TruSight Tumor 170 v2.0 Local App

User Guide

ILLUMINA PROPRIETARY

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

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

Introduction

The TruSight™ Tumor 170 Local App, a Docker-based software package, analyzes sequencing reads from DNA and RNA libraries prepared using the TruSight Tumor 170 sequencing panel. DNA libraries output variant call files for small variants and copy number variants. RNA libraries output fusions and splice variant call files. Results include user-friendly outputs for high confidence fusions and splice variants.

The application supports starting the secondary analysis workflow from an Illumina® sequencing run folder containing base call files (BCL) or from FASTQ files staged in a FASTQ folder. A TruSight Tumor 170 sample sheet is required for each analysis run.

This document provides an overview of the TruSight Tumor 170 Local App workflow, computing requirements, installation, analysis software, analysis methods, and analysis outputs.

Workflow Requirements

▶ Basic knowledge of the Linux operating system and the Docker software container platform

▶ Compatible with libraries prepared with the Illumina TruSight Tumor 170 Assay

▶ Compatible with NextSeq™ or HiSeq™ 2500 Rapid Run systems

    ▶ NextSeq:

        ▶ If sequencing DNA libraries only, do not include more than 8 samples per run.

        ▶ If sequencing RNA libraries only, do not include more than 16 samples per run.

        ▶ If sequencing both DNA and RNA libraries, the following sample combinations are recommended:

DNA (samples per run)	RNA (samples per run)
8	1–8
7	1–9
6	1–10
5	1–11
4	1–12
3	1–13
2	4–14
1	8–15

    ▶ HiSeq 2500 Rapid Run:

        ▶ Sequence 12 libraries (6 DNA and 6 RNA) per run to achieve maximum coverage for each sample.

        ▶ If sequencing DNA libraries only, do not include more than 6 samples per run.

        ▶ If sequencing RNA libraries only, do not include more than 12 samples per run.

▶ Sequencing settings

    ▶ 2 x 101 (recommended)

    ▶ Paired End (required)

    ▶ Sequencing run should have three libraries containing one of the following index sets to provide sufficient diversity; using fewer than three libraries is not recommended. It is also recommended that UP indexes are used with RNA libraries.

    Index set #1: UP01 (D702, D503), UP02 (D707, D504), UP03 (D717, D509)

    Index set #2: UP04 (D706, D510), UP05 (D712, D513), UP06 (D724, D515)

      Index set #3: UP07 (D705, D501), UP08 (D713, D502), UP09 (D715, D505)

      Index set #4: UP10 (D703, D506), UP11 (D710, D517), UP12 (D701, D518)

- ▶ Additional specifications can be found in the TruSight Tumor 170 data sheet
- ▶ SampleSheet.csv guidance
  - ▶ An example of the TruSight Tumor 170 Local App Sample Sheet (SampleSheet.csv) can be found on the Illumina support site.
  - ▶ Sample sheet must be named SampleSheet.csv
  - ▶ Manifest section must have PoolDNA and PoolRNA defined
  - ▶ The Reads section must be defined
  - ▶ Maximum number of samples per analysis run is 16
  - ▶ Data section, required columns and column headers
    - ▶ Sample ID
    - ▶ Sample_Name
    - ▶ Sample_Well
    - ▶ I7_Index_ID
    - ▶ index
    - ▶ I5_Index_ID
    - ▶ index2
    - ▶ Manifest
    - ▶ Pair_ID
  - ▶ Data section fields, optional unless noted otherwise
    - ▶ Sample IDs under column header Sample_ID (required)
    - ▶ Sample_Name
    - ▶ Sample_Plate
    - ▶ Sample_Well
    - ▶ Index_ID
    - ▶ I7_Index_ID under column header I7_Index_ID (required when input is BCL, and optional when input is FASTQ)
    - ▶ index
    - ▶ I5 _Index_ID under column header I5_Index_ID (required when input is BCL, and optional when input is FASTQ)
    - ▶ index2
    - ▶ Sample Project
    - ▶ Description
    - ▶ PoolDNA or PoolRNA (case-sensitive string) under column header Manifest (required)
    - ▶ Pair ID under column header Pair_ID (required)
      - ▶ Must begin with an alphanumeric character
      - ▶ Can contain up to 60 characters
      - ▶ Can contain alphanumeric characters, underscores, and dashes
      - ▶ Enter an alphanumeric character after and before underscores or dashes

- ▸ A Pair ID can be assigned to only one DNA sample and one RNA sample
- ▸ Assign a Pair ID to each sample
- ▸ Index guidance
  - ▸ Each sample index pair must be unique
  - ▸ Each sample index pair must be compatible with the TruSight Tumor 170 Assay
- ▸ Sample IDs on sample sheets configured using the Sample Sheet Template must meet the following criteria:
  - ▸ Fewer than 60 characters
  - ▸ Alphanumeric, underscore, and dash characters only
  - ▸ No duplicate Sample ID
  - ▸ Underscores and dashes must be preceded and followed by an alphanumeric character (no strings of multiple underscores or dashes)
  - ▸ Sample names cannot begin or end with an underscore or dash
- ▸ For DNA analysis
  - ▸ Minimum of 60 M reads (30 M sequencing clusters) is recommended
  - ▸ Paired-end samples with the same read lengths are required
- ▸ For RNA analysis
  - ▸ Minimum of 16 M reads (8 M sequencing clusters) is recommended
  - ▸ Paired-end samples with the same read lengths are required

Additional Resources

Visit the TruSight Tumor 170 support page on the Illumina website for documentation, software downloads, training resources, and information about compatible Illumina products.

