

TruSight Oncology 500 ctDNA Local App

User Guide

Introduction	3
Install TruSight Oncology 500 ctDNA Local App	6
Running TruSight Oncology 500 ctDNA Local App	7
Analysis Methods	9
Analysis Output	18
Uninstall TruSight Oncology 500 ctDNA Local App	27
Technical Assistance	28



This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY, AND WILL VOID ANY WARRANTY APPLICABLE TO THE PRODUCT(S).

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2020 Illumina, Inc. All rights reserved.

All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html.

Introduction

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

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.

The TruSight Oncology 500 ctDNA Local App uses the Illumina DRAGEN Bio-IT Platform, a specialized compute server and software package, to run the secondary analysis.

This document provides an overview of the TruSight Oncology 500 ctDNA 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 software container platform.

The TruSight Oncology 500 ctDNA Local App is the only secondary analysis software optimized for analyzing sequencing outputs generated by the TruSight Oncology 500 ctDNA assay. Modification may lead to inaccurate data and is a violation of the Illumina End-User License Agreement. Please refer to your end-user license agreement for complete terms.

Installation Requirements

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

Hardware

The TruSight Oncology 500 ctDNA Local App only runs on the Illumina DRAGEN Bio-IT Platform (server).

Software

The software installed by default on the DRAGEN Bio-IT server includes the following items:

- ▶ Linux CentOS 7.3 operating system, or later.
- ▶ Container engine Docker CE v18.09.
- ▶ DRAGEN Bio-IT software v3.4 (v3.4 is required).
- ▶ Other Linux tools.

Storage

- ▶ The DRAGEN Bio-IT server provides a 6.4 TB NVMe SSD that is optimized for analysis with the DRAGEN Bio-IT software used by the TruSight Oncology 500 ctDNA pipeline. The 6.4 TB NVMe SSD is located at the /staging directory and is suitable for storing only one or

two runs of the analysis pipeline.

- ▶ A NovaSeq sequencing run that uses an S4 flow cell can produce up to 3 TB of output.
- ▶ The TruSight Oncology 500 ctDNA Local App can produce an additional 3 TB of output.
- ▶ The analysis output is automatically written to the `/staging/TSO500_ctDNA_Analysis_<timestamp>` directory. This folder location ensures that the DRAGEN Bio-IT server-related processes read and write data on an optimized NVMe SSD.
- ▶ Network-attached storage is required for long-term storage of sequencing runs and TruSight Oncology 500 ctDNA Local App output.
- ▶ **Managing data storage is your responsibility.**
 - ▶ Illumina recommends developing a strategy to copy data from the DRAGEN Bio-IT server to network-attached storage.
 - ▶ Delete output data on the DRAGEN Bio-IT server as soon as possible.

Input Requirements

Compatibility with Illumina Systems

The TruSight Oncology 500 ctDNA Local App requires sequencing outputs from the NovaSeq 6000.

For optimal coverage, follow the sequencing settings as described in the TruSight Oncology 500 ctDNA Reference Guide (#1000000092559).

Sample Sheet Requirements

The TruSight Oncology 500 ctDNA Local App requires a sample sheet for each analysis. The sample sheet is a comma-separated values file (*.csv) that contains the information to set up and analyze a sequencing experiment. The file includes a list of samples, index sequences, and the analysis workflow.

Sample sheet examples are provided in the resource bundle of the TruSight Oncology 500 ctDNA software. The resource bundle is located in the `/staging/illumina/TSO500_ctDNA/resources/` directory. Use the sample sheet provided in the resource bundle as a template.

Create a Sample Sheet

Sample sheet examples are provided in the resource bundle of the TruSight Oncology 500 ctDNA software. The resource bundle is located in the `/staging/illumina/TSO500_ctDNA/resources/` directory. Use the sample sheet provided in the resource bundle as a template. Follow the steps below to create a TruSight Oncology 500 ctDNA sample sheet.



CAUTION

The analysis will fail if the sample sheet criteria are not met.

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

- 2 In the Settings section, enter the following parameters:
 - ▶ AdapterRead1–AGATCGGAAGAGCACACGTCTGAACTCCAGTCA.
 - ▶ AdapterRead2–AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT.
 - ▶ AdapterBehavior–trim.
 - ▶ MinimumTrimmedReadLength–35.
 - ▶ MaskShortReads–35.
 - ▶ OverrideCycles–U7N1Y143;I8;I8;U7N1Y143.
- 3 In the Settings section, enter parameters for AdapterRead1, AdapterRead2, AdapterBehavior, MinimumTrimmedReadLength, MaskShortReads, and OverrideCycles.
- 4 In the Data section, enter the following parameters. Do not leave blank rows between samples.
 - ▶ Sample_ID–The Sample_ID becomes a part of output file names. Sample IDs are not case sensitive. Enter a sample ID with the following characteristics:
 - ▶ 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.
 - ▶ Index_ID
 - ▶ Index
 - ▶ I7_Index_ID
 - ▶ Index2
 - ▶ I5_Index_ID
 - ▶ Sample_Type–Enter **DNA** for each sample.
 - ▶ Lane– The Lane column is required only if you are using the XP flow cell and an index pair is used for more than one sample. Use the lane column to specify sample location per lane. You must specify a lane for **each** sample. If a sample is in multiple lanes, you must use multiple rows for each sample (one row per lane).
 - ▶ Optional parameters– Additional optional parameters, such as "Sample_Name," are not required for analysis but may be added to the data section. Optional parameters must have the following characteristics:
 - ▶ ≤ 40 characters with no spaces.
 - ▶ Alphanumeric characters, underscores, and/or dashes only.

