

Local Run Manager RNA Amplicon Analysis Module

Workflow Guide

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

Local Run Manager Overview	3
Set Parameters	3
Analysis Methods	6
View Analysis Results	7
Analysis Report	8
Analysis Output Files	9
Revision History	17
Technical Assistance	18



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).

© 2019 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.

Local Run Manager Overview

The Local Run Manager RNA Amplicon analysis module analyzes RNA sequences for a set of predefined transcripts. These transcripts have been targeted by identifying splice junctions that are unique to that transcript. The workflow demultiplexes indexed reads, generates FASTQ files, aligns reads against amplicon sequences specified in the manifest file, counts amplicon fragments, performs differential expression analysis, and writes output files to the Analysis folder. Differential expression analysis is performed using DESeq2 in R.

The workflow requires that all samples in an analysis use the same manifest file. This workflow is designed specifically for RNA libraries prepared with the TruSeq Targeted RNA Expression Kit, AmpliSeq Library PLUS for Illumina (96) Library Prep Kit, and Custom prep kit with the ability to add ERCC targets to the selected panel for analysis.

Input Requirements

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

- ▶ **Manifest file**—The RNA Amplicon analysis module requires one manifest file per analysis. The encrypted manifest file is provided with your Targeted Oligo Pool (only applicable to some panels).

Uploading Manifests

To import a manifest file for all runs using the RNA Amplicon analysis module, use the Modules & Manifests option from the Tools menu. For more information, see the *Local Run Manager Software Guide (document # 1000000002702)*.

Alternatively, you can import a manifest for the current run using **Import Manifests** on the Create Run screen.

About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis parameters for the RNA Amplicon 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 Log in to Local Analysis Software.
- 2 Select **Create Run**, and then select **RNA Amplicon**.
- 3 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:
`~!@#\$\$%-_{}`.
- 4 **[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

- 1 Select your library prep kit from the Library Prep Kit drop-down list.
- 2 If applicable, select the number of index reads for the run.
The AmpliSeq Library PLUS for Illumina (96) Kit has hard-coded indexes and cannot be changed.

- 3 For Read Type, select from the following options:
 - ▶ **Single Read**
 - ▶ **Paired End**—Not available for TruSeq Targeted RNA Expression kits
- 4 To override the default read length value, enter the preferred read lengths for the run.
- 5 **[Optional]** For Custom Primers, specify any custom primer information to be used for the run by selecting the appropriate checkboxes.
Custom primer options may vary based on your instrument or Local Run Manager implementation.

Specify Module-Specific Settings

By default, the RNA Amplicon analysis module uses the Burrows-Wheeler Aligner (BWA) for alignment.

If you are using the AmpliSeq ERCC RNA Spike-In Mix for Illumina or the AmpliSeq ERCC RNA Companion Panel for Illumina, set the **ERCC Analysis** option.

When enabled, ERCC targets are merged into the input manifest for analysis. The gene name and amplicon sequence are checked to make sure that each ERCC target appears only once in the merged manifest.

Custom Analysis Settings

Custom analysis settings are intended for technically advanced users. If settings are applied incorrectly, serious problems can occur.

Add a Custom Analysis Setting

- 1 From the Module-Specific Settings section of the Create Run screen, select **Show advanced module settings**.
- 2 Select **Add custom setting**.
- 3 In the custom setting field, enter the setting name as listed in *Available Analysis Settings on page 4*.
- 4 In the setting value field, enter the setting value.
- 5 To remove a setting, select the **x** icon.

Available Analysis Settings

Setting Name	Value options	Description
FusionScoreCutoff	Any value 2–20 The default is 4.0.	Fusion score cutoff to determine whether a fusion is present. Used only when the manifest contains fusion targets.
FusionEmitScoreCutoff	Float The default is 1.0.	Fusion score cutoff to determine whether to output a fusion. Used only when the manifest contains fusion targets. Minimum value is 1.