Workflow Diagrams

Figure 1 TruSight Tumor 170 Software Workflow Starting at BCLs

For Research Use Only. Not for use in diagnostic procedures.
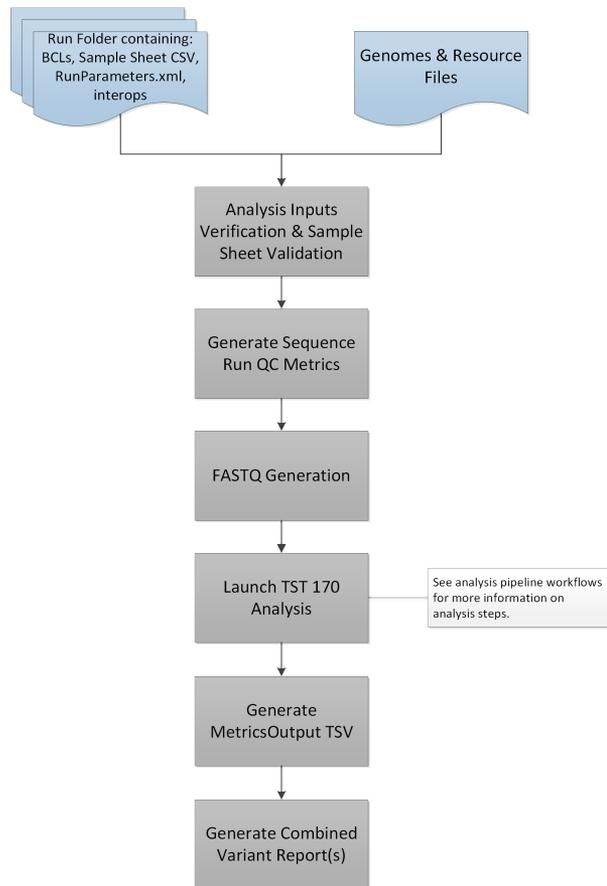
Figure 2 TruSight Tumor 170 Software Workflow Starting at FASTQ

For Research Use Only. Not for use in diagnostic procedures.
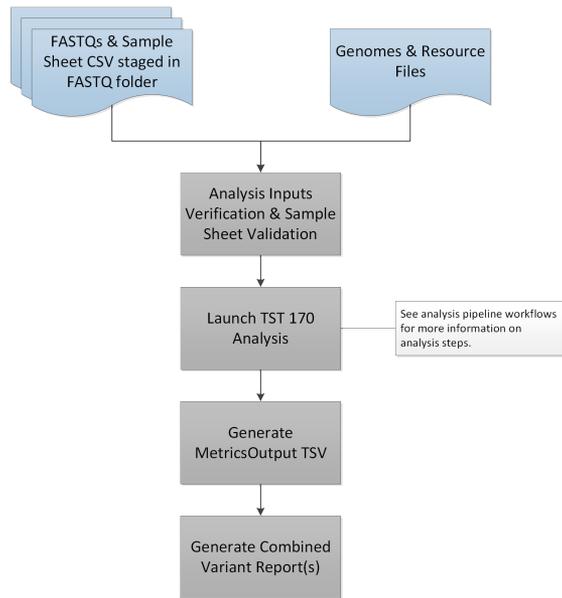
Figure 3 TruSight Tumor 170 DNA Workflow
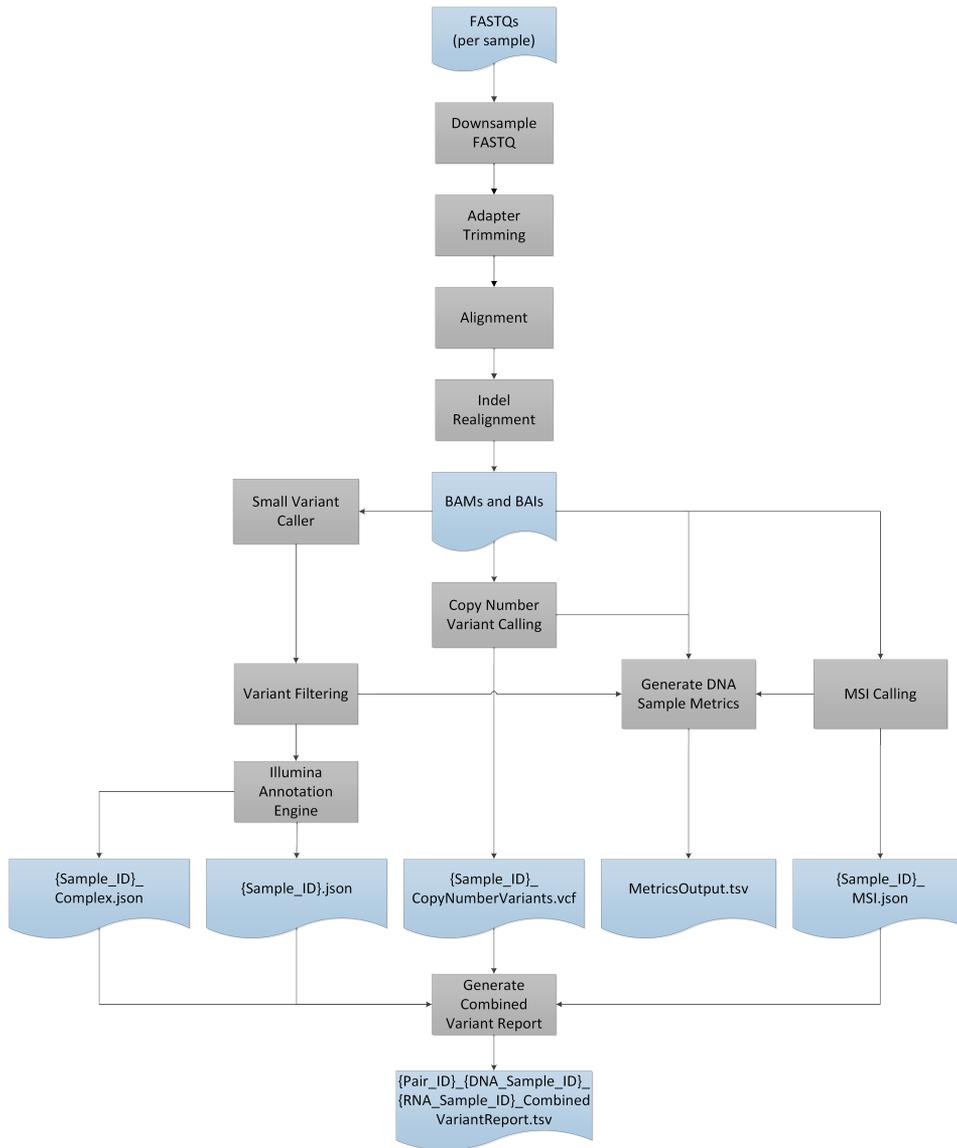
Figure 4 TruSight Tumor 170 RNA Workflow
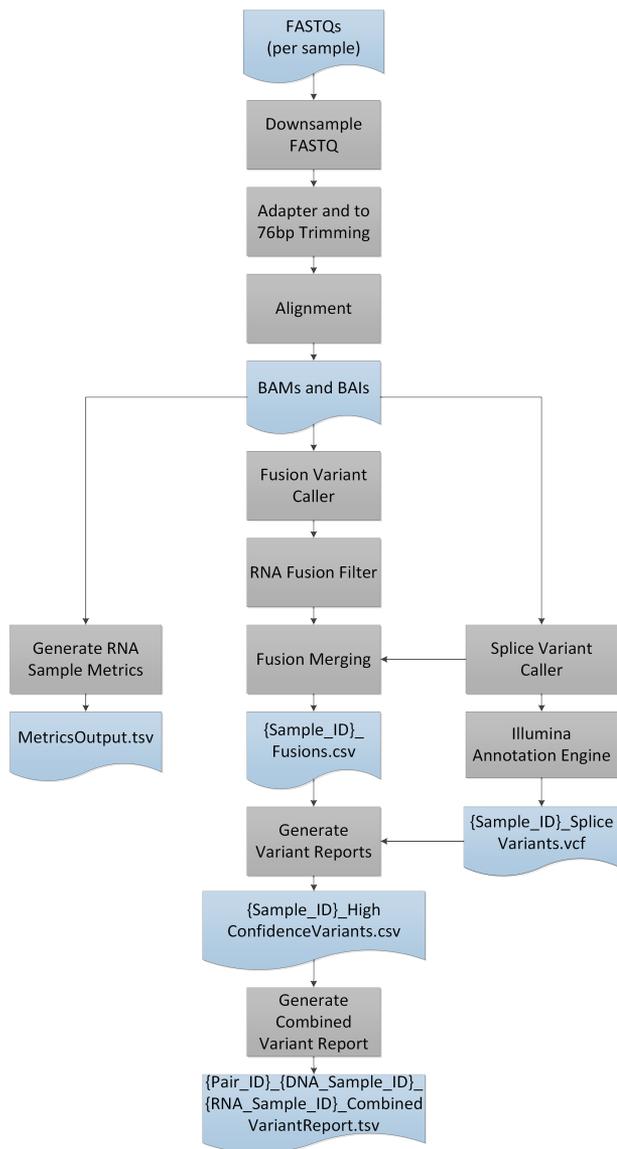
Computing Requirements

