

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.

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2016 Illumina, Inc. All rights reserved.

Illumina, BaseSpace, MiSeq, TruSeq, TruSight, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Revision History

Document #	Date	Description of Change
Document # 15042295 v05	January 2017	Added Microsoft Visual C++ 2013 to the list of computing requirements for installing MiSeq Reporter on an off-instrument computer. In the section on installing MiSeq Reporter off-instrument, removed information on installing Microsoft Visual C++ 2013 and NET Framework 4.5.1 because these files are no longer bundled with the MiSeq Reporter installation package.
Document # 15042295 v04	November 2015	Updated the computing requirements for installing MiSeq Reporter v2.6 on an off-instrument computer.
Document # 15042295 v03	October 2015	Updated the name of the workflow for letter designator TT to TruSight Tumor 15.
Document # 15042295 v02	September 2015	Updated directions for installing MiSeq Reporter v2.6 on an off-instrument computer.
Document # 15042295 v01	September 2015	Changed the name of the guide from the MiSeq Reporter User Guide to the MiSeq Reporter Software Guide. Added BWA-MEM aligner information. The original BWA aligner is renamed BWA-backtrack. In the BAM File Format section, revised the description of the alignment information in the file header, and updated the link for SAM format specifications. Updated computer requirements for MiSeq Reporter off-instrument as follows: <ul style="list-style-type: none"> • ≥ 8 GB RAM minimum; ≥ 16 GB RAM recommended to ≥ 16 GB RAM minimum; ≥ 32 GB RAM recommended • 64-bit Windows from English-US to English version with English-US regional settings • Internet Explorer 8 to Internet Explorer 11 Removed Chrome from the list of supported browsers for MiSeq Reporter off-instrument use. Added the letter designator TT for the TruSight Tumor Panel (15 Genes) Workflow. Changed references from primary analysis to analysis by RTA software.
Part # 15042295 Rev. E	December 2014	Added a note in the Demultiplexing section about the default index recognition for index pairs that differ by < 3 bases.

Document #	Date	Description of Change
Part # 15042295 Rev. D	September 2014	<p>Updated computing requirements for installing MiSeq Reporter on an off-instrument computer.</p> <p>Updated information on the ConvertMissingBclsToNoCalls to clarify the default setting.</p> <p>Updated the reference for a network Linux storage tech note to <i>Configuring MiSeq Reporter to Work with Samba Shares on a Linux Server</i> (part # 970-2014-027).</p>
Part # 15042295 Rev. C	February 2014	<p>Updated to changes introduced in MiSeq Reporter v2.4:</p> <ul style="list-style-type: none"> • Added the alignment method to the description of the BAM file header. • Added the command line and annotation algorithm to the description of VCF file header. • Added information on configuring the FileCopyWaitFinishTimeInSeconds parameter. <p>Updated information on the Starling variant caller.</p> <p>Removed the section on gVCF files. See the reference guide for your workflow for gVCF output information.</p> <p>Removed information on the ELAND alignment algorithm, which was deprecated in MiSeq Reporter v2.2. For more information, see the <i>MiSeq Sample Sheet Quick Reference Guide</i> (part # 15028392).</p>
Part # 15042295 Rev. B	August 2013	<p>Updated to changes introduced in MiSeq Reporter v2.3:</p> <ul style="list-style-type: none"> • Increased default for configuration setting MaximumHoursPerProcess from 1.5 to 72. • Changed letter designator for the TruSeq Amplicon workflow from C to TA. • Added description of genome VCF file, a file format optionally generated for the Enrichment, PCR Amplicon, and TruSeq Amplicon workflows.
Part # 15042295 Rev. A	May 2013	<p>Initial release.</p> <p>This guide provides information about the MiSeq Reporter web interface, how to view run results, how to requeue a run, and how to install and configure the software.</p> <p>For information about analysis workflows performed by MiSeq Reporter, see the workflow-specific reference guide. A reference guide for each analysis workflow is available for download from the Illumina website.</p>

Table of Contents

Revision History	iii
Table of Contents	v
Chapter 1 Getting Started	1
Introduction	2
Viewing MiSeq Reporter	3
MiSeq Reporter Concepts	5
MiSeq Reporter Interface	6
Requeue Analysis	11
Input File Requirements	12
Preinstalled Databases and Genomes	13
Chapter 2 Analysis Metrics and Procedures	15
Introduction	16
Analysis Metrics	17
Demultiplexing	19
FASTQ File Generation	20
Alignment	21
Variant Calling	22
Chapter 3 Folders, File Formats, and Settings	23
MiSeqAnalysis Folder	24
Folder Structure	25
Analysis File Formats	26
MiSeq Reporter Configurable Settings	32
Restarting the Service	34
Chapter 4 Installation and Troubleshooting	35
MiSeq Reporter Off-Instrument Requirements	36
Installing MiSeq Reporter Off-Instrument	37
Using MiSeq Reporter Off-Instrument	39
Troubleshooting MiSeq Reporter	40
Index	43
Technical Assistance	45

Getting Started

Introduction	2
Viewing MiSeq Reporter	3
MiSeq Reporter Concepts	5
MiSeq Reporter Interface	6
Requeue Analysis	11
Input File Requirements	12
Preinstalled Databases and Genomes	13



Introduction

The MiSeq[®] system provides on-instrument secondary analysis using the MiSeq Reporter software. MiSeq Reporter performs secondary analysis on the base calls and quality scores generated by Real-time Analysis (RTA) during the sequencing run.

MiSeq Reporter performs analysis based on the analysis workflow specified in the sample sheet. The analysis workflow is a series of steps specific to a type of analysis. Upon completion of analysis, MiSeq Reporter generates various types of information specific to the workflow. For most workflows, results appear on the MiSeq Reporter web interface in the form of graphs and tables for each run.

MiSeq Reporter runs as a Windows service and is viewed through a web browser.

About Windows Service Applications

Windows service applications perform specific functions without user intervention and continue to run in the background as long as Windows is running. Because MiSeq Reporter runs as a Windows service, it automatically begins MiSeq Reporter analysis when base calling is complete.

Sequencing During Analysis

The MiSeq system computing resources are dedicated to either sequencing or analysis. If a new sequencing run is started on the MiSeq before secondary analysis of an earlier run is complete, MiSeq Reporter analysis is stopped automatically.

To restart the analysis performed by MiSeq Reporter, use the Requeue feature on the MiSeq Reporter interface after the new sequencing run is complete. At that point, secondary analysis starts from the beginning.

Viewing MiSeq Reporter

The MiSeq Reporter interface can only be viewed through a web browser. To view the MiSeq Reporter interface during analysis, open any web browser on a computer with access to the same network as the MiSeq system. Connect to the HTTP service on port 8042 using one of the following methods:

- ▶ Connect using the instrument IP address followed by 8042.

IP Address	HTTP Service Port	HTTP Address
10.10.10.10, for example	8042	10.10.10.10:8042

- ▶ Connect using the network name for the MiSeq followed by 8042

Network Name	HTTP Service Port	HTTP Address
MiSeq01, for example	8042	MiSeq01:8042

- ▶ For off-instrument installations of MiSeq Reporter, connect using the method for locally installed service applications, **localhost** followed by 8042.

Off-Instrument	HTTP Service Port	HTTP Address
localhost	8042	localhost:8042

For more information, see *Installing MiSeq Reporter Off-Instrument* on page 37.

