

# TruSight Tumor 170 v1.0 Local App

# User Guide



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

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

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

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

© 2017 Illumina, Inc. All rights reserved.

Illumina, BaseSpace, TruSight, and the streaming bases design are registered or pending trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

# Table of Contents

Introduction	1
Workflow Requirements	1
Additional Resources	3
Workflow Diagrams	4
Computing Requirements	8
Install TruSight Tumor 170 Local App	8
Prerequisites	8
Install TruSight Tumor 170 Local App	9
Run TruSight Tumor 170 Local App in Docker	10
From BCL Files	10
From FASTQ Files	10
Analysis Methods	13
DNA Analysis	13
RNA Analysis	14
Analysis Output	17
Folders	17
Files	18
DNA Outputs	19
RNA Outputs	20
Generate Picard Metrics	24
Install Picard and Generate Files	24
Run Picard CollectHsMetrics	25
Uninstall TruSight Tumor 170 Local App	27
Technical Assistance	28

## 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, and published fusion results.

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, analysis outputs, and optional Picard metrics generation workflow and guide.

## Workflow Requirements

- ▶ Basic knowledge of the Linux operating system and the Docker software container platform
- ▶ Supports Illumina TruSight Tumor 170 Assay
- ▶ Compatible with NextSeq or HiSeq 2500 Rapid Run systems
  - ▶ NextSeq:
    - ▶ Sequence 16 libraries (eight DNA and eight RNA) per run to achieve maximum coverage for each sample.
    - ▶ If sequencing DNA libraries only, do not include more than 10 samples per run.
    - ▶ If sequencing RNA libraries only, do not include more than 16 samples per run.
  - ▶ HiSeq 2500 Rapid Run:
    - ▶ Sequence 12 libraries (six DNA and six RNA) per run to achieve maximum coverage for each sample.
    - ▶ If sequencing DNA libraries only, do not include more than six 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
- ▶ 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)
- ▶ 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:
  - ▶ Less 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 80 M reads (40 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.

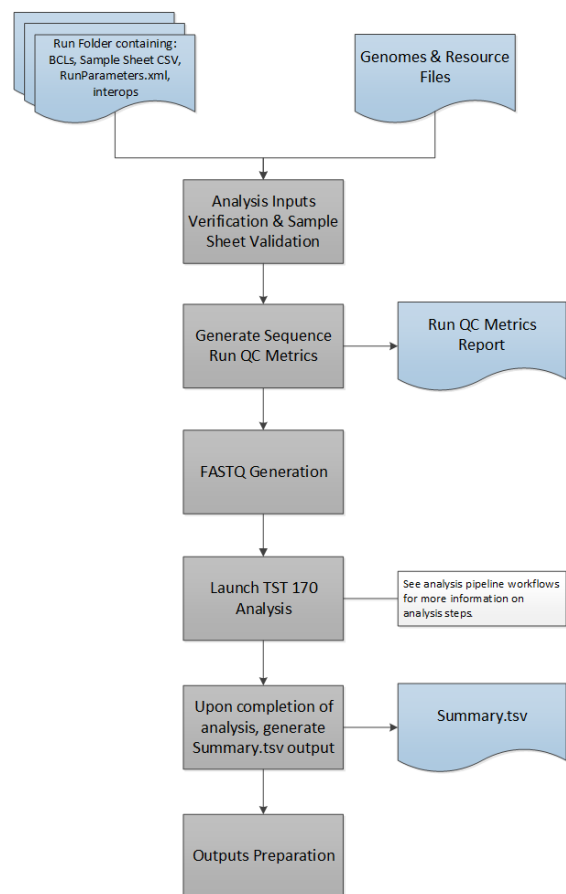
For information on installing the product, see [Downloads](#) on the support page.

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

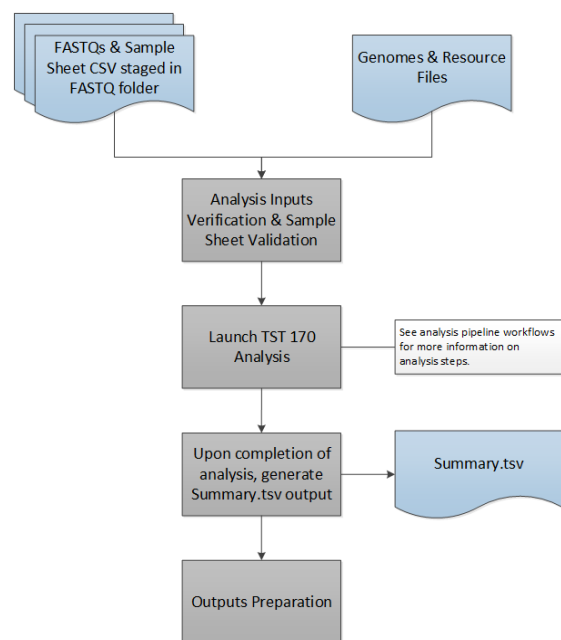
Resource	Description
<a href="#">TruSight Tumor 170 Local App Customer Release Notes</a> (document # 1000000036401)	Provides information on software enhancements, and limitations and considerations of the current release.
<a href="#">SampleSheet.csv</a> (document # 1000000038524)	Template used for running analysis.
<a href="#">Customer Software Installation Work Instructions for TruSight Tumor 170 Local App</a> (document # 1000000041749)	Provides instructions for installing the TruSight Tumor 170 Local App software.