Import Manifest Files for the Run

- 1 Make sure that the manifests you want to import are available locally or in an accessible network location.
- 2 Select **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 RNA Amplicon analysis module, use the Modules & Manifests option from the Tools drop-down menu on the navigation bar.

Specify Samples for the Run



Specify samples for the run using one of the following options:

- ▶ **Enter samples manually**—Use the blank table at the bottom of the Create Run screen.
- ▶ **Import sample sheet**—Use an external CSV file to specify samples for the run.

After the samples table is populated, you can export the sample information to an external file. Use this file as a reference when preparing libraries or importing the file for another run.

Enter Samples Manually

To manually enter sample information, you must first select a Library Prep Kit in the Run Settings section.

- 1 Adjust the samples table to an appropriate number of rows.
 - ▶ In the Rows field, use the up/down arrows or enter a number to specify the number of rows to add to the table. Select  to add the rows to the table.
 - ▶ Select  to delete a row.
 - ▶ Right-click on a row in the table and use the commands in the contextual menu.
- 2 Enter a unique sample ID in the Sample ID field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in the Sample ID field.
- 3 Enter a sample group name in the Sample Group field.
The sample group name is required to connect sample replicates with a common name.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in the Sample Group field.
- 4 **[Optional]** Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in the Sample Description field.
- 5 If you have a plated kit, select an index plate well from the Index well drop-down list and skip to step 8.
- 6 Specify an Index 1 sequence.
Select **Show Index Sequence/Show Index Names** to toggle between showing the name of the index and the index sequence.
- 7 Specify an Index 2 sequence.



NOTE

During analysis, the iSeq™ 100, MiniSeq™, and NextSeq™ Systems automatically reverse complement the i5 indexes in custom library prep kits. Make sure that the i5 indexes are in the forward orientation.

- 8 Select a manifest file from the Manifest drop-down list.
This manifest file will be used for all the samples defined in the table.
- 9 **[Optional]** Specify probe(s) to be used for normalization with one of the following methods.
 - ▶ Enter a probe name(s) in the Gene Normalization field. Separate each probe name with a semicolon.
 - ▶ Select a probe name from the Gene Normalization drop-down menu. This is only possible if the manifest file has fewer than 100 probes.

This name will be used for all the samples defined in the table.

- 10 **[Optional]** Enter a project name in the Sample Project field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in the Sample Project field.
- 11 **[Optional]** Select **Export Sample Sheet** to export the sample information in *.csv format.
The exported sample sheet can be used as a template or imported when creating new runs.
- 12 Select **Save Run**.

Import Sample Sheet

- 1 If you do not have a sample sheet to import, see *Enter Samples Manually* on page 5 for instructions on how to create and export a sample sheet. Edit the file as follows.
 - a Open the sample sheet in a text editor.
 - b Enter the sample information in the [Data] section of the file.
 - c Save the file. Make sure that the sample IDs are unique.
- 2 Select **Import Sample Sheet** at the top of the Create Run screen and browse to the location of the sample sheet.
Make sure that the information in the manifest and sample sheet is correct. Incorrect information can impact the sequencing run.



NOTE

During analysis, the iSeq™ 100, MiniSeq™, and NextSeq™ Systems automatically reverse complement the i5 indexes in custom library prep kits. If you are importing a sample sheet for a custom library prep kit, make sure that the i5 indexes are in the forward orientation.

- 3 When finished, select **Save Run**.

Analysis Methods

The RNA Amplicon 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 genome
- ▶ Performs differential expression analysis
- ▶ Performs fusion calling (when using encrypted manifest files)

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

Alignment

Clusters from each sample are aligned against amplicon sequences from that sample's manifest file using the BWA Aligner. Alignments with high number of mismatches are filtered from alignment results. Filtered alignments are written in alignment files as unaligned.

