

# Local Run Manager RNA Fusion

## Workflow Guide

|  |    |
|--|----|
| Overview                               | 3  |
| Install the RNA Fusion Analysis Module | 3  |
| Set Parameters                         | 4  |
| Analysis Methods                       | 6  |
| View Analysis Results                  | 7  |
| Analysis Report                        | 8  |
| Analysis Output Files                  | 9  |
| Revision History                       | 16 |
| 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, 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).

© 2018 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](http://www.illumina.com/company/legal.html).

## Overview

The Local Run Manager RNA Fusion analysis module aligns reads against the RNA Fusion reference genome using the STAR aligner, and then detects gene fusions using Manta. This workflow is designed specifically for RNA libraries prepared with the TruSight<sup>®</sup> RNA Fusion Panel kit.

The alignment and fusion calling algorithms that this module uses were developed specifically to enable discovery power. The algorithms inherently increase the possibility of calling unexpected fusions relative to algorithms that report only known fusions, such as *BCR-ABL1* or *EML4-ALK*.

In addition, the RNA Fusion reference genome was created to enable alignment on a desktop computer. The RNA Fusion reference genome is a compressed version of the human genome with intronic sequences removed. There may be slight differences in fusions reported from the RNA Fusion reference genome versus a whole genome alignment. For example, removing the intronic and intergenic sequences in the RNA Fusion reference genome removes some of the spatial context that can prevent read through transcripts from being reported as fusion events. Customers are encouraged to assess data thoroughly and establish their own acceptance criteria.

A known limitation of this software is that it does not report fusions, putative fusions, or read through transcripts where the two genes involved are located on the same chromosome, on the same strand, in the same orientation, in close proximity (~100 kbp). For example, this software does not report the *STIL-TAL1* fusion, which is caused by a 90 kb deletion on chromosome 1. Additionally, this software may not call rearrangements that include intronic regions within the gene fusion product.

Additional analysis tools are available in BaseSpace Sequence Hub.

## Input Requirements

In addition to sequencing data files generated during the sequencing run, such as base call files, the RNA Fusion analysis module requires a specific reference genome for alignment and fusion calling. Download the RNA Fusion Reference Genome installer from the TruSight RNA Fusion Panel support page on the Illumina website.

## About This Guide

This guide provides instructions for setting up run parameters for sequencing, and analysis parameters for the RNA Fusion analysis module. For information about the Local Run Manager dashboard and system settings for the Illumina MiniSeq<sup>®</sup> and MiSeq<sup>®</sup> sequencing systems, see the *Local Run Manager Software Guide* (document # 100000002702).

## Install the RNA Fusion Analysis Module

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

- 1 Right-click RNAFusion.Module.Installer.exe and select **Run as administrator**.
- 2 Click **Install**.
- 3 Log in using the user name and password for the admin account in Local Run Manager. The default admin credentials are the following:
  - ▶ User name: admin
  - ▶ Password: passwordOnly admin users can install analysis modules.
- 4 Click **Next**.

- When installation is complete, click **Finish**.  
The module installs some dependencies.
- When complete, click **Close**.

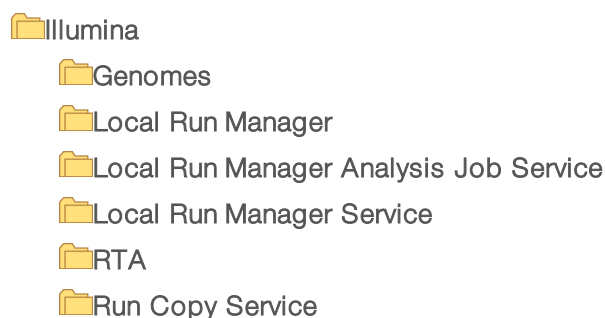
## Install the Reference Genome

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

- Unzip the Reference Genome installer.
- Double-click ReferenceGenomeInstaller.msi.
- Click **Next**.
- When complete, click **Close**.

## Installation Folder Structure

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



The RNA Fusion Analysis Module files are located under **C:\Illumina\Local Run Manager\Modules\RNAFusionWorkflow**.

The RNA Fusion reference genome files are located under **C:\Illumina\Genomes\HumanRNAFusion**.