Sample Sheet Tab

Row	Description
Investigator Name	[Optional] The name of the investigator.
Project Name	[Optional] A descriptive name of the run.
Experiment Name	[Optional] A descriptive name of the experiment.
Date	The date the sequencing run was performed.
Workflow	The analysis workflow for the run.
Assay	The name of the assay used to prepare your samples.
Chemistry	The chemistry name identifies recipe fragments used to build the run-specific recipe. For runs using the TruSeq Amplicon workflow or PCR Amplicon workflow, the name is amplicon. For all other workflows, the name is default or the field can be blank.
Manifests	The name of the manifest file that specifies alignments to a reference and targeted reference regions. This section is used with the TruSeq Amplicon workflow, Enrichment workflow, and PCR Amplicon workflow.
Reads	The number of cycles performed in Read 1 and Read 2. Index reads are not included in this section.
Settings	Optional run parameters used for modifying analysis results.

Row	Description
Data	The sample ID, sample name, index sequences, and path to the genome folder. Requirements vary by workflow.

For information about sample sheets and sample sheet settings, see the *MiSeq Sample Sheet Quick Reference Guide* (document # 15028392).

MiSeq Reporter Concepts

The following concepts and terms are common to MiSeq Reporter.

Concept	Description
Analysis Workflow	A secondary analysis procedure performed by MiSeq Reporter. The workflow for each run is specified in the sample sheet.
Manifest	The file that specifies a reference genome and targeted reference regions to be used in the alignment step. Manifests are not required for all workflows. For more information, see the workflow-specific reference guide.
Reference Genome	A FASTA format file that contains the genome sequences used during analysis. For some workflows, the reference genome is for alignment. For other workflows, the reference genome is used to generate supplementary data. The FASTA files can use the extension *.fa or *.fasta. They are contained in subfolders of the Genome Repository, which is specified in the MiSeq Reporter.config file. For more information, see <i>MiSeq Reporter Configurable Settings</i> on page 32 and <i>Preinstalled Databases and Genomes</i> on page 13.
Repository	A folder that holds the data generated during sequencing runs. Each run folder is a subfolder in the repository.
Run Folder	The folder structure populated by Real-time Analysis software (MiSeqOutput folder) or the folder populated by MiSeq Reporter (MiSeqAnalysis). For more information, see <i>MiSeqAnalysis Folder</i> on page 24.
Sample Sheet	A comma-separated values file (*.csv) that contains information required to set up and analyze a sequencing run, including a list of samples and their index sequences. The sample sheet must be provided during the run setup steps on the MiSeq. After the run begins, the sample sheet is renamed to SampleSheet.csv and copied to the run folders: MiSeqTemp, MiSeqOutput, and MiSeqAnalysis.

MiSeq Reporter Interface

When MiSeq Reporter opens in the browser, the main screen appears with an image of the instrument in the center. The Settings icon and Help icon are in the upper-right corner, and the Analyses tab is in the upper-left corner.


- ▶ **MiSeq Reporter Help**—Select the Help icon to open MiSeq Reporter documentation in the browser window.
- ▶ **Settings**—Select the Settings icon  to change the server URL and Repository path.
- ▶ **Analyses Tab**—Select Analyses to expand the tab. The Analyses tab shows a list of analysis runs that are either completed, queued for analysis, or currently processing.

Figure 1 MiSeq Reporter Main Screen

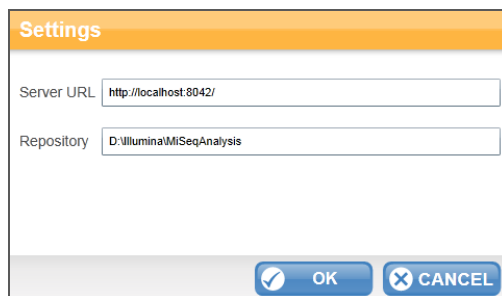


Server URL or Repository Settings

Select the Settings  icon. The Settings dialog box opens. Set the server URL and the repository path:

- ▶ **Server URL**—The server on which MiSeq Reporter is running.
- ▶ **Repository path**—Location of the analysis folder where output files are written.

Figure 2 Settings for Server URL and Repository



Typically, it is not necessary to change these settings unless MiSeq Reporter is running off-instrument. In this case, set the repository path to the network location of the MiSeqOutput folder.









For more information, see *Using MiSeq Reporter Off-Instrument* on page 39.

Analyses Tab

The Analyses tab lists the sequencing runs located in the specified repository. From this tab, you can open the results from any runs listed, or requeue a selected run for analysis.

- ▶ To refresh the list, select the **Refresh Analysis List** icon  in the upper-right corner.




Figure 3 Analyses Tab Expanded

Analyses				
Completed				
State	Type	Run	Completed On	Requeue
	G	YourExperimentHere	6/12/2012 1:50:45 PM	<input type="checkbox"/>
	A	Ecoli Assembly	6/20/2012 10:04:14 AM	<input type="checkbox"/>
	S	YourExperimentHere	8/6/2012 10:54:29 AM	<input type="checkbox"/>
	TA	2x151 DragonFruit v1 with Dual Indexing	6/21/2012 10:22:38 AM	<input type="checkbox"/>
	R	2X151RHo With Annotation	6/13/2012 10:23:02 AM	<input type="checkbox"/>
	M	Experiment01	6/13/2012 10:58:42 AM	<input type="checkbox"/>
	L	2X36 Dragonfruit v1 with Nextera	6/13/2012 5:00:38 PM	<input type="checkbox"/>
	P	PCRFast	8/6/2012 11:18:31 AM	<input type="checkbox"/>

The Analyses tab columns are State, Type, Run, Completed On, and Requeue:

- ▶ **State**—Shows the current state of the analysis using 1 of 3 status icons.

Table 1 State of Analysis Icons


Icon	Description
	Indicates that analysis by MiSeq Reporter completed successfully.
	Indicates that analysis by MiSeq Reporter is in progress.
	Indicates that analysis by MiSeq Reporter was not completed successfully.

- ▶ **Type**—Lists the analysis workflow associated with each run using a single letter designation. Letter designators for each workflow are standard in the MiSeq Reporter interface.

Table 2 Letter Designators for Analysis Workflows

Letter	Workflow
A	Assembly
E	Enrichment
G	GenerateFASTQ
L	Library QC
M	Metagenomics
P	PCR Amplicon
R	Resequencing
S	Small RNA
T	Targeted RNA
TA	TruSeq Amplicon
TT	TruSight Tumor 15

Letter	Workflow
U	Unknown This designator is used to represent a plug-in workflow

- ▶ **Run**—The name of the run as it is listed in the Experiment Name field of the sample sheet. If an experiment name was not included in the sample sheet before the sequencing run, this field lists the run folder name.
Alternatively, you can specify a different name for the run by editing the Experiment Name field in the sample sheet. For more information, see *Editing the Sample Sheet in MiSeq Reporter* on page 9.
- ▶ **Completed On**—The date that MiSeq Reporter analysis completed.
- ▶ **Requeue**—Select the checkbox to requeue a specific job for analysis. The **Requeue** button appears.
When analysis is queued, the run appears at the bottom of the Analyses tab and indicated as in-progress with the icon .

Analysis Information and Results Tabs

After selecting a run from the Analyses tab, information and results for that run appear in a series of tabs on the MiSeq Reporter interface.

Analysis results that appear on the Summary and Details tabs vary by workflow. For more information, see the workflow-specific reference guide. A reference guide for each workflow is available from the Illumina website.

Information on the Analysis tab, Sample Sheet tab, Logs tab, and Errors tab are similar for each workflow. All tabs are populated when analysis is complete.

