

Local Run Manager TruSight Tumor 15 Analysis Module v2.1

Workflow 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).

© 2021 Illumina, Inc. All rights reserved.

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

Revision History

Document	Date	Description of Change
Document # 1000000006976 v03	June 2021	Updated run name and description options to include special characters. Remove single mismatch in Demultiplexing
Document # 1000000006976 v02	March 2018	Updated Input Requirements to included Illumina Annotation Engine Added Install Illumina Annotation Engine topic. Updated Install Reference Genome topic to include requirement to install for when using Local Run Manager off-instrument. Updated FASTQ File Generation topic to remove outdated information. Updated Alignment topic to be version agnostic. Updated Low Coverage Amplicon table to include descriptions. Added Amplification Calling topic. Added Amplification VCF file topic. Updated Sample Metrics Report topic to notate that MixABases and MixBBases coverage refers to base coverage using collapsed reads.

Document	Date	Description of Change
Document # 1000000006976 v01	September 2016	<p>Updated the merged gVCF file format from *.vcf to *.sample.genome.vcf.</p> <p>Updated the location of the Analysis folder in the Folder Structure section.</p> <p>Added information that the Filtered Outputs Report is also available as a *.pdf and *.txt file, and removed the reference to an aggregate report in the Analysis folder.</p> <p>Updated the internal annotation algorithm from MARS to the Illumina Annotation Engine.</p> <p>Changed references to the Filtered gVCF File Report to the Filtered Outputs Report because the filtered report is not a gVCF file.</p> <p>Updated the description of *.ant files to indicate that they are for import to VariantStudio software.</p> <p>Updated the Variant Calling section, adding the use of a coordinate lookup table, and removing information on variant presence in both libraries, cumulative depth, and variant frequency in the merged VCF file.</p> <p>Removed the Read Stitching section.</p>
Document # 1000000006976 v00	January 2016	Initial release.

Table of Contents

Revision History	iii
Overview	1
Input Requirements	1
About This Guide	1
Install the TruSight Tumor 15 Analysis Module	1
Install the Reference Genome	2
Install the Illumina Annotation Engine	2
Installation Folder Structure	3
Set Parameters	3
Specify Module-Specific Settings	3
Specify Samples for the Run	4
Analysis Methods	5
Demultiplexing	5
FASTQ File Generation	5
Alignment	5
Read Stitching	6
Variant Calling	7
Illumina Annotation Engine	7
Amplification Calling	8
View Analysis Results	8
Analysis Report	9
Sample Information	9
Low Coverage Amplicons	10
Detected Variants	10
Tested Variant Sites with Flagged QC Metric(s)	10
Analysis Output Files	11
Demultiplexing File Format	12
FASTQ File Format	13
BAM File Format	13
VCF File Format	14
Per-Library and Merged gVCF Files	16

Run Metrics Report	17
Sample Metrics Report	18
Filtered Output Reports	18
Amplicon Coverage File	18
Supplementary Output Files	19
Analysis Folder	19
Technical Assistance	21

Overview

The Local Run Manager TruSight® Tumor 15 analysis module aligns reads against the reference specified in the manifest files using the banded Smith-Waterman algorithm. After alignment, the Somatic Variant Caller performs variant analysis. The results report somatic variants of a set of reference panel genes associated with cancer. The Local Run Manager TruSight Tumor 15 analysis module also reports amplifications for three genes: ERBB2, EGFR, and MET. This workflow is designed specifically for TruSight Tumor 15 libraries.

Input Requirements

In addition to sequencing data files generated during the sequencing run, such as base call files, the TruSight Tumor 15 analysis module requires the following files.

- **Manifest files (2)**—The TruSight Tumor 15 analysis module requires two assay-specific manifest files: a manifest file for mix A and a manifest file for mix B. Manifest files are included as part of in the analysis module.
- **Illumina Annotation Engine**—The TruSight Tumor 15 analysis module uses the Illumina Annotation Engine to annotate the variants listed in the VCF files. The Illumina Annotation Engine is included in the Local Run Manager software installation on-instrument, but needs to be installed separately when running the analysis module with Local Run Manager off-instrument.
- **Reference genome**—The TruSight Tumor 15 analysis module requires the hg19 reference genome for coordinates and chromosome mapping. The reference genome is included in the Local Run Manager software installation on-instrument, but needs to be installed separately when running the analysis module with Local Run Manager off-instrument.