Additional Resources

The TruSight Oncology 500 ctDNA Local App support pages on the Illumina [support website](#) provide additional system resources. These resources include software, training, compatible products, and the following documentation. Always check support pages for the latest versions.

Resource	Description
<i>TruSight Oncology 500 ctDNA Local App SampleSheet.csv</i>	Setup template for running the analysis.
<i>Illumina DRAGEN Computer Server Site Prep & Installation Guide</i>	Provides activation instructions for DRAGEN licenses.

Install TruSight Oncology 500 ctDNA Local App

Installation of the TruSight Oncology 500 ctDNA Local App requires root privileges.

Docker Considerations

Illumina recommends running Docker as a non-root user by adding the user to a docker group. It is possible to run the TruSight Oncology 500 ctDNA Local App as root, but not recommended. For more information, see the Docker website.

Installation Instructions

- 1 Contact Illumina Customer Support and confirm that the TSO500 DRAGEN license is enabled for your server. Follow the instructions for DRAGEN license installation provided by Illumina Customer Service and the Site Prep Guide.

- 2 Choose one of the following options to install the TSO500 DRAGEN license:

- ▶ If your server is connected to the Internet, run the following command:

```
sudo /opt/edico/bin/dragen_lic -i auto
```

- ▶ If your server is not connected to the internet, Illumina Customer Support provides a tso500.bin file. Store the tso500.bin file in a temporary location on the server (for example, /tmp), and then run the following command:

```
sudo /opt/edico/bin/dragen_lic -i /tmp/tso500.bin -f tso500
```



NOTE

Use the command `dragen_lic` at any point to display the DRAGEN licenses installed.

- 3 Download the installation script from the link provided in the email that you receive from Illumina. The link expires after 72 hours.

- 4 Store the `install_TSO500_ctDNA-<version>.run` script in the `/staging` directory. Illumina recommends storing and executing the self-extracting installation script from the `/staging` directory because the script requires approximately 63 GB of free space.

- 5 Update the permissions on the run script with the following command:

```
sudo chmod +x /staging/install_TSO500_ctDNA-<version>.run
```

- 6 Execute the installation script with the following command:

```
sudo TMPDIR=/staging /staging/install_TSO500_ctDNA-<version>.run
```

Run Self-Test Script

The self-test script performs the following tests, and runs for approximately 15 minutes.

- ▶ Checks if all of the required services are running.
- ▶ Checks if the proper resources are in place.
- ▶ Checks if the proper docker images have been installed.
- ▶ Checks if the TruSight Oncology 500 ctDNA pipeline can run successfully on a test data set.

- 1 Make sure the installation script has finished, and then execute the following self-test script:

```
./usr/local/bin/test_TSO500_ctDNA-<version>.sh
```

If the self-test prints a failure message, contact Illumina Technical Support and provide the output file found in `/staging/tmp/ test_TSO500_ctDNA-<timestamp>.tgz`.

- 2 **[Optional]** Only one user can run the self-test script. To run the self-test script for more than one user, perform one of the following options:

- ▶ Delete the `/staging/tmp` directory.
- ▶ Run the following command to allow all users to write to the `/staging/tmp` directory:

```
chmod 777 /staging/tmp
```

Running TruSight Oncology 500 ctDNA Local App

The TruSight Oncology 500 ctDNA Local App software is started with the Bash script called `TSO500_ctDNA.sh`, which is installed in the `/usr/local/bin` directory. The Bash script is executed on the command line and runs the software with Docker. For arguments, see [Shell Script Command Line Arguments on page 9](#). Analysis output is automatically written to the `/staging/TSO500_ctDNA_Analysis_<timestamp>` directory. This folder location ensures that the DRAGEN Bio-IT server-related processes read and write data on an optimized NVMe SSD.

You can start from BCL files in an Illumina sequencing run folder or from FASTQ files staged in a FASTQ folder. The following requirements apply for both methods:

- ▶ The full path to the sequencing system run folder `<FULL_PATH_TO_RUN_FOLDER>` location is selected using the `--runFolder` command.
- ▶ The sample sheet is in the run folder and is named `SampleSheet.csv`. If the sample sheet is not in the run folder or not named `SampleSheet.csv`, specify the full path to the sample sheet using the `--sampleSheet` override command.



CAUTION

Do not move files when the analysis is running. Moving files during the analysis may cause the analysis to fail or provide incorrect results.

**CAUTION**

Pressing Ctrl+C during a DRAGEN step will stop the currently running analysis. Additionally, Ctrl+C during a running analysis may cause an FPGA error. To recover from an FPGA error, shut down and restart the server.

