illumina®

DRAGEN TruSight Oncology 500 ctDNA Analysis Software v1.2 (Local)

User Guide

ILLUMINA PROPRIETARY
Document # 200017080 v00
April 2022

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

Table of Contents

Overview

The Illumina® DRAGEN™ TruSight™ Oncology 500 ctDNA Analysis Software supports analysis for 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), a Jensen-Shannon distance (sum JSD) score that can be used for evaluating microsatellite instability (MSI) status, small variant, copy number variation (CNV), and DNA fusion.

The secondary analysis software starts from a sequencing run folder containing base call files (BCL) or from FASTQ files staged in a FASTQ folder.

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software, computing requirements, installation, workflow, and analysis outputs. Use of the software requires basic knowledge of the Linux operating system and the Docker software container platform.

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software is the only local secondary analysis software optimized for analyzing sequencing outputs generated by the TruSight Oncology 500 ctDNA Assay. Modification might lead to inaccurate data and is a violation of the Illumina End-User License Agreement. Refer to your end-user license agreement for complete terms.g

Compatibility with Illumina Systems

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software requires sequencing outputs from the NovaSeq 6000 and is only compatible with sequencing read lengths of 2 x 151 bp.

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

Additional Resources

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software support pages on the Illumina support website provide additional resources. These resources include software, training, compatible products, sample sheets, and the following documentation. Always check support pages for the latest versions.

Document	Description
DRAGEN TruSight Oncology 500 ctDNA Analysis Software SampleSheet.csv	Setup template for running the analysis.

Document	Description
DRAGEN Server v3 Site Prep & Installation Guide (document # 1000000097923)	Instructions for preparing your site for installation and installing the Illumina DRAGEN server.
TruSight Oncology 500 ctDNA Reference Guide (document # 1000000092559)	Information on using the TruSight Oncology 500 ctDNA kit.
TruSight Oncology 500 Reference Guide (document # 1000000067621)	Information on using the TruSight Oncology 500 kit.

Installation Requirements

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

Hardware

- The DRAGEN TruSight Oncology 500 ctDNA Analysis Software only runs on the DRAGEN server.
- TSO 500 pipeline requires that mkfifo is enabled on the network-attached storage (NAS).

Software

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

- Linux CentOS 7.3 operating system, or later
- DRAGEN software v3.6 (v3.6.6 is required)

You will need to install Docker 18.09 or later.

Storage

- The Illumina DRAGEN server v3 provides a 6.4 TB NVMe SSD that is optimized for analysis with DRAGEN TruSight Oncology 500 ctDNA Analysis Software. 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 6000 sequencing run that uses an S4 flow cell can produce up to 3 TB of output.
  - The DRAGEN TruSight Oncology 500 ctDNA Analysis Software can produce an additional 3 TB of output.
- The analysis output is automatically written to the `/staging/DRAGEN_TSO500_ctDNA_Analysis_<timestamp>` directory. This folder location makes sure that the DRAGEN server-related processes read and write data on an optimized NVMe SSD.
- Network-attached storage is required for longterm storage of sequencing runs and DRAGEN TruSight Oncology 500 ctDNA Analysis Software output.
- Managing data storage is your responsibility.

- Illumina recommends developing a strategy to copy data from the DRAGEN server to network-attached storage.
- Delete output data on the DRAGEN server as soon as possible.

For optimal performance, run analysis on data stored locally on the DRAGEN server. Analysis of data stored on NAS can take longer and performance can be less reliable.

Install DRAGEN TruSight Oncology 500 ctDNA Analysis Software

Installing the DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software as root, but not recommended. For more information, see the Docker website.

