

Local Run Manager Amplicon DS Analysis Module Workflow Guide

Overview	3
Set Parameters	3
Analysis Methods	5
View Analysis Results	7
Analysis Report	7
Analysis Output Files	9
Custom Analysis Settings	15
Revision History	17
Technical Assistance	19

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.

Overview

The Local Run Manager Amplicon DS analysis module aligns reads against the reference specified in the manifest files using the banded Smith-Waterman algorithm. After alignment, the somatic variant caller performs variant analysis. This workflow is designed specifically for dual-strand targeted resequencing assays.

Compatible Library Types

The Amplicon DS 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 DS analysis module requires the following files.

- ▶ **Manifest files (2)**—The Amplicon DS analysis module requires 2 assay-specific manifest files: a manifest file for the forward pool and a manifest file for the reverse pool. Manifest files are available for download from the Illumina website.
- ▶ **Reference genome**—The Amplicon DS analysis module requires the hg19 reference genome for coordinates and chromosome mapping. The reference genome is included in the Local Run Manager software installation on-instrument, but needs to be installed separately when running the analysis module with Local Run Manager off-instrument.

Uploading Manifests

To import a manifest file for all runs using the Amplicon DS analysis module, use the Modules & Manifests option from the Tools drop-down 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 Amplicon DS 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 If needed, log in to Local Run Manager.
- 2 Select **Create Run**, and select **Amplicon DS**.
- 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 identify the run.
The run description can contain alphanumeric characters, spaces, and the following special characters:
`~!@#\$\$%_-_{}.

Specify Run Settings

- 1 Select the library prep kit from the Library Prep Kit drop-down list.
- 2 Specify the number of cycles for the run.
- 3 If applicable, select the number of index reads for the run.
Some library prep kits have hard-coded indexes and cannot be changed.
- 4 If applicable, select the read type for the run.
Some library prep kits have hard-coded read types and cannot be changed.
- 5 Specify the read lengths for the run.
- 6 **[Optional]** For Custom Primers, specify any custom primer information to be used for the run by selecting the appropriate checkboxes.
Custom primer options vary based on your instrument or Local Run Manager implementation.



NOTE

By default, the Amplicon DS analysis module is set to two index reads of eight cycles each and the read type Paired End.

Specify Module-Specific Settings

- 1 Select the **On/Off** toggle to enable or disable the Indel Repeat Filter Cutoff setting.
 - ▶ **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.



NOTE

By default, the Amplicon DS analysis module uses the Smith-Waterman algorithm for alignment and the Somatic Variant Caller for variant analysis.

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 Amplicon DS 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 the following options:

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

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

Enter Samples Manually

- 1 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 **X** 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 this field.
 - 3 Enter a sample name in the Sample Name field.
The sample name is required to connect strand A and strand B with a common name.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
 - 4 **[Optional]** Enter a sample description in the Description field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
 - 5 If you have a plated kit, select an index plate well from the Index well drop-down list and skip to step 6.
 - 6 Select a manifest file from the Manifest drop-down list.
 - 7 **[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.
 - 8 Select **Save Run**.

Import Sample Sheet

- 1 If you do not have a sample sheet to import, see [Enter Samples Manually on page 4](#) 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.
- 3 When finished, select **Save Run**.

Analysis Methods

The Amplicon DS 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.

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. In the Amplicon DS workflow, one FASTQ file is generated for each pool and each read per sample—one each from pool A Read 1, pool A Read 2, pool B Read 1, and pool B Read 2. FASTQ files are the primary input for alignment.

Alignment

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

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

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

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

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

Variant Calling

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




The somatic variant caller identifies SNPs in three steps:

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

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

- ▶ The variant is present in both pools
- ▶ Has a cumulative depth of 1000 or an average depth of 500x per pool
- ▶ Has a variant frequency of $\geq 3\%$ as reported in the merged VCF file

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 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.
 - ▶ From the left navigation bar, select a sample ID to view the report for another sample.
- 7 **[Optional]** Select the **Copy to Clipboard**  icon to copy the Analysis Folder file path.

