

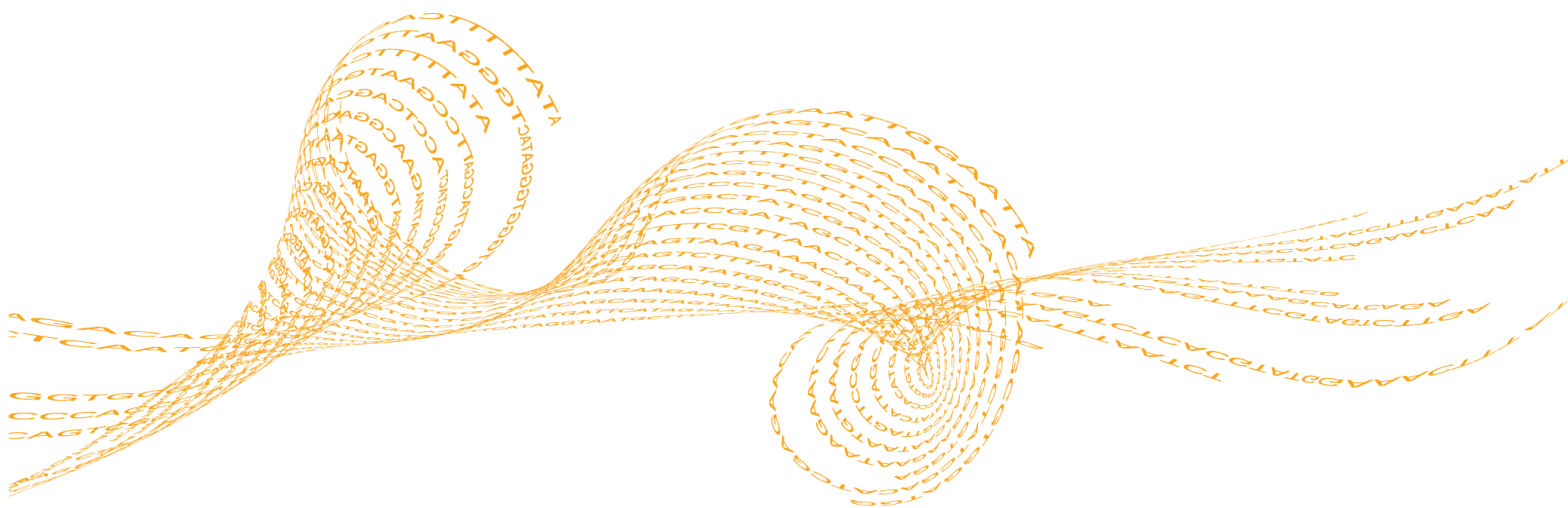
Local Run Manager

Targeted RNA Analysis Module

Workflow Guide

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

Overview	3
Set Parameters	4
Analysis Methods	6
View Analysis Results	7
Analysis Report	8
Analysis Output Files	11
Technical Assistance	17



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.

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

© 2016 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, ForenSeq, Genetic Energy, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiSeq, MiSeqDx, MiSeq FGx, NeoPrep, NextBio, Nextera, NextSeq, Powered by Illumina, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are 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.

Overview

The Local Run Manager Targeted RNA analysis module aligns reads specified in the manifest file, quantifies the relative expression of genes and isoforms between samples, and then compares abundance across samples. This workflow is designed specifically for RNA libraries prepared with the TruSeq Targeted RNA Expression Kit.

Input Requirements

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

- ▶ **Manifest file**—The Targeted RNA analysis module requires at least 1 manifest file. The manifest files are provided with your Targeted Oligo Pool.
- ▶ **Reference genome**—The Targeted RNA analysis module requires the reference genome specified in the manifest file. The reference genome sets the chromosome sizes in the BAM output files and provides variant annotations.

Uploading Manifests

To import a manifest for all runs using the Targeted RNA analysis module, use the Module Settings command from the Local Run Manager navigation bar. For more information, see the *Local Run Manager Software Guide (document # 1000000002702)*. Alternatively, you can import a manifest for the current run only using the **Import Manifests** command on the Create Run screen.

About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis parameters for the Targeted RNA analysis module. For information about the Local Run Manager dashboard and system settings, see the *Local Run Manager Software Guide (document # 1000000002702)*.

Set Parameters

- 1 Click **Create Run**, and select **Targeted RNA**.
- 2 Enter a run name that identifies the run from sequencing through analysis. Use alphanumeric characters, spaces, underscores, or dashes.
- 3 [Optional] Enter a run description to help identify the run. Use alphanumeric characters.

Specify Run Settings

- 1 Enter the number of cycles for the run, if other than the default setting of 51 cycles.
- 2 [Optional] Specify custom primers to be used for the run.