## Set Parameters

- Select **Create Run**, and select an analysis module.
- Enter a run name that identifies the run from sequencing through analysis.  
The run name can contain alphanumeric characters, spaces, and the following special characters:  
`~!@#\$%-\_{}.
- [Optional]** Enter a run description to further identify the run.  
The run description can contain alphanumeric characters, spaces, and the following special characters:  
`~!@#\$%-\_{}.

## Specify Run Settings

- Enter the number of cycles for the run, if other than the default setting of 2 × 76 cycles.



### NOTE

By default, the read type is set to Paired End and the number of Index Reads is set to 1. Read lengths are set to 76 cycles for Read 1 and Read 2 and 6 cycles for Index 1 Read.

## Specify Module-Specific Settings

The RNA Fusion analysis module uses STAR for alignment. For more information, see [STAR](#) on page 6.

To set module-specific settings, click **Show advanced module settings**.

**Table 1 Advanced Module Settings**

| Setting                     | Description  |
|-----------------------------|--|
| Minimum Breakpoint Distance | The minimum breakpoint distance between 2 genes to be considered for fusion. Any fusion with lesser breakpoint distance is filtered out. By default, set to 100,000.                           |
| Confidence Score Filter     | When switched on, the module only considers fusions as high confidence if they have the "Pass" filter <b>and</b> have a score greater than the threshold (set below). By default, switched on. |
| Confidence Score Threshold  | If the confidence score is below this threshold, the fusions are filtered out from high confidence fusion calls. By default, set to 0.6.   |

There are no applicable custom settings for this module.

## 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 or you can import a sample sheet created in Illumina Experiment Manager (IEM). After importing, the samples table is automatically populated.

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 add row icon to add a row.
  - ▶ Use the up/down arrows to add multiple rows. Click the add row icon.
  - ▶ Click **X** 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.  
Only use alphanumeric characters, dashes, or underscores. Do not include spaces.
- 3 **[Optional]** Enter a sample description in the Sample Description field.  
Use alphanumeric characters, dashes, underscores, or spaces.
- 4 Select an Index 1 adapter from the Index 1 (i7) drop-down list.
- 5 **[Optional]** Click Export to export sample information in \*.csv format.
- 6 Click **Save Run**.

### Import Samples Using a Template

- 1 If creating a sample information file, 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 new sample information file.
- 4 Select the sample information file and click **Open**.
- 5 Click **Save Run**.

## Import Samples using a Sample Sheet

To create a sample sheet in IEM, select the **FASTQ Only** application for your instrument, and then select **TruSeq LT** as the library prep kit. For more information, see *IEM TruSight RNA Pan-Cancer and TruSight RNA Fusion Quick Reference Card (document # 1000000014136)*.

- 1 Click **Import Samples** and browse to the sample sheet.
- 2 Select the sample sheet and click **Open**.
- 3 Click **Save Run**.

## Analysis Methods

The RNA Fusion 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 the RNA Fusion reference genome
- ▶ Detects gene fusions

## 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.  
If a collision is detected with one mismatch, the software switches to zero mismatches for the sample.

## 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 one sample, and the name of that sample is included in the FASTQ file name. FASTQ files are the primary input for alignment.

## STAR

Spliced Transcripts Alignment to a Reference (STAR) is a fast RNA-Seq read mapper, with support for splice-junction and fusion read detection.

STAR aligns reads by finding the Maximal Mappable Prefix (MMP) hits between reads (or read pairs) and the RNA Fusion reference genome, using a Suffix Array index. Different parts of a read can be mapped to different genomic positions, corresponding to splicing or RNA-fusions. The genome index includes known splice-junctions from annotated gene models, allowing for sensitive detection of spliced reads. STAR performs local alignment, automatically soft clipping ends of reads with high mismatches.

The following custom parameters are used with STAR to enable split-read (ie, fusion read) mapping:

- ▶ `outSAMmapqUnique 50`
- ▶ `outFilterType BySJout`
- ▶ `outSJfilterCountUniqueMin -1 2 2 2`
- ▶ `outSJfilterCountTotalMin -1 2 2 2`
- ▶ `outFilterIntronMotifs RemoveNoncanonical`
- ▶ `chimSegmentMin 12`
- ▶ `chimJunctionOverhangMin 12`
- ▶ `chimScoreDropMax 30`
- ▶ `chimSegmentReadGapMax 5`
- ▶ `chimScoreSeparation 5`

For more information, see [github.com/alexdobin/STAR](https://github.com/alexdobin/STAR).