The following are the minimum operating requirements per TruSight Tumor 170 analysis.

Software

The TruSight Tumor 170 Local App requires the Linux CentOS 7.3 operating system (kernel version 3.10.0-514.26.2.el7.x86_64) with Docker v1.12 or later.

Hardware

The TruSight Tumor 170 Local App requires the following hardware.

▶ Minimum:
   ▶ 64 GB RAM.

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

▶   64-bit CPU with 16 physical cores, 32 threads.

▶   1 TB HDD free space.

▶   64 GB IPC shared memory.

▶   Dual Intel Xeon 64-bit CPU. Use E5-2620 v4 (released Q1 2012) or newer. Processor must support the AVX instruction set.

▶   Recommended:

▶   64 GB RAM (2400 MHz DDR4 RDIMM ECC).

▶   1 TB HDD free space.

▶   Hardware RAID controller (1 GB cache).

▶   2 x 2 TB SATA 7.2k RPM HDD (enterprise grade) in RAID 1 configuration.

Install TruSight Tumor 170 Local App

Complete the prerequisite tasks before installing the TruSight Tumor 170 Local App.

Prerequisites

▶   Docker is installed and set up.

▶   You have sudo privileges. If you do not, contact your system administrator.

▶   You have reviewed the End User License Agreement (EULA) before downloading the software. Your download and use of the software is acknowledgment that you have reviewed and accepted the terms and conditions of the EULA.

The TruSight Tumor 170 Local App README and EULA can be found on the TruSight Tumor 170 Local App support page on the Illumina Support Site.

Install TruSight Tumor 170 Local App

1   Obtain support documentation from the TruSight Tumor 170 support page.

2   Contact your Illumina representative or Illumina Technical Support for access to the Docker image and resource file.

3   Verify the Docker image file by calculating the md5 hash value using the following command:

```
md5sum {DOWNLOAD_PATH}/TruSight_Tumor_170_Local_App_
    {AppVersionNumber}.tar
```

The calculated md5sum must match the value that you received when you requested the application.

4   Verify the resources file by calculating the md5 hash value using the following command:

```
md5sum {DOWNLOAD_PATH}/TruSight_Tumor_170_Resources_
    {AppVersionNumber}.tgz
```

The calculated md5sum must match the value that you received when you requested the application.

5   Extract the resources file using the following command:

```
tar -zxvf {DOWNLOAD_PATH}/TruSight_Tumor_170_Resources_
    {AppVersionNumber}.tgz -C {GENOME_RESOURCE_PATH}
```

6   Load the Docker image using the following command:

```
sudo docker load < {DOWNLOAD_PATH}/TruSight_Tumor_170_Local_App_
    {AppVersionNumber}.tar
```

7   List the Docker images using the following command, to confirm that the image was successfully loaded:

```
sudo docker images
```

8   Check the shared memory settings on your host, using the following command:

```
sudo cat /proc/sys/kernel/shmmax
```

The value must be greater than 63554432.

9   Check the shared memory setting within a Docker container, using the following command:

```
sudo docker run --entrypoint cat docker-
    oncology.dockerhub.illumina.com/zodiac/tst170localapp:
    {AppVersionNumber} /proc/sys/kernel/shmmax
```

The value must be 64 GB or higher.

Run TruSight Tumor 170 Local App in Docker

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

Consider the following when running the TruSight Tumor 170 Local App:

▶   Wait for any running TruSight Tumor 170 container to complete before launching a new analysis. Run the following command to check running containers before starting theTruSight Tumor 170 Local App:

```
sudo docker ps
```

▶   Paths in the command line must be absolute paths.

▶   The analysis output folder is saved to a user-specified location. To run multiple analyses, you must either specify a different folder location for each analysis or clear all files from the output folder after each analysis. For more information on outputs and reports, see *Analysis Output* on page 17.

▶   If the error, "Access to the path… is denied." is encountered, use the following command to run the docker image:

```
sudo docker run -t \
```

```
-v /etc/localtime:/etc/localtime:Z \
```

```
-v {RUN_PATH}:/data:Z \
```

```
-v {GENOME_RESOURCE_PATH}:/genomes:Z \
```

```
-v {ANALYSIS_OUTPUT_PATH}:/analysis:Z \
```

```
docker-oncology.dockerhub.illumina.com/zodiac/tst170localapp:
    {AppVersionNumber}
```

▶   If an error is encountered when STAR unloads the genome, check for hanging shared memory and clear it if necessary.