NOTE

By default, the Targeted RNA analysis module is set to the library type TruSeq Targeted RNA Expression and the read type Single Read. Read lengths are set to 51 cycles for Read 1, 6 cycles for Index 1 Read, and 8 cycles for Index 2 Read.

Specify Module-Specific Settings

By default, the Targeted RNA analysis module uses the banded Smith-Waterman algorithm for alignment.

No module-specific settings are required for the Targeted RNA analysis module.

Import Manifest Files for the Run

- 1 Make sure that the manifests you want to import are available in an accessible network location or on a USB drive.
- 2 Click **Import Manifests**.
- 3 Navigate to the manifest file and select the manifest that you want to add.



NOTE

To import manifests for any run using the Targeted RNA analysis module, use the Module Settings feature from the navigation bar.

Specify Samples for the Run


Specify samples for the run using the following options:

- ▶ **Enter samples manually**—Use the blank table on the Create Run screen.
- ▶ **Import samples**—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, and use the file as a reference when preparing libraries or import the file for another run.

Enter Samples Manually

- 1 Adjust the samples table to an appropriate number of rows.
 - ▶ Click the + icon to add a row.
 - ▶ Use the up/down arrows to add multiple rows. Click the + icon.

- ▶ Click the **x** icon to delete a row.
 - ▶ Right-click on a row in the table and use the commands in the drop-down menu.
- 2 Enter a unique sample ID in the Sample ID field.
Use alphanumeric characters, dashes, or underscores.
 - 3 Enter a sample name in the Sample Name field.
Use alphanumeric characters, dashes, or underscores.
 - 4 [Optional] Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, underscores, or spaces.
 - 5 Expand the Index 1 (i7) drop-down list and select an Index 1 adapter.
 - 6 Expand the Index 2 (i5) drop-down list and select an Index 2 adapter.
 - 7 Expand the Manifest drop-down list and select a manifest file.
 - 8 In the Gene Normalization field, enter a list of gene names to be used for normalization. Separate each gene name with a semi-colon.
 - 9 [Optional] Click the **Export**  icon to export sample information in *.csv format.
 - 10 When finished, click **Save Run**.

Import Samples

- 1 Click **Template**. 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 Click **Import Samples** and browse to the location of the sample information file.
- 4 When finished, click **Save Run**.

Analysis Methods

The Targeted RNA analysis module performs the following analysis steps and then writes analysis output files to the Alignment folder.

- ▶ Demultiplexes index reads
- ▶ Generates FASTQ files
- ▶ Aligns to a reference
- ▶ Performs differential expression analysis

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 or when there is up to a single mismatch per Index Read.

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. FASTQ files are the primary input for alignment.



Alignment

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

The banded Smith-Waterman algorithm performs local sequence alignments to determine similar regions between 2 sequences. Instead of looking at 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.

Additionally, the software generates target hits files that contain raw aligned replicate counts for each transcript.

View Analysis Results

- 1 From the Local Run Manager dashboard, click the run name.
- 2 From the Run Overview tab, review the sequencing run metrics.
- 3 [Optional] Click the **Copy to Clipboard**  icon for access to the output run folder.
- 4 Click the Sequencing Information tab to review run parameters and consumables information.
- 5 Click the Samples and Results tab to view the analysis report.
 - ▶ If analysis was repeated, expand the Select Analysis drop-down and select the appropriate analysis.
- 6 [Optional] Click the **Copy to Clipboard**  icon for access to the Analysis folder.

Analysis Report

The following results are provided on the Samples and Results tab. Use the toggle button to view results as Individual Samples or Sample Groups.

View Alignment Summary as Individual Samples

- 1 From the Samples and Results tab, click **View As Individual Samples**.
- 2 To filter the list, enter a Sample ID or Sample Group name in the Search field, and then press **Enter**. You can enter any part of the ID or name.
- 3 To view all, clear the Search field.
- 4 [Optional] Click **Export All (TSV)** to export the alignment summary results.

Alignment Summary in Individual Sample View

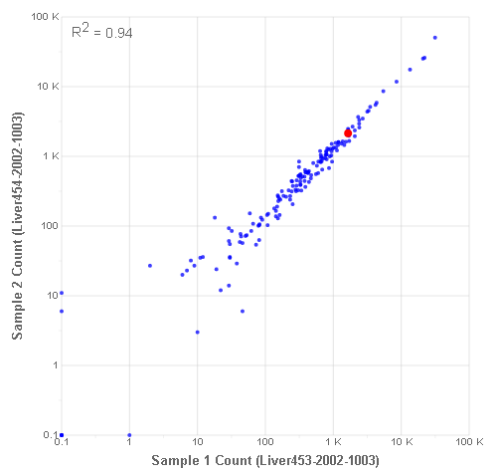
Table 1 Alignment Summary Table

Column Heading	Description
Sample Index	An sample number based on the order the sample was listed for the run.
Sample ID	The sample name provided when the run was created.
Sample Group	The number of biological replicate samples grouped by sample name.
Total Aligned Reads (R1)	The total count of reads passing filter that align for the sample.
Percent Aligned Reads (R1)	The percentage of reads passing filter that align for the sample.