About This Guide

This guide provides instructions for setting up run parameters for sequencing, and analysis parameters for the TruSight Tumor 15 analysis module. For an overview and instructions for using Local Analysis Software, see the *Local Run Manager Software Guide (document # 1000000002702)*.

Install the TruSight Tumor 15 Analysis Module

When installing a new version of the module, you do not need to uninstall the previous version.

1. Right-click TruSightTumor15.Module.Installer.exe and select **Run as administrator**.
2. Select **Install**.
3. Log in using the user name and password for the admin account in Local Run Manager. The default admin credentials are:
 - User name: admin
 - Password: passwordOnly admins can install analysis modules.
4. Select **Next**.
5. When installation is complete, select **Finish**.
The module installs some dependencies.
6. When complete, select **Close**.

Install the Reference Genome

If you are using the Local Run Manager off-instrument, the Reference Genome must be installed.

1. Unzip the Reference Genome installer.
2. Double-click ReferenceGenomeInstaller.msi.
3. Select **Next**.
4. When complete, select **Close**.

Install the Illumina Annotation Engine

The Illumina Annotation Engine must be installed when running the module with Local Run Manager off-instrument.

1. Unzip the Illumina Annotation Engine installer.
2. Double-click IlluminaAnnotationEngineInstaller.msi.
3. Click **Next**.
4. When complete, click **Close**.

Installation Folder Structure

By default, installation folders are located in C:\Illumina.

- 📁 **Illumina**
 - 📁 **Genomes**
 - 📁 **Local Run Manager**
 - 📁 **Local Run Manager Analysis Job Service**
 - 📁 **Local Run Manager Service**
 - 📁 **RTA**
 - 📁 **Run Copy Service**

The TruSight Tumor 15 Analysis Module and Illumina Annotation Engine (IAE) files are located under C:\Illumina\Local Run Manager\Modules\TruSightTumor15.

The TruSight Tumor 15 reference genome files are located under C:\Illumina\Genomes\Homo_sapiens\UCSC\hg19.

Set Parameters

1. Select **Create Run**, and select an analysis module.
2. Enter a run name that identifies the run from sequencing through analysis. Only use alphanumeric characters, spaces, underscores, or dashes.
3. [Optional] Enter a run description to further identify the run. Use alphanumeric characters and spaces.

Specify Module-Specific Settings

1. Click **Module-Specific Settings** and enter a maximum indel size. The default is 25 bp.
 - **Max Indel Size** —The maximum indel size must be an integer value from 25 bp to 70 bp. For more information on indel size, see the [Indel size is defined within the context of an individual amplicon. For example, a 16 bp insertion, followed by a 10 bp insertion in the same read is considered 26 bp in the maximum index size parameter. By default, the TruSight Tumor 15 analysis module is configured to report indels with a maximum size of 25 bp. You can optionally configure the maximum index size to report larger detected indels. The maximum index size can be set from 25 bp to 70 bp. on page 6.](#)

Specify Samples for the Run

Specify samples for the run using the following options:

- **Enter samples manually**—Use the blank table at the bottom of the Create Run screen.
- **Import sample sheet**—Navigate to an external file in a comma-separated values (*.csv) format. A template is available for download on the Create Run screen.

After you have populated the samples table, you can export the sample information to an external file. You can use this file as a reference when preparing libraries or import the file when configuring another run.

Enter Samples Manually

1. Click **Add Row** to adjust the samples table to an appropriate number of rows.
2. Enter a unique sample ID in the Sample ID field.
Use alphanumeric characters, dashes, or underscores.
3. [Optional] Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, underscores, or spaces.
4. Enter index adapters for Mix A as follows.
 - a. Select an Index 1 adapter from the Mix A Index 1 drop-down list.
 - b. Select an Index 2 adapter from the Mix A Index 2 drop-down list. Right-click in a table cell to use the Fill Down command.
 - c. [Optional] Enter a description for Mix A.
5. Enter index adapters for Mix B as follows.
 - a. Select an Index 1 adapter from the Mix B Index 1 drop-down list.
 - b. Select an Index 2 adapter from the Mix B Index 2 drop-down list. Right-click in a table cell to use the Fill Down command.
 - c. [Optional] Enter a description for Mix B.

 | The Report Definition field is populated automatically, by default.

