

Local Run Manager PCR Amplicon Analysis Module

Workflow Guide

Overview	3
Set Parameters	3
Analysis Methods	6
View Analysis Results	8
Analysis Report	8
Analysis Output Files	11
Custom Analysis Settings	14
Technical Assistance	16



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 PCR Amplicon analysis module aligns reads against the reference genomes specified in the sample sheet, and performs variant analysis for the regions specified in the manifest. This workflow is intended for analysis of PCR amplicons that have been fragmented using Nextera fragmentation.

Compatible Library Types

The PCR 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 PCR Amplicon analysis module requires the following files.

- ▶ **Manifest file**—The PCR Amplicon analysis module requires at least one manifest file.
- ▶ **Reference genome**—The PCR 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 file output.

Uploading Manifests

To import a manifest file for all runs using the PCR Amplicon 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 PCR 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 If needed, log in to Local Run Manager.
- 2 Select **Create Run**, and select **PCR 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 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 index reads.

- ▶ 0 for a run with no indexing
- ▶ 1 for a single-indexed run
- ▶ 2 for a dual-indexed run

Unsupported index reads for your library prep kit are automatically disabled.

- 3 Specify a read type: **Single Read** or **Paired End**.
If your library prep kit supports only one option, the read type is automatically selected.
- 4 Enter the number of cycles for the run.
- 5 **[Optional]** If using custom primers, specify their information.
Custom primer options may vary based on your instrument or Local Run Manager implementation.

Specify Module-Specific Settings

- 1 Select an alignment method from the Aligner drop-down list.
 - ▶ **BWA-MEM**—(Default) Optimized for Illumina sequencing data and reads ≥ 70 bp.
 - ▶ **BWA-Backtrack Legacy**—Use with legacy data or reads < 70 bp.
- 2 Select a variant calling method from the Variant Caller drop-down list.
 - ▶ **GATK**—(Default) Calls raw variants for each sample, analyzes variants against known variants, and then calculates a false discovery rate for each variant.
 - ▶ **Somatic**—Identifies variants at low frequency and minimizes false positives. Recommended for analysis of tumor samples.
- 3 Select the **On/Off** toggle to enable or disable the following settings.
 - ▶ **Export to gVCF**—Off by default. When enabled, gVCF files will output to the run folder.
 - ▶ **Flag PCR Duplicates**—On by default. When enabled, PCR duplicates are flagged in the BAM files and not used for variant calling. PCR duplicates are defined as two clusters from a paired-end run where both clusters have the exact same alignment position for each read.
 - ▶ **Reverse Complement**—(Only available with Nextera Mate Pair library prep kits) Off by default. When enabled, all reads are reverse-complemented as they are written to FASTQ files.

Specify Advanced Module Settings

- 1 Select **Show advanced module settings** to view the available module settings.
- 2 **[Optional]** Enter a value, **2–1,000**, to define the Variant Quality Filter.
The default value is 30. Variants with a variant quality score below the specified threshold are flagged as filtered in VCF files.

PCR Amplicon Manifest File Format

A manifest file is required input for the PCR Amplicon workflow. The manifest file can be generated using the Create Manifest feature in Local Run Manager. Manifest files with probes and targets are supported.



Specify a Manifest File

Specify a manifest file for the run using one of the following options:

- ▶ **Create a manifest**—Use the Create Manifest function at the top of the Create Run screen.
- ▶ **Import a manifest**—Navigate to a manifest file that is available locally or in a network-accessible location.

The created or imported manifest files will be available in the Manifest drop-down list in the sample table.

Create a Manifest File

- 1 Select **Create Manifest**.
- 2 Select a reference genome from the Genome drop-down list.
- 3 Adjust the manifest entry 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.
- 4 Enter a unique amplicon name in the Amplicon Name field.
Use alphanumeric characters, dashes, or underscores.
- 5 Select a chromosome from the Chromosome drop-down list.
- 6 Specify the chromosome start and end positions in the Amplicon Start and Amplicon End fields.
- 7 Specify the upstream and downstream probe lengths in the Upstream Probe Length and Downstream Probe Length fields.
- 8 Enter a unique name in the Manifest Name field.
Use alphanumeric characters only. Dashes, underscores, or special characters are not allowed.
- 9 Select **Save**.
The manifest file can be used in all future runs.

