



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) OR ANY USE OF SUCH PRODUCT(S) OUTSIDE THE SCOPE OF THE EXPRESS WRITTEN LICENSES OR PERMISSIONS GRANTED BY ILLUMINA IN CONNECTION WITH CUSTOMER'S ACQUISITION OF SUCH PRODUCT(S).

#### FOR RESEARCH USE ONLY

© 2011–2014 Illumina, Inc. All rights reserved.

**Illumina, IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIX, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iScan, iSelect, MiSeq, MiSeqDx, Nextera, NextSeq, NuPCR, SeqMonitor, Solexa, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode**, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks of Illumina, Inc. in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

## Read Before Using this Product

This Product, and its use and disposition, is subject to the following terms and conditions. If Purchaser does not agree to these terms and conditions then Purchaser is not authorized by Illumina to use this Product and Purchaser must not use this Product.

- Definitions.** "**Application Specific IP**" means Illumina owned or controlled intellectual property rights that pertain to this Product (and use thereof) only with regard to specific field(s) or specific application(s). Application Specific IP excludes all Illumina owned or controlled intellectual property that cover aspects or features of this Product (or use thereof) that are common to this Product in all possible applications and all possible fields of use (the "**Core IP**"). Application Specific IP and Core IP are separate, non-overlapping, subsets of all Illumina owned or controlled intellectual property. By way of non-limiting example, Illumina intellectual property rights for specific diagnostic methods, for specific forensic methods, or for specific nucleic acid biomarkers, sequences, or combinations of biomarkers or sequences are examples of Application Specific IP. "**Consumable(s)**" means Illumina branded reagents and consumable items that are intended by Illumina for use with, and are to be consumed through the use of, Hardware. "**Documentation**" means Illumina's user manual for this Product, including without limitation, package inserts, and any other documentation that accompany this Product or that are referenced by the Product or in the packaging for the Product in effect on the date of shipment from Illumina. Documentation includes this document. "**Hardware**" means Illumina branded instruments, accessories or peripherals. "**Illumina**" means Illumina, Inc. or an Illumina affiliate, as applicable. "**Product**" means the product that this document accompanies (e.g., Hardware, Consumables, or Software). "**Purchaser**" is the person or entity that rightfully and legally acquires this Product from Illumina or an Illumina authorized dealer. "**Software**" means Illumina branded software (e.g., Hardware operating software, data analysis software). All Software is licensed and not sold and may be subject to additional terms found in the Software's end user license agreement. "**Specifications**" means Illumina's written specifications for this Product in effect on the date that the Product ships from Illumina.
- Research Use Only Rights.** Subject to these terms and conditions and unless otherwise agreed upon in writing by an officer of Illumina, Purchaser is granted only a non-exclusive, non-transferable, personal, non-sublicensable right under Illumina's Core IP, in existence on the date that this Product ships from Illumina, solely to use this Product in Purchaser's facility for Purchaser's internal research purposes (which includes research services provided to third parties) and solely in accordance with this Product's Documentation, **but specifically excluding any use that** (a) would require rights or a license from Illumina to Application Specific IP, (b) is a re-use of a previously used Consumable, (c) is the disassembling, reverse-engineering, reverse-compiling, or reverse-assembling of this Product, (d) is the separation, extraction, or isolation of components of this Product or other unauthorized analysis of this Product, (e) gains access to or determines the methods of operation of this Product, (f) is the use of non-Illumina reagent/consumables with Illumina's Hardware (does not apply if the Specifications or Documentation state otherwise), or (g) is the transfer to a third-party of, or sub-licensing of, Software or any third-party software. All Software, whether provided separately, installed on, or embedded in a Product, is licensed to Purchaser and not sold. Except as expressly stated in this Section, no right or license under any of Illumina's intellectual property rights is or are granted expressly, by implication, or by estoppel.

**Purchaser is solely responsible for determining whether Purchaser has all intellectual property rights that are necessary for Purchaser's intended uses of this Product, including without limitation, any rights from third**

parties or rights to Application Specific IP. Illumina makes no guarantee or warranty that purchaser's specific intended uses will not infringe the intellectual property rights of a third party or Application Specific IP.