From BCL Files

The following inputs are required for running the TruSight Tumor 170 Local App using BCL files:

▶   <RUN_PATH>: Directory containing sequencing run folder with BCLs and SampleSheet.csv
▶   <GENOME_RESOURCE_PATH>: Genomes and Resource files location
▶   <ANALYSIS_OUTPUT_PATH>: Analysis output folder location

1   Run the TruSight Tumor 170 Local App using the following criteria:

```
sudo docker run -t \
```

```
-v /etc/localtime:/etc/localtime \
```

```
-v <RUN_PATH>:/data \

-v <GENOME_RESOURCE_PATH>:/genomes \

-v <ANALYSIS_OUTPUT_PATH>:/analysis \

docker-
    oncology.dockerhub.illumina.com/
    zodiac/tst170localapp:<AppVersionNumber>
```

The analysis logs and metrics output in the analysis output folder indicate analysis run and analysis status for each sample. For more information, see *Analysis Output* on page 17.

The analysis log prints "Completed analysis" when the analysis completes.

From FASTQ Files

The following inputs are required for running the TruSight Tumor 170 Local App using FASTQ files:

▶ {FASTQ_PATH}: Directory containing a folder with FASTQ files and SampleSheet.csv

▶ {GENOME_RESOURCE_PATH}: Genomes and Resource files location

▶ {ANALYSIS_OUTPUT_PATH}: Analysis output folder location

When starting from FASTQ files, the FASTQ files must be compatible with the TruSight Tumor 170 Local App. For more information, see *FASTQ Files Compatibility Requirements* on page 12.

1 Run the TruSight Tumor 170 Local App using the following criteria:

```
sudo docker run -t \

-v /etc/localtime:/etc/localtime \

-v {FASTQ_PATH}:/data \

-v {GENOME_RESOURCE_PATH}:/genomes \

-v {ANALYSIS_OUTPUT_PATH}:/analysis \

docker-oncology.dockerhub.illumina.com/zodiac/tst170localapp:
    {AppVersionNumber} -fastq
```

The analysis logs and metrics output in the analysis output folder indicate analysis run and analysis status for each sample. For more information, see *Analysis Output* on page 17.

The analysis log prints "Completed analysis" when the analysis completes.

FASTQ Files Compatibility Requirements

When starting from FASTQ files, the FASTQ files must meet the following requirements.

▶ Generated using bcl2fastq version 2.17.1.14, 2.18, or 2.19.1.

▶ The following adapters sequences are used:

    ▶ Adapter AGATCGGAAGAGCACACGTCTGAACTCCAGTCA

    ▶ AdapterRead2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

▶ Do not use FASTQ files with UMIs.

▶ Use the template sample sheet (SampleSheet.csv) to create sample sheets when running bcl2fastq to generate FASTQ files. FASTQ file names contain the unique sample ID from the Sample_ID column of the SampleSheet.csv.

▶ FASTQ file names must contain the Sample ID. By default, bcl2fastq generates FASTQ files with a file name that includes sample names if the sample names are specified in the SampleSheet.csv. When running bcl2fastq independently, leave the Sample_Name field blank so that bcl2fastq generates FASTQ files with the expected file name containing Sample IDs.

▶ Be sure that the memory settings specified in the minimum hardware specifications are used when running bcl2fastq. For more information, see *Hardware* on page 9.

▶ Leave the Sample Project column of the sample sheet blank to make sure that it is compatible with the TruSight Tumor 170 Local App workflow.

▶ On a NextSeq, reverse complement the index 2 sequences within the Data section of the sample sheet before running bcl2fastq.

▶ If the sample count is 4 or more, set --writing-threads to 4, otherwise, set --writing-threads to be equal to the sample count. If you specify more writing threads than samples, the extra threads do no work but can cost time due to context switching.

▶ Use the following arguments to generate FASTQ files using bcl2fastq version 2.17.1.14 and 2.18:

```
#{BCL2FASTQ_PATH}: Path to bcl2fastq executable

#{OUTPUT_DIRECTORY_PATH}: Absolute path for output generated fastqs

#{RUN_FOLDER_PATH}: Absolute path to run folder containing BCLs

#{SAMPLE_SHEET_PATH}: Absolute path to sample sheet (Typically
    contained within run folder)

{BCL2FASTQ_PATH} --output-dir '{OUTPUT_DIRECTORY_PATH}' --runfolder-
    dir '{RUN_FOLDER_PATH}' --sample-sheet '{SAMPLE_SHEET_PATH}' --
    ignore-missing-bcls --ignore-missing-filter --ignore-missing-
    positions --ignore-missing-controls --find-adapters-with-sliding-
    window --auto-set-to-zero-barcode-mismatches --adapter-stringency
    '0.9' --mask-short-adapter-reads '35' --minimum-trimmed-read-
    length '35' --loading-threads '4' --demultiplexing-threads '7' --
    processing-threads '32' --writing-threads '4'
```

▶ Use the following arguments to generate FASTQs using bcl2fastq version 2.19.1:

```
#{BCL2FASTQ_PATH}: Path to bcl2fastq executable

#{OUTPUT_DIRECTORY_PATH}: Absolute path for output generated fastqs

#{RUN_FOLDER_PATH}: Absolute path to run folder containing BCLs

#{SAMPLE_SHEET_PATH}: Absolute path to sample sheet (Typically
    contained within run folder)

{BCL2FASTQ_PATH) --output-dir '{OUTPUT_DIRECTORY_PATH}' --runfolder-
    dir '{RUN_FOLDER_PATH}' --sample-sheet '{SAMPLE_SHEET_PATH}' --
    ignore-missing-positions --ignore-missing-controls --find-
    adapters-with-sliding-window --auto-set-to-zero-barcode-
    mismatches --adapter-stringency '0.9' --mask-short-adapter-reads
    '35' --minimum-trimmed-read-length '35' --loading-threads '4' --
    processing-threads '32' --writing-threads '4'
```

▶ Do not merge FASTQ files, the TST170 analysis pipeline expects separate FASTQ files per sample (Number of FASTQ files = Total Lane Number x 2 Reads Per Lane).
  ▶ Eight FASTQ files per sample for NextSeq
  ▶ Four FASTQ files per sample for HiSeq 2500 Rapid Run

Analysis Methods

The TruSight Tumor 170 App workflow performs the following analysis steps and then writes analysis output files to the folder specified. For information on software components and file outputs for each analysis step, see the *File Overview* section of this guide.