Analysis Report

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

Sample Information

Column Heading	Description
Sample ID	The sample ID provided when the run was created.
Sample Name	The sample name provided when the run was created.
Run Folder	The name of the run folder.
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
Pool Name	The name of the file that specifies the reference for pool A (Pool_FPA) and pool B (Pool_FPB).
Number of Amplicon	The number of amplicon regions sequenced as specified in the manifest file. Listed for pool A and pool B.
Total Length of Amplicons	The total length in base pairs of sequenced amplicons in the target regions. Listed for pool A and pool B.

Read Level Statistics

Column Heading	Description
Pool_FPA or Pool_FPB	Read level statistics are listed separately for each pool.
Total Aligned Reads	The total number of reads that aligned to the reference for each read (Read 1 and Read 2) and the total of 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) and the percentage of Read 1 and Read 2 combined.

Base Level Statistics

Column Heading	Description
Pool_FPA or Pool_FPB	Base level statistics are listed separately for each pool.
Total Aligned Bases	The total number of bases that aligned to the reference for each read (Read 1 and Read 2) and the total of Read 1 and Read 2.
Percent Aligned Bases	The percentage of aligned bases averaged over cycles per read (Read 1 and Read 2) and the total of Read 1 and Read 2.
Percent Q30	The percentage of bases called with a quality score \geq Q30.
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

Row Heading	Description
Total Passing	The total number of variants passing filter for single nucleotide variations (SNVs), insertions, and deletions.
Percent Found in dbSNP	The percentage of variants called by the variant caller that are also present in dbSNP.
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. <ul style="list-style-type: none"> • 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

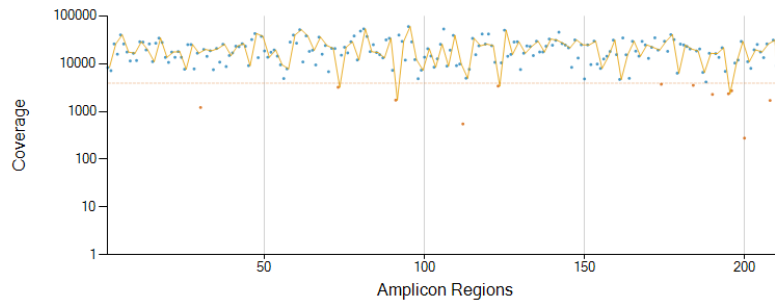
Column Heading	Description
Amplicon Mean Coverage	The total number of aligned bases divided by the targeted region size.
Uniformity of Coverage	The percentage of amplicon regions with coverage values greater than the low coverage threshold of $0.2 * \text{amplicon mean coverage}$.

Coverage by Amplicon Region Plot

The Coverage by Amplicon Region plot shows 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.

A plot is provided for the overall coverage, coverage for pool FPA, and pool FPB.

Figure 1 Coverage by Amplicon Region Plot (Example)



Analysis Output Files

The following analysis output files are generated for the Amplicon DS analysis module and provide analysis results for alignment and variant calling. Analysis output files are located in the Alignment_* folder unless otherwise specified.

File Name	Description
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_M#.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 is written to a demultiplexing (*.demux) file for each tile of the flow cell.

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

Demultiplexing files start with a header:

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

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

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

- ▶ In the file name, **F1** represents the flow cell number.
- ▶ In the file name, **L1** represents the lane number.
- ▶ Demultiplexing results in a table with one row per tile and one column per sample, including sample 0.
- ▶ The most commonly occurring sequences in index reads.

FASTQ File Format

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

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

The identifier is formatted as:

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

Example:

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

BAM File Format

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

BAM files use the file naming format of SampleName_S#.bam, in which # 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.
- ▶ **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 index files (*.bam.bai) provide an index of the corresponding BAM file.