Installation Instructions

1. Contact Illumina Customer Support and confirm that the TSO 500 DRAGEN license is enabled for your server. Follow the instructions for DRAGEN license installation provided by Illumina Customer Service. Instructions are also available in the DRAGEN Server v3 Site Prep & Installation Guide (document # 1000000097923).
2. Install TSO 500 DRAGEN using one of the following methods:
   - 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/license_TSO.bin -f tso500
     ```

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

3. Download and install Docker version 18.09, 19.03, or 20.10.12.
4. Download the installation script from the link provided in the email from Illumina. The link expires 7 days from generation.
5. Store the `DRAGEN_TSO500_ctDNA-<version>.run` script in the `/staging` directory.
   The script requires approximately 63 GB of free space. Illumina recommends storing and executing the self-extracting installation script from the `/staging` directory.
6. Update the permissions on the run script using the following command:

```
sudo chmod +x /staging/DRAGEN_TSO500_ctDNA-<version>.run
```

7. Execute the installation script using the following command:
   ```
   sudo TMPDIR=/staging /staging/DRAGEN_TSO500_ctDNA-<version>.run
   ```

Run Self-Test Script

The self-test script checks the following functions, and runs for approximately 30 minutes.

- If all of the required services are running.
- If the proper resources are in place.
- If the proper Docker images have been installed.
- 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 run the following self-test script:
   ```
   /usr/local/bin/test_DRAGEN_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_DRAGEN_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, use one of the following methods:
   - Delete the `/staging/tmp` directory.
   - Run the following command to allow all users to write to the `/staging/tmp` directory:
     ```
     chmod 777 /staging/tmp
     ```

When running analysis on DRAGEN server via SSH, proceed analysis execution commands with the no hang up command `nohup`. This command prevents analysis from terminating in the event disconnection from the DRAGEN server. All output from the analysis command are redirected to `nohup.out` in the directory.

If using MacOS, disable the ability to set environment variables automatically in Terminal settings, as this can cause the following error:

```
ERROR: locale::facet_S_create_c_locale name not valid
```

Uninstall DRAGEN TruSight Oncology 500 ctDNA Analysis Software

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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 `DRAGEN_TSO500_ctDNA.sh`, `test_DRAGEN_TSO500_ctDNA-<VERSION>.sh`, and uninstall_DRAGEN_TSO500_ctDNA-<VERSION>.sh.
- The resources found in `/staging/illumina/DRAGEN_TSO500_ctDNA`.

- The `tso500-ctdna:<VERSION>` Docker image.

1. To uninstall the DRAGEN TruSight Oncology 500 ctDNA Analysis Software, run the following command as root.

```
uninstall_DRAGEN_TSO500_ctDNA-<version>.sh
```

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

Running DRAGEN TruSight Oncology 500 ctDNA Analysis Software

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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 command-line options, refer to *Shell Script Command-Line Options* on page 10. Analysis output is automatically written to the `/staging/DRAGEN_TSO500_ctDNA_Analysis_<timestamp>` directory. This folder location makes sure that the DRAGEN Bio-IT server-related processes read and write data on an optimized NVMe SSD.

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

You can start from BCL files in the 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. Use the `--runFolder` command to specify the folder.
- 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.

1. Run the DRAGEN TruSight Oncology 500 ctDNA Analysis Software starting from BCL files using the following command:

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

2. Run the DRAGEN TruSight Oncology 500 ctDNA Analysis Software starting from FASTQ files using the following command:

```
TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_
FOLDER> --fastqFolder <FULL_PATH_TO_FASTQ_FOLDER>
```

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

Sample Sheet Requirements

Sample Sheet v1 Requirements

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software requires a sample sheet for each analysis. The file includes a list of samples, index sequences, and the analysis software.

Sample sheet examples are provided in the resource bundle of the DRAGEN TruSight Oncology 500 ctDNA software. The resource bundle is located in the `/staging/illumina/DRAGEN_TSO500_ctDNA/resources/` directory. Use the sample sheet v1 provided in the resource bundle as a template.

The sample sheet is made up of a list of samples and their index sequences. Delete index IDs you do not require. Different types of sequencing runs use different index adapters. Use the index IDs included in the DRAGEN TruSight Oncology 500 ctDNA Analysis Software resource bundle.

Create a Sample Sheet v1

Use the following steps to create a DRAGEN TruSight Oncology 500 ctDNA sample sheet.

The analysis fails 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 Data section, enter the following parameters. Do not leave blank rows between samples. Optional parameters must have the following characteristics:
   - ≤ 40 alphanumeric characters, underscores, and/or dashes.
   - Spaces and special characters are not allowed.

Sample Parameter	Required	Details
Sample_ID	Yes	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 alphanumeric characters, underscores, and/or dashes. Spaces and special characters are not allowed. • If you use an underscore or dash, enter an alphanumeric character before and after the underscore or dash. Example: Sample1-T5B1_022515.
Sample_Type	Yes	Enter `DNA` for each sample.
Sample_Name	No	Name of sample.
Index	Yes	Index 1 sequence.
Index2	Yes	Index 2 sequence.
Index_ID	Yes	Index ID.
I7_Index_ID	No	I7 index ID.
I5_Index_ID	No	I5 index ID.
Lane	No	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).

Sample Sheet v2 Parameters

Sample sheet examples are provided in the resource bundle of the DRAGEN 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 DRAGEN TruSight Oncology 500 ctDNA sample sheet.

The analysis fails if the sample sheet criteria are not met.

1. In the Header section, enter the following parameter:
   - FileFormatVersion—2.

2. The Settings and Data sections include the same parameters as listed in *Create a Sample Sheet v1 on page 6*, but with a `TSO500L` prefix (`TSO500L_Settings` and `TSO500L_Data`).
3. When starting analysis from BCL files, you must include the `BCLConvert_Data` section with the following parameters. The values for each of these parameters should match the parameters in the `TSO500L_Settings` section.
   - AdapterRead1
   - AdapterRead2
   - OverrideCycles
4. When starting analysis from BCL files, you must include the `BCLConvert_Settings` section with the following parameters. The values for each of these parameters should match the parameters in the `TSO500L_Data` section.
   - Sample_ID
   - Index
   - Index2
   - Lane (Optional)

Starting From BCL Files

The run folder contains data from the sequencing run. To run the DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software starting from BCL files using the following command:

```
DRAGEN_TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER>
```

Starting From FASTQ Files

The following inputs are required for running the DRAGEN TruSight Oncology 500 ctDNA Analysis Software using FASTQ files. For more information on FASTQ file generation, refer to *Illumina DRAGEN Bio-IT Platform Product Documentation*.

- 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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software using the following commands:

```
DRAGEN_TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_
FOLDER> --fastqFolder <FULL_PATH_TO_FASTQ_FOLDER>
```

FASTQ File Requirements

When starting from FASTQ files, use the following recommendations for FASTQ file organization and sample sheet considerations. FASTQ files must contain UMIs.

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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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.

Sample Sheet Considerations for FASTQ

FASTQ file names require a unique sample ID.

Shell Script Command-Line Options

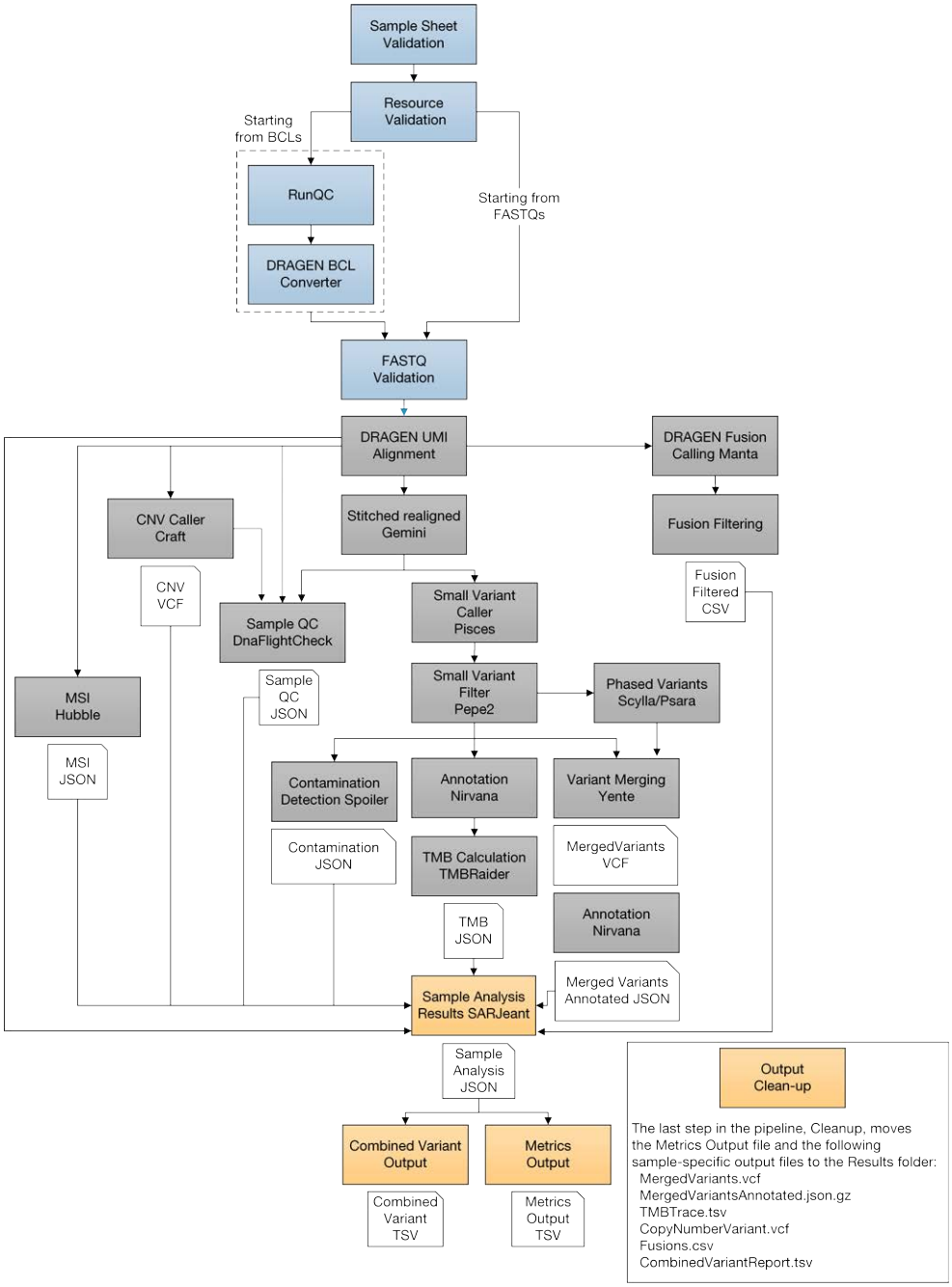
Argument	Required	Description
--runFolder	Yes	Full path to the sequencing system run folder that contains `SampleSheet.csv`, `RunInfo.xml`, `RunParameters.xml`, and the Data directory.
--analysisFolder	No	Full path to the alternative analysis folder. This folder must be on an NVMe SSD partition to achieve high performance.
--fastqFolder	No	Full path to an existing FASTQ folder. The `--runFolder` command is required with this option.
--sampleSheet	Yes	Full path to the sample sheet file. If the sample sheet is named `SampleSheet.csv`, this command is not required.
--sampleIDs	No	The comma-delimited sample IDs that are processed by the run. For example, `Sample_1,Sample_2`.
--version	No	Displays the version of the software, and then exits.

Analysis Methods

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

The DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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, refer to *Files* on page 21.

Figure 1 DRAGEN TruSight Oncology 500 ctDNA Workflow



Inputs Validation

Validates the existence and format of required analysis inputs.

FASTQ Validation

Although not an analysis step, the DRAGEN TruSight Oncology 500 ctDNA Analysis Software 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

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 13.

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 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. The limit of detection for VAF is .05% 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.
A, C, G, etc[1]	VarBias	Filtered variant candidate showed bias clustered at fragment ends.

[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 reduces false positives in a typical cell-free DNA sample from ~1500 per Mb to < 5 per Mb.

For more information on DNA alignment and read collapsing, refer to *Illumina DRAGEN Bio-IT Platform Product Documentation*.

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. The p-value is then converted to a variant quality score (AQ). The sample-specific error rate of each trinucleotide 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. In addition, positional information of mutant and WT allele in fragment will be extracted for each variant candidate. A Kolmogorov-Smirnov test will be applied to compute p-value between mutant and WT position. Variants with p-value < 0.05 and median difference > = 0.5 will be filtered and labeled VarBias. The net effect of the read collapsing and variant filtering significantly reduces false positives. For example, false positives in a typical cell-free DNA sample were reduced to < 5 per Mb from ~1500 per Mb.