6. [Optional] Click the **Export Samples** to export sample information to an external file.
7. Click **Save Run**.

Import Samples

1. Select **Template**. Make sure that the template file contains the correct column headings for import.
2. Enter the sample information in each column for the samples in the run, and then save the file.
3. Select **Import Samples** and browse to the location of the sample information file.
4. When finished, select **Save Run**.

Analysis Methods

The TruSight Tumor 15 analysis module performs the following analysis steps and then writes analysis output files to the Analysis folder.

- Demultiplexes index reads
- Generates FASTQ files
- Aligns to a reference
- Performs read stitching
- Identifies variants
- Runs the Illumina Annotation Engine
- Detects amplifications

Demultiplexing

Demultiplexing compares each Index Read sequence to the index sequences specified for the run. No quality values are considered in this step.

Index reads are identified using the following steps:

- Samples are numbered starting from 1 based on the order they are listed for the run.
- Sample number 0 is reserved for clusters that were not assigned to a sample.
- Clusters are assigned to a sample when the index sequence matches exactly.

FASTQ File Generation

After demultiplexing, the software generates intermediate analysis files in the FASTQ format, which is a text format used to represent sequences. FASTQ files 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 1 sample, and the name of that sample is included in the FASTQ file name. In the TruSight Tumor 15 workflow, 4 FASTQ files are generated per sample: 2 from the Mix A library (Read 1 and 2) and 2 from the Mix B library (Read 1 and 2). FASTQ files are the primary input for alignment.

Alignment

During the alignment step, the banded Smith-Waterman algorithm aligns clusters from each sample against amplicon sequences specified in the manifest file.

The banded Smith-Waterman algorithm performs local sequence alignments to determine similar regions between two sequences. Instead of comparing the total sequence, the Smith-Waterman algorithm compares segments of all possible lengths. Local alignments are useful for dissimilar sequences that are suspected to contain regions of similarity within the larger sequence. This process allows alignment across small amplicon targets, often less than 10 bp.

This assay sequences DNA using paired-end reads. Coverage is reported in two forms: per-amplicon and per-base. For per-amplicon coverage, 'read' coverage is reported. This form causes each read in a pair to be counted separately, and the total number of reads that are assigned to the amplicon are reported. For per-base coverage, 'collapsed read' or 'fragment' coverage is reported, where the reads in a pair are combined where they overlap. This type of read causes coverage to be counted only one time per pair. The information provides a more robust view of the data, as it allows combining two observations of the same underlying data (ie, a molecular fragment).

Each paired-end read is evaluated in terms of its alignment to the relevant probe sequences for that read.

- Read 1 is evaluated against the reverse complement of the Downstream Locus-Specific Oligos (DLSO).
- Read 2 is evaluated against the Upstream Locus-Specific Oligos (ULSO).
- If the start of a read matches a probe sequence with no more than one mismatch, the full length of the read is aligned against the amplicon target for that sequence.

Alignments that include more than three indels are filtered from alignment results. Filtered alignments are written in alignment files as unaligned and are not used in variant calling.

i | *Indel size* is defined within the context of an individual amplicon. For example, a 16 bp insertion, followed by a 10 bp insertion in the same read is considered 26 bp in the maximum index size parameter. By default, the TruSight Tumor 15 analysis module is configured to report indels with a maximum size of 25 bp. You can optionally configure the maximum index size to report larger detected indels. The maximum index size can be set from 25 bp to 70 bp.

Due to the way genes are tiled by amplicons, indels larger than 25 bp may not be detected as they may overlap amplicon boundaries.

To configure the maximum index size parameters, see [Specify Module-Specific Settings on page 3](#).

Read Stitching

A stitched read is a single read that has been bioinformatically stitched from a pair of reads. The ReadStitcher takes paired reads from a BAM, stitches them, and writes out a new BAM. Reads that cannot be stitched are left unchanged.

The stitched read contains the following:

- one consensus sequence of base calls.

- one consensus sequence of base call qualities.
- one consensus cigar string.
- one direction tag.

The direction tag gives the original directions of the contributing reads.

To save on future processing time, the ReadStitcher filters out reads that are not used downstream, such as PCR duplicates or poorly mapped reads. The pipeline performs sorting and indexing of output stitched BAM.