Fusion Calling

Two types of gene fusions are called by the workflow, fusions and putative fusions. Fusions are known fusions that are specifically targeted by a panel. Only one of the two fusion partner genes, called the driver gene, is targeted for a putative fusion. Two targets are placed on the 5' and 3' end of the driver gene. An imbalanced expression of the two targets indicates the presence of fusions with the driver gene.

Each fusion type produces one output file.



NOTE




Fusion calling is only performed in runs that use an encrypted manifest file. The fusion output files only contain fusion calls when using an encrypted manifest.

Differential Expression Analysis

Differential expression analysis is performed using DESeq2 in R. If the NormalizationGenes setting is set, estimateSizeFactors is called on the control genes. The normalized counts for the control and comparison groups are calculated from log2FoldChange and baseMean in the DESeq2 results.

For more information, visit the [DESeq2 page](#) on the Bioconductor website.

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 requeued, 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

The following results are provided on the Samples and Results tab.

Sample Information

Column Heading	Description
Sample ID	The sample ID provided when the run was created.
Total PF Reads	The total number of reads passing filter.
Percent Q30 Bases	The percentage of bases called with a quality score \geq Q30.

Amplicon Summary

Column Heading	Description
Manifest	The name of the file that specifies the reference and targeted reference regions.
Number of Amplicon Regions	The number of amplicon regions sequenced.
Total Length of Target Regions	The total length in base pairs of sequenced amplicons in the target regions.

Read Level Statistics

Column Heading	Description
Total Aligned Reads	The total number of reads that aligned to the reference for each read (Read 1 and Read 2).
Percent Aligned Reads	The percentage of reads that aligned to the reference for each read (Read 1 and Read 2).

Base Level Statistics

Column Heading	Description
Percent Q30 Bases	The percentage of bases called with a quality score \geq Q30.
Total Aligned Bases	The total number of bases that aligned to the reference for each read (Read 1 and Read 2).
Percent Aligned Bases	The percentage of aligned bases averaged over cycles per read (Read 1 and Read 2).
Percent Mismatches	The average percentage of mismatches across both reads 1 and 2 over all cycles. Numbers are calculated per read.

Coverage Summary

Column Heading	Description
Amplicon Mean Coverage	The total number of aligned bases divided by the targeted region size.

Analysis Details – Settings

Column Heading	Description
Run Folder	The Run ID of the analyzed run.
Analysis Folder	The file location of the analysis information.
Reference Genome	The reference genome used for the analysis.
Manifest	The manifest file used for the run.

Analysis Details – Software Versions

Column Heading	Description
RNA Amplicon Workflow	Software version of the RNA Amplicon app.
BWA Aligner	Software version of the Burrows-Wheeler Aligner.
bammetrics	Software version of the BamMetrics workflow.
SAMtools	Software version of the Sequence Alignment/Map (SAM) format.

Analysis Output Files

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

File Name	Description
DemuxSummary*	Intermediate files containing demultiplexing summary 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. Not applicable for runs that use an encrypted manifest file.
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.
fusions.csv	Contains fusion calls based on the fusion target type.
fusions.vcf.gz	Contains information about detected gene fusion events.
putativefusions.csv	Contains fusion call information based on the Balance5p3p and ExpressionControl target types
exonvariants.csv	Contains summary information for RnaExonVariant and RnaExonVariantWildType targets.
exonvariants.vcf.gz	Contains detected DEL structural variant information.

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:IndexSequences or Pairs

Example:

```
@M00376:11:000000000-A26YW:1:1101:15415:1774 1:N:0:ATCACG+TGTTCTCT
```

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.

BAM files use the file naming format, **SampleName.SampleGroup_S#.bam**. The variable, #, is the sample number determined by the order that samples are listed for the run.

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 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.
- ▶ **MD**—Mismatching positions/bases (BWA only).

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.



NOTE

BAM files are not available for runs that performed fusion calling.

Target Hits File Format

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.
Target ID	The unique identifier for the target.
Transcript	Transcript targeted by the assay.
Exons	Left and right exons of the targeted region.
Off Target	TRUE if the entry is a predicted off-target. FALSE otherwise.