In addition to the evaluation of the quality scores, certain regions covered in the product manifest are excluded from analysis due to high background noise. All excluded variants are identified in the VCF using a flag. The block list of excluded sites can be obtained on request from your local Illumina representative.

Some regions are known to be difficult to sequence. One example region is the TERT promoter region. Although sequencing can occur at the TERT promoter region, this location might result in low coverage due to the GC rich content of the sequenced region.

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 software uses Scylla to detect specific, clinically relevant mutations in EGFR exon 19 and mutations in RET 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.

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 2 EGFR Variants

Chromosome	Position	Reference Allele	Alternative Allele
chr7	55242482	CATCTCCGAAAGCCAACAAGGAAAT	C
chr7	55242466	GAATTAAGAGAAGCAACAT	G
chr7	55242465	GGAATTAAGAGAAG	AATTC
chr7	55242465	GGAATTAAGAGAAGCAAC	AAT
chr7	55242469	TTAAGAGAAGCAACATCTC	T
chr7	55242467	AATTAAGAGAAGCAACATC	A
chr7	55242469	TTAAGAGAAG	C
chr7	55242467	AATTAAGAGAAGCAACATC	T

Chromosome	Position	Reference Allele	Alternative Allele
chr7	55242465	GGAATTAAGA	G
chr7	55242467	AATTAAGAGAAGCAACATCTC	TCT
chr7	55242467	AATTAAGAGAAGCAAC	T
chr7	55242464	AGGAATTAAGAGAAGC	A
chr7	55242466	GAATTAAGAGAAGCAA	G
chr7	55242464	AGGAATTAAGAGA	A
chr7	55242469	TTAAGAGAAGCAA	T
chr7	55242465	GGAATTAAGAGAAGCAACATC	AAT
chr7	55242469	TTAAGAGAAGCAACATCT	CAA
chr7	55242463	AAGGAATTAAGAGAAG	A
chr7	55242468	ATTAAGAGAAGCAACATCT	A
chr7	55242462	CAAGGAATTAAGAGAA	C
chr7	55242465	GGAATTAAGAGAAGCAA	AATTC
chr7	55242469	TTAAGAGAAGCAA	C
chr7	55242467	AATTAAGAGAAGCAAC	A
chr7	55242469	TTAAGAGAAGCAACATCTCC	CA
chr7	55242468	ATTAAGAGAAG	GC
chr7	55242465	GGAATTAAGAGAAGCA	G
chr7	55242468	ATTAAGAGAAGCAAC	GCA
chr7	55242465	GGAATTAAGAGAAGCAACA	G
chr7	55249011	AC	CCAGCGTGGAT

Regions of the RET gene in exons 11, 13, and 15 are evaluated and added if detected from Phased Variant Calling.

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:

$$TMB = Eligible\ Variants \div Effective\ Panel\ Size$$

The eligible variants and effective panel size of the TMB calculation are detailed in Table 3.

Table 3 TMB Calculation Details

Calculation Value	Description
Eligible variants (numerator)	• Variants in the coding region (RefSeq Cds). • Variant frequency ≥ 0.2% and below 40%. • Coverage ≥ 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)	• Total coding region with coverage ≥ 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.

Max Somatic VAF

Max somatic VAF is a surrogate variant for tumor fraction in ctDNA. The TMB step will flag the variants by potential somatic status using database and VAF information. For the remaining variants, potential clonal hematopoiesis mutations in TP53, DNMT3A, CBL, TET2 are removed. The 4th variant allele frequency of the remaining variants are reported as the MaxSomaticVAF for each sample.

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). In total, there are 2408 selected MSI sites with 6 or 7 single nucleotide repeat in the panel. For MSI sites with 500 or more spanning duplex collapsed 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 with p-value < = 0.01 and the JS distance difference is > = 0.02, 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 final MSI score aggregates all JS distance across all unstable sites. 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 DRAGEN 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 DRAGEN 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 ≥ 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 DRAGEN TruSight Oncology 500 ctDNA Analysis Software generates an analysis output in a specified location with the folder name `DRAGEN_TSO500_ctDNA_Analysis_<timestamp>`.