## Workflow Diagrams

**Figure 1** TruSight Tumor 170 Software Workflow Starting at BCLs

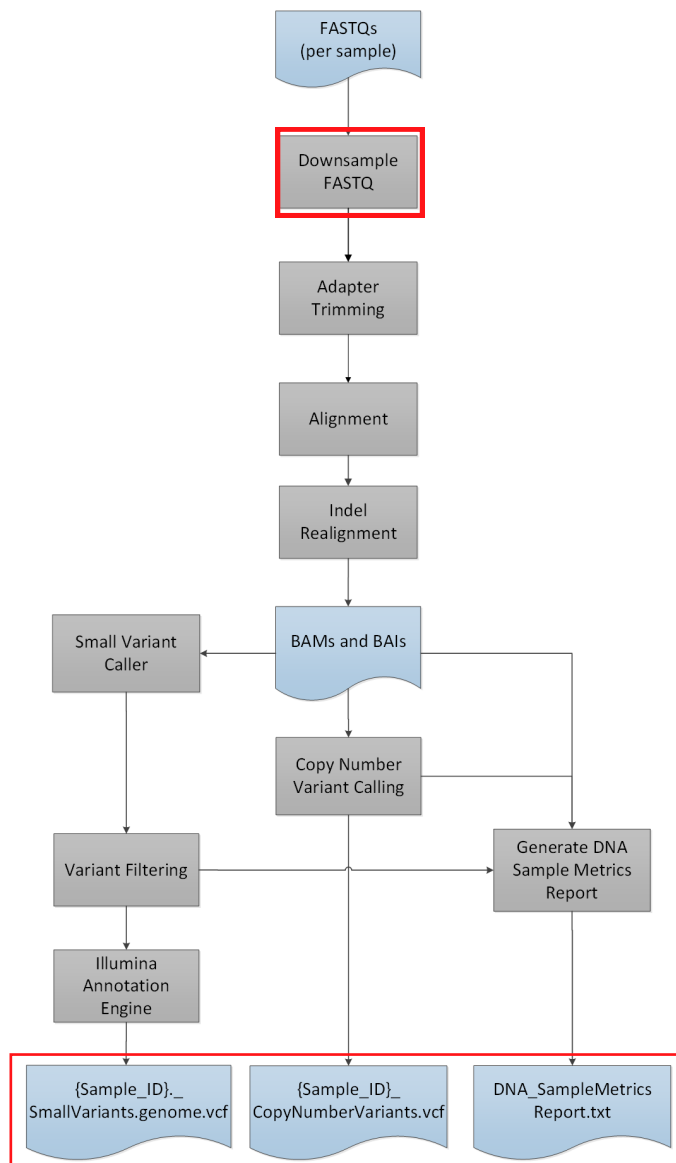


**Figure 2** TruSight Tumor 170 Software Workflow Starting at FASTQ

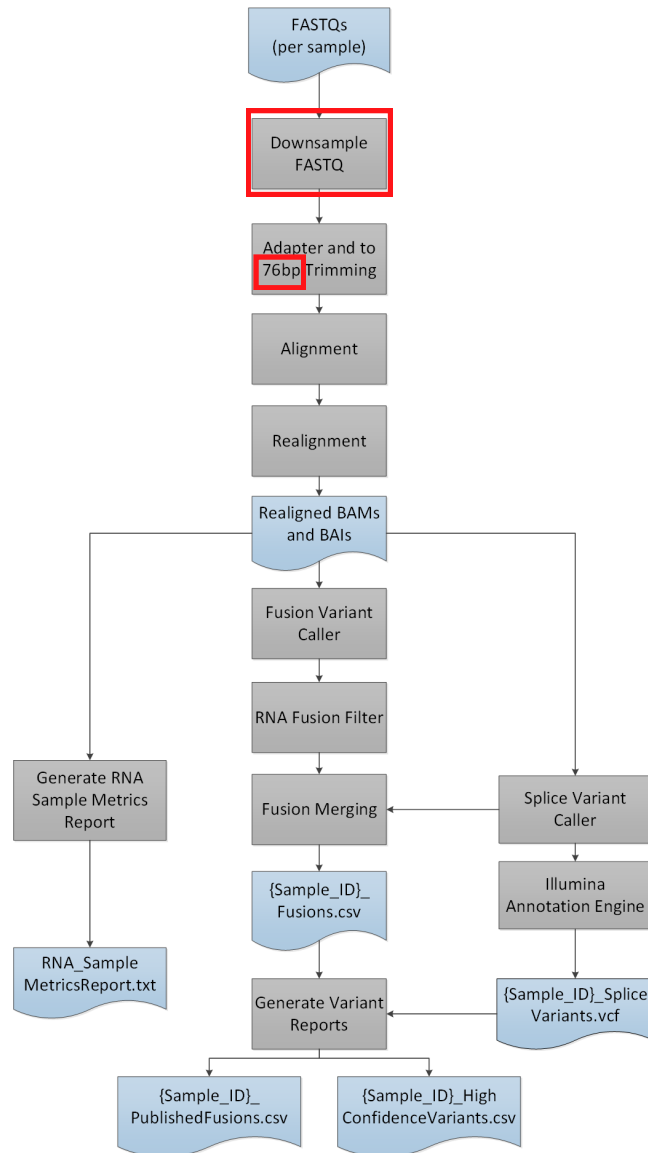




**Figure 3** TruSight Tumor 170 DNA Workflow



**Figure 4** TruSight Tumor 170 RNA Workflow



## Computing Requirements

The following are the minimum operating requirements per TruSight Tumor 170 analysis, which supports a maximum of 16 libraries (DNA and RNA combined) per run.

### Software

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.

Other recommended kernel versions:

- ▶ 2.6.32-696.6.3.el6.x86\_64
- ▶ 4.4.0-67-generic

### Hardware

TruSight Tumor 170 Local App requires the following hardware.

- ▶ Minimum:
  - ▶ 64 GB RAM
  - ▶ 64-bit CPU with 16 physical cores, 32 threads
  - ▶ 1 TB HDD free space
  - ▶ 64 GB IPC shared memory
- ▶ Recommended:
  - ▶ 64 GB RAM, 2400 MHz DDR4 RDIMM ECC
  - ▶ Dual Intel Xeon 64-bit CPU; E5-2620 v4 or better
  - ▶ 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

Before installing the TruSight Tumor 170 Local App make sure that you have completed the prerequisite tasks.

### Prerequisites

- ▶ Make sure that you have sudo privileges. If you do not, contact your system administrator.
- ▶ Docker is installed and set up. For more information, see [www.docker.com](http://www.docker.com).
- ▶ AWS Command Line Interface is installed and configured. For more information, see [aws.amazon.com/cli/](http://aws.amazon.com/cli/).
- ▶ For faster download speeds, consider increasing the maximum download threads for s3 before downloading, using the following command:
 

```
aws configure set default.s3.max_concurrent_requests 40
```
- ▶ Review and accept 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](#).

## Install TruSight Tumor 170 Local App

- 1 Obtain support documentation from the [TruSight Tumor 170 support page](#).
- 2 Contact your Illumina representative 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 provided from the [support page](#).
- 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 provided from the [support page](#).
- 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. **18446744073692774399**
- 9 Check the shared memory setting within a Docker container, using the following command:  

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

The value must be 64 GB or higher.

```
sudo docker run --entrypoint cat docker-oncology.dockerhub.illumina.com/zodiac/tst170localapp:1.0.1.0 /  
proc/sys/kernel/shmmax
```

18446744073692774399

```
hexiaoyu@login:~/panel599_20190128$ docker images
```

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
python	latest	2fbd95050b66	3 days ago	927MB
truSight-oncology-500	1.3.1.3	2f0996374f88	7 weeks ago	1.95GB
ubuntu	16.04	a51deb77e1eb	2 months ago	116MB
hello-world	latest	4ab4c602aa5e	4 months ago	1.84kB
docker-oncology.dockerhub.illumina.com/zodiac/tst170localapp	1.0.1.0	f68fb9586f83	7 months ago	3.43GB
rpasmanabhan9/qiaseq-dna	pipeline	5b058ecce311	7 months ago	24.4GB
fred2/optitype	latest	a80c2891a121	10 months ago	1.78GB
wliao/cnvator	latest	7e928b3e6784	15 months ago	1.43GB
rpasmanabhan9/qiaseq-dna	latest	f9a8e6ccb1db	19 months ago	16.2GB

```
hexiaoyu@login:~/panel599_20190128$
```

## 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:

- ▶ It is recommended to 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 TruSight Tumor 170 Local App:  

```
sudo docker ps
```
- ▶ Paths in the command line must be absolute paths.
- ▶ Analysis Outputs folder is saved to the specified outputs location. 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/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/tst170localapp:<AppVersionNumber>
```

The analysis logs and Summary.tsv outputs 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 11](#).

- 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/tst170localapp:{AppVersionNumber} -
fastq
```