Column Heading	Description
Target Type	The type of amplicon.
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
Control Group	The sample group used as the control in the analysis.
Comparison Group	The sample group 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.
Target ID	The unique identifier for the target.
Control Group	The sample group used as the control in the analysis.
Comparison Group	The sample group being compared.
Control Raw Mean Count	The mean of counts for the control group across replicates.
Comparison Raw Mean Count	The mean of counts for the comparison group across replicates.
Control Normalized Mean Count	The mean raw read counts for the control group after library size normalization.
Comparison Normalized Mean Count	The mean raw read counts for the comparison group after library size normalization.
Fold Change	The ratio of mean normalized counts for sample 2 divided by mean normalized counts for sample 1 ($mnc2/mnc1$).
$\log_2(\text{Fold Change})$	The $\log_2(\text{ratio of } mnc2/mnc1)$.
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.
Transcript	Transcript targeted by the assay.
Exons	Left and right exons of the target region.
Target Type	The type of amplicon.

Exon Variant File Format

exonvariants.csv

A comma-separated values (CSV) format file containing summary information for RnaExonVariant and RnaExonVariantWildType targets.

The file contains the following columns:

Column Heading	Description
Target ID	A unique alphanumeric identifier for the target.
Gene	The gene in which the target is located.
Transcript ID	The transcript identifier of the affected transcript.
Chromosome	The chromosome on which the variant is located.
Position	The position of the variant.
Exon	The number of the affected exon for the affected transcript.
Supporting Reads	The number of reads supporting the variant.
Wild Type	Boolean value that indicates if the variant matches the "wild-type" genotype.
Ratio to Wild Type	Calculated with the following equation: $\frac{\text{Read count}}{\text{Mean read count of all Wild Type RNAExonVariants for the gene}}$ Blank if there are no Wild Type targets for the gene.
Norm Count within Gene	Calculated with the following equation: $\frac{\text{Read count}}{\text{Sum of read counts of all RNAExonVariants of the gene}}$ Value is 0 if Read count =0 regardless of the value of the denominator.
Contains DEL Variants	Boolean value that indicates the presence of deletions in the CIGAR string.
Present	Boolean value that indicates if Norm Count within Gene is greater than zero.

exonvariants.vcf.gz

This file contains detected DEL structural variants. These are reported from RnaExonVariant and RnaExonVariantWildType targets whose CIGAR string contains D operations and which have non-zero read counts.

The file is in VCF 4.1 format and compressed by the gzip compression software. In addition to a format-compliant header, the file contains the following columns:

Column Heading	Description
CHROM	The chromosome on which the variant occurs.
POS	Position, given as the base before the left-most deleted position.
REF	The reference base preceding the deletion. It is taken from the full amplicon sequence in the manifest and will be set to N if not available.
ALT	The alternative allele (deletion) at the variant position.
ID	An identifier for the variant, formatted as: <Gene>.<Transcript ID>.<Chromosome>:<1-based Start>-<1-based End>.
QUAL	Will always be "." (missing).

Column Heading	Description
FILTER	Will always be "PASS".
INFO	Consists of the following sub-fields: <ul style="list-style-type: none"> • SVTYPE— "DEL" • READS— The number of reads supporting this variant • NORMREADSWITHINGENE— The within-gene normalized number of reads that support this splice variant • ANT— Annotation of the breakpoint formatted as: "Gene Transcript Strand", where strand is either + or -.
FORMAT	Consists of the following sub-fields: <ul style="list-style-type: none"> • DP— The number of reads supporting this variant • NR— The within-gene normalized number of reads that support this splice variant

Output Fusion Files

Fusion calling is only performed in runs that use an encrypted manifest file. The fusion output files only contain fusion calls when using an encrypted manifest.

fusions.csv

The fusions.csv file is generated from fusion calls based on the Fusion target type. The file contains the columns listed below. The rows are sorted by score in descending order. Entries with scores below 2 are not included.