Tab Name	Description
Summary Tab	Contains a summary of analysis results in graphs for mismatches, phasing and prephasing, alignment, and clusters passing filter, for example.
Details Tab	Contains details of analysis results in tables and graphs for samples, coverage, Q-scores, variants, and targets, for example.
Analysis Tab	Contains logistical information about the run.
Sample Sheet Tab	Contains run parameters specified in the sample sheet, and provides tools to edit the sample sheet and requeue the run.
Logs Tab	Lists every step performed during analysis. These steps are recorded in log files located in the Logs folder. A summary is written to AnalysisLog.txt, which is an important file for troubleshooting purposes.
Errors Tab	Lists any errors that occurred during analysis. A summary is written to AnalysisError.txt, which is an important file for troubleshooting purposes.

Analysis Info Tab

Row	Description
Investigator	[Optional] The name of the investigator.
Read Cycles	Represents the number of cycles in each read, including notation for any index reads. For example, 151, 8(I), 8(I), 151, indicates a first read of 151 cycles, 2 reads of 8 cycles, and a final read of 151 cycles.
Start Time	The clock time that analysis by MiSeq Reporter was started.
Completion Time	The clock time that analysis by MiSeq Reporter was completed.
Data Folder	The root level of the output folder produced by Real-time Analysis software (MiSeqOutput), which contains all primary and secondary analysis output for the run.
Analysis Folder	The full path to the Alignment folder in the MiSeqAnalysis folder (Data\Intensities\BaseCalls\Alignment).
Copy Folder	The full path to the Queued subfolder in the MiSeqAnalysis folder.

Editing the Sample Sheet in MiSeq Reporter

You can edit the sample sheet for a specific run from the Sample Sheet tab on the MiSeq Reporter web interface. A mouse and keyboard are required to edit the sample sheet.

- ▶ To edit a row in the sample sheet, click any field in the row and make required changes.
- ▶ To add a row to the sample sheet, click the row above the intended location of the new row and select **Add Row**.



- ▶ To delete a row from the sample sheet, click anywhere in the row and select **Delete Row**.



- ▶ After editing the sample sheet, select **Save and Requeue** to save changes and initiates secondary analysis with the edited sample sheet.



- ▶ If a change to the sample sheet was made in error, click an adjacent tab before saving any changes. A warning appears that states changes were not saved. Click **Discard** to undo any changes or **Save** to save and requeue analysis.



Saving Graphs as Images

MiSeq Reporter provides the option to save an image of graphs shown on the Summary or Details tabs. Right-click any location on the Summary tab or the graphs location on the Details tab, and then left-click **Save Image As**. When prompted, name the file and browse to a location to save the file.

All images are saved in a JPG (*.jpg) format. Graphs are exported as a single graphic for all graphs shown on the tab. A mouse is required to use this option.

Requeue Analysis

To requeue a run for analysis, use the **Requeue** feature from the MiSeq Reporter Analyses tab. Make sure that a sequencing run on the MiSeq is not currently in progress.

Each time analysis is requeued, the following folders and files are created:

- ▶ A new Alignment folder is created with a sequential number appended to the folder name, such as Alignment2.
MiSeqAnalysis \<RunFolderName> \Data \Intensities \BaseCalls \Alignment2
- ▶ Existing intermediate analysis files written in FASTQ file format are overwritten with new analysis files. FASTQ files are written to the BaseCalls folder.
MiSeqAnalysis \<RunFolderName> \Data \Intensities \BaseCalls.



NOTE

If changes were made to the sample sheet, make sure that the file is named SampleSheet.csv and saved to the root level of the analysis folder.

- 1 From the MiSeq Reporter web interface, click **Analyses**.
- 2 Locate the run from the list of available runs on the Analyses tab, and click the Requeue checkbox next to the run name.
If the run is not listed, confirm that the correct repository is specified using the Settings icon. For more information, see *Server URL or Repository Settings* on page 6.

Figure 4 Requeue Button

Analyses				
Completed				
State	Type	Run	Completed On	Requeue
✓	G	YourExperimentHere	6/12/2012 1:50:45 PM	<input type="checkbox"/>
✓	A	Ecoli Assembly	6/20/2012 10:04:14 AM	<input type="checkbox"/>
✓	S	YourExperimentHere	8/6/2012 10:54:29 AM	<input type="checkbox"/>
✓	C	2x151 DragonFruit v1 with Dual Indexing T12SD	6/21/2012 10:22:38 AM	<input checked="" type="checkbox"/>
✓	R	2X151RHo With Annotation	6/13/2012 10:23:02 AM	<input type="checkbox"/>
✓	M	Experiment01	6/13/2012 10:58:42 AM	<input type="checkbox"/>

- 3 Click **Requeue**. The State icon to the left of the run name changes to show that analysis is in progress .
 - ▶ If analysis does not start, make sure that the following input files are present in the analysis run folder: SampleSheet.csv, RTAComplete.txt, and RunInfo.xml.
 - ▶ During analysis, a status bar with elapsed time appears on the Analysis Info tab. To stop analysis, select the stop analysis icon next to the status bar on the Analysis Info tab.

Input File Requirements

MiSeq Reporter requires the following files generated during the sequencing run to perform analysis or to requeue analysis. Files, such as *.bcl, *.filter, and *.locs, are required to perform analysis.

There is no need to move or copy files to another location before analysis begins. Required files are copied automatically to the MiSeqAnalysis folder during the sequencing process.

File Name	Description
RTAComplete.txt	A marker file that indicates RTA processing is complete. The presence of this file triggers MiSeq Reporter to queue analysis.
SampleSheet.csv	Provides parameters for the run and subsequent analysis. At the start of the run, the sample sheet is copied to the root level of the run folder and renamed SampleSheet.csv.
RunInfo.xml	Contains high-level run information, such as the number of reads and cycles in the sequencing run, and whether a read is indexed.

Required Files

MiSeq Reporter requires the following files generated during the sequencing run to perform secondary analysis.

File Type	Path and File Name Example	Description
*.bcl files	Data\Intensities\BaseCalls\L001\C1.1\s_1_3.bcl	Base calls for lane 1, cycle 1, tile 3
*.filter files	Data\Intensities\BaseCalls\L001\s_1_0003.filter	Filter results file for lane 1, tile 3
*.locs files	Data\Intensities\L001\s_1_3.locs	Location file for lane 1, tile 3

Preinstalled Databases and Genomes

For most workflows, a reference is required to perform alignment. The MiSeq includes several preinstalled databases and genomes.

Preinstalled	Description
Databases	<ul style="list-style-type: none"> • miRbase for human • dbSNP for human • RefGene for human
Genomes	<ul style="list-style-type: none"> • <i>Arabidopsis thaliana</i> • cow (<i>Bos taurus</i>) • <i>E. coli</i> strain DH10b • human (<i>Homo sapiens</i>) build hg19 • mouse (<i>Mus musculus</i>) • rat (<i>Rattus norvegicus</i>) • yeast (<i>Saccharomyces cerevisiae</i>) • <i>Staphylococcus aureus</i>

The reference genome used for analysis by MiSeq Reporter is specified for each sample in the sample sheet (SampleSheet.csv). The full path to the folder containing the whole genome FASTA file must be specified in the sample sheet.



NOTE

Enter the full path (UNC path) to the GenomeFolder in the sample sheet. Do not enter the path using a mapped drive.



NOTE

Introduced in MiSeq Reporter v2.1, you can specify genome references for multiple species in the same sample sheet for all workflows *except* the Small RNA workflow.

Available Genomes

In addition to the preinstalled genomes, genome sequence files and reference annotation for other commonly used model organisms are available from the Illumina iGenomes page. Go to my.illumina.com/Message/iGenome. A MyIllumina login is required.

The sequence and annotation files for each iGenome are provided in a compressed file format, *.tar.gz. Refer to the iGenomes Overview for installation instructions.

Custom Genomes

You can upload your own reference in FASTA format to the MiSeq computer. The reference must have a *.fa or *.fasta extension and be stored in a single folder.

