for each sample.	Pisces Psara
SmallVariantFilter	*_genome.vcf *_genome.vcf.errorRates	Genome VCF files from Pepe for each sample.	Pepe
Annotation CvAnnotation	*.json.gz	Annotation and complex variant annotation output from Illumina Annotation Engine for each sample.	Illumina Annotation Engine
CvDnaAlignment MarkDuplicates Hypothesis Caller	*.bam, *.bam.bai - *.markduplicates.bam, *.markduplicates.bam.bai *.vcf	CvDnaAlignment–Aligned BAM and BAI files from BWA for each sample. MarkDuplicates–Aligned BAM and BAI files with flagged duplicates from SuperDuper for each sample. Hypothesis Caller–Complex variant VCF files from CoVaSe for each sample.	SuperDuper BWA, SAMtools CoVaSe

Document # 1000000067616 v00

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

14

System Name / Output Folder Names	File Output Type	Content	Software Component
Tmb	*.tmb.json *.tmb.tsv	Tumor mutational burden files for each sample.	TmbRaider
Msi	*.msi.json	Microsatellite files from Hubble for each sample.	Hubble
Contamination	*.contamination.json	Metric files from Spoiler for each sample.	Spoiler
DnaQcMetrics	*.aligned.metrics.json *.collapsed.metrics.json *.stitched.metrics.json	Metrics files from DnaFlightCheck for each sample.	DnaFlightCheck

Folders

This section diagrams each output folder generated during analysis.

> **NOTE**
>
> The software only produces FastqGeneration and RunQC folders if the app starts from BCL files.

📁 **Logs_Intermediates**
- 📁 **Annotation**
- 📁 **CollapsedReads**
- 📁 **Contamination**
- 📁 **CvAnnotation**
- 📁 **CvDnaAlignment**
- 📁 **DnaAlignment**
- 📁 **DnaQCMetrics**
- 📁 **DnaRealignment**
- 📁 **FastqGeneration**
  - 📁 Sample1
    - 📄 Sample1_S1_L001_R1_001.fastq.gz
    - 📄 Sample1_S1_L001_R2_001.fastq.gz
    - 📄 Sample1_S1_L002_R1_001.fastq.gz
    - 📄 Sample1_S1_L002_R2_001.fastq.gz
    - 📄 Sample1_S1_L003_R1_001.fastq.gz
    - 📄 Sample1_S1_L003_R2_001.fastq.gz
    - 📄 Sample1_S1_L004_R1_001.fastq.gz
    - 📄 Sample1_S1_L004_R2_001.fastq.gz
- 📁 **FastqValidation**
- 📁 **HypothesisCaller**

📁 IndelRealignment

📁 MarkDuplicates

📁 MetricsAggregation

📁 Msi

📁 ResourceVerification

📁 RunQC

📁 SampleSheetValidation

📁 SmallVariantFilter

📁 StitchedReads

📁 StitchedReads

📁 Tmb

📁 VariantCaller

📄 TruSight-Oncology-500-pipeline-<datetime>.log

📁 Results

📄 <sample-id>_BiomarkerReport.txt

📄 MetricsReport.tsv

Metrics Report

The metrics report is a final combined metrics report with key analysis metrics, and metadata in a *.tsv file. Sample metrics within the report indicate guideline lower limits (LSL) and upper limits (USL) for each sample in the run. The metrics report also displays guideline run quality thresholds.

Header

Metric	Description
Report Date	Date the report was generated.
Report Time	Time the report was generated.
Pipeline Version	Pipeline version number.

Run Metrics

Metric	Description	Recommended Threshold
PCT_PF_READS (%)	Percent of reads passing filter.	≥ 80%
PCT_Q30_R1 (%)	Percent of base calls with a quality score of Q30 or higher for Read 1.	≥ 80%
PCT_Q30_R2 (%)	Percent of base calls with a quality score of Q30 or higher for Read 2.	≥ 80%

DNA Library QC Metrics for Contamination

Metric	Description	Recommended Threshold
CONTAMINATION_P_VALUE	Probability that the distribution of SNPS across the chromosomes align with a binomial distribution, which can indicate highly-rearranged genomes.	0.000-0.049
CONTAMINATION_SCORE	Score of contamination.	0-3106

