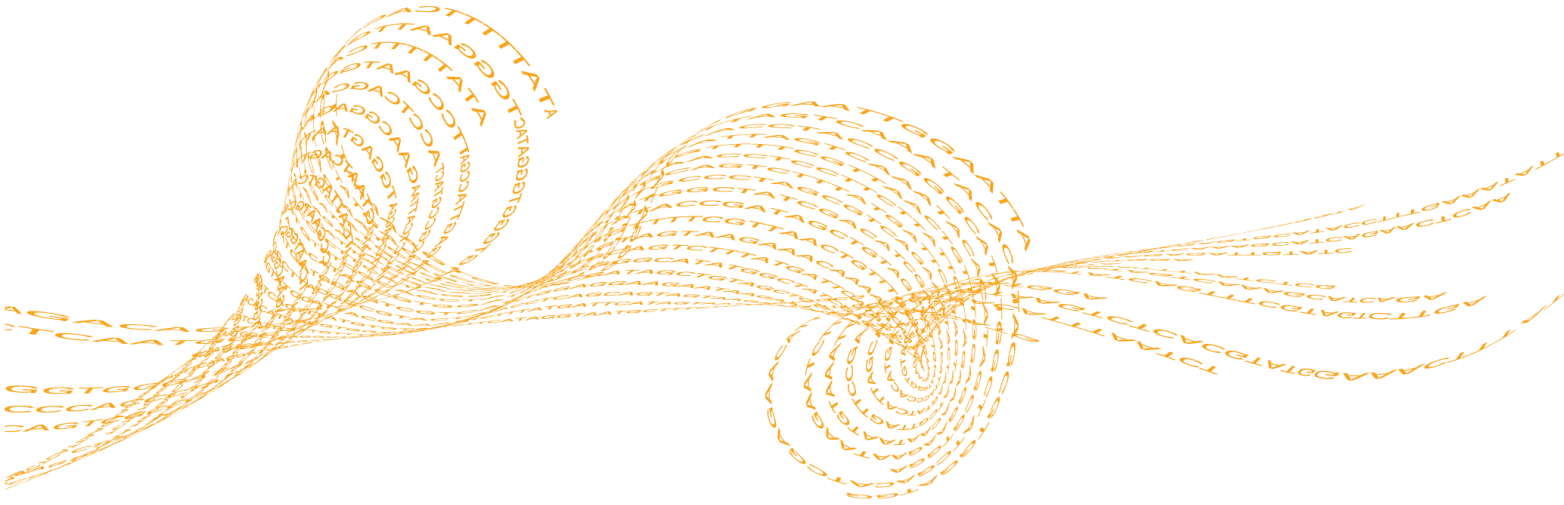
Local Run Manager
Amplicon Analysis Module
Workflow Guide

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



ILLUMINA PROPRIETARY
Document # 1000000002708 v00
January 2016

illumina®

Overview

The Local Run Manager Amplicon analysis module aligns reads against the reference specified in the manifest file using the banded Smith-Waterman algorithm, and then performs variant analysis.

Compatible Library Types

The Amplicon analysis module is compatible with specific library types represented by library kit categories on the Create Run screen. For a current list of compatible library kits, see the Local Run Manager support page on the Illumina website.

Input Requirements

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

- **Manifest file**—The Amplicon analysis module requires at least 1 manifest file. The manifest file is provided with either your custom assay (CAT) or from the Illumina website depending on the library kit used for the run.
- **Reference genome**—The Amplicon analysis module requires the reference genome specified in the manifest file. The reference genome provides variant annotations and sets the chromosome sizes in the BAM output files.

Uploading Manifests

To import a manifest for all runs using the Amplicon 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 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. Click **Create Run**, and select **Amplicon**.
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. From the Library Kit drop-down list, select from the following library kit categories.
   - TruSight Amplicon Panels
   - TruSeq Amplicon
2. Specify the number of cycles for the run.
3. [Optional] Specify any custom primers to be used for the run.

NOTE
By default, the Amplicon analysis module is set to 2 index reads of 8 cycles each and the read type Paired End.