You can upload several single FASTA files *or* a single multi-FASTA file (recommended), but not a combination of both. To upload files, use the Manage Files feature in CS.



NOTE

The chromosome name, which is the section of the > line up to any white space, must not contain the following characters:

- ? () [] / \ = + < > : ; " ' , * ^ | &

For best results, use only alpha-numeric characters as chromosome names.

Illumina recommends the use of a simple text editor, such as Notepad to make sure that no illegal or invisible characters are added to the file.

Analysis Metrics and Procedures

Introduction	16
Analysis Metrics	17
Demultiplexing	19
FASTQ File Generation	20
Alignment	21
Variant Calling	22



Introduction

During the sequencing run, Real-time Analysis (RTA) generates data files that include analysis metrics used by MiSeq Reporter for secondary analysis. The following metrics appear in reports from MiSeq Reporter software:

- ▶ Clusters passing filter
- ▶ Base call quality scores
- ▶ Phasing and prephasing values

MiSeq Reporter performs secondary analysis using a series of analysis procedures, which include demultiplexing, FASTQ file generation, alignment, and variant calling.

Table 3 Analysis Procedures

Analysis Procedure	Description
Demultiplexing	Performed for all workflows if the run has index reads and the sample sheet lists multiple samples. For indexed libraries containing either 1 or 2 indexes, demultiplexing separates data from pooled samples based on short index sequences from different libraries.
FASTQ File Generation	Performed for all workflows. FASTQ files are the primary input for the alignment step. FASTQ files contain non-indexed reads for each sample, excluding reads identified as inline controls and reads that did not pass filter.
Alignment	Performed for workflows that require alignment against a reference. Alignment compares sequences against the reference specified in the sample sheet and assigns a score based on regions of similarity. MiSeq Reporter uses an alignment method best-suited for the workflow. Aligned reads are written to files in BAM file format.
Variant Calling	Performed for workflows that require variant identification as a final output. Variant calling records SNPs and other structural variants in a standardized and parsable text file. MiSeq Reporter uses variant calling algorithms best-suited for the workflow. Variant calls are written to files in VCF file format.

Analysis Metrics

Real-time Analysis software filters data and calculates statistical estimates to measure data quality. These metrics are later included with secondary analysis results. Metrics that appear in secondary analysis reports are clusters passing filter, base call quality scores, and phasing and prephasing values.

Clusters Passing Filter

The software performs base calling of raw data to remove any reads that do not meet the overall quality as measured by the Illumina chastity filter. The chastity of a base call is calculated as the ratio of the brightest intensity divided by the sum of the brightest and second brightest intensities.

Clusters pass filter (PF) when no more than 1 base call in the first 25 cycles has a chastity of < 0.6.

Quality Scores

A quality score, or Q-score, is a prediction of the probability of an incorrect base call. A higher Q-score implies that a base call is more reliable and less likely to be incorrect.

Based on the Phred scale, the Q-score serves as a compact way to communicate small error probabilities. Given a base call, X, the probability that X is not true, $P(\sim X)$, results in a quality score, $Q(X)$, according to the relationship:

$$Q(X) = -10 \log_{10}(P(\sim X))$$

where $P(\sim X)$ is the estimated probability of the base call being wrong.

The following table shows the relationship between the quality score and error probability.

Quality Score $Q(X)$	Error Probability $P(\sim X)$
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

For more information on the Phred quality score, see en.wikipedia.org/wiki/Phred_quality_score.

During the sequencing run, base call quality scores are calculated after cycle 25 and results are recorded in base call (*.bcl) files, which contain the base call and quality score per cycle.

ASCII Format for Quality Scores

During analysis, base call quality scores are written to FASTQ files in an encoded ASCII format (the value + 33). The ASCII format is illustrated in the following table.

Table 4 ASCII Codes for Q-Scores 0–40

Symbol	ASCII Code	Q-score	Symbol	ASCII Code	Q-score
!	33	0	6	54	21
"	34	1	7	55	22
#	35	2	8	56	23
\$	36	3	9	57	24
%	37	4	:	58	25

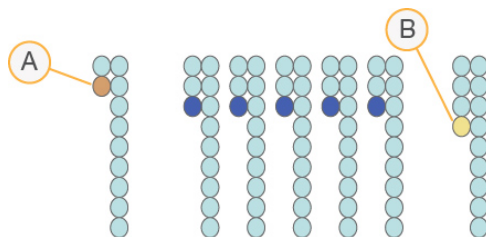
Table 4 ASCII Codes for Q-Scores 0–40

Symbol	ASCII Code	Q-score	Symbol	ASCII Code	Q-score
&	38	5	;	59	26
'	39	6	<	60	27
(40	7	=	61	28
)	41	8	>	62	29
*	42	9	?	63	30
+	43	10	@	64	31
,	44	11	A	65	32
-	45	12	B	66	33
.	46	13	C	67	34
/	47	14	D	68	35
0	48	15	E	69	36
1	49	16	F	70	37
2	50	17	G	71	38
3	51	18	H	72	39
4	52	19	I	73	40
5	53	20			

Phasing and Prephasing

During the sequencing reaction, each DNA strand in a cluster extends by 1 base per cycle. A small portion of strands can become out of phase with the current incorporation cycle. Phasing occurs when a base falls behind. Prephasing occurs when a base jumps ahead. Phasing and prephasing rates indicate an estimate of the fraction of molecules that became phased or prephased in each cycle.

Figure 5 Phasing and Prephasing



- A Read with a base that is phasing
- B Read with a base that is prephasing

The number of cycles performed in a read is 1 more cycle than the number of cycles analyzed. For example, a paired-end 150-cycle run performs 2 151-cycle reads (2×151) for a total of 302 cycles. At the end of the run, 2×150 cycles are analyzed. The 1 extra cycle for Read 1 and Read 2 is required for prephasing calculations. Phasing and prephasing results are recorded in the file named `phasing.xml`, which is located in the folder `Data \ Intensities \ BaseCalls \ Phasing`.

Phasing and prephasing calculations use statistical averaging over many clusters and sequences to estimate the correlation of signal between different cycles. Therefore, phasing estimates tend to be more accurate for tiles with larger numbers of clusters and a mixture of different sequences. Samples containing only a few different sequences do not produce reliable estimates. Sequencing into adapters or other highly homogeneous samples are expected to result in poor phasing estimates.

Demultiplexing

For runs with multiple samples and index reads, demultiplexing compares each Index Read sequence to the index sequences specified in the sample sheet. No quality values are considered in this step.

Demultiplexing separates data from pooled samples based on short index sequences that tag samples from different libraries. Index reads are identified using the following steps:

- ▶ Samples are numbered starting from 1 based on the order they are listed in the sample sheet.
- ▶ Sample number 0 is reserved for clusters that were not successfully assigned to a sample.
- ▶ Clusters are assigned to a sample when the index sequence matches exactly or there is up to a single mismatch per Index Read.



NOTE

Illumina indexes are designed so that any index pair differs by ≥ 3 bases, allowing for a single mismatch in index recognition. Index sets that are not from Illumina can include pairs of indexes that differ by < 3 bases. In such cases, the software detects the insufficient difference and modifies the default index recognition (mismatch=1). Instead, the software performs demultiplexing using only perfect index matches (mismatch=0).

When demultiplexing is complete, 1 demultiplexing file named `DemultiplexSummaryF1L1.txt` is written to the Alignment folder with the following information:

- ▶ In the file name, **F1** represents the flow cell number.
- ▶ In the file name, **L1** represents the lane number, which is always L1 for MiSeq.
- ▶ A table of demultiplexing results with 1 row per tile and 1 column per sample, including sample 0.
- ▶ The most commonly occurring sequences for the index reads.