## Manta

Manta calls structural variants (SVs) from mapped paired-end sequencing reads. Manta discovers candidate SVs from discordant pair and split-read alignments, followed by local assembly and realignment to refine candidates.

The module uses Manta to detect gene fusions from reads aligned by STAR, which appear like translocations in the RNA alignments.



- ▶ Read counts across the fusion and alignment qualities.
- ▶ Genome-wide realignment of fusion contigs to filter candidates that can be explained by a local alignment elsewhere in the genome.
- ▶ Length of coverage around the breakpoints, indicating presence of stable fusion transcripts.

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


## RNA Fusion Reference Genome

The RNA Fusion reference genome is a modified version of the hg19 genome. Download the installer from the TruSight RNA Fusion Panel support page on the Illumina website.

## View Analysis Results

- 1 From the Local Run Manager dashboard, select the run name.
- 2 From the Run Overview tab, review the sequencing run metrics.
- 3 **[Optional]** Select the **Copy to Clipboard**  icon to copy the output run folder file path.
- 4 If you want to change the file location of the analysis data for any future runs, 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.

- 5 Select the Sequencing Information tab to review run parameters and consumables information.
- 6 Select the Samples and Results tab to view the analysis report.
  - ▶ If analysis was requeued, select the appropriate analysis from the Select Analysis drop-down list.
  - ▶ From the left navigation bar, select a sample name to view the report for another sample.
- 7 **[Optional]** Select the Copy to Clipboard  icon to copy the Analysis folder file path.

## Analysis Report

The following results are provided on the Samples and Results tab. Results can be viewed for each sample.

### Run Information and Library Information

The RNA Fusion Report includes run information and library information for the sample.

**Table 2 Run Information and Library Information**

| Section                        | Description   |
|--------------------------------|---|
| Run Information                | Total of PF reads (entire run)—Total number of reads in the run passing filter  |
|                                | % Bases $\geq$ Q30—Percent of bases with a quality score $\geq$ 30  |
| Library Information for sample | % Aligned reads—Percent of reads mapping to the reference genome  |
|                                | Number of reads—Total number of pass filter reads   |
|                                | Number of genes at 1X coverage—Number of genes that have at least 1x coverage by aligned RNA Fusion reads along the exons |
|                                | % Aligned to rRNA—Percent of reads that align to ribosomal RNA repeats  |

## Report Tables

Each report includes a main table with high confidence fusion calls and two supplementary tables.

**Table 3 Table Descriptions**

| Table                        | Description  |
|------------------------------|--|
| High Confidence Fusion Calls | Fusions called with both a filter value of Pass and, if enabled, a score greater than the threshold value. For more information, see <a href="#">Filtering on page 12</a> and <a href="#">Scoring on page 13</a> . |
| Low Confidence Fusion Calls  | Fusions called that <i>do not</i> have a filter value of 'Pass' and, if enabled, <i>do not</i> have a score greater than the threshold value.  |
| Recurrent Fusions Not Called | Recurrent gene fusions that are defined by the Mitelman database that were not detected. The TruSight RNA Fusion Panel targets both genes of the fusions listed in this table.                                     |



Each table includes the following columns. You can sort the data by clicking any column title or use search to search for a disease association or gene fusion.

Click **Export (csv)** or **Export (pdf)** to export a table. The exported file contains all results, regardless of the filters used.

**Table 4 Column Heading Descriptions for High and Low Confidence Fusion Calls Tables**

| Column Heading          | Description   |
|-------------------------|---|
| Disease Association*    | Diseases the fusion has been linked with in the Mitelman database.  |
| Gene Fusion             | Gene names of the fused genes.  |
| Cytogenetic Coordinates | Genomic coordinates of the fused genes.   |
| Fusion Supporting Reads | Reads specifically supporting the fusion transcripts.<br>The sum of PairedAlt and SplitAlt reads in the fusions.csv file. For more information, see <a href="#">Analysis Output Files on page 9</a> . |
| Gene 1 Reference Reads  | Reads aligning across the breakpoint of Gene 1 that do not support the fusion (ie, number of reads from the wild type or endogenous transcript).  |
| Gene 2 Reference Reads  | Reads aligning across the breakpoint of Gene 2 that do not support the fusion (ie, number of reads from the wild type or endogenous transcript).  |