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 PCR 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 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  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.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
- 4 **[Optional]** Enter a sample description in the Sample 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 Select a reference genome from the Genome Folder drop-down list.
- 8 **[Optional]** Enter a project name in the Sample Project field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
- 9 **[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.
- 10 Select **Save Run**.

Import Sample Sheet

- 1 If you do not have a sample sheet to import, see 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 sample sheet is correct. Incorrect information can impact the sequencing run.
- 3 When finished, select **Save Run**.

Sample Sheet Fields

Manual editing of the sample sheet is intended for technically advanced users. If settings are applied incorrectly, serious problems can occur.

Visit the Local Run Manager support page for available sample sheet settings. Settings must be entered as specified to avoid analysis failure.

Analysis Methods

The PCR Amplicon analysis module performs the following analysis steps and then writes analysis output files to the folder.

- ▶ Demultiplexes index reads
- ▶ Generates FASTQ files
- ▶ Aligns to a reference
- ▶ Identifies variants
- ▶ Reports statistics

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.

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

Reads are aligned against the entire reference genome using the Burrows-Wheeler Aligner (BWA), which aligns relatively short nucleotide sequences against a long reference sequence. BWA automatically adjusts parameters based on read lengths and error rates, and then estimates insert size distribution.

Paired-End Evaluation

For paired-end runs, the top-scoring alignment for each read is considered. Reads are flagged as an unresolved pair under the following conditions:

- ▶ If either read did not align, or the paired reads aligned to different chromosomes.
- ▶ If 2 alignments come from different amplicons or different rows in the Targets section of the manifest.

Bin/Sort

The bin/sort step groups reads by sample and chromosome, and then sorts by chromosome position. Results are written to 1 BAM file per sample.

Indel Realignment

Reads near detected indels are realigned to remove alignment artifacts.

Variant Calling

SNPs and short indels are identified using the Genome Analysis Toolkit (GATK), by default. 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.




Variant Annotation

Variant analysis is performed only for the amplicon regions specified in the manifest file.

Statistics Reporting

Statistics are summarized and reported, and written to the Alignment folder.

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.
- 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	Description
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Run Folder	The folder location of the run.
Total PF Reads	The number of reads passing filter.
Percent Q30	The percentage of reads with a quality score of 30 (Q30) or greater.
Median Read Length	The median fragment length for the sample.
Adapters Trimmed	States whether adapter trimming was specified for the run.

Enrichment Summary

Column	Description
Target Manifest	The specified manifest file.
Total Length of Targeted Reference	The length of the targeted reference region specified in the manifest file.
Padding Size	The length of sequence immediately upstream and downstream of the enrichment targets that is included for a padded target.

Read Level Enrichment

Column	Description
Total Aligned Reads	The total number of reads passing filter that aligned to the reference genome.
Percent Aligned Reads	The percentage of reads that aligned to the reference genome [$100 \times (\text{Total Aligned Reads} / \text{Total PF Reads})$].
Targeted Aligned Reads	The total aligned bases in the target region.
Read Enrichment	The percentage of aligned bases in the target region [$100 \times (\text{Target Aligned Reads} / \text{Total Aligned Reads})$].
Padded Target Aligned Reads	Number of reads that aligned to the padded target region.
Padded Read Enrichment	The percentage of aligned bases in the padded target region [$100 \times (\text{Padded Target Aligned Reads} / \text{Total Aligned Reads})$].

Base Level Enrichment

Column	Description
Total Aligned Bases	The total number of bases passing filter that aligned to the reference genome.
Percent Aligned Bases	The percentage of bases that aligned to the reference genome.
Targeted Aligned Bases	The total aligned bases in the target region.
Base Enrichment	The percentage of aligned bases in the target region [$100 \times (\text{Target Aligned Bases} / \text{Total Aligned Bases})$].
Padded Target Aligned Bases	Number of bases that aligned to the padded target region.
Padded Base Enrichment	The percentage of aligned bases in the padded target region [$100 \times (\text{Padded Target Aligned Bases} / \text{Total Aligned Bases})$].