Variant Calling

Developed by Illumina, the Somatic Variant Caller identifies variants present at low frequency in the DNA sample.

The somatic variant caller identifies SNPs in three steps:

- Considers each position in the reference genome separately
- Counts bases at the given position for aligned reads that overlap the position
- Computes a variant score that measures the quality of the call using Poisson model.

Variants are first called for each separately. Then, variants from each are combined into a single output file. If a variant meets the criteria, the variant is marked as PASS in the variant call (VCF).

Illumina Annotation Engine

Illumina Annotation Engine (IAE) annotates variants, populating the following INFO fields in the VCF file.

Column	Description
Transcript	Ensembl variant Transcript ID.
Consequence	Variant consequence value from the Sequence Ontology Project.
Canonical	True when the variant transcript is canonical.
HDVSc	Human Genome Variation Society (HGVS) coding nomenclature.
HGVsp	Human Genome Variation Society (HGVS) protein nomenclature.
HGNC	Human Gene Nomenclature Committee (HGNC) gene symbol.
Exons	The exons affected by the variant. For example, 4-6/7 indicates that exons 4, 5, and 6 were affected and that this transcript contains 7 exons in total.
COSMIC ID	Variant ID in Catalogue of Somatic Mutations in Cancer (COSMIC) database.

Amplification Calling

TruSight Tumor 15 incorporates an algorithm to call CNVs that is designed for tumor samples without matched normals. It can detect amplifications in three genes: EGFR, ERBB2, and MET above a 1.6 fold change.

Amplification output for TruSight Tumor 15 is reported in the amplifications VCF file. Amplifications are reported using the semiquantitative AmpScore metric. A higher AmpScore reflects both a higher confidence in the likelihood of an amplification and a higher expected fold change. Each amplification gene is reported in one of the following four groups, with thresholds as shown in [Table 1](#). For more information on the amplification VCF file, see [Amplification VCF File on page 17](#).

- Samples with AmpScore ≥ 100 are considered to be amplified, and are reported as ALT=<DUP>, FILTER=PASS (ie, $X \geq \text{Amplification Threshold}$)
- Samples with borderline AmpScores are noted in the VCF as ALT=<DUP> FILTER=NoCall calls. These samples are not determined to be amplified or normal. (ie, $\text{Amplification Threshold} > X \geq \text{Positive Threshold}$)
- Samples with normal read depth are called as ALT=., FILTER=PASS (ie, $\text{Positive Threshold} > X \geq \text{Negative Threshold}$)
- Samples with low AmpScore are called as ALT=, FILTER=NoCall. These samples are suspected deletions. (ie, $\text{Negative Threshold} \geq X$)

i | A gene amplification in a sample with an AmpScore much greater than 100 can cause secondary amplifications to not be detected, or cause normally present genes to be called low confidence deletions.

Table 1 Amplification Thresholds

Gene	Amplification Threshold	Positive Threshold	Negative Threshold
EGFR	100	91.51	65.40
ERBB2	100	93.40	81.13
MET	100	97.17	72.64

View Analysis Results

1. From the Local Analysis Software dashboard, select the run name.
2. From the Run Overview tab, review the sequencing run metrics.
3. To change the analysis data file location for future requeues of the selected run, select the **Edit** icon, and edit the output run folder file path.

The file path leading up to the output run folder is editable. The output run folder name cannot be changed.

4. [Optional] Select the **Copy to Clipboard** icon to copy the output run folder file path.
5. Select the Sequencing Information tab to review run parameters and consumables information.
6. Select the Samples & Results tab to view the analysis report.
 - If analysis was requested, select the appropriate analysis from the Select Analysis drop-down list.
7. [Optional] Select the **Copy to Clipboard** icon to copy the Analysis Folder file path.

Analysis Report

Analysis results are summarized on the Samples & Results tab. The report displayed in the Samples & Results tab is a combination of the Filtered Variant Summary, Amplicon Coverage Report, and Sample and Run Metrics Reports for each sample analyzed.

- The Filtered Variant Summary is also available in a PDF, *.vcf, and *.txt file format for each sample.
- The Amplicon Coverage Report is also available in a PDF file format for each sample.
- A table summarizing Run and Sample Info and Metrics is included at the bottom of each PDF report and the Samples & Results tab.
- The PDF file for the Amplicon Coverage Report contains a Low Coverage Amplicons table.
 - The Low Coverage Amplicons table is not available on the Samples & Results tab.