The analysis logs and Summary.tsv outputs 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
- ▶ Generated using only the following bcl2fastq arguments and commands:
  - ▶ 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 filenames will contain the unique sample ID from the Sample\_ID column of the SampleSheet.csv.
  - ▶ FASTQ filenames must contain the Sample ID. By default, bcl2fastq generates FASTQ files with a filename that includes sample names if they are specified in the SampleSheet.csv. When running bcl2fastq independently, leave the Sample\_Name field blank so that bcl2fastq will generate FASTQ files with the expected filename containing Sample IDs.
  - ▶ Be sure that the memory settings specified in the minimum hardware and software specifications are used when running bcl2fastq.
  - ▶ FASTQ files will be generated in the directory specified as output directory when running bcl2fastq.
    - ▶ Leave the Sample Project column of the sample sheet blank to ensure compatibility with 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'
```

- It is not required, nor recommended to 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 workflow evaluates DNA and RNA targeting variants in cancer-related genes.

## DNA Analysis

### Inputs Validation

Validates the existence and format of required analysis inputs.

### RunQC

RunQC is only used when running the TruSight Tumor 170 Local App using BCL files to generate the Run Metrics Report. This report provides run metrics and suggested values to determine if run quality results are within an acceptable range using InterOp files from the sequencing run folder. For Read 1 and Read 2, the report provides the average percentage of bases  $\geq$  Q30, which is a quality score (Q-score) measurement. A Q-score is a prediction of the probability of a wrong base call.

### FASTQ Generation

FASTQ generation is only done 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 10*.

### Down-sampling

TS0 500 里面已经去掉该过程

The TruSight Tumor 170 analysis pipeline down-samples each sample to 100 million DNA reads, where 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 DNA library.

### Alignment

The Isaac Aligner performs alignment of DNA reads to the human hg19 genome. The inputs are adapter trimmed FASTQ files and the outputs are BAM and BAM index files. COSMIC indels are provided as priors during initial alignment by Isaac.

For more information on the Isaac Aligner, see

<https://academic.oup.com/bioinformatics/article/29/16/2041/199472/Isaac-ultra-fast-whole-genome-secondary-analysis>.

### Isaac Aligner

The Isaac Aligner software uses the following steps to align DNA sequencing data with read lengths of 32–150 bp, single or paired-end, and low error rates.

- **Candidate mapping positions**—Identifies the complete set of relevant candidate mapping positions using a 32-mer seed-based search.



- ▶ **Mapping selection**—Selects the best mapping among all candidates.
- ▶ **Alignment score**—Determines alignment scores for the selected candidates based on a Bayesian model.
- ▶ **Alignment output**—Generates final output in a sorted and duplicate-marked BAM file, realigned indels, and a summary file.

## 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.

For more information, see [github.com/Illumina/Pisces/wiki](https://github.com/Illumina/Pisces/wiki).

## 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 and outputs are gVCF files.

## Copy Number Variants Calling

The CRAFT copy number variant caller performs amplification, reference, and deletion calling for target amplification genes within the assay, based on calculated fold change value. Fold change for a given gene is defined as the median normalized bin count value for the 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.

## 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.

## RNA Analysis

### Inputs Validation

Validates the existence and format of required analysis inputs.

### RunQC

RunQC is only used when running the TruSight Tumor 170 Local App using BCL files to generate the Run Metrics Report. This report provides run metrics and suggested values to determine if run quality results are

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

## FASTQ Generation

FASTQ generation is only done 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 10*.

## Down-sampling

The TruSight Tumor 170 analysis pipeline down-samples each sample to 300 million RNA reads, where 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 and gene transcripts. The alignment step identifies candidates for single-end realignment and reruns the candidates through STAR. The two STAR outputs are merged into a single BAM file per library. The inputs are trimmed FASTQ files and the outputs are BAM files.

## Fusion Calling

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

For more information on Manta, see

<http://biorxiv.org/content/biorxiv/suppl/2015/08/10/024232.DC1/024232-1.pdf>

## RNA Fusion Filtering

The RNAFusionFilter performs fusion filtering. The inputs are BAM and VCF files, and the outputs are CSV files.

## Fusion Merging

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

## 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 TSV file containing intergenic variants which are used in fusion merging.

## Annotation

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

## RNA Sample Metrics Reporting

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 TruSightTumor170\_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 describes the content of each output folder generated during analysis when using either BCL or FASTQ workflows.

### TruSightTumor170\_Analysis\_{Date}\_{Time}

- ▶ **FASTQs**—contains original sample FASTQ files
- ▶ **DNA\_IntermediateFiles**
  - ▶ **Alignment**
    - ▶ Sample specific BAM and BAI files (iSAAC outputs)