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

To view analysis output, do as follows.

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.

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 information on all detected small variants and biomarkers into a single TSV file.
  - Sample1_MergedSmallVariantsAnnotated.json.gz—Contains information on all small variants, including phased variants.
  - Sample1_CopyNumberVariants.vcf —Contains information on copy number variants in a VCF file.
  - Sample1_MergedSmallVariants.genome.vcf—Contains information on all small variants, including phased variants in a VCF file.
  - 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 includes 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

The `MetricsOutput.tsv` 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 are listed as NA in the following situations:

- 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 and CONTAMINATION_P_VALUE	The contamination score based on VAF distribution of SNPs. The contamination p-value used to assess highly rearranged genomes. A p-value ≤ 0.049 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 OR Contamination Score > 1672 AND Contamination p-value ≤ 0.049	All
MEDIAN_EXON_COVERAGE	Median exon fragment coverage across all exon bases.	≥ 1300	Small variant, TMB, fusion, MSI
PCT_EXON_1000X	Percent exon bases with 1000X fragment coverage.	≥ 80.0	Small variant, TMB
GENE_SCALED_MAD	The median of absolute deviations normalized by gene fold change.	≤ 0.059*	CNV
MEDIAN_BIN_COUNT_CNV_TARGET	The median raw bin count per CNV target.	≥ 6.0	CNV

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

Merged Variant Genome VCF

The `{SAMPLE_ID}_MergedSmallVariants.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.VarPosFilter.vcf`. The small variant genome VCF 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 includes phased EGFR variants. The FILTER column in the genome VCF determines the variant status. Refer to Table 4 for more information.

Table 4 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: • Fails AQ, LQ, or BFQ threshold. or • 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	Blocklist	An excluded list of sites (referred to as blacklist in some files). Refer to *Small Variant Filtering* on page 13 for more information.
A, C, G, etc	VarBias	Filtered variant candidate showed bias clustered at fragment ends.

Merged Small Variants Annotated File

The `{Sample_ID}_MergedSmallVariantsAnnotated.json.gz` 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

The `{Sample_ID}.tmb.json` file contains the tumor mutational burden (TMB) 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.

Metric	Description
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 nonsynonymous positions in the panel that have sufficient coverage for confident variant calling.
AdjustedNonsynonymousTmbPerMb	TMB value, adjusted for bias, for nonsynonymous 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.