DNA Analysis

Inputs Validation

Validates the existence and format of required analysis inputs.

RunQC

The software performs the RunQC analysis step when you run the TruSight Tumor 170 Local App from BCL files. Using InterOp files from the sequencing run folder, the RunQC report provides run metrics 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. Outputs from the RunQC analysis are generated in the MetricsOutput.tsv file. For more information, see *Metrics Output* on page 20.

FASTQ Generation

FASTQ files are only generated when running the TruSight Tumor 170 Local App using BCL files. BCL files are demultiplexed and the software generates intermediate analysis files in the FASTQ format. FASTQ files use a text format to represent sequences. They contain reads for each sample and the associated quality scores. Any controls used for the run and clusters that did not pass filter are excluded. Each FASTQ file contains reads for only one sample, and the name of that sample is included in the FASTQ file name. For more information about FASTQ files requirements, see *From FASTQ Files* on page 12.

Adapter Trimming

Adapter trimming is performed on the FASTQ files corresponding to each DNA library.

Alignment

The alignment analysis uses the BWA-MEM alignment algorithm to align DNA sequences in FASTQ files to the hg19 genome. The alignment analysis generates BAM files (*.bam) and BAM index files (*.bam.bai). A BAM file is the compressed binary version of a SAM file that is used to represent aligned sequences.

Indel Realignment

Reads near detected indels are realigned to remove alignment artifacts.

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

Small Variant Calling

Pisces performs somatic variant calling to identify variants at low frequency in DNA samples.

Variant Filtering

The DNA Variant Filter software performs post processing on the small variant calling VCF, with the intended result of background polishing and quality score adjustment. The inputs and outputs are gVCF files.

Annotation

The Illumina Annotation Engine performs annotation of small variants. The inputs are VCF files. The outputs are JSON files.

Complex Variant Calling

BWA-MEM aligns any previously generated FASTQ files to a complex variant reference sequence data set. The data set represents 29 complex indels of the EGFR gene. The complex variant caller uses EGFR variant reference sequences for the following tasks.

▶ Creating alternative alignments and performing duplicate marking on the resulting alignments

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

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

Copy Number Variant Calling

The CRAFT copy number variant caller performs amplification, reference, and deletion calling for target amplification genes within the assay. This calling is based on the calculated fold change value. Fold change is defined as the median normalized bin count value for a target gene over the median bin count value of the entire panel. The inputs are in BAM format and the outputs are VCF files. Deletion is marked with the LowValidation filter in the VCF file because they have only been tested with *in silico* data sets.

Determining Microsatellite Instability

Microsatellite instability (MSI) is determined from the BAM file, and then an MSI metric file is created. Using information entropy metrics, the software assesses microsatellite sites for evidence of instability relative to a set of baseline normal samples. The proportion of unstable MSI sites to total assessed MSI sites is reported as a sample-level microsatellite score.

DNA Sample Metrics Reporting

This component performs DNA sample metric calculation and reporting. The inputs are CRAFT normalized BinCount.txt files, BAM files, and gVCF files. The output is a text file that contains sample metrics for all samples analyzed in the run (DNA and RNA).

RNA Analysis

Inputs Validation

Validates the existence and format of required analysis inputs.

RunQC

The software performs the RunQC analysis step when you run the TruSight Tumor 170 Local App from BCL files. Using InterOp files from the sequencing run folder, the RunQC report provides run metrics 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. Outputs from the RunQC analysis are generated in the MetricsOutput.tsv file. For more information, see *Metrics Output* on page 20.

FASTQ Generation

FASTQ files are only generated when running the TruSight Tumor 170 Local App using BCL files. BCL files are demultiplexed and the software generates intermediate analysis files in the FASTQ format. FASTQ files use a text format to represent sequences. They contain reads for each sample and the associated quality scores. Any controls used for the run and clusters that did not pass filter are excluded. Each FASTQ file contains reads for only one sample, and the name of that sample is included in the FASTQ file name. For more information about FASTQ files requirements, see *From FASTQ Files* on page 12.

Down-sampling

The TruSight Tumor 170 analysis pipeline down-samples each sample to 30 million RNA reads. This number represents the total number of single reads (ie, R1 + R2, from all lanes). When using the recommended sequencing configurations or plexing, the samples can have fewer reads than the down-sampling limit. In these cases, the FASTQ files are left as-is.

Adapter Trimming

Adapter trimming is performed on the down-sampled FASTQ files corresponding to each RNA library. Reads are then trimmed to 76 base pairs for further processing.

Alignment

The STAR Aligner aligns RNA reads to the human hg19 genome. The inputs are trimmed FASTQ files and the outputs are BAM files.

Fusion Calling

Manta[1] performs the fusion calling. The inputs are BAM files and the outputs are VCF files.

1. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics.* 2016;32:1220-1222.

RNA Fusion Filtering

Candidate fusion calls are evaluated against a range of quality filters during this step. The inputs are BAM and VCF files, and the outputs are CSV files.

Splice Variant Calling

Splice variant calling is performed using the RNA Splice Variant Calling software. The inputs are BAM files and SJ.out.tab files from STAR. The outputs are VCF files, which are the final output, and a TSV file containing intergenic variants, which are used in fusion merging.

Fusion Merging

Fusions identified using Manta are combined with intergenic calls that are identified using the RNA Splice Variant Calling software. Also, the total number of supporting reads (including duplicates) for each precise fusion event is calculated. The inputs are intergenic splice variant TSV files, fusion filter CSV files, and BAM files. The outputs are CSV files.

Annotation

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

RNA Sample Metrics Output

This component performs RNA sample metric calculation and reporting. The inputs are BAM files and the output is a text file.

Analysis Output

When the analysis run is completed, the TruSight Tumor 170 Local App generates an analysis output folder in the user-specified location with the folder name Analysis_{Date}_{Time}.

To view analysis output:

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 the output folders generated during analysis. The TruSight Tumor 170 Local App generates different outputs depending on whether analysis begins from either BCL or FASTQ files, and whether DNA and RNA samples are analyzed together. The output folders are always present in the top-level output directory.

The following example illustrates a single DNA library (Sample1) and a single RNA library (Sample2) associated together in a single pair (Pair1).