**Table 5 Column Heading Descriptions for Recurrent Fusions Not Called Table**

| Column Heading                | Description   |
|-------------------------------|---|
| Disease Association*          | Diseases the fusion has been linked with in the Mitelman database.  |
| Gene Fusion                   | Gene names of the fused genes.  |
| Cytogenetic Coordinates       | Genomic coordinates of the fused genes.   |
| Gene 1 Whole Gene Read Counts | Read counts covering the whole gene of the potential fusion partner (no fusion supporting reads were identified). |
| Gene 2 Whole Gene Read Counts | Read counts covering the whole gene of the potential fusion partner (no fusion supporting reads were identified). |

\*The disease associations are defined by data collected from the Mitelman Database on August 26, 2015. The RNA Fusion module does not retrieve updates to the database and users cannot update the database. The RNA Fusion module reports only the most frequently observed disease association of the fusion from scientific literature recorded in the Mitelman database as of August 26, 2015. Disease associations for fusions with undefined disease associations are reported as "NA" (not available). The disease association that the RNA Fusion module provides, via the Mitelman Database, is for research use only and must not be used for any clinical decisions. The TruSight RNA Fusion System is classified as Research Use Only.

"Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2015/08/26). Mitelman F, Johansson B and Mertens F (Eds.), <http://cgap.nci.nih.gov/Chromosomes/Mitelman>"

## Analysis Output Files

The following analysis output files are generated for the RNA Fusion 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.   |

| File Name                                 | Description  |
|---|--|
| Counts files in GeneCounts folder (*.csv) | Contains read counts information for a given sample. |
| fusions.csv                               | Contains information for all fusion candidates.      |

## 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=AAAAAAAAAA9#:<#<;<<<????#=#
```

## FASTQ File Names

FASTQ files are named with the sample name and the sample number. The sample number is a numeric assignment based on the order that the sample is listed for the run. For example:

```
...\samplename_S1_L001_R1_001.fastq.gz
```

- ▶ **samplename**—The sample name listed for the sample. If a sample name is not provided, the file name includes the sample ID.
- ▶ **S1**—The sample number based on the order that samples are listed for the run starting with 1. In this example, S1 indicates that this sample is the first sample listed for the run.

**NOTE**

Reads that cannot be assigned to any sample are written to a FASTQ file for sample number 0, and excluded from downstream analysis.

- ▶ **L001**—The lane number.
- ▶ **R1**—The read. In this example, R1 means Read 1. For a paired-end run, a file from Read 2 includes R2 in the file name. When generated, the Index Reads are I1 or I2.
- ▶ **001**—The last segment is always 001.

FASTQ files are compressed in the GNU zip format, as indicated by \*.gz in the file name. FASTQ files can be uncompressed using tools such as gzip (command-line) or 7-zip (GUI).

## BAM File Format

A BAM file (\*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences. SAM and BAM formats are described in detail at [samtools.github.io/hts-specs/SAMv1.pdf](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. In multinode mode, the S# is set to S1, regardless the order of the sample.

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. Alignments correlate to the RNA Fusion reference genome.

For descriptions of possible tags in the BAM files, see [github.com/alexdobin/STAR](https://github.com/alexdobin/STAR).

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

## GeneCounts Files

The GeneCounts folder contains the following files for each sample.

| File                    | Description  |
|-------------------------|--|
| Contigs.csv             | Counts of reads to different contigs of the reference genome |
| Counts.csv              | Read counts per gene   |
| Counts.csv.coverage.txt | Read coverage information per gene                           |
| Counts.csv.geneinfo.txt | Gene length (exonic) and gene IDs                            |

## Fusions File Format

The fusions.csv file located in the MantaFusions folder contains the following information for all fusion candidates.

| Column Heading     | Description  |
|--------------------|--|
| Name               | Manta name of the fusion candidate.  |
| Chr1               | Chromosome of fusion end 1 (5' for stranded data).   |
| Pos1               | Position of end 1.   |
| Strand1            | Strand of RNA fusion transcript at end 1<br>"." for unstranded.  |
| Chr2               | Chromosome fusion end 2 (3' for stranded data).  |
| Pos2               | Position of end 2.   |
| Strand2            | Strand of RNA fusion transcript at end 2<br>"." for unstranded.  |
| Depth              | Total read depth for the fusion candidate.   |
| PairedRef          | Read pairs supporting the reference alleles.   |
| PairedAlt          | Read pairs supporting the fusion.  |
| SplitRef           | Split reads supporting the reference alleles.  |
| SplitAlt           | Split reads supporting the fusion.   |
| Ref1               | Reads supporting the reference allele at end 1.  |
| Ref2               | Reads supporting the reference allele at end 2.  |
| Gene1              | Names of all annotated genes (exons) overlapping fusion end 1.   |
| Gene2              | Names of all annotated genes (exons) overlapping fusion end 2.   |
| Cis                | States whether the fusion is in cis orientation (same chromosome, same strand).  |
| BreakpointHomology | Length of perfect homology around 2 fusion breakpoints.  |
| Contig             | Sequence of the fusion-spanning contig.  |
| ContigAlign1       | Length of fusion contig alignment at fusion end 1.   |
| ContigAlign2       | Length of fusion contig alignment at fusion end 2.   |
| ContigLocalAlign   | Length of longest alternative nonfusion alignment of the contig.   |
| CoverageGene1      | Length of genomic region covered by nonfusion reads after fusion end 1.  |
| CoverageGene2      | Length of genomic region covered by nonfusion reads after fusion end 2.  |
| CandidateReads     | Number of fusion candidate reads.  |
| RepeatOverlap      | Indicates if the fusion call overlaps a genomic repeat.  |
| Filter             | "Pass" or the name of the threshold filter triggered by the fusion candidate. For more information, see <i>Filtering</i> on page 12. |
| Score              | Score of the fusion candidate. For more information, see <i>Scoring</i> on page 13.  |

## Filtering

A fusion candidate can have the following filter values. A fusion candidate without a "Pass" filter can have multiple nonpassing threshold filters.

**Table 6 Threshold Filters**

| Filter     | Description   |
|------------|---|
| Pass       | Meets the threshold filters detailed in <i>Scoring</i> .      |
| Imprecise  | A low-resolution candidate, not an assembled fusion call.     |
| Intragenic | The presumed fusion is between portions within the same gene. |

| Filter           | Description   |
|------------------|---|
| LocalContigAlign | Contig realignment found a nonfusion alignment for this contig.   |
| LowFusionRatio   | Few strong evidence reads compared to wild-type reads.  |
| LowFusionReads   | Few strong evidence reads.  |
| NoGeneCoverage   | Length of read coverage around the fusion breakpoint is small.  |
| NonExonic        | Fusion breakpoint does not fall within an exon.   |
| NoReferenceReads | No reads on either side of the presumed breakpoint are marked as reference (structurally normal) reads.   |
| 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). |

## Scoring

The score is calculated as a weighted average of the individual features of each fusion candidate, using the formula  $\max(\min(f, f_{\text{Min}}, f_{\text{Max}}) - f_{\text{Min}}) * \text{Coef}$ . The score is between 0–1.

| Feature   | Scored Range      | Max(Coefficient) |