The `{Sample_ID}_TMB_Trace.tsv` 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.

Column	Description
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

The `{SampleID}_SampleAnalysisResults.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

The `{Sample_ID}_CopyNumberVariants.vcf` contains CNV calls for DNA libraries of the CNV genes targeted by the DRAGEN 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:

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

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. <DEL> 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

The {Sample1}_MergedSmallVariants.genome.vcfcontains 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

The `{Sample_ID}_Fusions.csv` 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.

Fusion Object Field	Description
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

The `{SampleID}_CombinedVariantOutput.tsv` file contains the variants and biomarkers in a single file. The output contains the following variant types and biomarkers:

- 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

- Combined variant output produces small variants with blank fields in the following situations:
  - The variant has been matched to a canonical RefSeq transcript on an overlapping gene not targeted by TruSight Oncology 500 Assay.
  - The variant is located in a region designated iSNP, indel, or Flanking in the `TST500_Manifest.bed` file located in the resources folder.
- **Small Variants**—All variants with the FILTER field marked as PASS in the merged genome VCF and which have a canonical RefSeq transcript (recorded in the MergedSmallVariantsAnnotated.json) 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**—Copy number variants must meet the following conditions:
  - FILTER field marked as PASS.

  - ALT field is <DUP> or <DEL>.
- **Fusion Variants**—Fusion variants must meet the following conditions:
  - Passing Variant Call (KeepFusion field is true).
  - Contains at least one gene on the fusion allow list.
  - Genes separated by a dash (-) indicate that the fusion directionality could be determined. Genes separated by a slash (/) indicate that the fusion directionality could not be determined.
- **Biomarkers TMB/MSI**—Always present when DNA sample is processed.

Block List

The following table lists the genes that have associated block listed sites. For the exact location of the block listed site, contact your local Illumina Technical Support.

Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites
ABL1	5	ETV4	502	LATS2	4	PTPRS	14
AKT2	5	ETV5	11	LoH	85	PTPRT	2
AKT3	20	ETV6	187	LRP1B	3	QKI	2
ALK	90	EWSR1	364	LZTR1	1	RAD21	1
ANKRD11	6	EZH2	2	MAGI2	2	RAD50	5
ANKRD26	9	FANCA	1	MALT1	4	RAD51	18
AR	81	FANCD2	11	MAP2K2	1	RAD51B	8
ARID1A	40	FANCG	10	MAP2K4	5	RAF1	98
ARID1B	87	FANCI	1	MAP3K1	8	RANBP2	12
ARID2	1	FANCL	1	MAP3K14	2	RARA	2
ASXL1	3	FAT1	2	MAP3K4	10	RASA1	1
ASXL2	5	FBXW7	4	MAPK1	6	RB1	5
ATM	2	FGF1	25	MAPK3	6	RBM10	13
ATR	3	FGF10	17	MCL1	1	RECQL4	3
ATRX	17	FGF14	15	MDC1	23	REL	3
AURKA	1	FGF19	102	MDM2	53	RET	3
AXIN2	4	FGF2	26	MDM4	67	RFWD2	22

Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites
AXL	74	FGF23	38	MED12	28	RICTOR	1
BBC3	2	FGF3	60	MGA	6	ROS1	287
BCL10	2	FGF4	25	MLL	9	RPS6KA4	3
BCL2L11	16	FGF5	14	MLLT3	18	RPS6KB1	109
BCOR	2	FGF6	9	MRE11A	5	RUNX1	3