📁 Analysis_ YYYYDDMM_HHMMSS
    📁 Logs_Intermediates
        📁 Logs
            📁 DNA—Contains component logs for each step.
            📁 RNA—Contains component logs for each step.
            📁 Fastq—Contains FASTQ logs.
            📁 Reports
                📄 CombInedVariantReport_{Date-TIme}.log
                📄 MetricsReport_{Date-TIme}.log
            📄 AnalysisLog_{Date-TIme}.log
        📁 Fastqs
            📁 Sample1

📄 Sample1_S1_L001_R1_001.fastq.gz

📄 Sample1_S1_L001_R2_001.fastq.gz

📄 Sample1_S1_L002_R1_001.fastq.gz

📄 Sample1_S1_L002_R2_001.fastq.gz

📄 Sample1_S1_L003_R1_001.fastq.gz

📄 Sample1_S1_L003_R2_001.fastq.gz

📄 Sample1_S1_L004_R1_001.fastq.gz

📄 Sample1_S1_L004_R2_001.fastq.gz

📁 Sample2

📄 Sample2_S2_L001_R1_001.fastq.gz

📄 Sample2_S2_L001_R2_001.fastq.gz

📄 Sample2_S2_L002_R1_001.fastq.gz

📄 Sample2_S2_L002_R2_001.fastq.gz

📄 Sample2_S2_L003_R1_001.fastq.gz

📄 Sample2_S2_L003_R2_001.fastq.gz

📄 Sample2_S2_L004_R1_001.fastq.gz

📄 Sample2_S2_L004_R2_001.fastq.gz

📁 DNA_IntermediateFiles

📁 **Alignment**—Contains BAM and BAI files from alignment and indel realignment.

📁 **Annotations**—Contains small variants and complex variant annotation JSON files.

📁 **CNV**—Contains bincounts and fold change values (from CRAFT).

📁 **MSI**

📄 SampleID1_MSI.json

📁 RNA_IntermediateFiles

📁 **Alignment**—Contains BAM and BAI files and a splice junction file.

📁 **Annotations**—Contains splice variant annotation JSON files.

📁 **FusionsCalling**

📄 SampleID2_Fusions.csv

📁 **SpliceVariantCalling**

📄 SampleID2_SpliceVariants.vcf

📁 Reports

📄 MetricsOutput.tsv

📄 Pair1_SampleID1_SampleID2_CombinedVariantReport.tsv

📁 Sample_ID1

📄 SampleID1_CopyNumberVariants.vcf

📄 SampleID1_SmallVariants.genome.vcf

📄 SampleID1_Complex.vcf

📁 **Sample_ID2**

📄 SampleID2_HighConfidenceVariants.csv

📄 SampleSheet.csv

📄 RunQC.json

File Overview

This section describes the summary output files generated during analysis.

Reports

File Output Type	Content
MetricsOutput.tsv	Analysis status and QC metrics for each sample in the analysis.
{Pair_ID}_{DNA_Sample_ID}_{RNA_Sample_ID}_CombinedVariantReport.tsv	For each Pair ID specified in the sample sheet, a report is created which contains the following information: • Passing small variants • Passing complex variants • MSI • Fusions (high confidence only) • Splice Variants (high confidence only) • Copy Number Variants

DNA

System Name / Output Folder Names	File Output Type	Content	Software Component
{Sample_ID}	*_SmallVariants.genome.vcf *_Complex.vcf	Small variant and complex variant genome VCF files for each sample.	Pisces Psara Pepe
	*_CopyNumberVariants.vcf	Copy number variants for each sample	CRAFT
Logs_Intermediates/Fastqs	*.fastq.gz — six per sample (HiSeq), eight per sample (NextSeq)	Original FASTQ files.	bcl2fastq
Logs_Intermediates/DNA_IntermediateFiles/Alignment	*.bam *.bam.bai	Aligned BAM and BAI files from BWA for each sample.	BWA-MEM SAMtools
	*._realigned.bam *._realigned.bam.bai	BAM and BAI files from Hygea for each sample.	Hygea
	*._Complex.bam *._Complex.bam.bai	Aligned Complex BAM and BAI files for each sample.	Complex Variant Caller
Logs_Intermediates/DNA_IntermediateFiles/Annotation	*_Complex.json *.json	Annotated complex and small variant JSONs for each sample.	Illumina Annotation Engine

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

System Name / Output Folder Names	File Output Type	Content	Software Component
Logs_Intermediates/ DNA_ IntermediateFiles/ MSI	*_MSI.json	Microsatellite results JSON for each sample.	Hubble

RNA

System Name / Output Folder Names	File Output Type	Content	Software Component
{Sample_ID}	*_HighConfidenceVariants.csv	High confidence fusion and splice variants for each sample.	Not applicable
Logs_Intermediates/ Fastqs	*.fastq.gz – two per sample	Original FASTQ files.	bcl2fastq
Logs_Intermediates/ RNA_ IntermediateFiles/ Alignment	*.bam *.bam.bai	Aligned BAM and BAI files from BWA for each sample.	STAR SAMTools SuperDuper
Logs_Intermediates/ RNA_ IntermediateFiles/ FusionCalling	*_Fusions.csv *_fusionfilter.csv	Fusion and filtered fusion CSV files for each sample. Raw and combined fusion candidates.	Manta RNAFusionFilter
Logs_Intermediates/ RNA_ IntermediateFiles/ SpliceVariantCalling	*_SpliceVariants.vcf *.tsv *_fusions.tsv	Splice variant VCF and TSV for each sample.	Splice Variant Caller
Logs_Intermediates/ RNA_ IntermediateFiles/ Annotation	*_SpliceVariants.json	Annotated splice variants for each sample.	Illumina Annotation Engine

SampleSheet.csv

File name: SampleSheet.csv

The TruSight Tumor 170 specific sample sheet is a comma-delimited file that stores the information to set up and analyze a sequencing experiment. The sample sheet includes the list of samples to be analyzed, the unique index of each sample, the Pair ID for each sample, and whether a sample contains DNA or RNA. For information on populating the sample sheet, see *Workflow Requirements* on page 3.

Metrics Output

The following topics describe the sections of the Metric Output file.

Analysis Details

The analysis details contain the report date and time, and the version of the pipeline used in the analysis.

Run Metrics Output

The run metrics section is only generated when inputs are BCLs from a sequencing run folder.

Metric	Description	Recommended Threshold
PCT_PF_READS (%)	Percent of reads passing filter.	≥ 80.0
PCT_Q30_R1 (%)	Percent of base calls with a quality score of Q30 or higher for Read 1.	≥ 80.0
PCT_Q30_R2 (%)	Percent of base calls with a quality score of Q30 or higher for Read 2.	≥ 80.0

Sample Status

Metric	Description
COMPLETED_ALL_STEPS	"True" indicates that the sample has completed all analysis steps.
FAILED_STEPS	Indicates if a sample failed an analysis step.
STEPS_NOT_EXECUTED	Indicates any skipped analysis steps.