Column	Description
Target ID	The unique identifier for the fusion target.
Target Number	Target number from the manifest. Negative numbers indicate off-targets while positive numbers indicate on-targets.
Gene A	Gene symbol of fusion partner A.
Gene B	Gene symbol of fusion partner B.
Transcript ID A	Transcript ID of fusion partner A.
Transcript ID B	Transcript ID of fusion partner B.
Exon A	Exon(s) of fusion partner A.
Exon B	Exon(s) of fusion partner B.
Breakpoint A	The breakpoint of fusion on gene A with genomic coordinates <chrom>:<position>.
Breakpoint B	The breakpoint of fusion on gene B with genomic coordinates <chrom>:<position>.
Supporting Reads	The target read count.
Score	The score is calculated using the read count of the target and the total read count of the sample: $\frac{10^6 \times \text{targetreadcount}}{\text{totalreadcount}} - 32.74276$ <p style="text-align: center;">254.8784</p> <ul style="list-style-type: none"> • Where <i>totalreadcount</i> is the number of reads mapped to all the targets in the manifest. The normalized read count is centered by 32.74276 and scaled by 254.8784 to produce a z-score-like score, where the two constants were learned from control samples without known fusions. • The higher the score, the stronger the evidence of a fusion being present in the sample in question. Potential fusions with scores below 1 are not included in the fusions.csv file and potential fusions with scores lower than the 4 are filtered as LowScore.

Column	Description
Filter	<ul style="list-style-type: none"> • OffTarget— The target is an off-target. • LowScore— The score is less than the FusionScoreCutoff setting. • PASS— The target is an on-target and the score is at least the FusionScoreCutoff setting.
Fusion Call	<ul style="list-style-type: none"> • NoCall— The score is less than the FusionScoreCutoff setting. • Present— The target read count is greater than 0 and the score is at least the Fusion ScoreCutoff setting.

fusions.vcf.gz

This file contains information about detected gene fusion events.

The file is in VCF 4.1 format and compressed by gzip compression software. In addition to a format-compliant header, the file contains the following columns:

Column Heading	Description
CHROM	The chromosome on which the fusion occurs.
POS	The position of the gene fusion.
REF	The reference base preceding the fusion.
ALT	The alternative allele at the variant position.
ID	A unique identifier for the fusion event, formatted as: <Target ID>.<Target Number>.<A or B> for breakpoint A or B.
QUAL	Will always be "." (missing).
FILTER	<p>Contains information on whether the variant passed quality filtering (PASS) or the name of the filter that caused it to fail quality filtering.</p> <ul style="list-style-type: none"> • Entries with scores \geq FusionScoreCutoff (default value is 4.0) are marked as PASS and called Present. • Entries with scores $<$ FusionScoreCutoff and \geq FusionEmitScoreCutoff are filtered as LowScore and called NoCall. • Entries with scores $<$ FusionEmitScoreCutoff (default value is 1.0) are not included in the report. • The thresholds apply to both on- and off-targets. • Targets are checked for OffTarget and LowScore. The rest will be marked as PASS.

Column Heading	Description
INFO	<p>Consists of the following sub-fields:</p> <ul style="list-style-type: none"> • SVTYPE— BND <ul style="list-style-type: none"> • Two "SVTYPE=BND" records are created for each fusion call, one for each breakpoint) • MATEID— <Target ID>:<B or A> for breakpoint A or B. • READS— The target read count. • SCORE— The score is calculated using the read count of the target and the total read count of the sample: $\frac{\left[\frac{10^6 \times \text{targetreadcount}}{\text{totalreadcount}} - 32.74276 \right]}{254.8784}$ <ul style="list-style-type: none"> • Where <i>totalreadcount</i> is the number of reads mapped to all the targets in the manifest. The normalized read count is centered by 32.74276 and scaled by 254.8784 to produce a z-score-like score, where the two constants were learned from control samples without known fusions. • The higher the score, the stronger the evidence of a fusion being present in the sample in question. Potential fusions with scores below 1 are not included in the fusions.csv file and potential fusions with scores lower than the 4 are filtered as LowScore. • ANT— Annotation from the manifest for breakpoint A or B. Formatted as Gene Transcript Exon Strand, where multiple exon numbers are joined by a comma (,) and strand is either + or -.
FORMAT	<p>Consists of the following sub-fields:</p> <ul style="list-style-type: none"> • DP— The number of reads supporting this fusion. • SC— The score of this fusion.