For more information on sample metrics and run details, see the [Sample Metrics Report on page 18](#) and the [Run Metrics Report on page 17](#).

Sample Information

Table 2 Results Summary Header Table

Column Heading	Description
Sample ID	The sample ID provided when the run was created.
Run Name	The run name provided when the run was created.

Low Coverage Amplicons

Table 3 Low Coverage Amplicons Table*

Column Heading	Description
Gene	The gene covered by the low coverage amplicon.
Amplicon Name	Amplicon Target ID as defined in the manifest file.
Chromosome	The chromosome covered by the amplicon.
Coordinate Start	The start position of the amplicon as defined in the manifest file.
Coordinate End	The end position of the amplicon as defined in the manifest file.
Coverage	Per-amplicon read coverage: total number of reads assigned to the amplicon.

*Only available in the PDF file.

Detected Variants

Table 4 Detected Variants Table

Column Heading	Description
Gene	The gene where the SNV, insertion, or deletion is detected.
Variant of Interest	Human Genome Variation Society (HGVS) protein notation.
Coordinate	The chromosome and coordinate where the detected variant is located.
Nucleotide Change	HGVS nucleotide notation.
Consequence	Consequence on protein function.
Frequency	Fraction of reads in which the variant is detected.
Coverage	Collapsed read count for the detected variant.

Tested Variant Sites with Flagged QC Metric(s)

Table 5 Tested Variant Sites with Flagged QC Metric(s) Table

Column Heading	Description
Gene	The gene where the no call is located.
Coordinate	The chromosome and coordinate where the detected variant is located.

Column Heading	Description
QC Metric(s) Flagged	<p>The reason for the no call.</p> <p>Low Variant Frequency—The variant frequency is between 1% and 2.6%. Identical to LowVariantFreq in the VCF File Filter entry.</p> <p>Low Coverage—The depth of coverage is below 250. Identical to LowDP in the VCF File Filter entry.</p> <p>Low Genotype Quality—The genotyping quality is <q30. Identical to q30 in the VCF File Filter entry.</p> <p>Indel Reference Repeat—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8. Identical to R8 in the VCF File Filter entry.</p> <p>Strand Bias—The strand bias is more than the given threshold. Identical to SB in the VCF File Filter entry.</p>

Analysis Output Files

The following analysis output files are generated for the TruSight Tumor 15 analysis module and provide analysis results for alignment and variant calling. Analysis output files are located in the Analysis folder or one of its subfolders.

File Name	Description
FASTQ (*.fastq.gz)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step. FASTQ files are located in the FastQ subfolder.
Alignment files in the BAM format (*.bam)	Contains aligned reads for a given library. BAM files are located in the Libraries subfolder.
Per-library variant call files in the VCF format (*.vcf)	Contains information about variants found at specific positions in a reference genome. Per-library variant call files are located in the Libraries subfolder.
Per-library call files in the genome VCF format (*.genome.vcf)	Contains the genotype for each position, whether called as a variant or called as a reference. Per-library variant call files in the genome VCF format are located in the Libraries subfolder.
Merged variant call files in the VCF format (*.sample.genome.vcf)	Contains selected specific coordinates from the gVCF files for the Mix A and Mix B libraries for a final merged VCF file for the sample.

File Name	Description
RunMetricsReport.txt	The Run Metrics Report shows run metrics and suggested values to determine if run quality results are within an acceptable range. For Read 1 and Read 2, shows 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.
SampleMetricsReport.txt	Provides calculations from the gVCF file for each sample in Mix A and Mix B.
AmpliconCoverageReport.pdf	Shows the per-amplicon read coverage for each amplicon defined in the TruSight Tumor 15 Manifests. A graph of read-coverage for each amplicon and a table identifying low-coverage amplicons is included in this file.
FilteredVariantSummary.*	Sample reports for a subset of variants listed in the TruSight Tumor 15 Report Definition File. Reports are in *.vcf, *.pdf, and *.txt formats.
AmpliconCoverage_M#.tsv	Contains information about coverage per amplicon per sample for each manifest provided. M# represents the manifest number.
*.ant and *.ant.idx files	Accompanying annotation files of merged variant call files and filtered VCF files. For import into VariantStudio software.
Amplifications.vcf	Contains the AmpScore for the 3 target genes for each sample.