Starting From BCL Files

The run folder contains data from the sequencing run. To run the TruSight Oncology 500 ctDNA Local App using BCL files, make sure that the run 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.
CopyComplete.txt file	Indicates that the sequencing run is finished, and data has completed copying.

- 1 Run the TruSight Oncology 500 ctDNA Local App using the following argument:

```
TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER>
```

Starting From FASTQ Files

The following inputs are required for running the TruSight Oncology 500 ctDNA Local App App using FASTQ files.

- ▶ Full path to an existing FASTQ folder <FULL_PATH_TO_FASTQ_FOLDER> location.
- ▶ FASTQ files conform to the folder structure in *FASTQ File Organization on page 9*.
- ▶ Full path to an existing run folder.

- 1 Run the TruSight Oncology 500 ctDNA Local App using the following arguments:

```
TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER> --fastqFolder <FULL_PATH_TO_FASTQ_FOLDER>
```

FASTQ File Requirements

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

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 TruSight Oncology 500 ctDNA Local App requires separate FASTQ files per sample. Do not merge FASTQ files.

The NovaSeq 6000 Sequencing System 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 18](#).

Sample Sheet Considerations for FASTQ

FASTQ file names require a unique Sample ID. FASTQ files must contain UMIs.

Shell Script Command Line Arguments

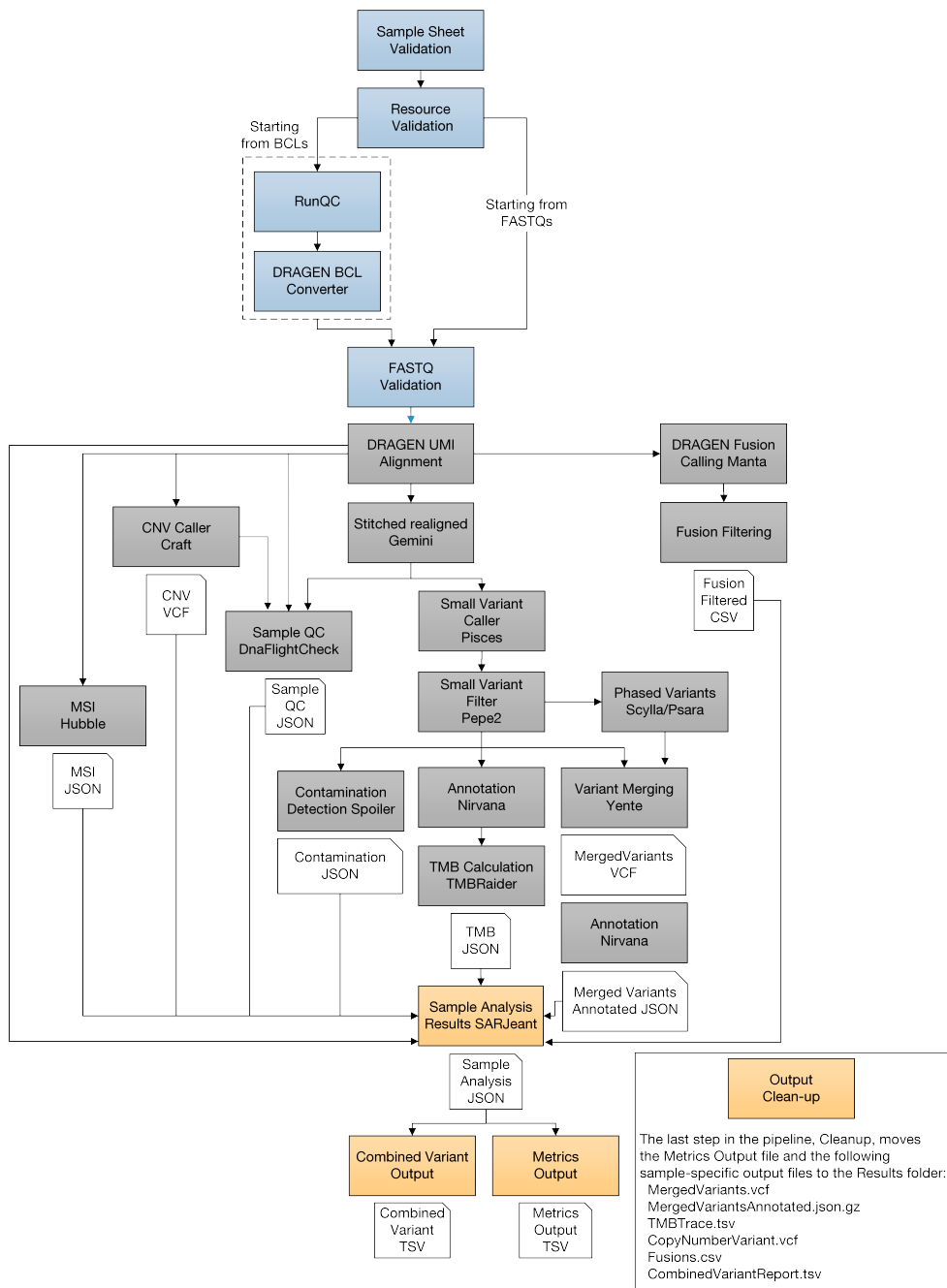
Argument	Description
--runFolder	[Required] Full path to the sequencing system run folder that contains SampleSheet.csv, RunInfo.xml, RunParameters.xml, and the Data directory.
--analysisFolder	[Optional] Full path to the alternative analysis folder. This folder must be on an NVMe SSD partition to achieve high performance.
--fastqFolder	[Optional] Full path to an existing FASTQ folder. The --runFolder command is required with this argument.
--sampleSheet	[Optional] Full path to the SampleSheet.csv file. If your sample sheet is not named SampleSheet.csv, this command is <i>required</i> .
--sampleIDs	[Optional] Provides the comma delimited Sample IDs that are processed by the run. For example, "Sample_1,Sample_2".
--version	[Optional] Displays the version of the software, and then exits.