putativefusions.csv

The putativefusions.csv file is generated from fusion calls based on the Balance5p3p and ExpressionControl target types.

Two Balance5p3p targets are designed for the driver gene, one on the 3' end and one on the 5' end. Reads aligned to these targets and the ExpressionControl targets are used to calculate the imbalance described below. The Imbalance score reflects the magnitude of 5'/3' imbalance for the targeted driver gene. A higher imbalance score may suggest a potential fusion event involving that driver gene, however additional testing would be necessary to confirm the partner gene and fusion break point. Currently, there are Balance5p3p targets available for four driver genes (ALK, RET, ROS1, NTRK1), all contained in the Focus Panel.

The file contains the following columns.

Column	Description
Gene	Gene symbol of the targeted fusion partner
Transcript ID	Transcript ID of the targeted fusion partner
Target ID 5'	Target ID of the amplicon on the 5' of the targeted fusion partner
Target ID 3'	Target ID of the amplicon on the 3' of the targeted fusion partner
Exon 5'	Exon(s) of the amplicon on the 5' of the targeted fusion partner
Exon 3'	Exon(s) of the amplicon on the 3' of the targeted fusion partner
Supporting Reads 5'	Read count of the 5' target of the gene.
Supporting Reads 3'	Read count of the 3' target of the gene.
Expression Control Reads	Number of reads aligned to targets of type ExpressionControl.
Imbalance	Calculated by the following equation: $\frac{(\text{SupportingReads3}' - \text{SupportingReads5}')}{\text{ExpressionControlReads}}$

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
WorkflowLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages.
WorkflowError.txt	Processing log that lists messages or errors that occurred during analysis.
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.
[Workflow]RunStatistics.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

Folder Data

Folder Intensities

Folder BaseCalls

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

Folder L001 — Contains *.locs files, 1 for each tile.

Folder RTA Logs — Contains log files from RTA software analysis.

Folder Alignment_Imported_#

Folder Time Stamp — Contains *.bam files (if applicable) and files specific to the analysis module.

Folder Fastq

Sample1_S1_L001_R1_001.fastq.gz

Sample2_S1_L001_R1_001.fastq.gz

Undetermined_S0_L001_R1_001.fastq.gz

Folder InterOp — Contains binary files used by Sequencing Analysis Viewer (SAV).

Folder Logs — Contains log files describing steps performed during sequencing.

 **Queued**—A working folder for software; also called the copy folder.

 CompletedJobInfo.xml

 [WorkflowName]RunStatistics

 RTAComplete.txt

 RunInfo.xml

 runParameters.xml

Alignment Folders

When analysis begins, the Local Run Manager creates an Alignment folder named **Alignment_#**, where # is a sequential number.

If you created the run by importing the information for a run that has already been analyzed, the Alignment folder is named **Alignment_Imported_#**.

Revision History

Document	Date	Description of Change
Document # 1000000048048 v02	March 2019	Added ERCC analysis to procedure and related information to output files section.
Document # 1000000048048 v01	April 2018	Added i5 index sequence instruction when creating custom prep kit sample sheets for use on the iSeq, MiniSeq, and NextSeq systems.
Document # 1000000048048 v00	January 2018	Initial release.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

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

Illumina Customer Support Telephone Numbers

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

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

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.