    - ▶ Sample specific realigned BAM and BAI files (Hygea outputs)
    - ▶ DNA probe BED file (for Picard)
    - ▶ DNA target BED file (for Picard)
  - ▶ **VariantCalling**—DNA interval file (consumed by Pisces)
  - ▶ **Annotation**—contains sample specific JSON files
  - ▶ **CNV**
    - ▶ **Fold Change File**—tab-delimited file containing the gene name, fold change, and Q-score for amplification for each of the 59 amplification genes targeted by the TruSight Tumor 170 panel.
    - ▶ **Normalized Bin Count File**—tab-delimited file containing the normalized bin count for each genomic region of interest used for amplification calling, and metadata for each region.
    - ▶ **Raw Bin Count File**—tab-delimited file containing the raw bin count for each genomic region of interest used for amplification calling, and metadata for each region.
- ▶ **DNA\_{Sample\_ID}** (per sample)
  - ▶ {Sample ID}\_CopyNumberVariants.vcf
  - ▶ {Sample ID}\_SmallVariants.genome.vcf
- ▶ **DNA\_SampleMetricsReport.txt** see *DNA Sample Metrics Report* on page 20
- ▶ **Logs**
  - ▶ AnalysisLog-YYYYMMDD-HHMMSS .txt
  - ▶ **FASTQ\_Logs**—contains the following:
    - ▶ Sample specific FASTQ logs
    - ▶ FASTQ component logs
    - ▶ Intermediate SampleSheet.csv (For BCL only inputs)
  - ▶ **DNA\_Logs**—contains component logs.
  - ▶ **RNA\_Logs**—contains component logs.
- ▶ **RNA\_IntermediateFiles**
  - ▶ **Alignment**

- ▶ Sample specific second pass BAM and BAI files (Chimera Realignment outputs)
- ▶ RNA probe BED file (for Picard)
- ▶ RNA target BED file (for Picard)
- ▶ **Annotation** — Contains sample specific JSON files
- ▶ **RNA\_{Sample\_ID}** (per sample)
  - ▶ {Sample ID}\_Fusions.csv
  - ▶ {Sample ID}\_SpliceVariants.vcf
  - ▶ {Sample ID}\_HighConfidenceVariants.csv
  - ▶ {Sample ID}\_PublishedFusions.csv
- ▶ **RNA\_SampleMetricsReport.txt**—see *RNA Sample Metrics Report* on page 23
- ▶ **RunMetricsReport.txt**—see *Run Metrics Report* on page 18
- ▶ **SampleSheet.csv**—see *SampleSheet.csv* on page 19
- ▶ **Summary.tsv**—see *Summary.tsv* on page 18

## Files

This section describes the summary output files generated during analysis.

### Run Metrics Report

File name: RunMetricsReport.txt

This report contains sequencing run metrics and threshold guidance. The RunMetricsReport.txt is only generated when inputs are BCLs from a sequencing run folder.

Run Metric	Description	Ex. Calculated Value	Threshold
Reads PF (%):	Percentage of clusters passing filter	93.0	≥80.0
Q30+(R1): Q30 score for Read 1	Percentage of bases with a quality score of 30 or higher for Read 1	93.2	≥80.0
Q30+(R2): Q30 score for Read 2	Percentage of bases with a quality score of 30 or higher for Read 2	87.5	≥80.0

### Summary.tsv

File name: Summary.tsv

A tab-delimited summary report is generated for each analysis showing the analysis status for each sample analyzed. The Summary column in the output displays Analysis Completed when the sample completes analysis. When there is an error, the Summary column displays the analysis step(s) that failed or were skipped.

## 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 file includes a list of samples and the index sequences, as well as the workflow to be employed for each sample in the analysis. For information on populating the sample sheet, see [Workflow Requirements on page 1](#).

## DNA Outputs

### Files

This section describes the output files generated during a DNA sample analysis.

#### 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 Variant Quality	Filtered variant 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 $\geq$ to 0.5 (applies to variant allele calls only)
Low Variant Frequency	Frequency is < 0.026 (applies to variant allele calls only)
Indel Reference Repeat	Repeat is $\geq$ to 8 (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)

If available, variant calls are annotated using the Illumina Annotation Engine with the following information:

- ▶ HGNC Gene
- ▶ Transcript (using the canonical Ensembl transcript)
- ▶ Exon (using the canonical Ensembl transcript)
- ▶ Consequence (using the canonical Ensembl transcript)
- ▶ HGVS Coding Sequence Name (using the canonical Ensembl transcript)
- ▶ HGVS Protein Sequence Name (using the canonical Ensembl transcript)
- ▶ COSMIC ID

#### 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. The CNV call indicates fold change results for each gene classified as reference, deletion, or amplification. The value in the QUAL column of the VCF is a Phred transformation of the p-value where  $Q = -10 \times \log_{10}(p\text{-value})$ . 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, and <DEL> indicates the detected fold change (FC) is less than a predefined deletion cutoff for that gene. This cutoff can vary from gene to gene.

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

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

Given tumor purity X%, for a reported fold change Y, the copy number  $n$  can be calculated by:  $n = [(200 * Y) - 2 * (100 - X)] / X$

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

## DNA Sample Metrics Report

File name: DNA\_SampleMetricsReport.txt

Review this report to assess DNA sample data quality using the following metrics.

Metric	Description	Passing Quality Threshold	Indication When Failed
Coverage_MAD	Median of absolute deviations from the median of the normalized count of each CNV target region	≤0.2	Less confidence in Copy Number Variant Calling
Median_BinCount_CNV_Targets	Median of the raw read counts (normalized by region size) of each CNV target region	≥1	Less confidence in Copy Number Variant Calling
Median_Insert_Size	Median of fragment sizes calculated from read alignments	≥79	Less confidence in Small Variant Calling
PCT_ExonBases_100X	Percentage of exonic bases in gVCF that have greater than or equal to 100x coverage	≥0.95	Less confidence in Small Variant Calling