Analysis Methods

The TruSight Oncology 500 ctDNA analysis module evaluates ctDNA libraries targeting variants of cancer-related genes in the TruSight Oncology 500 ctDNA panel.

The TruSight Oncology 500 ctDNA Local App workflow performs the analysis steps in the following sections, and then writes analysis output files to the folder specified. For information on software components and file outputs for each analysis step, see *This section describes the summary output files generated during analysis.* on page 20.

Figure 1 TruSight Oncology 500 ctDNA Workflow



Inputs Validation

Validates the existence and format of required analysis inputs.

FASTQ File 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. The files contain sequencing reads for each sample and the associated quality scores. Controls used for the run, and clusters that did not pass filter, are excluded from the file. Each FASTQ file contains reads for only one sample and includes the name of that sample in the file name.

Unique molecular identifier (UMI) sequences for each read are retrieved during the FASTQ generation step, and the combination for read1 and read2 is added to the end of the sequence identifier.

The DRAGEN Bio-IT BCL conversion software uses the SampleSheet.csv generated by the analysis module for demultiplexing.

For more information, see [FASTQ File Requirements on page 8](#).

FASTQ Validation

Although not an analysis step, the TruSight Oncology 500 ctDNA Local App confirms that the FASTQ files are named appropriately and uniquely, that there are files for each sample listed in the sample sheet, and that the UMI sequences are valid.

Resource Verification

Although not an analysis step, the software confirms the existence and format of the required analysis inputs in the resource folder. The resource folder includes the following files.

- ▶ Manifest
- ▶ hg19_decoy reference genome
- ▶ Complex variant reference sequences
- ▶ Baseline definitions for variant calling and filtering
- ▶ Databases that the analysis steps require
- ▶ SampleSheet.csv template for reference purposes
- ▶ Test data for running a sanity check test

DNA Alignment and Read Collapsing

The alignment step uses DRAGEN Aligner with UMI collapsing to align DNA sequences in FASTQ files to the hg19_decoy genome. This step combines sets of reads (ie, families) that are grouped together based on genomic locations and UMI tags into representative sequences. This process accurately removes duplicate reads and sequencing errors without losing the signal of very low frequency (< 1%) sequence variations.

This alignment step generates BAM files (*.bam) and BAM index files (*.bam.bai) that are saved to the alignment folder. A BAM file is the compressed binary version of a SAM file that is used to represent aligned sequences.

Read collapsing adds the following BAM tags:

- ▶ RX/XU—UMI combination. RX is duplicated from XU to satisfy the BAM/SAM format.
- ▶ XV—Number of reads in the family on one strand.
- ▶ XW—Number of reads in the duplex-family or 0 if not a duplex family.

Indel Realignment and Read Stitching

To improve small variant calling results, the Gemini software component performs local indel realignment, paired read stitching, and read filtering. 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 and increase indel calling sensitivity. Gemini inputs a single BAM file and the reference genome FASTA used to align it, and then 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.

The following BAM tags are added in this step:

- ▶ XD—For successfully stitched reads only. The directional support string indicates forward, reverse, and stitched positions.
- ▶ XR—For successfully stitched reads only. The read pair orientation is either forward-reverse (FR) or reverse-forward (RF).

Small Variant Calling

Pisces software performs somatic variant calling to identify variants candidates in DNA samples. Pisces detects small variant candidates in the BAM files that are generated from the indel realignment and read stitching analysis step.

For each variant candidate, Pisces adds a US field under the Format column in the genome.vcf that contains values for the mutant support of the following read type counts. The first six values represent mutant support. The next six values represent total support.

- ▶ Duplex-stitched
- ▶ Duplex-nonstitched
- ▶ Simplex-forward-stitched
- ▶ Simplex-forward-nonstitched
- ▶ Simplex-reverse-stitched
- ▶ Simplex-reverse-nonstitched

The small variant calling genome.vcf in this step only collects candidates, and then outputs corresponding read support information. The final variant call is determined in *Small Variant Filtering* on page 14.

Small Variant Filtering

The software component, Pepe, performs post-processing on the small variant calling genome VCFs to polish backgrounds and adjust quality scores. The software filters out variants when their error rates do not meet quality thresholds. This analysis step produces genome VCF files and associated error rate files. The minimum read depth for reference calls is 1000, and the limit of detection for VAF is 0.5% at the minimum read depth.