DNA Library QC Metrics for Small Variant Calling

Metric	Description	Recommended Threshold
PCT_CHIMERIC_READS (%)	Percentage of unique reads that are part of a chimeric alignment (for example, half of the read maps to one location and the other half maps to a nonsequential location) out of the total number of mapped unique reads.	≤ 3.5
MEDIAN_INSERT_SIZE (bp)	Median of fragment sizes calculated from read alignments.	≥ 70
PCT_EXON_250X (%)	Percentage of exonic bases in gVCF that have greater than or equal to 250X coverage.	≥ 81.3

DNA Library QC Metrics for Copy Number Variant Calling

Metric	Description	Recommended Threshold
PCT_CHIMERIC_READS (%)	Percentage of unique reads that are part of a chimeric alignment (for example, half of the read maps to one location and the other half maps to a nonsequential location) out of the total number of mapped unique reads.	≤ 3.5
COVERAGE_MAD	Median of absolute deviations from the median of the normalized count of each CNV target region.	≤ 0.2
MEDIAN_BIN_COUNT_CNV_TARGET	Median of the raw read counts (normalized by region size) of each CNV target region.	≥ 1

DNA Library QC Metrics for MSI

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

DNA Expanded Metrics

Metric	Description
MEAN_TARGET_COVERAGE (Count)	Median fragment coverage of target bases.
PCT_ALIGNED_READS (%)	Percentage of reads that are mapped.
MEDIAN_BIN_COUNT_CNV_TARGET	Median of the raw read counts (normalized by region size) of each copy number variant target region. This metric indicates the quality of copy number variant (amplification) calls.
PCT_PF_UQ_READS (%)	Percentage of unique reads passing filter.
PCT_READ_ENRICHMENT (%)	Percentage of reads enriched for the target regions.
PCT_TARGET_0.4X_MEAN (%)	Percentage of target bases with target coverage greater than 0.4 times the mean.
PCT_TARGET_250X (%)	Percent target bases with greater than 250X coverage.
TOTAL_PF_READS (Count)	Total number of reads passing filter.

RNA Library QC Metrics

Metric	Description	Recommended Threshold
MEDIAN_INSERT_SIZE (bp)	Median of fragment sizes calculated from read alignments.	≥ 80
MEDIAN_CV_GENE_500X	The median of the CV value of target region coverage across all genes with coverage above 500X.	≤ 0.93
TOTAL_ON_TARGET_READS (Count)	Total number of reads that map to the target regions.	≥ 9,000,000

RNA Expanded Metrics

Metric	Description	Recommended Threshold
PCT_CHIMERIC_READS (%)	Percentage of unique reads that are part of a chimeric alignment (for example, half of the read maps to one location and the other half maps to a nonsequential location) out of the total number of mapped unique reads.	Not applicable
TOTAL_PF_READS (Count)	Total number of reads passing filter.	Not applicable
PCT_ON_TARGET_READS	Percentage of reads that cross any part of the target region out of the total number of reads. A read that partially maps to a target region is counted as on target.	Not applicable
SCALED_MEDIAN_GENE_COVERAGE	The median of median gene coverage scaled by gene length. Median coverage is calculated for each gene, and then scaled by the respective gene's length, and finally divided by the total length of all targeted genes.	Not applicable

Combined Variant Report

The combined variant report combines selected results from the following output files into a single document for quick results review.

▶ Annotation.json

▶ Annotation_Complex.json

▶ MSI.json

▶ CopyNumberVariants.vcf

▶ HighConfidenceVariants.csv

The following topics describe the sections of the Combined Variant Report file.

Analysis Details

The analysis details contain the Pair ID and Sample IDs of the sample(s) in this pair, the report date, the report time, and the pipeline version.

Sequencing Run Details

The sequencing run details contain run information extracted from the RunInfo.xml and RunParameters.xml files. The sequencing run details are only displayed when analysis begins from BCL inputs.

MSI

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

Small Variants

When available, the Illumina Annotation Engine provides the following fields for each small variant.

Metric	Description
Gene (HGVS)	Human Genome Variation Society (HGVS) gene notation.
Chromosome	The chromosome where the SNV, insertion, or deletion is detected.
Genomic Position	The chromosome coordinate where the SNV, insertion, or deletion is detected.
Reference Call	The hg19 reference call for the respective chromosome:position.
Alternative Call	The observed variant call for the respective chromosome:position.
P-Dot Notation	HGVS protein change notation.
C-Dot Notation	HGVS nucleotide change notation.
Consequence	Consequence on protein function.
Allele Frequency	Fraction of reads in which the variant is detected.

Metric	Description
COSMIC ID	Catalogue of Somatic Mutations in Cancer (COSMIC) ID for this variant.
Affected Exon	The exons affected by the variant. For example, 4-6/7 indicates that exons 4, 5, and 6 are affected, and that the respective transcript contains a total of seven exons.
Depth	Unique read count for the detected variant.
GQX Value	Genotype quality.

EGFR Complex Variants

The annotations used for small and complex variants are the same. See *Small Variants* on page 23 for an explanation of the annotated fields for EGFR Complex Variants.

Fusions

See *High Confidence Variants* on page 26 for an explanation of the variants provided in this section.

Splice Variants

See *High Confidence Variants* on page 26 for an explanation of the variants provided in this section.

Amplifications

See *Copy Number Variants VCF* on page 24 for an explanation of the variants provided in this section.

DNA Outputs

Small Variants Genome VCF

File name: {Sample_ID}_SmallVariants.genome.vcf

This file contains small variant calls for each target coordinate in the TruSight Tumor 170 DNA panel.

The following variant call filters are applied during the variant call and filter step.

Filter Name	Description
Low Quality	Quality score is < 20 (applies to reference and variant allele calls).
Low Depth	Coverage is < 100 for variant calls and < 250 for reference calls.
Strand Bias	Strand bias score is ≥ to 0.5 (applies to variant allele calls only).
Low Variant Frequency	Frequency is < 0.02 (applies to variant allele calls only).
Indel Reference Repeat	Partial or total repeats of the variant allele (max repeat length 5) in the reference are ≥ 9 (applies to insertion and deletion variant calls only).
Adjusted Quality	Baseline quality score is < 10 (applies to variant allele calls only).
Low Variant Support	Variants with supporting reads < 7 (applies to variant allele calls only).