PCT_Chimeric_Reads	Percentage of unique reads that are part of a chimeric alignment (ie, one half of the read maps to one location and the other half maps to a nonsequential position) out of total mapped unique reads	N/A	Not used to evaluate sample data quality

## RNA Outputs

### Files

This section describes the top level output files generated during an RNA sample analysis.

### Fusions CSV

File name: {Sample\_ID}\_Fusions.csv

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

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 the filters described below are triggered, the Filter column will display 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 non-uniquely 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. (Confidence filter)
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 1kb if unannotated). (Confidence filter)
LowQ	Fusion supporting reads (unique) < 5 (+ 1 for every 10 million reads over 16 million reads). (Confidence filter)
LowDupReads	Fusion supporting reads (duplicate) < 5. (Confidence filter)
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  $\geq 0.45$  and none of the confidence filters are triggered.

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



#### NOTE

When using Microsoft Excel to view this report, genes that can be converted to dates (such as MARCH1) are automatically converted 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 may 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 $\geq$ 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 = 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, gene 1 reference reads, gene 2 reference reads, score.

- ▶ 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.
- ▶ Fusions with overlapping genes are denoted by a semicolon. 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, Reference Reads, and score.
- ▶ The splice variants are sorted by, in ascending order, Gene, then by Transcript, then by Breakpoint Start, and then by Breakpoint End.

## Published Fusions

File name: {SampleID}\_PublishedFusions.csv

Contains a list of published fusions from the Mitelman Reference Database (“Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2016). Mitelman F and Johansson B and Mertens F (Eds.) <http://cgap.nci.nih.gov/Chromosomes/Mitelman>”) that are targeted by the TruSight Tumor 170 panel.

- ▶ When Observed=True, the fusion was called either with high or low confidence in the sample.
- ▶ When Observed=False the fusion was not detected in the sample.
- ▶ The Published Fusion report shows Mitelman fusions that have at least one gene targeted by the TruSight Tumor 170 panel in the gene pair.
- ▶ The gene fusion column does not indicate the directionality of the fusion. For any fusion of interest, search for gene1/gene2 and gene2/gene1.

For each fusion, the following information is captured:

- ▶ Mitelman ID(s) - It is possible for a fusion to have multiple Mitelman IDs because they can have various clinical associations or interpretations

- ▶ Breakpoint 1: Gene A breakpoint from the {Sample\_ID}\_Fusions.csv
- ▶ Breakpoint 2: Gene B breakpoint from the {Sample\_ID}\_Fusions.csv
- ▶ Fusion supporting reads: Alt Split + Alt Pair values from the {Sample\_ID}\_Fusions.csv
- ▶ Gene 1 Reference Reads: Ref A Split + Ref A Pair values from the {Sample\_ID}\_Fusions.csv
- ▶ Gene 2 Reference Reads: Ref B Split + Ref B Pair values from the {Sample\_ID}\_Fusions.csv
- ▶ Score: Score value from the {Sample\_ID}\_Fusions.csv
- ▶ Filter: Pass if Keep Fusion=True else Filter = Filter column values from the {Sample\_ID}\_Fusions.csv

The fusions are sorted by Gene 1, then by Gene 2, then by Breakpoint 1, and then by Breakpoint 2.

## RNA Sample Metrics Report

File name: RNA\_SampleMetricsReport.txt

Review this report to assess RNA sample data quality by evaluating the following metrics.

Metric	Description	Output File Evaluated	Quality Threshold
Median_Insert_Size	The median of fragment sizes calculated from read alignments	Fusions.csv, HighConfidenceVariants.csv, SpliceVariants.vcf, PublishedFusions.csv	≥63
Median_CV_Coverage_1000x	The median of the CV value of target region coverage across all genes with coverage above 1000	Fusions.csv, HighConfidenceVariants.csv, SpliceVariants.vcf, PublishedFusions.csv	≤0.88
PCT_Chimeric_Reads	Percentage of reads that are part of a chimeric alignment (ie, one half of the read maps to one location and the other half maps to a nonsequential position) out of total mapped reads	Not used to evaluate sample data quality	N/A
PCT_On_Target_Reads	Percentage of reads that map to target regions out of total PF reads	Not used to evaluate sample data quality	N/A

## Generate Picard Metrics

TruSight Tumor 170 Local App outputs can be used to generate Picard metrics.

For a description of Picard Metrics generated by the Picard HsMetrics tool, see [broadinstitute.github.io/picard/picard-metric-definitions.html#HsMetrics](https://broadinstitute.github.io/picard/picard-metric-definitions.html#HsMetrics).



### NOTE

There is no correlation between the calculated Picard metrics and the quality of the results generated from TruSight Tumor 170 Local App. These metrics are intended for information purposes only.

## Install Picard and Generate Files

Use the following procedure to install supporting environments and generate files for running Picard CollectHsMetrics.



### NOTE

Memory settings must be set to -Xmx2g to boost Java Virtual Machine (JVM) memory to accommodate large files when running Picard tools.

For more information on installing and setting up the environment for Picard see, [broadinstitute.github.io/picard/](https://broadinstitute.github.io/picard/)

- 1 Download and install Picard and supporting environments using the following command line:  
#Linux command to download Picard v2.9.4 executable  