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

- ▶ Background noise
- ▶ Trinucleotide change
- ▶ Read support type

Table 1 Filter Information for Small Variant Filtering

ALT	FILTER	Note
.	PASS	WT
., A, C, G, etc ¹	LowDP ²	Reference or filtered variant candidate with depth <1000.
A, C, G, etc ¹	PASS	PASS variants.
A, C, G, etc ¹	LowSupport, LowVarSupport	Filtered variant candidate with low quality score.
A, C, G, etc ¹	Blacklist	Position with high background noise. Not available for variant detection.

¹ Etc refers to other variants types not mentioned in the table.

² For reference positions, a coverage below 1000X directly translates into LowDP, but if a position has a PASS filter for variant calls, LowDP is not applied. (This depends on the LQ/AQ thresholds for COSMIC or non-COSMIC variants, and allele frequencies.)

The net effect of the read collapsing and TMB analysis steps (add xref) reduces false positives in a typical cell-free DNA sample from ~1500 per Mb to < 5 per Mb.

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 tri-nucleotide change is estimated from different support categories in each sample by using all the positions with an allele frequency less than 1%. For each variant candidate, a likelihood ratio score (LQ) is computed by the corresponding error rate of the observed total and mutant read. A bias score (BFQ) is computed on each variant candidate to evaluate the imbalance of mutant vs total read support between different support groups.

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. For non-COSMIC variant, threshold for BFQ is < 20. The net effect of the read collapsing and variant filtering reduces false positives in a typical cell-free 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 CNV genes within the assay. The CRAFT software component counts the 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 caused by the following potential variables: sequencing depth, target size, PCR duplicates, probe efficiency, GC bias, and DNA type. A collection of normal cell-free DNA and contrived samples is used to correct for some of these variables. The inputs are collapsed reads in BAM format and the outputs are VCF files. Amplification and deletion are annotated as DUP and DEL 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.

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 VCF is generated.

Variant Merging

The software merges the phased variants with the other small variants generated from the small variant filtering step. In this process, exact duplicates that match chromosome, position, Ref, and Alt are removed.

Annotation

The Illumina Annotation Engine, Nirvana, performs annotation of small variants. The inputs are gVCF files and the outputs are annotated JSON files. Files are annotated after the small variant filter analysis step, and then again after the variant merging step.

Tumor Mutational Burden

The tumor mutational burden (TMB) analysis step generates TMB metrics from the annotated small variant JSON file and the VCF file that is generated from the small variant filtering analysis step. To remove germline variants from the TMB calculation, the software uses a

combination of public database filtering and a post-database filtering strategy that uses allele frequency information and variants in close proximity. Database filtering uses the GnomAD exome, genome, and 1000 genomes database.

The TMB is calculated using the following equation:

$$TMB = \text{Eligible Variants} \div \text{Effective Panel Size}$$

The eligible variants and effective panel size of the TMB calculation are detailed in [Table 2](#).

Table 2 TMB Calculation Details

Calculation Value	Description
Eligible variants (numerator)	<ul style="list-style-type: none"> • Variants in the coding region (RefSeq Cds). • Variant frequency $\geq 0.2\%$ and below 40%. • Coverage $\geq 1000X$. • SNVs and Indels (MNVs excluded). • Nonsynonymous and synonymous variants. • Variants with cosmic count > 50 are excluded. • Mutations in TET2, TP53, DNMT3A, and CBL are excluded.
Effective panel size (denominator)	<ul style="list-style-type: none"> • Total coding region with coverage $\geq 1000X$.

Outputs are captured in a *_TMB_Trace.tsv file that contains the 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 (MSI) step assesses microsatellite sites for evidence of microsatellite instability, relative to a set of baseline normal samples that are based on Jensen-Shannon (JS) distance (an information entropy metric). For MSI sites with 20 or more spanning reads, JS distance is calculated using a test sample vs baseline normal samples, and then any two baseline normal samples. If the JS distance is significantly higher in the test sample vs baseline normal, the site is considered unstable. The percentage of unstable MSI sites to total assessed MSI sites is reported as a sample-level microsatellite score. The score can be interpreted to indicate the sample as MSI-high or microsatellite stable (MSS). The input is the BAM file from the DNA alignment and read collapsing step, and the output is an MSI metric file.

Contamination Detection

The contamination analysis step detects foreign human DNA contamination in the VCF files that are generated during the small variant filtering step. The software determines whether a sample has foreign DNA from the combination of the contamination p-value and the contamination scores.

The contamination score is the sum of all the log likelihood scores across all positions. The contamination p-value measures the uniformity of low frequency SNPs across the genome by identifying samples with unexpectedly low allele frequency that are a result of copy number events.

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

- ▶ Estimates the error rate per sample.
- ▶ Counts mutation support.
- ▶ Counts total depth.

Results

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

The CombinedVariantOutput.tsv file is created for each sample. It contains analysis details, sequencing run details, TMB, MSI, small variants, gene copy number variation, and fusions.

Fusion Calling

DRAGEN Manta performs the fusion calling. The inputs are BAM files and the outputs are VCF files.