Other demultiplexing files are generated for each tile of the flow cell. For more information, see *Demultiplexing File Format* on page 26.

FASTQ File Generation

MiSeq Reporter 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 their quality scores, excluding reads identified as inline controls and clusters that did not pass filter.

FASTQ files are the primary input for alignment. The files are written to the BaseCalls folder (Data\Intensities\BaseCalls) in the MiSeqAnalysis folder, and then copied to the BaseCalls folder in the MiSeqOutput folder. Each FASTQ file contains reads for only 1 sample, and the name of that sample is included in the FASTQ file name. For more information, see *FASTQ File Names* on page 27.

FASTQ Config Settings

Some default settings for FASTQ file generation can be changed by editing the following settings in the MiSeq Reporter configuration file (C:\Illumina\MiSeq Reporter\MiSeq Reporter.exe.config):

- ▶ **ConvertMissingBclsToNoCalls**—By default, FASTQ files include all tiles. During FASTQ file generation, MiSeq Reporter treats *.bcl files that are missing or corrupt as no-calls (Ns), and logs a warning in the Analysis.Error.txt file for the affected cycle and tile. You can override this default setting by changing the value to 0 (false), so that the software logs a fatal error and aborts analysis when encountering a missing or invalid base call.
- ▶ **CreateFastqForIndexReads**—By default, FASTQ files are not generated for index reads. You can override this setting by changing the value to 1 (true).
- ▶ **FilterNonPFReads**—By default, FASTQ files only include clusters passing filter. You can override this setting by changing the value to 0 (false).

For more information, see *MiSeq Reporter Configurable Settings* on page 32.

Quality Trimming

FASTQ file generation optionally performs quality trimming of the 3' portion of nonindex reads with low quality scores. This step is performed by default during alignment using BWA. For workflows that do not use BWA, use the **QualityScoreTrim** sample sheet setting to include trimming during FASTQ file generation. For more information, see the *MiSeq Sample Sheet Quick Reference Guide* (document # 15028392).

Alignment

Alignment is a way of identifying optimal matches between read sequences and the sequence of a reference genome. Aligned sequences are assigned a score based on their similarity to the reference.

Alignment results are written to Binary Alignment/Map (BAM) files. BAM files are the primary input for variant calling. For more information, see *BAM File Format* on page 27.

Alignment Methods

For workflows that include alignment, reads are aligned against the reference specified in the sample sheet or in a manifest file. MiSeq Reporter uses one of the following alignment methods best-suited for the workflow: Smith-Waterman or BWA, or Bowtie.

Smith-Waterman Algorithm

The banded Smith-Waterman algorithm performs local sequence alignments to determine similar regions between 2 sequences. Instead of looking at 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.

BWA-Backtrack

Formerly referred to as BWA, BWA-backtrack is an earlier version of the BWA (Burrows-Wheeler Aligner) algorithm that aligns sequencing read lengths in < 70 bp segments. Use this version for very short reads, or when consistency is required with previous study data.

BWA 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.

When using BWA-backtrack for alignment, GATK is used for variant calling, by default.

BWA-MEM

BWA-MEM is the new version of the Burrows-Wheeler Alignment algorithm, and is the default BWA method for MiSeq Reporter. Optimized for longer read lengths of ≥ 70 bp, BWA-MEM has a significant positive impact on detection of variants, especially insertions and deletions.

When using BWA-MEM for alignment, GATK is used for variant calling, by default.

Bowtie

Bowtie is a short-read aligner that quickly aligns large sets of short sequences. For more information, see bowtie-bio.sourceforge.net.

Variant Calling

Variant calling records single nucleotide polymorphisms (SNPs), insertions/deletions (indels), and other structural variants in a standardized variant call format (VCF). For more information, see *VCF File Format* on page 28.

For each SNP or indel call, 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.

Variant Callers

For workflows that include variant calling, variants are detected using one of the following variant callers best-suited for the workflow: GATK, the somatic variant caller, or Starling.

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.

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. Variant scores are computed using a Poisson model that excludes variants with a quality score below Q20.
- ▶ For indels, the somatic variant caller analyzes how many alignments covering a given position include a particular indel compared to the overall coverage at that position. The somatic variant caller does not perform an indel realignment step included in other variant callers, such as GATK.

For more information, see the *Somatic Variant Caller Tech Note* available on the Illumina website.

Starling

Starling calls both SNPs and small indels, and summarizes depth and probabilities for every site in the genome. The output files Starling produces includes 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.

Starling can be used as an optional alternative variant caller to GATK.

Folders, File Formats, and Settings

MiSeqAnalysis Folder	24
Folder Structure	25
Analysis File Formats	26
MiSeq Reporter Configurable Settings	32
Restarting the Service	34

























MiSeqAnalysis Folder

The MiSeqAnalysis folder is the main run folder for MiSeq Reporter. The relationship between the MiSeqOutput and MiSeqAnalysis run folders is summarized as follows:

- ▶ During sequencing, Real-time Analysis (RTA) populates the MiSeqOutput folder with files generated during image analysis, base calling, and quality scoring.
- ▶ Except for focus images and thumbnail images, RTA copies files to the MiSeqAnalysis 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 run folders.
- ▶ MiSeq Reporter monitors the MiSeqAnalysis folder and begins secondary analysis when the file RTAComplete.xml appears.
- ▶ As secondary analysis continues, MiSeq Reporter writes analysis output files to the MiSeqAnalysis folder, and then copies the files to the MiSeqOutput folder.

Folder Structure

-  **Data**
 -  **Intensities**
 -  **Basecalls**
 -  **Alignment**—Contains *.bam and *.vcf files, if applicable.
 -  **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 MiSeq Reporter; also called the copy folder.
-  AnalysisError.txt
-  AnalysisLog.txt
-  CompletedJobInfo.xml
-  QueuedForAnalysis.txt
-  [Workflow]RunStatistics
-  RTAComplete.xml
-  RunInfo.xml
-  runParameters.xml
-  SampleSheet.csv

When using BaseSpace for secondary analysis without replicating analysis locally, the local MiSeqAnalysis folder is empty.

Alignment Folder Contents

Most secondary analysis files are written to the Alignment folder. Each time that analysis is requested, MiSeq Reporter creates an Alignment folder named **AlignmentN**, where N is a sequential number.

Log files from analysis algorithms, such as BWA or GATK, are written to Data\BaseCalls\Alignment\Logging.

Analysis File Formats

Analysis results are written to file formats specific to their function and purpose.

Analysis Step	Format	Purpose
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	*.bam	Compressed binary files containing sequence alignment data. BAM files are the primary input for the variant calling step.
Variant Calling	*.vcf	Text files containing SNPs, indels, and other structural variants.

Other file formats used in analysis results are *.txt, *.xml, *.htm, and *.png. Many of these files contain information that appears in tables, graphs, and charts on the MiSeq Reporter web interface.

Demultiplexing File Format

For multiple sample indexed runs, 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.

Demultiplexing files are binary files written to the L001 folder in Data\Intensities\BaseCalls\L001. The 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.

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:

- ▶ Identifier
- ▶ Sequence
- ▶ Plus sign (+)
- ▶ Quality scores in an ASCII encoded format

The identifier is formatted as **@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber** as shown in the following example:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```


FASTQ File Names

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

Data\Intensities\BaseCalls\samplename_S1_L001_R1_001.fastq.gz

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



NOTE

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

- ▶ **L001**—The lane number. This segment is always L001 with the single-lane flow cell.
- ▶ **R1**—The read. In this example, R1 means Read 1. For a paired-end run, a file from Read 2 includes R2 in the file name.
- ▶ **001**—The last segment is always 001.