```
wget https://github.com/broadinstitute/picard/releases/download/2.9.4/picard.jar
```
- 2 Generate sequence dictionary file using hg19 genome.fa reference genome sequence using Picard CreateSequenceDictionary tool, where:
  - ▶ R= Absolute path to reference genome sequence file. The hg19 genome.fa file is required.
  - ▶ O= Path to output sequence dictionary file. A file name with a .dict extension can be specified. By default the base name of the input reference with .dict extension is used.For example  
#Generate sequence dictionary for a reference sequence  

```
java -Xmx2g -jar picard.jar CreateSequenceDictionary R=genomes/Homo_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=hg19.dict
```

  - ▶ For information on the Picard CreateSequenceDictionary tool, see [broadinstitute.github.io/picard/command-line-overview.html#CreateSequenceDictionary](https://broadinstitute.github.io/picard/command-line-overview.html#CreateSequenceDictionary)
  - ▶ The hg19 genome.fa reference genome sequence file is contained within TruSight\_Tumor\_170\_Resources\_{AppVersionNumber}.tgz
    - ▶ When extracted, the hg19 genome.fa file is located in the genomes folder: genomes/Homo\_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
    - ▶ For information on extracting files to the genomes folder, see *Install TruSight Tumor 170 Local App* on page 9.
- 3 Retrieve the absolute directory path to the Target and Probe BED Files for use with the INPUT variable in step 4.
  - ▶ TruSight Tumor 170 Target BED Files
    - ▶ For DNA located in DNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name DNA\_PicardTarget.bed
    - ▶ For RNA located in RNA\_IntermediateFiles/Alignment under the Analysis Output folder with file name RNA\_PicardTarget.bed.

- ▶ TruSight Tumor 170 Probe BED Files
  - ▶ For DNA located in DNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name DNA\_PicardProbe.bed
  - ▶ For RNA located in RNA\_IntermediateFiles/Alignment under the Analysis Output folder with file name RNA\_PicardProbe.bed.
- 4 Run Picard BedToIntervalList tool to generate Interval files compatible with Picard CollectHsMetrics for generating Picard metrics, where:
  - ▶ INPUT: Absolute path to Target or Probe BED files. See step 3.
  - ▶ SEQUENCE\_DICTIONARY: Absolute path to the sequence dictionary file generated in step 2.
  - ▶ OUTPUT: Path to generated output interval files.

For example