BCORL1	1	FGF7	9	MSH3	10	SDHA	18
BCR	64	FGF8	30	MSH6	2	SDHB	3
BIRC3	1	FGF9	21	MSI	148	SDHD	17
BLM	4	FGFR1	26	MST1	18	SETBP1	7
BMPR1A	4	FGFR2	144	MYB	402	SETD2	26
BRAF	283	FGFR3	1	MYC	78	SF3B1	1
BRCA1	49	FGFR4	36	MYCL1	28	SH2B3	4
BRCA2	21	FLCN	2	MYCN	69	SH2D1A	2
BRD4	16	FLI1	36	MYOD1	3	SLIT2	1
CARD11	4	FLT1	91	NAB2	10	SLX4	2
CASP8	2	FLT4	3	NCOA3	28	SMARCA4	4
CBL	8	FOXA1	48	NCOR1	9	SMC1A	1
CCND1	25	FOXL2	4	NF1	3	SMC3	8
CCND3	49	FOXO1	2	NKX2-1	4	SMO	2
CCNE1	72	FOXP1	3	NOTCH1	4	SOX10	7
CD74	50	FUBP1	1	NOTCH3	7	SOX17	1
CDH1	4	GATA4	6	NOTCH4	9	SOX9	14
CDK12	3	GATA6	12	NPM1	5	SPEN	4
CDK4	46	GEN1	1	NRAS	29	STAG1	5
CDK6	13	GID4	3	NRG1	47	STAG2	2
CDK8	4	GNAQ	4	NTRK1	134	STAT4	1
CDKN2B	2	GNAS	11	NTRK2	145	STAT5A	1
CEBPA	12	GPR124	3	NTRK3	13	STAT5B	4

Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites
CHD2	5	GRM3	1	NUTM1	134	SUFU	5
CHD4	12	H3F3A	1	PAK1	68	SUZ12	9
CHEK1	75	H3F3C	2	PAK3	8	TAF1	9
CHEK2	64	HGF	1	PALB2	1	TBX3	1
chrY	93	HIST1H1C	2	PARK2	23	TCEB1	1
CIC	2	HLA-A	72	PARP1	2	TCF3	2
CREBBP	4	HNF1A	2	PAX3	156	TCF7L2	6
CSNK1A1	4	HNRNPK	9	PAX7	5	TERT	2
CTNNB1	1	HOXB13	1	PAX8	275	TET1	1
CUL3	1	HSP90AA1	4	PBRM1	3	TET2	23
CUX1	9	ICOSLG	6	PDCD1	2	TFE3	299
DAXX	5	IFNGR1	2	PDGFRA	5	TFRC	33
DDR2	1	iIndel	91	PDGFRB	2	TGFBR1	6
DDX41	1	INHBA	4	PDK1	1	TGFBR2	2
DIS3	2	INPP4A	1	PDPK1	6	TMEM127	5
DNAJB1	6	INPP4B	1	PGR	5	TMPRSS2	236
DNMT1	1	IRS1	9	PHF6	2	TOP2A	1
DNMT3A	4	IRS2	19	PHOX2B	15	TP53	22
DOT1L	2	iSNP	4	PIK3C2G	2	TRAF7	4
E2F3	70	JAK2	4	PIK3CA	18	TSC1	4
EGFR	304	JUN	7	PIK3CB	42	TSC2	1
EIF4E	12	KAT6A	5	PIK3R1	6	U2AF1	1
EML4	9	KDM5A	7	PIK3R2	2	VEGFA	7
EP300	1	KDM5C	2	PLCG2	3	WISP3	2
ERBB2	14	KDM6A	2	PLK2	2	WT1	10
ERBB3	62	KDR	1	PMAIP1	7	XIAP	1
ERCC1	53	KIF5B	7	PMS2	1	XPO1	2
ERCC2	57	KIT	5	POLE	3	XRCC2	1

Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites	Gene	Number of Block List Sites
ERCC3	4	KMT2B	51	PPARG	446	YAP1	1
ERCC5	4	KMT2C	118	PRDM1	1	ZBTB7A	11
ERG	2	KMT2D	108	PRKCI	2	ZFHX3	56
ESR1	32	KRAS	44	PRKDC	5	ZNF703	7
ETS1	45	LAMP1	64	PTCH1	13	ZRSR2	2
ETV1	862	LATS1	1	PTEN	41	N/A	N/A

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 software 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.tsv`. 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 Technical Support Telephone Numbers

Region	Toll Free	International
Australia	+61 1800 775 688	
Austria	+43 800 006249	+43 1 9286540
Belgium	+32 800 77 160	+32 3 400 29 73
Canada	+1 800 809 4566	
China		+86 400 066 5835
Denmark	+45 80 82 01 83	+45 89 87 11 56
Finland	+358 800 918 363	+358 9 7479 0110
France	+33 8 05 10 21 93	+33 1 70 77 04 46
Germany	+49 800 101 4940	+49 89 3803 5677
Hong Kong, China	+852 800 960 230	
India	+91 8006500375	
Indonesia		0078036510048
Ireland	+353 1800 936608	+353 1 695 0506
Italy	+39 800 985513	+39 236003759
Japan	+81 0800 111 5011	
Malaysia	+60 1800 80 6789	
Netherlands	+31 800 022 2493	+31 20 713 2960
New Zealand	+64 800 451 650	
Norway	+47 800 16 836	+47 21 93 96 93
Philippines	+63 180016510798	
Singapore	1 800 5792 745	
South Korea	+82 80 234 5300	

Region	Toll Free	International
Spain	+34 800 300 143	+34 911 899 417
Sweden	+46 2 00883979	+46 8 50619671
Switzerland	+41 800 200 442	+41 56 580 00 00
Taiwan, China	+886 8 06651752	
Thailand	+66 1800 011 304	
United Kingdom	+44 800 012 6019	+44 20 7305 7197
United States	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

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.

illumina®