Small Variants Summary

Row	Description
Small Variants Summary	The total number of Single Nucleotide Variants, insertions, and deletions passing the quality filters.
Percent found in dbSNP	The percentage of SNVs, insertions, and deletions that were found in the dbSNP.
Het/Hom Ratio	Number of heterozygous SNVs/Number of homozygous SNVs, insertions, and deletions.
Ts/Tv Ratio	The number of Transition SNVs, insertions, and deletions that pass the quality filters divided by the number of Transversion SNVs, insertions, and deletions that pass the quality filters. Transitions are interchanges of purines (A, G) or of pyrimidines (C, T). Transversions are interchanges of purine and pyrimidine bases (for example, A to T).

Variants by Sequence Context

Row	Description
In Genes	The number of variants that fall into a gene. Genes include exons, introns and UTR regions.
In Exons	The number of variants that fall into an exon. Exons include coding and UTR regions.
In Coding Regions	The number of variants that fall into a coding region.
In UTR Regions	The number of variants that fall into an untranslated region (UTR). UTR regions include 5' and 3' UTR regions.
In Splice Site Regions	The number of variants that fall into a splice site region. Splice site regions include regions annotated as splice acceptor, splice donor, splice site or splice region.

Variants by Consequence

Row	Description
Frameshifts	The number of variants that cause a frameshift.
Non-synonymous	The number of variants that cause an amino acid change in a coding region.
Synonymous	The number of variants that are within a coding region, but do not cause an amino acid change.
Stop Gained	The number of variants that cause an additional stop codon.
Stop Lost	The number of variants that cause the loss of a stop codon.

Coverage Summary

Report	Description
Mean Region Coverage Depth	The total number of aligned bases to the targeted region divided by the targeted region size.
Uniformity of Coverage (Pct > 0.2*mean)	The percentage of targeted base positions in which the read depth is greater than 0.2 times the mean region target coverage depth.
Target Coverage at 1X	Percentage targets with coverage greater than 1X.
Target Coverage at 10X	Percentage targets with coverage greater than 10X.
Target Coverage at 20X	Percentage targets with coverage greater than 20X.
Target Coverage at 50X	Percentage targets with coverage greater than 50X.
Mean Coverage by Targeted Region	A line graph showing the moving average of all coverage values by targeted region.
Depth of Coverage in Targeted Regions	A histogram showing the number of targeted bases covered by the depth of sequencing coverage. The table below the histogram shows the number of targeted bases covered at sequencing depths 1X, 10X, 20X, 30X, 50X, 100X, 150X, and 200X, total targeted bases covered, and the target coverage percentage.

Fragment Length Summary

Report	Description
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Minimum	Minimum length of the sequenced fragment.
Maximum	Maximum length of the sequenced fragment.
Standard Deviation	Standard deviation of the sequenced fragment length.
Fragment Length Distribution	A histogram showing the number of fragments by fragment length.

Additional Information

Report	Description
Duplicate Information	Percentage of duplicate paired reads. Duplicates are only reported for paired-end reads and if PCR duplicate flagging was selected in the settings.
Settings	Settings (reference genome, depth threshold, PCR duplicates, targeted regions, base padding, and Picard HS Metrics) specified during analysis set-up.
Software Versions	Software versions of Isis (Analysis Software), SAMtools, BWA (Aligner), Somatic Variant Caller, and IAS (Annotation Service).

Analysis Output Files

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

File Name	Description
Alignment files in the BAM format (*.bam)	Contains aligned reads for a given sample.
Variant call files in the VCF format (*.vcf)	Contains information about variants found at specific positions in a reference genome.

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_S#.bam. The variable, #, is the sample number determined by the order that samples are listed for the run. In multinode mode, the S# is set to S1, regardless of the order of the sample.

BAM files contain a header section and an alignment section:

- ▶ **Header**—Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- ▶ **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and 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.
- ▶ **MD**—Mismatching positions/bases (BWA only).
- ▶ **MQ**—Mapping quality.
- ▶ **AS**—Paired-end alignment quality.
- ▶ **NM**—Edit distance tag, which records the Levenshtein distance between the read and the reference.
- ▶ **XS**—Suboptimal alignment score.