Candidate fusion calls are generated using Manta, a structural variant detection method developed by Illumina. Manta first scans the genome to discover evidence of possible structural variants (SV) and large indels based on split- and spanning-reads. The evidence is enumerated into a graph with edges that connect all regions of the genome that have a possible breakend association. Manta then analyzes individual graph edges to discover and score SVs associated with the edges. The substeps of this process include the following items:

- ▶ Inference of SV candidates associated with the edge.
- ▶ Attempted assembly of the SVs breakends.
- ▶ Scoring, genotyping, and filtering of the SV.
- ▶ Output to VCF.

Fusion Filtering

In the TruSight Oncology 500 ctDNA analysis module, Manta calls hundreds to thousands of fusion candidates in a single sample. Most fusion candidates (~99%) are false positives.

The fusion filtering tool, DNA Fusion Filter (DNAFF), distinguishes true fusion calls from the false positives. DNAFF filters perform the following functions:

- ▶ Removes spurious fusions including fusions with only one supporting read pair and fusions that overlap with repeat regions, which are more likely to have sequencing errors.
- ▶ Filters nonconfident supporting reads for all fusion candidates based on the following criteria:
 - ▶ Filter reads with low-sequence identity with the fusion contig.

- ▶ Filter erroneous reads, which are reads that do not support the fusion. For example, reads that have suspicious supplementary alignment.
- ▶ Deduplicate reads based on UMI information.
- ▶ Applies the following rules for the final fusion output:
 - ▶ At least one fusion breakpoint must fall within the 23 target genes.
 - ▶ If the fusion gene pair has been reported in COSMIC, it must have ≥ 2 supporting reads.
 - ▶ If the fusion gene pair has not been reported in COSMIC, it must have ≥ 3 supporting reads.

Quality Control

The TruSight Oncology 500 ctDNA includes several quality control analyses.

Run QC

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

DNA QC Metrics

This step performs DNA sample metric calculation and reporting. The inputs are the BAM file from the DNA alignment and read collapsing step, the BAM file from the indel realignment and read stitching step, and CRAFT normalized BinCount.tsv files. The output is a JSON file.

Analysis Output

When the analysis run completes, the TruSight Oncology 500 ctDNA Local App generates an analysis output folder in the user specified location with the folder name TSO500_ctDNA_Analysis_YYMMDD-HHMMSS.

Within the analysis folder, each analysis step generates a subfolder within the Logs_Intermediates 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.

Folders

This section diagrams each output folder generated during analysis.


 **Results**—Contains the final result files from the pipeline.


 MetricsOutput.tsv—Contains summary metrics for all samples.


 **Sample1**


 Sample1_CombinedVariantOutput.tsv—Contains a TSV containing all detected


small variants and biomarkers into a single file.


 Sample1_MergedSmallVariantsAnnotated.json.gz—Contains information on all small variants, including phased variants.






















 Sample1_CopyNumberVariants.vcf —Contains a VCF file that contains copy number variants.

 Sample1_MergedSmallVariants.genome.vcf—Contains a VCF that contains information on all small variants, including phased variants.

 Sample1_Fusions.csv—Contains all DNA fusions.

 Sample1_TMB_Trace.tsv—Contains a trace file with information regarding how the TMB value was calculated.

 **Logs_Intermediates**—Contains all intermediate files for each step of the pipeline. Each output folder contains a dsdm.json file that contains all executed steps, output logs, and any intermediate files.

-  **AlignCollapseFusionCaller**
-  **Annotation**
-  **Cleanup**
-  **CnvCaller**
-  **CombinedVariantOutput**
-  **Contamination**
-  **DnaFusionFiltering**
-  **DnaQCMetrics**
-  **FastqGeneration (only when app starts from BCLs)**
-  **FastqValidation**
-  **MergedAnnotation**
-  **MetricsOutput**
-  **Msi**
-  **PhasedVariants**
-  **ResourceVerification**
-  **RunQc (only when app starts from BCLs)**
-  **SampleAnalysisResults**
-  **SamplesheetValidation**
-  **SmallVariantFilter**
-  **StitchedRealigned**
-  **Tmb**

VariantCaller

VariantMatching

All logs in Logs_Intermediates are generated from the running docker container. Inputs to the running docker container (for example, the run folder, sample sheet, and FASTQ folder) are mapped from native locations on the server to the following locations in the container:

Input	Running Docker Container Location
Run folder	/opt/illumina/run-folder
Sample sheet	/opt/illumina/SampleSheet.csv
FASTQ folder	/opt/illumina/fastq-folder
Resources	/opt/illumina/resources
Analysis output folder	/opt/illumina/analysis-folder

The paths in the log messages refer to paths within the running docker container, not paths on the server.

Files

This section describes the summary output files generated during analysis.

Metrics Output

File name—MetricsOutput.tsv

The metrics output file is a final combined metrics report with sample status, key analysis metrics, and metadata in a *.tsv file. Sample metrics within the report indicate guideline-suggested lower specification limits (LSL) and upper specification limits (USL) for each sample in the run.

Run Metrics

Run metrics from the analysis module indicate the quality of the sequencing run.