Demultiplexing File Format

The process of demultiplexing reads the index sequence attached to each cluster to determine from which sample the cluster originated. The mapping between clusters and sample number is written to a demultiplexing (*.demux) file for each tile of the flow cell.

The demultiplexing file naming format is `s_1_X.demux`, where X is the tile number.

Demultiplexing files start with a header:

- Version (4 byte integer), currently 1
- Cluster count (4 byte integer)

The remainder of the file consists of sample numbers for each cluster from the tile.

When the demultiplexing step is complete, the software generates a demultiplexing file named `DemultiplexSummaryF1L1.txt`.

- In the file name, **F1** represents the flow cell number.

- In the file name, **L1** represents the lane number.
- Demultiplexing results in a table with one row per tile and one column per sample, including sample 0.
- The most commonly occurring sequences in index reads.

FASTQ File Format

FASTQ is a text-based file format that contains base calls and quality values per read. Each record contains 4 lines:

- The identifier
- The sequence
- A plus sign (+)
- The Phred quality scores in an ASCII + 33 encoded format

The identifier is formatted as:

@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber

Example:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#
```

BAM File Format

A BAM file (*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences up to 128 Mb. SAM and BAM formats are described in detail at samtools.github.io/hts-specs/SAMv1.pdf.

BAM files use the file naming format of `SampleName_S#.bam`, where # is the library number determined by the order that libraries are listed in the run setup.

BAM files contain a header section and an alignment section:

- **Header**—Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and the match descriptor string.

The alignments section includes the following information for each read or read pair:

- **RG:** Read group, which indicates the number of reads for a specific sample.

- **BC:** Barcode tag, which indicates the demultiplexed sample ID associated with the read.
- **SM:** Single-end alignment quality.
- **NM:** Edit distance tag, which records the Levenshtein distance between the read and the reference.
- **XU:**—Identifies the UMI BAM tag
- **XB:**—Identifies the cell-barcode BAM tag

BAM index files (*.bam.bai) provide an index of the corresponding BAM file.

VCF File Format

Variant Call Format (VCF) is a common file format developed by the genomics scientific community. It contains information about variants found at specific positions in a reference genome. VCF files end with the .vcf or .vcf.gz suffixes.

The VCF file header includes the VCF file format version and the variant caller version and lists the annotations used in the remainder of the file. The last line in the header contains the column headings for the data lines. Each of the VCF file data lines contains information about one variant.

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For SNPs, this position is the reference base with the variant; for indels or deletions, this position is the reference base immediately before the variant.
REF	The reference genotype. For example, a deletion of a single T is represented as reference TT and alternate T. An A to T single nucleotide variant is represented as reference A and alternate T.
ALT	The alleles that differ from the reference read. For example, an insertion of a single T after a single A is represented as reference A and alternate AT. An A to T single nucleotide variant is represented as reference A and alternate T.
QUAL	A Phred-scaled quality score assigned by the variant caller. Higher scores indicate higher confidence in the variant and lower probability of errors. For a quality score of Q, the estimated probability of an error is $10^{-(Q/10)}$. For example, the set of Q30 calls has a 0.1% error rate. Many variant callers assign quality scores based on their statistical models, which are high in relation to the error rate observed.

VCF File Annotations

Heading	Description
FILTER	<p>If all filters are passed, PASS is written in the filter column.</p> <p>q30—Quality score less than 30.</p> <p>LowVariantFreq—The variant frequency is less than the given threshold.</p> <p>R8—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8.</p> <p>SB—The strand bias is more than the given threshold.</p>
INFO	<p>Possible entries in the INFO column include:</p> <p>ANT—Consequence as predicted by Illumina Annotation Engine (IAE).</p> <p>DP—The depth (number of base calls aligned to a position and used in variant calling).</p>