|---|-------------------|------------------|
| Split Reads   | 0–10              | 0.15             |
| Paired Reads  | 0–5               | 0.15             |
| Alt/Ref Reads   | 0–0.1             | 0.1              |
| Fusion Contig Align Length (bp)<br>(computed for each break-end)              | 12–76             | 0.2              |
| Alternative Local Contig Align Fraction                                       | 0.5–1             | -0.3             |
| Break-end Homology (bp)   | 2–20              | -0.2             |
| Fusion Length (bp)<br>(for cis fusions)                                       | 2,000,000–100,000 | -0.1             |
| Is Cis?   | —                 | -0.2             |
| Coverage After Fusion (bp)<br>(computed for the break-end with less coverage) | 0–500             | 0.2              |

Fusions must pass the following threshold filters to have a "Pass" filter value.

- ▶ Split Reads + Paired Reads  $\geq 3$
- ▶ Alt/Ref Reads  $\geq 0.01$
- ▶ Fusion Contig Align Length (bp)  $> 16$
- ▶ Break-end Homology (bp)  $\leq 10$
- ▶ Alternative Local Contig Align Fraction  $< 0.8$
- ▶ Coverage after fusion (bp)  $\geq 100$

## Supplementary Output Files

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

| File Name            | Description   |
|----------------------|---|
| AdapterTrimming.txt  | Lists the number of trimmed bases and percentage of bases for each tile. This file is present only if adapter trimming was specified for the run. |
| AnalysisError.txt    | Processing log that lists any errors that occurred during analysis. This file is present only if errors occurred.                                 |
| AnalysisLog.txt      | Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages.      |
| CompletedJobInfo.xml | Written after analysis is complete, contains information about the run, such as date, flow cell ID, software version, and other parameters.       |
| DemuxSummaryF1L1.txt | Reports demultiplexing results in a table with 1 row per tile and 1 column per sample.  |
| FastqSummaryF1L1.txt | FASTQ summary file with the sample number, tiles, number of raw reads, and number of reads passing filter.  |
| SampleSheetUsed.csv  | A copy of the sample sheet.   |



































## 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 and assigns a quality score to each base for each cycle.
- ▶ 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

-  **Alignment\_Imported\_X\*** or **Alignment\_X\*** (\* indicates different analyses performed)
  -  **Import Date/Time**
    -  **Fastq**
      -  **Reports**—Contains run information.
        -  **[SampleName]**
          -  samplename\_S1\_L001\_R1\_001.fastq.gz
          -  samplename\_S1\_L001\_R2\_001.fastq.gz
        -  **Stats**
          -  AdapterTrimming.txt
          -  ConversionStats.xml
          -  DemultiplexingStats.xml
          -  DemuxSummaryF1L1.txt
          -  FastqSummaryF1L1.txt
          -  Stats.json
        -  **Undertermined\_S0**
          -  Undetermined\_S0\_L001\_R1\_001.fastq.gz
          -  Undetermined\_S0\_L001\_R2\_001.fastq.gz
    -  **Logging**—Contains log files describing steps performed during alignment.
    -  **Report**—Contains files for the sample analysis report.
    -  **samples**
      -  **[SampleName]**
        -  **Align**—Contains \*.bam and \*.out files.
          -  **GeneCounts**—Contains contigs and counts files.
          -  **MantaFusions**—Contains the fusions.csv file.
      -  AnalysisError.txt
      -  AnalysisLog.txt
      -  Checkpoint.txt
      -  CompletedJobInfo.xml
      -  SampleSheetUsed.csv
    -  **Config**—Contains instrument configuration files.
    -  **Data**—Contains \*.bcl files and scatter plots from the sequencing run.
    -  **Images**—Contains images generated during the sequencing run (if appropriate).
    -  **InterOp**—Contains binary reporting files used for Sequencing Analysis Viewer.
    -  **Logs**—Contains log files describing steps performed by the instrument for each cycle.

-  **Recipe**—Contains a \*.xml file of the sequencing recipe used.
-  **RTALogs**—Contains logs of RTA events.
-  **Thumbnail\_Images**—Contains smaller versions of files in the Images folder.
-  RTAComplete.txt
-  RTAConfiguration.xml
-  RTAReadComplete.txt
-  RunCompletionStatus.xml
-  RunInfo.xml
-  RunParameters.xml

## Alignment Folders

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

## Revision History

| Document                        | Date            | Description of Change   |
|---------------------------------|-----------------|---|
| Document #<br>1000000010786 v01 | March<br>2018   | Applied latest branding and formatting.   |
| Document #<br>1000000010786 v00 | October<br>2016 | Initial release.<br>RNA Fusion specific content was removed from LRM Off-Instrument Guide and placed here to create a separate RNA Fusion Module guide. |



## 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           | 400.066.5835    |                |
| Denmark         | +45 80820183    | +45 89871156   |
| Finland         | +358 800918363  | +358 974790110 |
| France          | +33 805102193   | +33 170770446  |
| Germany         | +49 8001014940  | +49 8938035677 |
| Hong Kong       | 800960230       |                |
| Ireland         | +353 1800936608 | +353 016950506 |
| Italy           | +39 800985513   | +39 236003759  |
| Japan           | 0800.111.5011   |                |
| Netherlands     | +31 8000222493  | +31 207132960  |
| New Zealand     | 0800.451.650    |                |
| Norway          | +47 800 16836   | +47 21939693   |
| Singapore       | +1.800.579.2745 |                |
| 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  
www.illumina.com

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

© 2018 Illumina, Inc. All rights reserved.

**illumina**<sup>®</sup>