Review the following metrics to assess run data quality:

Metric	Description	Recommended Threshold
PCT_PF_READS	Percentage of reads on the sequencing flow cell that pass the filter.	≥ 55.0
PCT_Q30_R1	Percentage of bases with a quality score ≥ 30 from Read 1.	≥ 80.0
PCT_Q30_R2	Percentage of bases with a quality score ≥ 30 from Read 2.	≥ 80.0

The values in the Run Metrics section will be listed as NA in the following two cases:

- ▶ If the analysis was started from FASTQ files.
- ▶ If the analysis was started from BCL files and the InterOp files are missing or corrupt.

Sample QC Metrics

Review the following metrics to assess sample data quality:

Metric	Description	Recommended Threshold	Variant Class
CONTAMINATION_SCORE	The contamination score based on VAF distribution of SNPs.	Contamination Score \leq 1672	All
and		OR	
CONTAMINATION_P_VALUE	The contamination p-value used to assess highly rearranged genomes. A p-value less than 0.05 suggests that the sample has large-scale rearrangements that could lead to high-contamination scores without actual sample contamination. The contamination p-value is only needed when the contamination score is above USL.	Contamination Score $>$ 1672 AND Contamination p-value \leq 0.049	
MEDIAN_EXON_COVERAGE	Median exon fragment coverage across all exon bases.	\geq 1300	Small variant, TMB, fusion
PCT_EXON_1000X	Percent exon bases with 1000X fragment coverage.	\geq 80.0	Small variant, TMB
GENE_SCALED_MAD	The median of absolute deviations normalized by gene fold change.	\leq 0.059*	CNV
MEDIAN_BIN_COUNT_CNV_TARGET	The median raw bin count per CNV target.	\geq 6.0	CNV
USABLE_MSI_SITES	The number of MSI sites usable for MSI calling.	\geq 40	MSI

* The recommended threshold of 0.059 for GENE_SCALED_MAD only applies to real cell-free DNA.

Merged Variant Genome VCF

File name: {SAMPLE_ID}_MergedSmallVariants.genome.vcf

The merged variant genome VCF is combined from the small variant genome VCF and the EGFR complex variant VCF.

The small variant genome VCF file name is SmallVariantFilter/{sample_ID}/{sample_ID}_SmallVariants.genome.vcf. The small variant genome VCF generated by Pepe includes the variant call status for all targeted intervals, left-padded by 25 bp.

The EGFR complex variant VCF file name is PhasedVariants /{sample_ID}/{sample_ID}.Complex.vcf. The EGFR complex variant VCF generated by Scylla includes phased EGFR variants. The FILTER column in the genome VCF determines the variant status. See [Table 3](#) for more information.

Table 3 FILTER Information for Genome VCF

ALT	FILTER	Note
.	PASS	WT.
., A, C, G, etc	LowDP	No call (DP < 1000).
A, C, G, etc	PASS	PASS variants.
A, C, G, etc	LowSupport	Filtered variant candidate with the following conditions: <ul style="list-style-type: none"> • Fails AQ, LQ, or BFQ threshold. or <ul style="list-style-type: none"> • The stitched support for indel or variant in the homopolymer context is zero.
A, C, G, etc	LowVarSupport	Filtered variant candidate (mutant support < threshold).
A, C, G, etc	Blacklist	Position with high background noise. Not available for variant detection.

Merged Small Variants Annotated file

File name: {Sample_ID}_MergedSmallVariantsAnnotated.json.gz

The merged small variants annotated file provides variant annotation information for all non-reference positions in the merged genome VCF, including pass and non-pass variants.

The variant consequence definition can be found on the Sequence Ontology website.

TMB JSON

File name: {Sample_ID}.tmb.json

The TMB json file contains the tumor mutational burden metrics for each DNA sample. Illumina recommends using the TmbPerMb, which represents both the synonymous and nonsynonymous mutations as part of the numerator in the TMB calculation.

Metric	Description
TmbPerMb	TMB value that is determined using all eligible variants regardless of consequence.
AdjustedTmbPerMb	TMB value, adjusted for bias, for the total number of positions in the panel that have sufficient coverage for confident variant calling.
NonsynonymousTmbPerMb	TMB value for non-synonymous positions in the panel that have sufficient coverage for confident variant calling.
AdjustedNonsynonymousTmbPerMb	TMB value, adjusted for bias, for non-synonymous positions in the panel that have sufficient coverage for confident variant calling.
TotalRegionSizeMb	Total size of the panel in megabases.
CodingRegionSizeMb	Coding region panel size in megabases.

TMB Trace

For TMB calculation only. Do not use this file for variant interpretation.

File name: {Sample_ID}_TMB_Trace.tsv

The TMB trace file provides comprehensive information on how the TMB value is calculated for the 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.

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. If the region overlaps with multiple genes, gene names are concatenated with a : character.
VariantType	Type of the variants: SNV, insertion, deletion, and MNV.
CosmicIDs	COSMIC IDs. If there are multiple COSMIC IDs, they are concatenated with a ; character.
MaxCosmicCount	Maximum COSMIC study count .
AlleleCountsGnomadExome	Variant allele count in gnomAD exome database.
AlleleCountsGnomadGenome	Variant allele count in gnomAD genome database.
AlleleCounts1000Genomes	Variant allele count in 1000 genomes database.
MaxDatabaseAlleleCounts	Maximum variant allele count over the exome, genome, and genomes databases.
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.
IncludedInTMBNumerator	TRUE if variant is used in the TMB calculation.