FORMAT	<p>The format column lists fields separated by colons. For example, GT:GQ. Available fields include:</p> <p>GQ—Genotype quality.</p> <p>GQX—Genotype quality. GQX is the minimum of the GQ value and the QUAL column. In general, these values are similar; taking the minimum makes GQX the more conservative measure of genotype quality.</p> <p>GT—Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available.</p> <p>NL—Noise level; an estimate of base calling noise at this position.</p> <p>AD—Allele depth.</p> <p>DP—Total depth used for variant calling.</p> <p>NC—Fraction of bases that were uncalled or with basecall quality below the minimum threshold.</p> <p>SB—Strand bias at this position. Larger negative values indicate less bias; values near 0 indicate more bias. Used with the somatic variant caller and GATK.</p> <p>VF—Variant frequency; the percentage of reads supporting the alternate allele.</p>
SAMPLE	The sample column gives the values specified in the FORMAT column.

Genome VCF Files

Genome VCF (gVCF) files are VCF v4.1 files that follow a set of conventions for representing all sites within the genome in a reasonably compact format. The gVCF files include all sites within the region of interest in a single file for each sample.

The gVCF file shows no-calls at positions with low coverage, or where a low-frequency variant (< 2.6%) occurs often enough (> 1%) that the position cannot be called to the reference. A genotype (GT) tag of `./.` indicates a no-call.

For more information, see sites.google.com/site/gvcf-tools/home/about-gvcf.

Per-Library and Merged gVCF Files

The TruSight Tumor 15 workflow generates three sets of variant call files.

- **Per-library VCF and gVCF files**—Contains variants called in either library. Per-library files are written to the Libraries folder.
- **Merged gVCF files**—Contain variants called from both libraries. Merged files are written to the Analysis folder.
- **Amplification VCF files**—Contain copy number variants called from both libraries. Amplification *.vcf are written to the Analysis folder.

Per-library files include both VCF (*.vcf) and gVCF (*.genome.vcf) files, and use the following naming convention, where S# represents the order that the sample is listed for the run:

- **Reports for all sites**—SampleName_S#.genome.vcf
- **Reports variants only**—SampleName_S#.vcf

Merged gVCF files use the following naming convention:

- **Reports for all sites**—SampleName.sample.genome.vcf

For filtered gVCF files, see [Filtered Output Reports on page 18](#).

Amplification VCF files use the following naming convention:

- **Reports for all amplification genes**—SampleName_Amplifications.vcf

Per-Library VCF Files

Variants are called in the Mix A library and the Mix B library to produce an independent set of VCF files for each library. The set of per-library VCF files include both VCF and gVCF files.

Variants are listed in the VCF file using the following criteria:

- Include variants that were flagged as filtered
- Exclude variants with a variant frequency of less than 2.6%
- Variants that pass filters include PASS in the FILTER column
- Variants that fail filters include the filter name in the FILTER column
- Filter variants due to probe bias (PB) when the variant frequency differs significantly between libraries

Merged gVCF Files

The software selects specific coordinates from the gVCF files generated for Mix A and Mix B to create a final merged gVCF file for the sample.

Merged gVCF files are written to the Analysis folder.

Amplification VCF File

Amplification calls are reported in an amplification VCF file. This file uses the same file headings as the other VCF formatted files output by TruSight Tumor 15. However, amplification VCF files only report on copy number variation information for the three TruSight Tumor 15 amplification genes, EGFR, ERBB2, and MET.

Amplification VCF Headings

Column Heading	Description
CHROM	The chromosome of the amplification gene.
POS	The start position of the first probe for this gene.
ALT	. (period): No amplification called <DUP>: Amplification above typical copy number : Deletion below typical copy number
FILTER	Pass: Confident Call NoCall: Call does not pass the threshold for confidence. For more information, see Amplification Calling on page 8 .
INFO	SVTYPE: The type of structural variation, either DUP or DEL END: The end position of the last probe for this gene ANT: The gene annotation (ie, name) FORMAT: The amplification VCF contains a single FORMAT field, AmpScore. For more information on AmpScore, see Amplification Calling on page 8 .

Run Metrics Report

Table 6 Run Metrics Report Table

Column Heading	Description
Metric	Type of metric.
Reads PF (%)	The percentage of reads passing filter.
Q30+(R1)	The percentage of reads in Read 1 with a quality score of 30(Q30) or greater.
Q30+(R2)	The percentage of reads in Read 2 with a quality score of 30(Q30) or greater.
Value	Percentage of reads.

Sample Metrics Report

Table 7 Sample Metrics Results Table

Column Heading	Description
Sample ID	The sample ID provided when the run was created.
MixABases \geq 250x (%)	For the sample, the number of bases in library A that have \geq 250 coverage.*