DNA Library QC Metrics for Small Variant Calling and TMB

Metric	Description	Recommended Threshold
MEDIAN_INSERT_SIZE	Median insert size from the sequence reads.	≥ 70 bp
MEDIAN_EXON_COVERAGE	The median fragment coverage of exonic bases in the panel.	≥ 150
PCT_EXON_50X	Percent of exon bases with coverage ≥ 50X.	90-100%

DNA Library QC Metrics for MSI

Metric	Description	Recommended Threshold
Usable MSI Sites	Total number of usable microsatellite sites.	≥ 40

DNA Expanded Metrics

Metric	Description
TOTAL_PF_READS	Total number of reads passing filter.
MEAN_FAMILY_SIZE	Average number of fragments across all UMI families.
MEDIAN_TARGET_COVERAGE	Median fragment coverage of target bases.
PCT_CHIMERIC_READS	Percentage of reads that are chimeric reads.
PCT_EXON_100X	Percent of exon bases with coverage ≥ 100X.
PCT_READ_ENRICHMENT	Percent of reads enriched for the target regions.
PCT_USABLE_UMI_READS	Percent of reads that contain usable UMI information.
MEAN_TARGET_COVERAGE	Mean fragment coverage of target bases.
PCT_ALIGNED_READS	Percent of reads that are mapped.
PCT_CONTAMINATION_EST	Percent of foreign DNA estimated to be in the sample.
PCT_PF_UQ_READS	Percent unique reads passing filter.
PCT_TARGET_0.4X_MEAN	Percent target bases with target coverage greater than 0.4 times the mean.
PCT_TARGET_100X	Percent target bases with greater than 100X coverage.
PCT_TARGET_250X	Percent target bases with greater than 250X coverage.

*DNA Expanded Metrics are for information only. They do not indicate the quality of DNA libraries.

Biomarker Report

The biomarker report is a final report per sample that includes TMB and MSI metrics in a *.txt file.

Header

Header	Description
Sample ID	The sample ID from the sample sheet provided when the run was created.
Report Date	Date the report was generated.
Report Time	Time the report was generated.
Pipeline Version	Pipeline version number.

Sample Status

Metric	Description
Completed all steps	"True" indicates that the sample has completed all analysis steps.
Failed steps	Indicates if a sample failed an analysis step.
Steps not Executed	Indicates any skipped analysis steps.

TMB

Metric	Description
Total TMB	Tumor mutational burden.
Nonsynonymous TMB	TMB score from non-synonymous variants.
Coding Region Size in Megabases	Coding region in megabases.
Number of Passing Eligible Variants	Total number of qualified variants for total TMB calculation.
Number of Passing Eligible Nonsynonymous Variants	Total number of qualified non-synonymous variants that meet criteria for inclusion in non-synonymous TMB score calculation.

MSI

Metric	Description
Usable MSI Sites	Total number of usable microsatellite sites.
Total Microsatellite Sites Unstable	Total number of unstable microsatellite sites.
Percent Unstable Sites	Percent of unstable microsatellite sites.

Uninstall TruSight Oncology 500 Local App

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

▶ TruSight Oncology 500 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 dockers 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.

```
# delete the downloaded tarball
$ rm ${APP_PATH}/${VERSION}/TruSight_Oncology_500_${VERSION}.tgz
# delete any remaining assets that were uncompressed from the above
   tarball
$ rm ${APP_PATH}/${VERSION}/TruSight_Oncology_500_Docker.tgz
$ rm ${APP_PATH}/${VERSION}/TruSight_Oncology_500.img
$ rm ${APP_PATH}/${VERSION}/TruSight_Oncology_500_Resources.tgz
$ rm ${APP_PATH}/${VERSION}/TruSight_Oncology_500.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 out and standard error log files in the folder ./{analysisFolder}/Logs_Intermediates/.
Samples	Open the final sample biomarker report log file ./{analysisFolder}/Results/{Sample_ID}_BiomarkerReport.txt. 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}/.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Website: www.illumina.com
Email: techsupport@illumina.com

Illumina Customer Support Telephone Numbers

Region	Toll Free	Regional
North America	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgium	+32 80077160	+32 34002973
China	400.066.5835	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.

Document # 1000000067616 v00

20

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

Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

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

illumına®