Sample Analysis Results Json

File name: {SampleID}_SampleAnalysisResults.json

The Sample Analysis Results Json (SARJ) file is an aggregated results file created for each sample. The SARJ file is used for the generation of downstream outputs. The file contains passing variants and passing variant annotations.

Copy Number Variants VCF

File name: {Sample_ID}_CopyNumberVariants.vcf

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

The value in the QUAL column of the copy number variants VCF is a Phred transformation of the p-value represented by the following equation:

$$Q = -10x \log_{10}(p\text{-value})$$

The p-value is derived from the t-test between the fold change of the gene against the 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. indicates the (FC) is less than a predefined deletion cutoff for that gene. This cutoff can vary from gene to gene.

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

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

$$n = \frac{[(200 \times Y) - 2 \times (100 - X)]}{X}$$

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

Merged Small Variant VCF

File name—{Sample1}_MergedSmallVariants.genome.vcf

This file contains both phased variants and all other small variants. The header sections from both the phased variant (complex) VCF and the small variant VCF are included in this merged VCF. Variants that are found for both phased variants and small variants are only displayed as phased variants.

All pass variant calls are annotated using the Illumina Annotation Engine with the following information (using the RefSeq transcript).

- ▶ HGNC Gene
- ▶ Transcript
- ▶ Exon
- ▶ Consequence
- ▶ HGVS Coding Sequencing Name

- ▶ HGVS Protein Sequence Name
- ▶ COSMIC ID

Fusions CSV

File name: {Sample_ID}_Fusions.csv

This file contains all candidate fusions identified by the analysis pipeline.

The fusion columns are described in the following table. If you use Microsoft Excel to view this file, genes that are convertible to dates (for example, MARCH1) automatically convert to dd-mm format (1-Mar).

Fusion Object Field	Description
Sample	Input sample ID.
Name	Fusion name.
Chr1	The chromosome of the first breakpoint.
Pos1	The position one base upstream of the first breakpoint.
Chr 2	The chromosome of the second breakpoint.
Pos2	The position one base upstream of the second breakpoint.
Direction	The directionality of the fusion pieces.
Alt	The number of reads supporting the fusion call.
Depth	Total number of reads at the fusion breakpoints.
VAF	Variant allele frequency.
Gene1	The gene associated with the start of the fusion (semicolon delimited).
Gene2	The gene associated with the end of the fusion (semicolon delimited).
Contig	The sequence of the fusion.
Filter	Indicates whether the fusion has passed all of the fusion filters.
Is_Cosmic_GenePair	Indicates whether the gene pair has been reported by COSMIC (True/False).

The following table lists the meaning of the values in the direction column. The values are in the format used by Samtools.

Direction	VCF Format	Description
L1R2	t[p[The left of breakend1 is joined with the right of breakend2.
L1rL2	t]p]	The left of breakend1 is joined with the reverse complement of the left of breakend2.
L2R1]p]t	The left of breakend2 is joined with the right of breakend1.
rR2R1]p[t	The reverse complement of the right of breakend2 is joined with the right of breakend1.

Combined Variant Output

File name: {SampleID}_CombinedVariantOutput.tsv

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

- ▶ Small Variants (including EGFR complex variants)
- ▶ Copy number variants
- ▶ TMB
- ▶ MSI
- ▶ Fusions

Analysis Details

The following information is in the combined variant output analysis details section:

- ▶ Sample ID
- ▶ Output date
- ▶ Output time
- ▶ Pipeline Version (Docker image version number)

Sequencing Run Details

If the analysis starts from the run folder, the following information is in the combined variant output:

- ▶ Run name
- ▶ Run date
- ▶ Sample index ID
- ▶ Instrument ID
- ▶ Instrument control software version
- ▶ Instrument type
- ▶ RTA version
- ▶ SBS Reagent Cartridge Lot Number
- ▶ Cluster 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.

Uninstall TruSight Oncology 500 ctDNA Local App

The TruSight Oncology 500 ctDNA Local App installation includes an uninstall script called `uninstall_TSO500_ctDNA-<VERSION>.sh`, which is installed in `/usr/local/bin`.

Executing the uninstall script removes the following assets.

- ▶ All scripts (`TSO500_ctDNA.sh`, `test_TSO500_ctDNA-<VERSION>.sh`, and `uninstall_TSO500_ctDNA-<VERSION>.sh`).
 - ▶ The resources found in `/staging/illumina/TSO500_ctDNA`.
 - ▶ The `tso500-ctdna:<VERSION>` docker image.
- 1 To uninstall the TruSight Oncology 500 ctDNA Local App, run the following command as root.

```
uninstall_TSO500_ctDNA-<version>.sh
```

Do not uninstall Docker or DRAGEN Bio-IT server. Docker or DRAGEN Bio-IT server can be removed by removing the associated RPM packages.

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, China	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	
South Korea	+82 80 234 5300	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan, China	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 from support.illumina.com.



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.
© 2020 Illumina, Inc. All rights reserved.

illumina[®]