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

BAM File Format

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

BAM files are written to the alignment folder in Data\Intensities\BaseCalls\Alignment. BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed in the sample sheet.

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. Alignment methods include banded Smith-Waterman, Burrows-Wheeler Aligner (BWA), and Bowtie. The term Isis indicates that an Illumina alignment method is in use, which is the banded Smith-Waterman method.
- ▶ **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags.


```
GA23_40:8:1:10271:11781 64 chr22 17552189 8 35M * 0 0
TACAGACATCCACCACCACCCAGCTAATTTTTG
IIIII>FA?C::B=:GGGB>GGGEGIIIIHI3EEE#
BC:Z:ATCACG XD:Z:55 SM:I:8
```

The read name maps to the chromosome and start coordinate **chr22 17552189**, with alignment quality **8**, and the match descriptor CIGAR string **35M**.

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 in the sample sheet.

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 as the annotator, the Illumina internal annotation algorithm is in use to annotate the VCF file. The VCF header also contains the command line call used by MiSeq Reporter to run the variant caller. The command-line call specifies all parameters used by the variant caller, including the reference genome file and .bam file. The last line in the header is column headings for the data lines. For more information, see *VCF File Annotations* on page 30.

```
##fileformat=VCFv4.1
##FORMAT=<ID=GQX,Number=1,Type=Integer>
##FORMAT=<ID=AD,Number=.,Type=Integer>
##FORMAT=<ID=DP,Number=1,Type=Integer>
##FORMAT=<ID=GQ,Number=1,Type=Float>
##FORMAT=<ID=GT,Number=1,Type=String>
##FORMAT=<ID=PL,Number=G,Type=Integer>
##FORMAT=<ID=VF,Number=1,Type=Float>
##INFO=<ID=TI,Number=.,Type=String>
##INFO=<ID=GI,Number=.,Type=String>
##INFO=<ID=EXON,Number=0,Type=Flag>
##INFO=<ID=FC,Number=.,Type=String>
##INFO=<ID=IndelRepeatLength,Number=1,Type=Integer>
##INFO=<ID=AC,Number=A,Type=Integer>
##INFO=<ID=AF,Number=A,Type=Float>
##INFO=<ID=AN,Number=1,Type=Integer>
##INFO=<ID=DP,Number=1,Type=Integer>
##INFO=<ID=QD,Number=1,Type=Float>
##FILTER=<ID=LowQual>
##FILTER=<ID=R8>
##annotator=MARS
##CallSomaticVariants_cmdline=" -B D:\Amplicon_DS_Soma2\121017_
M00948_0054_000000000-
A2676_Binf02\Data\Intensities\BaseCalls\Alignment3_Tamsen_
SomaWorker -g [D:\Genomes\Homo_sapiens
\UCSC\hg19\Sequence\WholeGenomeFASTA,] -f 0.01 -fo False -b 20 -q
100 -c 300 -s 0.5 -a 20 -F 20 -gVCF
True -i true -PhaseSNPs true -MaxPhaseSNPLength 100 -r D:
\Amplicon_DS_Soma2\121017_M00948_0054_000000000-A2676_Binf02"
##reference=file:///d:\Genomes\Homo_
sapiens\UCSC\hg19\Sequence\WholeGenomeFASTA\genome.fa
##source=GATK 1.6
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT 10002 - R1
```

VCF File Data Lines—Contains information about a single variant. Data lines are listed under the column headings included in the header.

VCF File Headings

The VCF file format is flexible and extensible, so not all VCF files contain the same fields. The following tables describe VCF files generated by MiSeq Reporter.

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 relative 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. Configure cutoff using the MinimumCoverageDepth sample sheet setting. • LowGQ—The genotyping quality (GQ) is below a cutoff. Configure cutoff using the VariantMinimumGQCutoff sample sheet setting. • LowQual—The variant quality (QUAL) is below a cutoff. Configure using the VariantMinimumQualCutoff sample sheet setting. • LowVariantFreq—The variant frequency is less than the given threshold. Configure using the VariantFrequencyFilterCutoff sample sheet setting. • R8—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8. This filter is configurable using the IndelRepeatFilterCutoff setting in the config file or the sample sheet. • SB—The strand bias is more than the given threshold. This filter is configurable using the StrandBiasFilter sample sheet setting; available only for somatic variant caller and GATK. <p>For more information about sample sheet settings, see <i>MiSeq Sample Sheet Quick Reference Guide (document # 15028392)</i>.</p>
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). In regions of high coverage, GATK down-samples the available reads. • 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.

Heading	Description
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. • VF—Variant frequency; the percentage of reads supporting the alternate allele.
SAMPLE	The sample column gives the values specified in the FORMAT column.

MiSeq Reporter Configurable Settings

Typically, you do not need to change configurable settings. However, if you want to customize analysis results, you can edit settings in `MiSeq Reporter.exe.config` located in the MiSeq Reporter installation folder, `C:\Illumina\MiSeqReporter`, by default. Always restart the service after modifying the config file.

The editable portion of this file is contained between the `<appSettings>` tags, which show key/value pairs for the parameter settings applied.

```
<appSettings>
  <add key="Repository" value="D:\Illumina\MiSeqAnalysis" />
  <add key="GenomePath" value="C:\Illumina\MiSeqReporter\Genomes" />
  <add key="TempFolder" value="D:\Illumina\MiSeqAnalysis\Temp" />
  <add key="EnableHTTPService" value="1"/>
  <add key="ClientSettingsProvider.ServiceUri" value="" />
  <add key="CopyToRTAOutputPath" value="1"/>
  <add key="DemuxMaxSequencesToReport" value="100"/>
  <add key="MaximumHoursPerProcess" value="24"/>
</appSettings>
```

Available Configurable Settings


The following configurable settings are used in `MiSeq Reporter.exe.config`.

Setting Name	Values and Description
AdapterTrimmingStringency	0.9 (default) The minimum match rate allowed in adapter trimming. The default setting trims sequences with > 90% sequence identity with the adapter.
ConvertMissingBclsToNoCalls	1 (true; default) 0 (false) If set to true, any missing or invalid *.bcl files cause MiSeq Reporter to log an error and flag the tile as having no-calls (Ns) for the affected cycle. If set to false, any missing or truncated *.bcl files cause MiSeq Reporter to log an error and abort analysis.
CopyToRTAOutputPath	1 (true; default) 0 (false) If set to true, copy all alignment data to the <code><OutputDirectory></code> specified in the <code>RTAConfiguration.xml</code> file, which is located in <code>Data\Intensities</code> .
CreateFastqForIndexReads	0 (false; default) 1 (true) If set to false, FASTQ files are not generated for index reads. If set to true, FASTQ files are generated for index reads.
EnableHTTPService	1 (true; default) 0 (false) Determines whether MiSeq Reporter provides the web interface.