Specify Module-Specific Settings

1. Expand the Variant Caller drop-down list and select a variant calling method.
   - **Starling**—(Default) Calls SNPs and small indels, and summarizes depth and probabilities for every site in the genome.
   - **Somatic**—Identifies variants at low frequency and minimizes false positives. Recommended for analysis of tumor samples.
   - **GATK**—Calls raw variants for each sample, analyzes variants against known variants, and then calculates a false discovery rate for each variant.
2. If using the Somatic Variant Caller, specify the following settings.
   - **Variant Frequency**—Set to a threshold of 0.05 by default. Variants with a frequency below the specified threshold are not reported in VCF files.
   - **Indel Repeat Filter Cutoff**—On by default. When enabled, indels are filtered when the reference has a 1-base or 2-base motif over 8 times next to the variant.
3. Click **Show advanced module settings** and click the **On/Off** toggle to specify the following analysis settings.
   - **Read Stitching**—On by default. When enabled, reads that overlap ≥ 10 bases between Read 1 and Read 2 are stitched together to form a single read.
   - **Variant Quality Filter**—Set to 30 by default. Variants with a variant quality score below the specified threshold are flagged as filtered in VCF files. Possible values are 2–1000.

NOTE
By default, the Amplicon analysis module uses the Smith-Waterman algorithm for alignment.

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 Amplicon 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 [Optional] Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, underscores, or spaces.

4 Expand the Index 1 (i7) drop-down list and select an Index 1 adapter.

5 Expand the Index 2 (i5) drop-down list and select an Index 2 adapter.

6 Expand the Manifest drop-down list and select a manifest file.

7 Expand the Genome Folder drop-down list and select a reference genome.

8 [Optional] Click the **Export** icon to export sample information in *.csv format.

9 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 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
- Identifies variants

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.

Read Stitching

The Amplicon analysis module performs read stitching by default.

When enabled, paired-end reads that overlap are stitched to form a single read in the FASTQ file. At each overlap position, the consensus stitched read has the base call and quality score of the read with higher Q-score.

For each paired read, a minimum of 10 bases must overlap between Read 1 and Read 2 to be a candidate for read stitching. The minimum threshold of 10 bases minimizes the number of reads that are stitched incorrectly due to a chance match. Candidates for read stitching are scored as follows:

- For each possible overlap of 10 base pairs or more, a mismatch score is calculated. Perfectly matched overlaps have a MismatchRate of 0, resulting in a score of 1.
- If the best overlap has a score of ≥ 0.9 ***and*** the score is ≥ 0.1 higher than any other candidate, then the reads are stitched together at this overlap.
- Paired-end reads that cannot be stitched are converted to 2 single reads in the FASTQ file.

Although the stitched reads are aligned as a single sequence, the stitched read is split into individual alignments in the BAM file.

Alignment

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

The banded Smith-Waterman algorithm performs local sequence alignments to determine similar regions between 2 sequences. Instead of comparing the total sequence, the Smith-Waterman algorithm compares segments of all possible lengths. Local alignments are useful for dissimilar sequences that are suspected to contain regions of similarity within the larger sequence. This process allows alignment across small amplicon targets, often less than 10 bp.

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

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

Alignments that include more than 3 indels are filtered from alignment results. Filtered alignments are written in alignment files as unaligned and are not used in variant calling.

Variant Calling

Variant calling records single nucleotide polymorphisms (SNPs), insertions/deletions (indels), and other structural variants in a standardized variant call format (VCF).

For each SNP or indel called, the probability of an error is provided as a variant quality score. Reads are realigned around candidate indels to improve the quality of the calls and site coverage summaries.

The Amplicon analysis module provides the option of using Starling, Somatic, or GATK for variant calling.

Starling

Starling calls both SNPs and small indels, and summarizes depth and probabilities for every site in the genome. Starling produces a VCF file for each sample that contains variants.

Starling treats each insertion or deletion as a single mismatch. Base calls with more than 2 mismatches to the reference sequence within 20 bases of the call are ignored. If the call occurs within the first or last 20 bases of a read, the mismatch limit is increased to 41 bases.

Somatic Variant Caller

Developed by Illumina, the somatic variant caller identifies variants present at low frequency in the DNA sample and minimizes false positives.

The somatic variant caller identifies SNPs in 3 steps:

- Considers each position in the reference genome separately
- Counts bases at the given position for aligned reads that overlap the position
- Computes a variant score that measures the quality of the call using a Poisson model. Variants with a quality score below Q20 are excluded.

The somatic variant caller analyzes how many alignments covering a given position include a particular indel compared to the overall coverage at that position.

GATK

The Genome Analysis Toolkit (GATK) calls raw variants for each sample, analyzes variants against known variants, and then calculates a false discovery rate for each variant. Variants are flagged as homozygous (1/1) or heterozygous (0/1) in the VCF file sample column. For more information, see www.broadinstitute.org/gatk.

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.
   - From the left navigation bar, select a sample name to view the report for another sample.
6. [Optional] Click the **Copy to Clipboard** icon for access to the Analysis folder.