Copy Number Variants VCF

File name: {Sample_ID}_CopyNumberVariants.vcf

Contains CNV calls for DNA libraries of the amplification genes targeted by the TruSight Tumor 170 panel. Each copy number variant is reported with a corresponding fold change. This value is derived from the normalized read depth of the gene in the sample relative to the normalized read depth of diploid regions from the same sample. The value in the QUAL column of the VCF is a Phred transformation of the p-value:

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

The p-value is derived from the t-test between the fold change of the gene against rest of the genome. Higher Q-scores indicate higher confidence in the CNV call.

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

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

Given tumor purity, the copy number of the gene in the tumor can be inferred from the reported fold change of the gene.

Given tumor purity X%, for a reported fold change Y, the copy number *n* can be calculated 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 fold change of 2.2x indicates that 10 copies of MET DNA are observed.

RNA Outputs

Fusions CSV

File name: {Sample_ID}_Fusions.csv

This file contains all candidate fusions identified by the RNA analysis pipeline. Candidate fusions from the RNA splice variant caller are listed in this output with relevant supporting information. Two key output columns in the file describe the candidate fusions: Filter and KeepFusion.

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

Filter	Description
Imprecise	A low-resolution candidate, not an assembled fusion call. (Confidence filter)
RepeatOverlap	The fusion is tagged as overlapping with a repeat region. Only used as a confidence filter for nonuniquely mapping fusion candidates, otherwise information only. (Confidence filter)
WeakBreakend	The read/alignment evidence on one side of the fusion is weak. Usually this filter indicates that the reads only overlap the fusion by a few base pairs. Alternatively, it can indicate too much homology (no unique sequence). (Confidence filter)
Homology	The fusion contig is a substring of another fusion contig. (Information only)
DuplicateContig	The two contigs of the fusion are the same contig. (Confidence filter)
ContigIntragenic	The realignment of half-contigs produces alignments that map to the same gene on both sides (or within 1 kb if unannotated). (Confidence filter)

Filter	Description
LowQ	Fusion supporting reads (unique) < 5 (+ 1 for every 10 million reads over 16 million reads). (Confidence filter)
LowDupReads	Fusion supporting reads (duplicate) < 5. (Information only)
NonExonic	Fusion breakpoint does not fall within an exon. (Information only)
LocalContigAlign	Contig realignment found a nonfusion alignment for this contig. (Information only)
LowFusionRatio	Few strong evidence reads compared to wild type reads. (Information only)
NoReferenceReads	No reads on either side of the presumed breakpoint are marked as reference (structurally normal) reads. (Information only)

The KeepFusion column of the output has a value of True when the RNAFusionFilter score is ≥ 0.45 and none of the confidence filters are triggered.

The directionality of a fusion can be determined from the fusionfilter.csv file. If either the Strand1 or Strand2 column of the output has a value of + or −, then the gene in the Gene1 column is in the 5′ position, and the gene in the Gene2 column is in the 3′ position.

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

> **NOTE**
>
> When using Microsoft Excel to view this report, genes that are convertible to dates (such as MARCH1) automatically convert to dd-mm format (1-Mar) by Excel.

Splice Variants VCF

File name: {Sample_ID}_SpliceVariants.vcf

Contains all candidate splice variants targeted by the TruSight Tumor 170 panel identified by the RNA analysis pipeline.

The following filters can be applied for each variant call.

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

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

▶ HGNC Gene

▶ Transcript

▶ Exons

▶ Introns

▶ Canonical

▶ Consequence

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

High Confidence Variants

File name: {Sample_ID}_HighConfidenceVariants.csv

Contains the high confidence fusion and splice variant calls.

High confidence fusions are fusions targeted with passing filter (KeepFusions column displays True in the {Sample_ID}_Fusions.csv). There are a total of 55 genes targeted for fusion calls. The fusion is reported if at least one gene of the gene pair is in the targeted gene list.

► Each reported gene fusion has the following information:
  ► Breakpoint 1
  ► Breakpoint 2
  ► Fusion Supporting Reads
► The order of gene1/gene2 fusion reported in the Gene Fusion column does not indicate directionality of the fusion. For any fusion of interest, search for the gene pair in both orientations.
► Semicolons denote fusions with overlapping genes. For example, Gene1;Gene2/Gene 3 indicates possible fusion event Gene1/Gene3 and Gene2/Gene3.
► The fusion variants are sorted, in ascending order, by Gene 1, then by Gene 2, then by Gene 1 Breakpoint, and then by Gene 2 Breakpoint.

High confidence splice variant calls are splice variants targeted by the panel (total of three genes) with a passing filter.

► Each reported splice variant has the following information:
  ► Affected Exon(s)
  ► Transcript
  ► Breakpoint Start
  ► Breakpoint End
  ► Splice Supporting Reads
► The splice variants are sorted by, in ascending order, Gene, then by Transcript, then by Breakpoint Start, and then by Breakpoint End.

Uninstall TruSight Tumor 170 Local App

1 Delete the TruSight Tumor 170 Local App tar file using the following command:

```
rm -f TruSight_Tumor_170_Local_App_{App_version_number}.tar
```

2 List the Docker images using the following command to determine the image ID that is installed:

```
sudo docker images
```

3 Delete the Docker image using the image ID using the following command:

```
sudo docker rmi -f {image_ID}
```

4 Delete the resources tarball (.tgz file and unzipped genomes directory) using the following commands:

```
rm -f TruSight_Tumor_170_Resources_{App_version_number}.tgz

rm -rf {GENOME_RESOURCES_PATH}
```

Where {GENOME_RESOURCES_PATH} is the path to the genomes directory after unzipping the *.tgz file.

Revision History

Document	Date	Description of Change
Document # 1000000079441 v02	August 2019	Updated CPU hardware requirements. Updated the file names in the RNA summary output files.
Document # 1000000079441 v01	July 2019	Updated the gene fusion and splice variant information that is included in the high confidence variant RNA output file.
Document # 1000000079441 v00	June 2019	Initial release

Technical Assistance

For technical assistance, contact Illumina Technical Support.

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

Illumina Customer Support Telephone Numbers

Region	Toll Free	Regional
North America	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgium	+32 80077160	+32 34002973
China	400.066.5835	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
South Korea	+82 80 234 5300	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

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

Product documentation—Available for download from support.illumina.com.

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

illumina®