Select Replicate Pair

- 1 Select a sample name from each drop-down list to view the relative abundance of each RNA transcript between the selected samples.
 - ▶ Data points close to the diagonal line represent transcripts with similar abundance.
 - ▶ Points distant from the line represent transcripts expressed at different levels.

Figure 1 Replicate Pair Graph (Example)



- 2 Hover over a point in the graph to view the unique assay ID for the transcript.

- Click an individual assay ID point on the graph to view the corresponding transcript in the Replicate Results table.
- From the Replicate Results table, select another gene or transcript from the drop-down list. Highlight the resulting row in the table to view the corresponding location on Replicate Pair graph.
- [Optional] Click **Export All (TSV)** to export the differential expression results.

Replicate Results

When viewing results as individual samples, the Targeted RNA analysis module provides a replicate results table for each target in the selected comparison.

Table 2 Replicate Results Table

Column Heading	Description
Assay ID	A unique ID for each amplicon.
Gene	The gene targeted by the assay.
Transcript	The transcript targeted by the assay.
Left Exon	The left exon of targeted region.
Right Exon	The right exon of targeted region.
Sample 1 Count	The raw read count for sample 1.
Sample 2 Count	The raw read count for sample 2.
Log Ratio	The percentage of reads passing filter that align for the group.

View Alignment Summary as Sample Group

- From the Samples and Results tab, click **View As Sample Groups**.
- To filter the alignment summary list, enter a Sample Group name in the Search field, and then press **Enter**. You can enter any part of the name.
- To view all, clear the Search field.
- [Optional] Click **Export All (TSV)** to export the alignment summary results.

Alignment Summary in Sample Group View

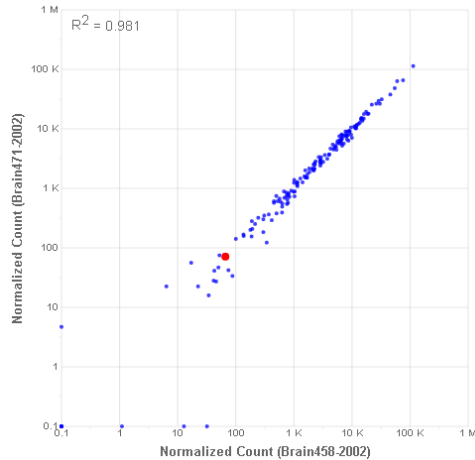
Table 3 Alignment Summary Table

Column Heading	Description
Group Index	A unique number for the group.
Sample Group	The sample name provided when the run was created.
# Replicates	The number of biological replicate samples grouped by sample name.
Total Aligned Reads (R1)	The total count of reads passing filter that align for the group.
Percent Aligned Reads (R1)	The percentage of reads passing filter that align for the group.

Select Comparison

- Select a pair of sample groups from the Select Comparison drop-down list to view the normalized read counts between the selected sample groups on the Normalized Read Count Comparison graph.

Figure 2 Normalized Read Count Comparison Graph (Example)



- 2 Hover over a point in the graph to view the unique assay ID for the transcript.
- 3 Click an individual assay ID point on the graph to view the corresponding transcript in the Differential Expression Results table.
- 4 From the Differential Expression Results table, select another gene or transcript from the drop-down list. Highlight the resulting row in the table to view the corresponding location on Normalized Read Count Comparison graph.
- 5 [Optional] Click **Export All (TSV)** to export the differential expression results.

Differential Expression Results

When viewing results as sample groups, the Targeted RNA analysis module provides a differential expression table for each target in the selected comparison.

Table 4 Differential Expression Results Table

Column Heading	Description
Assay ID	A unique ID for each amplicon.
Gene	The gene targeted by the assay.
Transcript	The transcript targeted by the assay.
Left Exon	The left exon of targeted region.
Right Exon	The right exon of targeted region.
Normalized Count 1	The mean raw read counts for sample 1 after library size normalization.
Normalized Count 2	The mean raw read counts for sample 2 after library size normalization.
Fold Change	The ratio of mean normalized counts for sample 2 divided by mean normalized counts for sample 1.
P-Value	The statistical significance of the differential expression.
Q-Value	The p-value adjusted for false discovery rate (FDR) using the Benjamini-Hochberg method.

Analysis Output Files

The following analysis output files are generated for the Targeted RNA analysis module and provide analysis results for alignment. Analysis output files are located in the Alignment folder.