Analysis Report

Analysis results are summarized on the Samples and Results tab. The report is also available in a PDF file format for each sample and as an aggregate report in the Analysis folder.

Sample Information

Table 1 Sample Information Table

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 ≥ Q30.

Amplicon Summary

Table 2 Amplicon Summary Table

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 Amplicon Regions	The total length in base pairs of sequenced amplicons in the target regions.

Read Level Statistics

Table 3 Read Level Statistics Table

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

Table 4 Base Level Statistics Table

Column Heading	Description
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).
Mismatch Rate	The percentage of bases that did not align to the reference averaged over cycles per read (Read 1 and Read 2).

Small Variants Summary

Table 5 Small Variants Summary Table

Row Heading	Description
Total Passing	The total number of variants passing filter for single nucleotide variations (SNVs), insertions, and deletions.
Het/Hom Ratio	The ratio of the number of heterozygous SNPs and number of homozygous SNPs detected for the sample.
Ts/Tv Ratio	The ratio of transitions and transversions in SNPs. • Transitions are variants of the same nucleotide type (pyrimidine to pyrimidine, C and T; or purine to purine, A and G). • Transversions are variants of a different nucleotide type (pyrimidine to purine, or purine to pyrimidine).

Coverage Summary

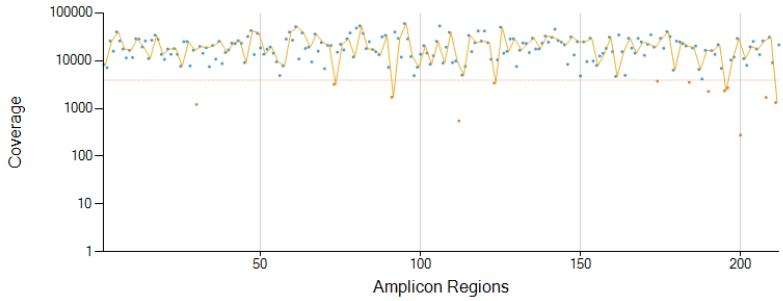
Table 6 Coverage Summary Table

Column Heading	Description
Amplicon Mean Coverage	The total number of aligned bases divided by the targeted region size.
Uniformity of Coverage	The percentage of amplicon regions with coverage values greater than the low coverage threshold of 0.2 * amplicon mean coverage.

Coverage by Amplicon Region Plot

The Coverage by Amplicon Region plot show the coverage across amplicon regions. Regions with coverage values lower than the coverage threshold are highlighted in red. The average of all values is indicated by an orange line.

Figure 1 Coverage by Amplicon Region Plot (Example)



Analysis Output Files

The following analysis output files are generated for the Amplicon analysis module and provide analysis results for alignment and variant calling. 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.
Per-Pool variant call files in the VCF format (*.vcf)	Contains variants called at each position from either the forward pool or the reverse pool.
Variant call files in the genome VCF format (*.genome.vcf)	Contains the genotype for each position, whether called as a variant or called as a reference.
Consensus variant call files in the VCF format (*.vcf)	Contains variants called at each position from both pools.
AmpliconCoverage_M1.tsv	Contains information about coverage per amplicon per sample for each manifest provided. M# represents the manifest number.

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

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.

VCF File Format

VCF is a widely used file format developed by the genomics scientific community that contains information about variants found at specific positions in a reference genome.

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

VCF File Header—Includes the VCF file format version and the variant caller version. The header lists the annotations used in the remainder of the file. If MARS is listed, the Illumina internal annotation algorithm annotated the VCF file. The VCF header includes

the reference genome file and BAM file. The last line in the header contains the column headings for the data lines.

VCF File Data Lines—Each data line contains information about a single variant.

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For SNPs, this position is the reference base with the variant; for indels or deletions, this position is the reference base immediately before the variant.
ID	The rs number for the SNP obtained from dbSNP.txt, if applicable. If there are multiple rs numbers at this location, the list is semicolon delimited. If no dbSNP entry exists at this position, a missing value marker ('.') is used.
REF	The reference genotype. For example, a deletion of a single T is represented as reference TT and alternate T. An A to T single nucleotide variant is represented as reference A and alternate T.
ALT	The alleles that differ from the reference read. For example, an insertion of a single T is represented as reference A and alternate AT. An A to T single nucleotide variant is represented as reference A and alternate T.
QUAL	A Phred-scaled quality score assigned by the variant caller. Higher scores indicate higher confidence in the variant and lower probability of errors. For a quality score of Q, the estimated probability of an error is $10^{-(Q/10)}$. For example, the set of Q30 calls has a 0.1% error rate. Many variant callers assign quality scores based on their statistical models, which are high in relation to the error rate observed.