```
#Generate DNA Target Intervals File
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=DNA_PicardTarget.bed
SEQUENCE_DICTIONARY=hg19.dict OUTPUT=DNA_PicardTarget.intervals
#Generate DNA Probe Intervals File
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=DNA_PicardProbe.bed
SEQUENCE_DICTIONARY=hg19.dict OUTPUT=DNA_PicardProbe.intervals
#Generate RNA Target Intervals File
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=RNA_PicardTarget.bed
SEQUENCE_DICTIONARY=hg19.dict OUTPUT=RNA_PicardTarget.intervals
#Generate RNA Probe Intervals File
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=RNA_PicardProbe.bed
SEQUENCE_DICTIONARY=hg19.dict OUTPUT=RNA_PicardTarget.intervals
```

## Run Picard CollectHsMetrics

Picard metrics can be generated using individual command lines or by running a bash script.

### Command Line

Run Picard CollectHsMetrics using individual command lines to generate metrics.

- 1 Retrieve the absolute directory path for sample BAM files from TruSight Tumor 170 Assay App output for use in step 4.
  - ▶ For DNA samples, the Isaac aligned BAMs are located in the DNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name {SampleID}.bam. Do not use realigned BAMs.
  - ▶ For RNA samples, the STAR aligned BAMs are located in the RNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name {SampleID}.bam.
- 2 Retrieve the absolute directory path to hg19 genome.fa file from the genomes folder for use in step 4.
  - ▶ The hg19 genome.fa file is located in the genomes folder: genomes/Homo\_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
- 3 Retrieve absolute output paths to interval files generated in step 4 *Install Picard and Generate Files on page 24*.
- 4 Run Picard CollectHsMetrics using individual command lines, where:
  - ▶ {Picard CollectHsMetrics Command} (eg CollectHsMetrics for Picard v2.9.4): Picard tool that calculates a set of specific metrics from a BAM file.
  - ▶ BI (or BAIT\_INTERVALS): Absolute path to probe interval file that contains information about the probes. See step 3.

- ▶ TI (or TARGET\_INTERVALS): Absolute path to target interval file that contains information about the target regions. See step 3.
- ▶ R: Absolute path to reference genome sequence file. The hg19 genome.fa file is required. See step 2.
- ▶ I: Absolute path to sample BAM files. See step 1.
- ▶ O: Path to output file for writing metrics. Users can specify name of the file.
- ▶ VALIDATION\_STRINGENCY=SILENT: Flag to set the validation stringency to SILENT. This flag can improve performance when processing a BAM file in which variable-length data (read, qualities, tags) do not otherwise need to be decoded.

#Example command line below For DNA analyses using Picard v2.9.4:

```
java -Xmx2g -jar picard.jar CollectHsMetrics BI=DNA_PicardProbe.intervals
    TI=DNA_PicardTarget.intervals R=genomes/Homo_
    sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=DNA-Picard-
    HsMetrics.txt I=DNA_IntermediateFiles/Alignment/dnasample.bam
    VALIDATION_STRINGENCY=SILENT
```

#Example command line below For RNA analyses using Picard v2.9.4:

```
java -Xmx2g -jar picard.jar CollectHsMetrics BI=RNA_PicardProbe.intervals
    TI=RNA_PicardTarget.intervals R=genomes/Homo_
    sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=RNA-Picard-
    HsMetrics.txt I=RNA_IntermediateFiles/Alignment/rnasample.bam
    VALIDATION_STRINGENCY=SILENT
```

## Bash Script

Run Picard CollectHsMetrics using a bash script to define paths to the required files to generate metrics in a single execution.

- 1 Retrieve the absolute directory path for sample BAM files from TruSight Tumor 170 Assay App output for use in step 4.
  - ▶ For DNA samples, the Isaac aligned BAMs are located in the DNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name {SampleID}.bam. Do not use realigned BAMs.
  - ▶ For RNA samples, the STAR aligned BAMs are located in the RNA\_IntermediateFiles/Alignment in the Analysis Output folder with file name {SampleID}.bam.
- 2 Retrieve the absolute directory path to hg19 genome.fa file from the genomes folder for use in step 4.
  - ▶ The hg19 genome.fa file is located in the genomes folder: genomes/Homo\_
 sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
- 3 Retrieve absolute output paths to interval files generated in step 4 *Install Picard and Generate Files on page 24*.
- 4 Use the following template to create the bash script. See steps 1, 2, and 3 for details on the variables needed for the script.
 

```
#!/bin/bash
#Path to location of picard.jar file executable
PICARDPATH=picard.jar
#Path to genomes folder
GENOMESPATH=genomes/Homo_
    sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
#Path to DNA Probe Interval Files
DNAPROBEINVERVALS=DNA_PicardProbe.intervals
#Path to DNA Target Interval
```

```

DNATARGETINTERVALS=DNA_TargetProbe.intervals File
#Path to BAM file
DNABAMPATH=AnalysisOutput/DNA_IntermediateFiles/Alignment/dnasample.bam
#Path to RNA Probe Interval File
RNAPROBEINTERVALS=RNA_PicardProbe.intervals
#Path to RNA Target Interval
RNATARGETINTERVALS=RNA_TargetProbe.intervals File
#Path to bam file
RNABAMPATH=AnalysisOutput/RNA_IntermediateFiles/Alignment/rnasample.bam

java -Xmx2g -jar $PICARDPATH CollectHsMetrics BI=$DNAPROBEINTERVALS
    TI=$DNATARGETINTERVALS R=$GENOMESPATH O=DNA-Picard-HsMetrics.txt
    I=$DNABAMPATH VALIDATION_STRINGENCY=SILENT

java -Xmx2g -jar $PICARDPATH CollectHsMetrics BI=$RNAPROBEINTERVALS
    TI=$RNATARGETINTERVALS R=$GENOMESPATH O=RNA-Picard-HsMetrics.txt
    I=$RNABAMPATH VALIDATION_STRINGENCY=SILENT

```

## 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.

# Technical Assistance

For technical assistance, contact Illumina Technical Support.

Website: [www.illumina.com](http://www.illumina.com)  
Email: [techsupport@illumina.com](mailto: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	<b>400.635.9898</b>	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

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

**Product documentation**—Available for download in PDF from the Illumina website. Go to [support.illumina.com](http://support.illumina.com), select a product, then select **Documentation & Literature**.



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](mailto:techsupport@illumina.com)

[www.illumina.com](http://www.illumina.com)

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

© 2017 Illumina, Inc. All rights reserved.

**illumina®**