Setting Name	Values and Description
FilterNonPFReads	1 (true; default) 0 (false) Determines whether those clusters that fail the chastity filter are filtered from all FASTQ files.
GATKDownsampleDepth	5000 (default) When using GATK for variant calling, reads in regions of high depth are (optionally) randomly down-sampled. <ul style="list-style-type: none"> • Set to a higher value to retain more reads. • Set to 0 to disable down-sampling. <i>Disabling down-sampling can lead to increased run time and memory use on high-coverage runs.</i>
IndelRepeatFilterCutoff	8 (default) By default, indels are flagged as filtered if the reference has a 1- or 2-base motif repeated 8 or more times next to the variant.
MaximumGigabytesPerProcess	Varies The maximum gigabytes of memory allowed for a child process. By default, this threshold is adjusted automatically based on the memory available on the system.
MaximumHoursPerProcess	72 (default) The maximum number of hours to allow a child process to run.
MaximumMegabasesAssembly	550 (default) The maximum number of megabases to assemble. Larger values require more RAM. Assembly of reads from longer runs requires more memory than assembly of reads from shorter runs. If the process terminates due to memory requirements, consider lowering the MaximumMegabasesAssembly value.
MinimumAlignReadLength	21 (maximum; default) 8 (min) The minimum length of a non-indexed read to align using BWA.
NMaskShortAdapterReads	10-base (default) The number of bases from the start of the adapter that triggers N-masking of the entire read.
RetainTempFiles	0 (false; default) 1 (true) If set to true, temporary files are retained. Retaining temporary files requires large amounts of disk space. Use this setting for troubleshooting only.
VariantFilterQualityCutoff	30 (default) for GATK and somatic variant caller 20 (default) for Starling SNPs with variant quality scores below this threshold are flagged as filtered in the *.vcf files.

Restarting the Service

After updating MiSeq Reporter.exe.config, restart the service to enable changes.

- 1 From the Control Panel, select **Administrative Tools | Services**.
- 2 Select **MiSeq Reporter service**, and then click the **Restart Service** icon .

Installation and Troubleshooting

MiSeq Reporter Off-Instrument Requirements	36
Installing MiSeq Reporter Off-Instrument	37
Using MiSeq Reporter Off-Instrument	39
Troubleshooting MiSeq Reporter	40



MiSeq Reporter Off-Instrument Requirements

Installing a copy of MiSeq Reporter on an off-instrument Windows computer allows secondary analysis of sequencing data while the MiSeq performs a subsequent sequencing run.

For more information, see *Installing MiSeq Reporter Off-Instrument* on page 37.

Computing Requirements

MiSeq Reporter software requires the following computing components:

- ▶ 64-bit Windows OS (Vista, Windows 7, Windows Server 2008 64-bit, English version with English-US regional settings)
- ▶ ≥ 16 GB RAM minimum; ≥ 32 GB RAM recommended
- ▶ ≥ 1 TB disk space
- ▶ Quad core processor (2.8 GHz or higher)
- ▶ Microsoft .NET 4.5.1 (Microsoft .NET 4.0 for MiSeq Reporter v2.5 and earlier)
- ▶ Visual C++ 2013

Supported Browsers

MiSeq Reporter can be viewed with the following web browsers:

- ▶ Safari 5.1.7, or later
- ▶ Firefox 13.0.1, or later
- ▶ Internet Explorer 11, or later

Downloading and Licensing

- 1 Download a second copy of the MiSeq Reporter software from the Illumina website. A MyIllumina login is required.
- 2 Accept the end-user licensing agreement (EULA) when prompted during installation. No license key is required as this additional copy is free of charge.

Installing MiSeq Reporter Off-Instrument

[Optional] To install MiSeq Reporter on an off-instrument Windows computer, perform the following steps:

- 1 Set up **Log on as a service** permission, and then run the installation wizard.
- 2 Configure the software to point to the appropriate Repository and GenomePath.

Set Up User or Group Accounts on Windows 7

To configure user or group accounts to enable **Log on as a service** permission, you must have administrator rights to the computer. If you do not have administrator rights or need assistance setting up a user or group account, contact your local facility administrator.

- 1 From the Windows **Start** menu, select **Control Panel**, and then click **System and Security**.
- 2 Click **Administrative Tools**, and then double-click **Local Security Policy**.
- 3 From the Security Settings tree on the left, double-click **Local Policies** and then click **User Rights Assignments**.
- 4 In the details pane on the right, double-click **Log on as a service**.
- 5 In the Properties dialog box, click **Add User or Group**.
- 6 Enter the name of the user or group account for this computer. Click **Check Names** to validate the account.
- 7 Click **OK** through any open dialog boxes and then close the control panel.

Run the MiSeq Reporter Installation Wizard

- 1 Download and unzip the MiSeq Reporter installation package from the Illumina website.
- 2 Browse to the unzipped directory.
- 3 Double-click the setup.exe file.
- 4 Click **Next** through the prompts in the installation wizard.
- 5 When prompted, specify the user name and password for an account with **Log on as a service** permission, as set up in the previous step.
- 6 Continue through any remaining prompts.

Configure MiSeq Reporter

To configure MiSeq Reporter to locate the run folder and reference genome folder, edit the configuration file in a text editor, such as Notepad.

- 1 Navigate to the installation folder (C:\Illumina\MiSeq Reporter, by default) and open the file MiSeq Reporter.exe.config in a text editor.
- 2 Locate the **Repository** tag and change the **value** to the default data location on the off-instrument computer.

```
<add key="Repository" value="E:\Data\Repository" />
```

Alternatively, this location can be a network location accessible from the off-instrument computer.

- 3 Locate the **GenomePath** tag and change the **value** to the location of the folder containing reference genomes files in FASTA format.

```
<add key="GenomePath" value="E:\MyGenomes\FASTA" />
```


Start the MiSeq Reporter Service

After completing the installation, the MiSeq Reporter service starts automatically. If the service does not start, start it manually using the following instructions, or reboot the computer.

- 1 From the Windows **Start** menu, right-click **Computer** and select **Manage**.
- 2 From the Computer Management tree on the left, double-click **Services and Applications** and then click **Services**.
- 3 Right-click **MiSeq Reporter** and select **Properties**.
- 4 On the General tab, make sure that the **Startup Type** is set to **Automatic**, and then click **Start**.
- 5 On the Log On tab, set the **user name** and **password** for a Services account that has permissions to write to the server. Illumina recommends the **Local System** account for most users. For assistance or site-specific network requirements, contact the local facility administrator.
- 6 Click **OK** through any open dialog boxes and then close the Computer Management window.
- 7 After starting the MiSeq Reporter service, connect to the software locally using localhost:8042 in a web browser.

Using MiSeq Reporter Off-Instrument

To use MiSeq Reporter off-instrument, make sure that folders containing run data and reference genomes are accessible.

- 1 If you are not using a network location for sequencing data and reference genomes, copy the following folders to your local computer:
 - ▶ Copy run data from the MiSeq computer in D:\MiSeqOutput\ - ▶ Copy reference genomes from the MiSeq computer in C:\Illumina\MiSeq Reporter\Genomes.
- 2 Open a web browser to localhost:8042, which opens the MiSeq Reporter web interface.
- 3 If the location of the run data differs from the location specified in MiSeq Reporter.exe.config, change the path using the **Settings**  icon.



NOTE

Specifying the repository path in Settings is temporary. The next time you restart your computer, the path defaults to the Repository location specified in MiSeq Reporter.exe.config.

- 4 Select **Analyses** on the left-side of the web interface to view the runs available in the specified Repository location.
- 5 Before you requeue analysis using an off-instrument installation of MiSeq Reporter, update the path of the GenomeFolder in the sample sheet to the new location. After updating the GenomeFolder path, click **Save and Requeue**. For more information, see *Editing the Sample Sheet in MiSeq Reporter* on page 9.

Troubleshooting MiSeq Reporter

MiSeq Reporter runs as Windows service application. User accounts must be configured to enable **Log on as a service** permission before installing MiSeq Reporter. For more information, see *Set Up User or Group Accounts on Windows 7* on page 37.

For more information, see msdn.microsoft.com/en-us/library/ms189964.aspx.


Service Fails to Start

If the service fails to start, check the Window Event Log and view the details of the error message.

- 1 Open the **Control Panel** and select **Administrative Tools**.
- 2 Select **Event Viewer**.
- 3 In the Event Viewer window, select **Windows Logs | Application**. The error listed in the event log describes any syntax errors in MiSeq Reporter.exe.config. Incorrect syntax in the MiSeq Reporter.exe.config file can cause the service to fail.