MixBBases \geq 250x (%)	For the sample, the number of bases in library B that have \geq 250 coverage.*
MixAOnTarget (%)	For the sample, percentage of reads in library A that aligned to the manifest.
MixBOnTarget (%)	For the sample, percentage of reads in library B that aligned to the manifest.

*This refers to per-base coverage using collapsed reads.

Filtered Output Reports

The TruSight Tumor 15 workflow provides variant calls for all genes specified in the manifest files. The workflow also filters gVCF file information for a subset of variants. The subset is listed in the TruSight Tumor 15 Report Definition File and is included with the Local Run Manager software installer. When a variant from this list is detected, it is added to a filtered report that is provided in the *.vcf, *.pdf, and *.txt formats.

Filtered files are written to the Analysis folder and use the following naming conventions:

- SampleName_FilteredVariantSummary.vcf, .pdf, or .txt

Amplicon Coverage File

An amplicon coverage file is generated for each manifest file. The M# in the file name represents the manifest number as it is listed in the samples table for the run.

- Each file includes a header row that contains the sample IDs associated with the manifest.
- Each data row lists the amplicon name followed by the noncollapsed read coverage value for each amplicon library.
- The amplicon coverage files are named AmpliconCoverage_M1.tsv and AmpliconCoverage_M2.tsv.
- The coverage values from these files are used to generate the Amplicon Coverage Report for each sample.

Supplementary Output Files

The following output files provide supplementary information, or summarize run results and analysis errors. Although these files are not required for assessing analysis results, they can be used for troubleshooting purposes. All files are located in the Analysis folder unless otherwise specified.

File Name	Description
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages. Located in the root level of the run folder.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file will be empty if no errors occurred. Located in the root level of the run folder.
CompletedJobInfo.xml	Written after analysis is complete. Contains information about the run, such as date, flow cell ID, software version, and other parameters. Located in the root level of the run folder.
DemultiplexSummaryF1L1.txt	Reports demultiplexing results in a table with one row per tile and one column per sample. Located in the analysis folder in \Fastq\Stats.
AmpliconRunStatistics.xml	Contains summary statistics specific to the run. Located in the root level of the run folder.

Analysis Folder

The analysis folder holds the files generated by the Local Run Manager software.

The relationship between the output folder and analysis folder is summarized as follows:

- During sequencing, Real-Time Analysis (RTA) populates the output folder with files generated during image analysis, base calling, and quality scoring.
- RTA copies files to the analysis folder in real time. After RTA assigns a quality score to each base for each cycle, the software writes the file RTAComplete.txt to both folders.
- When the file RTAComplete.txt is present, analysis begins.
- As analysis continues, Local Run Manager writes output files to the analysis folder, and then copies the files back to the output folder.

Folder Structure

 **Analysis**—Contains *.bam and *.vcf files, and files specific to the analysis module.

 **Data** **Intensities** **BaseCalls** **L001**—Contains one subfolder per cycle, each containing *.bcl files. Sample1_S1_L001_R1_001.fastq.gz Sample2_S1_L001_R1_001.fastq.gz Undetermined_S0_L001_R1_001.fastq.gz **L001**—Contains *.locs files, 1 for each tile. **RTA Logs**—Contains log files from RTA software analysis. **InterOp**—Contains binary files used by Sequencing Analysis Viewer (SAV). **Logs**—Contains log files describing steps performed during sequencing. **Queued**—A working folder for software; also called the copy folder. AnalysisError.txt AnalysisLog.txt CompletedJobInfo.xml QueuedForAnalysis.txt AmpliconRunStatistics.xml RTAComplete.txt RunInfo.xml runParameters.xml

Analysis Folders

Each time that analysis is requeued, the Local Run Manager creates an Analysis folder named **Analysis_N**, where N is a sequential number.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

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

Illumina Technical Support Telephone Numbers

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

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

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

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



Illumina

5200 Illumina Way

San Diego, California 92122 U.S.A.

+1.800.809.ILMN (4566)

+1.858.202.4566 (outside North America)

techsupport@illumina.com

www.illumina.com

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

© 2021 Illumina, Inc. All rights reserved.

illumina[®]