VCF File Format

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

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

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For SNVs, this position is the reference base with the variant. For indels, this position is the reference base immediately preceding the variant.
ID	The rs number for the SNP obtained from dbSNP.txt, if applicable. If multiple rs numbers exist at this location, semicolons delimit the list. If a dbSNP entry does not exist 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	<p>If all filters are passed, PASS is written in the filter column.</p> <ul style="list-style-type: none"> • 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. • PB—Detected probe-pool bias. The variant is not found, or found with low frequency, in 1 of 2 probe pools.

Heading	Description
INFO	<p>Possible entries in the INFO column include:</p> <ul style="list-style-type: none"> • 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	<p>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:</p> <ul style="list-style-type: none"> • 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. • PB— Probe-pool bias. Values closer to 0 indicate more bias toward 1 probe pool and less confidence in a variant call.
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.

Per-Pool and Consensus VCF Files

The Amplicon DS workflow generates two sets of variant call files.

- ▶ **Per-pool VCF files**— Contain variants called in either the forward pool or the reverse pool. Per-pool files are written to the VariantCallingLogs folder.
- ▶ **Consensus VCF files**— Contain variants called from both pools. Consensus files are written to the Alignments folder.

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

- ▶ **Reports for all sites**— SampleName_*S#*.genome.vcf

- ▶ **Reports variants only**—SampleName_S#.vcf

Per-Pool VCF Files

Variants are called in the forward pool and the reverse pool to produce an independent set of VCF files for each pool.

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

- ▶ Include variants that were flagged as filtered
- ▶ Exclude variants with a variant frequency of less than 3%
- ▶ Variants that pass filters include PASS in the FILTER column
- ▶ Variants that fail filters include the filter name in the FILTER column
- ▶ Filter variants due to probe bias (PB) when the variant frequency differs significantly between pools

Consensus VCF Files

The software compares the per-pool VCF files and combines the data at each position to create a consensus VCF file for the sample.

Variant calls from each pool are merged into consensus VCF files using the following criteria.

Criteria	Result
A reference call in each pool	Reference call
A reference call in 1 pool and a variant call in the other pool	Filtered variant call
Matching variant calls with similar frequencies in each pool	Variant call
Matching variant calls with significantly different frequencies in each pool	Filtered variant call
Unmatched variant calls in each pool	Filtered variant call

Metrics from each pool are merged using the following values.

Metric	Value
Depth	Addition of depths from both pools
Variant Frequency	Total variant counts divided by total coverage depth
Q-Score	Minimum value of both pools

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 three 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 will be empty if no errors occurred. Located in the root level of the run folder.
CompletedJobInfo.xml	Written after analysis is complete. Contains information about the run, such as date, flow cell ID, software version, and other parameters. Located in the root level of the run folder.
DemultiplexSummaryF1L1.txt	Reports demultiplexing results in a table with one row per tile and one column per sample.
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.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

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


 **Data**

 **Intensities**

 **BaseCalls**

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

 Sample1_S1_L001_R1_001.fastq.gz

 Sample2_S1_L001_R1_001.fastq.gz

 Undetermined_S0_L001_R1_001.fastq.gz

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

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

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

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

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

 AnalysisError.txt

 AnalysisLog.txt

 CompletedJobInfo.xml

 QueuedForAnalysis.txt

 [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_#**.

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 the Available Analysis Settings section.
- 4 In the setting value field, enter the setting value.
- 5 To remove a setting, select .

Available Analysis Settings

- ▶ **Variant Frequency**—Filters variants with a frequency less than the specified threshold.

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

- ▶ **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 the Somatic Variant Caller, the default value is 30.

- ▶ **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 the Somatic Variant Caller, the default value is 30.

Revision History

Document	Date	Description of Change
Document # 1000000003341 v01	June 2018	Updated to match updates in user interface. Corrections in Analysis Reports and Analysis Output sections.
Document # 1000000003341 v00	January 2016	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	
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.



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