Files Failed to Copy

If files fail to copy to the intended location, check the following settings:

- 1 Check the path to the specified repository folder or MiSeqOutput folder:
 - ▶ If you are using MiSeq Reporter off-instrument, check the repository location using Settings  on the MiSeq Reporter web interface.
 - ▶ If you are using MiSeq Reporter on-instrument, check the MiSeqOutput folder location on the MCS Run Options screen, Folder Settings tab.

Use the full UNC path, such as \\server1\Runs. Because MiSeq Reporter runs as a Windows service, it does not recognize user-mapped drives, such as Z:\Runs.
- 2 Confirm that you have write-access to the output folder location. If you need assistance, contact your facility administrator.
- 3 If you use a network Linux storage location, and MiSeq Reporter analysis files fail to transfer there, see the technote *Configuring MiSeq Reporter to Work with Samba Shares on a Linux Server (document # 970-2014-027)* for assistance. The technote is on the Documentation and Literature page of support.illumina.com.
- 4 Make sure that copying is not disabled in the <appSettings> section of the MiSeq Reporter.exe.config file. Make sure that the value is set to **1**.


```
<add key="CopyToRTAOutputPath" value="1"/>
```
- 5 Check if the files failed to copy because of a timeout error.
 - ▶ Open the AnalysisError.txt file, located in the root level of the MiSeqAnalysis folder.
 - ▶ If there is a timeout error, the file contains the message


```
Copy thread has taken too long (over 1800 seconds) -aborting.
```

 Use the procedure *Configuring File Copy Timeout* to increase the file copy timeout value.

If you continue to receive timeout errors after adjusting the parameter value, a network problem can be the cause of file copy delays. Consult your IT department.

Configuring File Copy Timeout

File copy timeout length is determined by the `FileCopyWaitFinishTimeInSeconds` parameter setting in the `MiSeq Reporter.exe.config` file.

- 1 Open the `MiSeq Reporter.exe.config` file and make sure that the file contains the string `<add key="FileCopyWaitFinishTimeInSeconds" value="1800"/>`.
For more information on the `MiSeq Reporter.exe.config` file, see *MiSeq Reporter Configurable Settings* on page 32.
- 2 If the string is not in the `MiSeq Reporter.exe.config` file, add it under `<appSettings>`.
- 3 Configure the `FileCopyWaitFinishTimeInSeconds` parameter value according to the recommendation of your IT department.
The `FileCopyWaitFinishTimeInSeconds` value is in seconds. The default value is 1800, which is equivalent to 30 minutes.
- 4 Restart the service to enable changes.
For more information, see *Restarting the Service* on page 34.



NOTE

Setting the `FileCopyWaitFinishTimeInSeconds` value too high can delay MiSeq Reporter analysis.

Viewing Log Files for a Failed Run

Viewing logs files can help identify specific errors for troubleshooting purposes.

- 1 To view the log files using the MiSeq Reporter web browser interface, select the run in the `Analyses` tab.
- 2 Select the `Logs` tab to view a list of every step that occurred during analysis. Log information is recorded in `AnalysisLog.txt`, which is located in the root level of the `MiSeqAnalysis` folder.
- 3 Select the `Errors` tab to view a list of errors that occurred during analysis. Error information is recorded in `AnalysisError.txt`, which is located in the root level of the `MiSeqAnalysis` folder.

*

*.bam 27
 *.bam.bai 27
 *.bcl files 12
 *.demux 26
 *.fastq.gz 27
 *.filter files 12
 *.locs.files 12
 *.vcf 28

A

AdapterTrimmingStringency 32
 alignment
 BWA 21
 scores 21
 Smith-Waterman 21
 analysis
 during sequencing 2
 analysis folder 9, 24
 analysis tab 8
 AnalysisError.txt 41
 AnalysisLog.txt 41
 ASCII codes 17

B

BAM files
 file format 27
 in alignment 21
 BAM index files 27
 base call files 12
 bcl files 12
 BWA-backtrack 21
 BWA-MEM 21

C

CD coding region 30
 clusters passing filter 17
 computing requirements 36
 configurable settings 32
 ConvertMissingBclsToNoCalls 20, 32
 copy folder 9
 CopyToRTAOutputPath 32
 CreateFastqForIndexReads 20, 32
 customer support 45

D

data folder 9
 databases, pre-installed 13
 dbsnp database 13
 demultiplexing 19, 26
 DemultiplexSummaryF1L1.txt 19
 details tab 8
 documentation 45
 DP depth 30

E

editing the sample sheet 9
 EnableHTTPService 32
 error probability 17
 errors tab 8

F

FASTQ files
 config settings 20
 file format 26
 file naming 27
 generation 20
 quality trimming 20
 FASTQ files for index reads 32
 files fail to copy 40-41
 filter files 12
 FilterNonPFReads 20, 32

G

GATK 22
 GATKDownsampleDepth 32
 genome path 32, 37
 GI gene ID 30
 GNU zip format 27
 GT genotype 30

H

help, technical 45

I

icons, state of analysis 6
 iGenomes 13
 IndelRepeatFilterCutoff 30, 32
 input files 12
 installation, off-instrument 37
 IP address, MiSeq Reporter 3

L

license (EULA) 36
 Linux 40
 local security policy 37
 Local System account 38
 localhost 3
 locs files 12
 log files 41
 log on as a service 37
 logs tab 8
 LowDP 30
 LowGQ 30
 LowVariantFreq 30

M

manifest file
 definition 5

- in sample sheet 3
- MaxGigabytesPerProcess 32
- MaxHoursPerProcess 32
- MaxMegabasesAssembly 32
- MinimumAlignReadLength 32
- MinimumCoverageDepth 30
- miRbase database 13
- MiSeq Reporter.exe.config 32
- MiSeqAnalysis folder 24
- MiSeqOutput folder 24

N

- NL noise level 30
- NMaskShortAdapterReads 32

P

- passing filter (PF) 17
- phasing 18
- Phred scale 17
- prephasing 18

Q

- Q-scores 17
- q20 30
- quality score 22
- quality scores 17
- QualityScoreTrim 20

R

- r8s 30
- read cycles 9
- reference genome
 - file format 5
- reference genomes
 - custom genomes 13
 - file format 13
 - pre-installed 13
- refGene database 13
- repository path 6, 32, 37
- requeue analysis 6, 9, 11
- RetainTempFiles 32
- RTAComplete.txt 12
- run folder
 - definition 5
 - relationship 24
- RunInfo.xml 12

S

- SAM tools 27
- sample number 0 19, 27
- sample sheet
 - definition 5
 - editing 9
- sample sheet tab 8
- SampleSheet.csv 12
- SB strand bias 30
- sb0.5 30
- server URL 6
- service fails to start 40
- Smith-Waterman 21
- SNPs 22
- somatic variant caller 22
- Starling 22
- StrandBiasFilter 30
- summary tab 8

T

- technical assistance 45
- TI transcript ID 30
- timeout error 40-41
- troubleshooting
 - files fail to copy 40-41
 - log files 41
 - service fails to start 40

V

- variant caller
 - GATK 22
 - somatic variant caller 22
 - Starling 22
- VariantFilterQualityCutoff 30, 32
- VariantFrequencyFilterCutoff 30
- VariantMinimumGQCutoff 30
- VCF files
 - annotations 30
 - file format 28
 - filter annotations 30
 - in variant calling 22
- VF variant frequency 30
- viewing MiSeq Reporter 3

W

- Windows service
 - about 2
 - Log on as service 40
 - restart the service 34
- workflows
 - letter designators 6

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 5 Illumina General Contact Information

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

Table 6 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