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

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"> • ForcedReport—Filter if the variant would normally fail emit filters. Is printed to VCF because it is in one of the genotypes of interest VCF files. • LowDP—Applied to sites with depth of coverage below a cutoff. • LowGQ—The genotyping quality (GQ) is below a cutoff. • LowVariantFreq—The variant frequency is less than the given threshold. • MultiAllelicSite—Filter if the variant is in a multiallelic site that breaks ploidy assumptions. Only applicable to germline variant calling. • q{threshold}—Quality below {threshold}. • R8—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8. • R{thresholdM}x{thresholdN}—Filter if the variant is in a repeat region, where a repeat is defined as any region where the reference has motif up to length thresholdM that repeats thresholdN or more times. Only applicable to indels that contain the repeat motif, and are under the cutoff frequency. • SB—The strand bias is more than the given threshold.
INFO	<p>Possible entries in the INFO column include:</p> <ul style="list-style-type: none"> • AF1000G—The allele frequency from all populations of 1000 genomes data. • AA—The inferred allele ancestral (if determined) to the chimpanzee/human lineage. • clinvar—Clinical significance. Format: GenotypeIndex Significance. • cosmic—The numeric identifier for the variant in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Format: GenotypeIndex Significance. • CSQ—Consequence type as predicted by IAE. Format: GenotypeIndex HGNC Transcript ID Consequence. • CSQR—Predicted regulatory consequence type. Format: GenotypeIndex RegulatoryID Consequence • DP—The total depth (number of base calls aligned to a position and used in variant calling). • EVS—Allele frequency, coverage, and sample count taken from the Exome Variant Server (EVS). Format: AlleleFreqEVS EVSCoverage EVSSamples. • GMAF—Global minor allele frequency (GMAF); technically, the frequency of the second most frequent allele. Format: GlobalMinorAllele AlleleFreqGlobalMinor. • phyloP—PhyloP conservation score. Denotes how conserved the reference sequence is between species throughout evolution. • RefMinor—Denotes positions where the reference base is a minor allele and is annotated as though it was a variant.
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—Allele Depth; if the GT is 0/0, the AD is the reference count. If the GT is 0/1 or 1/1, the AD is of the form X,Y, where X is the reference allele count and Y is the alternative allele count. If the GT is 1/2, the AD is of the form Y,Z, where Y and Z are the alternative allele 1 and 2 counts. • DP—Total depth used for variant calling. • GQ—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. • SB—Strand bias at this position. Larger negative values indicate less bias; values near 0 indicate more bias. • VF—Variant frequency; if the GT is 0/0, the VF is the nonreference frequency. If the GT is 0/1 or 1/1, the VF is the frequency of the variant allele. If the GT is 1/2, the VF is the frequency of the two variant alleles, together.
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.

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.


File Name	Description
AdapterTrimming.txt	Lists the number of trimmed bases and percentage of bases for each tile. This file is present only if adapter trimming was specified for the run.
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.
DemultiplexSummaryF1L#.txt	Reports demultiplexing results in a table with one row per tile and one column per sample. The # represents lane 1, 2, 3, or 4 of the flow cell. Located in the Alignment folder.
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.
PCRAmpliconRunStatistics.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.

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 .
- 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 .

Sample Sheet Settings for Analysis

Parameter	Description
Adapter	Specify the 5' portion of the adapter sequence to prevent reporting sequence beyond the sample DNA. Illumina recommends adapter trimming for Nextera libraries and Nextera Mate Pair libraries. To specify 2 or more adapter sequences, separate the sequences by a plus (+) sign. For example: CTGTCTTATACACATCT+AGATGTGTATAAGAGACAG
AdapterRead2	Specify the 5' portion of the Read 2 adapter sequence to prevent reporting sequence beyond the sample DNA. Use this setting to specify a different adapter other than the one specified in the Adapter setting.
QualityScoreTrim	If set to a value > 0, then the 3' ends of non-indexed reads with low quality scores are trimmed. Trimming is automatically applied by default at a value of 15 when using BWA for alignment.

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.

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

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

© 2018 Illumina, Inc. All rights reserved.

illumina[®]