File Name	Description
Demultiplexing (*.demux)	Intermediate files containing demultiplexing results.
FASTQ (*.fastq.gz)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step.
Alignment files in the BAM format (*.bam)	Contains aligned reads for a given sample.
TargetHitsPerSample_M#.tsv	Contains the raw aligned replicate counts for each transcript.
TargetedRNASeqGene-Expression.tsv	Contains the genes used for normalization and normalization results.
TargetedRNASeqGene-Expression_M#.tsv	Contains sample correlation and differential expression results.

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 are written to 1 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 1 row per tile and 1 column per sample, including sample 0.
- ▶ The most commonly occurring sequences in index reads.

FASTQ File Format

FASTQ file 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 quality scores in an ASCII 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 <https://samtools.github.io/hts-specs/SAMv1.pdf>.

BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed for the run.

BAM files contain a header section and an alignments 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 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.
- ▶ **AS:** Paired-end alignment quality.
- ▶ **NM:** Edit distance tag, which records the Levenshtein distance between the read and the reference.
- ▶ **XN:** Amplicon name tag, which records the amplicon tile ID associated with the read.

BAM files are suitable for viewing with an external viewer such as IGV or the UCSC Genome Browser.

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

Target Hits File Format

The target hits file, TargetHitsPerSample_M#.tsv, is a tab-delimited file that contains the raw aligned replicate counts for each transcript. An output file is created for each manifest.

Column Heading	Description
Gene Name	The name of the gene.
Amplicon ID	The amplicon identifier constructed from the gene name, transcript ID, left exon, right exon, and assay ID.

Column Heading	Description
Assay ID	The unique identifier for the probe set.
Sample ID	Aligned count for all transcripts for this sample. There is a column for each sample using this manifest.

Gene Expression File Format

The gene expression file, *TargetedRNASeqGeneExpression.tsv*, is a tab-delimited text file that is organized into the following sections: Sample Correlation and Differential Expression. This file is the final result of the Targeted RNA workflow.

Sample Correlation Section

Column Heading	Description
Sample Name 1	The first sample (one or more replicates) being compared.
Sample Name 2	The second sample (one or more replicates) being compared.
R ²	The square of the correlation coefficient.

Differential Expression Section

Column Heading	Description
Gene Name	The name of the gene.
Amplicon ID	The amplicon identifier constructed from the gene name, transcript ID, left exon, right exon, and assay ID.
Assay ID	The unique identifier for the probe set.
Sample Name 1	The first sample being compared, which can be 1 or more replicates.
Sample Name 2	The second sample being compared, which can be 1 or more replicates.
Raw Mean Counts 1	The mean of counts for sample 1 across replicates.
Raw Mean Counts 2	The mean of counts for sample 2 across replicates.
Normalized Mean Counts 1	The mean raw read counts for sample 1 after library size normalization.
Normalized Mean Counts 2	The mean raw read counts for sample 2 after library size normalization.
Fold Change	The ratio of mean normalized counts for sample 2 divided by mean normalized counts for sample 1 (mnc2/mnc1).
log ₂ (Fold Change)	The log ₂ (ratio of mnc2/mnc1).

Column Heading	Description
P-value	The statistical significance of the differential expression.
Q-value	The p-value adjusted for false discovery rate (FDR) using the Benjamini-Hochberg method.

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 Alignment 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 is present only if 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 1 row per tile and 1 column per sample.
ErrorsAndNoCallsByLaneTileReadCycle.csv	A comma-separated values file that contains the percentage of errors and no-calls for each tile, read, and cycle.
Mismatch.htm	Contains histograms of mismatches per cycle and no-calls per cycle for each tile.
TargetedRNARunStatistics.xml	Contains summary statistics specific to the run. Located in the root level of the run folder.
Summary.xml	Contains a summary of mismatch rates and other base calling results.
Summary.htm	Contains a summary web page generated from Summary.xml.

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.xml to both folders.
- ▶ When the file RTAComplete.xml 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

📁 Data

📁 Intensities

📁 BaseCalls

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

📁 **L001**—Contains one subfolder per cycle, each containing *.bcl files.

📄 Sample1_S1_L001_R1_001.fastq.gz

📄 Sample2_S2_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

📄 [WorkflowName]RunStatistics

📄 RTAComplete.xml

📄 RunInfo.xml

📄 runParameters.xml

Alignment Folders

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

Notes

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 5 Illumina General Contact Information

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

Table 6 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Japan	0800.111.5011
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
China	400.635.9898	Singapore	1.800.579.2745
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	Taiwan	00806651752
Hong Kong	800960230	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000
Italy	800.874909		

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

Product documentation—Available for download in PDF from the Illumina website. Go to 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

www.illumina.com