VCF File Annotations

Heading	Description
FILTER	If all filters are passed, **PASS** is written in the filter column. • **LowDP**—Applied to sites with depth of coverage below a cutoff. • **LowGQ**—The genotyping quality (GQ) is below a cutoff. • **LowQual**—The variant quality (QUAL) is below a cutoff. • **LowVariantFreq**—The variant frequency is less than the given threshold. • **R8**—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8. • **SB**—The strand bias is more than the given threshold. Used with the Somatic Variant Caller and GATK.

Heading	Description
INFO	Possible entries in the INFO column include: • **AC**—Allele count in genotypes for each ALT allele, in the same order as listed. • **AF**—Allele Frequency for each ALT allele, in the same order as listed. • **AN**—The total number of alleles in called genotypes. • **CD**—A flag indicating that the SNP occurs within the coding region of at least 1 RefGene entry. • **DP**—The depth (number of base calls aligned to a position and used in variant calling). • **Exon**—A comma-separated list of exon regions read from RefGene. • **FC**—Functional Consequence. • **GI**—A comma-separated list of gene IDs read from RefGene. • **QD**—Variant Confidence/Quality by Depth. • **TI**—A comma-separated list of transcript IDs read from RefGene.
FORMAT	The format column lists fields separated by colons. For example, GT:GQ. The list of fields provided depends on the variant caller used. Available fields include: • **AD**—Entry of the form X,Y, where X is the number of reference calls, and Y is the number of alternate calls. • **DP**—Approximate read depth; reads with MQ=255 or with bad mates are filtered. • **GQ**—Genotype quality. • **GQX**—Genotype quality. GQX is the minimum of the GQ value and the QUAL column. In general, these values are similar; taking the minimum makes GQX the more conservative measure of genotype quality. • **GT**—Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available. • **NL**—Noise level; an estimate of base calling noise at this position. • **PL**—Normalized, Phred-scaled likelihoods for genotypes. • **SB**—Strand bias at this position. Larger negative values indicate less bias; values near 0 indicate more bias. Used with the Somatic Variant Caller and GATK. • **VF**—Variant frequency; the percentage of reads supporting the alternate allele.
SAMPLE	The sample column gives the values specified in the FORMAT column.

Genome VCF Files

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

The gVCF file shows no-calls at positions with low coverage, or where a low-frequency variant (< 3%) occurs often enough (> 1%) that the position cannot be called to the reference. A genotype (GT) tag of **./.** indicates a no-call.

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

Amplicon Coverage File

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

Each file includes a header row that contains the sample IDs associated with the manifest. Under the header row are 3 columns that list the following information:

- The Target ID as it is listed in the manifest.
- The coverage depth of reads passing filter.
- The total coverage depth.

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.
AmpliconRunStatistics.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.

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, click **Show advanced module settings**.
2. Click **Add custom setting**.
3. In the custom setting field, enter the setting name as listed in the Available Analysis Settings section.
4. In the setting value field, enter the setting value.
5. To remove a setting, click the **x** icon.

Available Analysis Settings

- **Variant Frequency**—Filters variants with a frequency less than the specified threshold. If using the Somatic Variant Caller, adjust the value for this setting on the Create Run screen.

Setting Name	Setting Value
VariantFrequencyFilterCutoff	Enter a threshold value. With the Somatic Variant Caller, the default value is 0.05. With GATK or Starling, the default value is 0.20.

- **Indel Repeat Cutoff**—Filters insertions and deletions when the reference has a 1-base or 2-base motif over 8 times (by default) next to the variant. If using the Somatic Variant Caller, enable or disable this setting on the Create Run screen.

Setting Name	Setting Value
IndelRepeatFilterCutoff	Enter a threshold value. The default value is 8.

- **Variant Genotyping Quality**—Filters variants with a genotype quality (GQ) less than the specified threshold.

Setting Name	Setting Value
VariantMinimumGQCutoff	Enter a value less than 99. With GATK or Somatic Variant Caller, the default value is 30. With Starling, the default value is 20.

- **Variant Quality Cutoff**—Filters variants with a quality (QUAL) less than the specified threshold. QUAL indicates the confidence of the variant call.

Setting Name	Setting Value
VariantMiniumQualCutoff	Enter a threshold value. With GATK or Somatic Variant Caller, the default value is 30. With Starling, the default value is 20.

Notes

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 7 Illumina General Contact Information

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

Table 8 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