- 3 **Regulatory.** This Product has not been approved, cleared, or licensed by the United States Food and Drug Administration or any other regulatory entity whether foreign or domestic for any specific intended use, whether research, commercial, diagnostic, or otherwise. This Product is labeled For Research Use Only. Purchaser must ensure it has any regulatory approvals that are necessary for Purchaser's intended uses of this Product.
- 4 **Unauthorized Uses.** Purchaser agrees: (a) to use each Consumable only one time, and (b) to use only Illumina consumables/reagents with Illumina Hardware. The limitations in (a)-(b) do not apply if the Documentation or Specifications for this Product state otherwise. Purchaser agrees not to, nor authorize any third party to, engage in any of the following activities: (i) disassemble, reverse-engineer, reverse-compile, or reverse-assemble the Product, (ii) separate, extract, or isolate components of this Product or subject this Product or components thereof to any analysis not expressly authorized in this Product's Documentation, (iii) gain access to or attempt to determine the methods of operation of this Product, or (iv) transfer to a third-party, or grant a sublicense, to any Software or any third-party software. Purchaser further agrees that the contents of and methods of operation of this Product are proprietary to Illumina and this Product contains or embodies trade secrets of Illumina. The conditions and restrictions found in these terms and conditions are bargained for conditions of sale and therefore control the sale of and use of this Product by Purchaser.
- 5 **Limited Liability. TO THE EXTENT PERMITTED BY LAW, IN NO EVENT SHALL ILLUMINA OR ITS SUPPLIERS BE LIABLE TO PURCHASER OR ANY THIRD PARTY FOR COSTS OF PROCUREMENT OF SUBSTITUTE PRODUCTS OR SERVICES, LOST PROFITS, DATA OR BUSINESS, OR FOR ANY INDIRECT, SPECIAL, INCIDENTAL, EXEMPLARY, CONSEQUENTIAL, OR PUNITIVE DAMAGES OF ANY KIND ARISING OUT OF OR IN CONNECTION WITH, WITHOUT LIMITATION, THE SALE OF THIS PRODUCT, ITS USE, ILLUMINA'S PERFORMANCE HEREUNDER OR ANY OF THESE TERMS AND CONDITIONS, HOWEVER ARISING OR CAUSED AND ON ANY THEORY OF LIABILITY (WHETHER IN CONTRACT, TORT (INCLUDING NEGLIGENCE), STRICT LIABILITY OR OTHERWISE).**
- 6 **ILLUMINA'S TOTAL AND CUMULATIVE LIABILITY TO PURCHASER OR ANY THIRD PARTY ARISING OUT OF OR IN CONNECTION WITH THESE TERMS AND CONDITIONS, INCLUDING WITHOUT LIMITATION, THIS PRODUCT (INCLUDING USE THEREOF) AND ILLUMINA'S PERFORMANCE HEREUNDER, WHETHER IN CONTRACT, TORT (INCLUDING NEGLIGENCE), STRICT LIABILITY OR OTHERWISE, SHALL IN NO EVENT EXCEED THE AMOUNT PAID TO ILLUMINA FOR THIS PRODUCT.**
- 7 **Limitations on Illumina Provided Warranties. TO THE EXTENT PERMITTED BY LAW AND SUBJECT TO THE EXPRESS PRODUCT WARRANTY MADE HEREIN ILLUMINA MAKES NO (AND EXPRESSLY DISCLAIMS ALL) WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THIS PRODUCT, INCLUDING WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR ARISING FROM COURSE OF PERFORMANCE, DEALING, USAGE OR TRADE. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, ILLUMINA MAKES NO CLAIM, REPRESENTATION, OR WARRANTY OF ANY KIND AS TO THE UTILITY OF THIS PRODUCT FOR PURCHASER'S INTENDED USES.**
- 8 **Product Warranty.** All warranties are personal to the Purchaser and may not be transferred or assigned to a third-party, including an affiliate of Purchaser. All warranties are facility specific and do not transfer if the Product is moved to another facility of Purchaser, unless Illumina conducts such move.
  - a **Warranty for Consumables.** Illumina warrants that Consumables, other than custom Consumables, will conform to their Specifications until the later of (i) 3 months from the date of shipment from Illumina, and (ii) any expiration date or the end of the shelf-life pre-printed on such Consumable by Illumina, but in no event later than 12 months from the date of shipment. With respect to custom Consumables (i.e., Consumables made to specifications or designs made by Purchaser or provided to Illumina by, or on behalf of, Purchaser), Illumina only warrants that the custom Consumables will be made and tested in accordance with Illumina's standard manufacturing and quality control processes. Illumina makes no warranty that custom Consumables will work as intended by Purchaser or for Purchaser's intended uses.
  - b **Warranty for Hardware.** Illumina warrants that Hardware, other than Upgraded Components, will conform to its Specifications for a period of 12 months after its shipment date from Illumina unless the Hardware includes Illumina provided installation in which case the warranty period begins on the date of installation or 30 days after the date it was delivered, whichever occurs first ("Base Hardware Warranty"). "Upgraded Components" means Illumina provided components, modifications, or enhancements to Hardware that was previously acquired by Purchaser. Illumina warrants that Upgraded Components will conform to their Specifications for a period of 90 days from the date the Upgraded Components are installed. Upgraded Components do not extend the warranty for the Hardware unless the upgrade was conducted by Illumina at Illumina's facilities in which case the upgraded Hardware shipped to Purchaser comes with a Base Hardware Warranty.
  - c **Exclusions from Warranty Coverage.** The foregoing warranties do not apply to the extent a non-conformance is due to (i) abuse, misuse, neglect, negligence, accident, improper storage, or use contrary to the Documentation or Specifications, (ii) improper handling, installation, maintenance, or repair (other than if performed by Illumina's personnel), (iii) unauthorized alterations, (iv) Force Majeure events, or (v) use with a third party's good not provided by Illumina (unless the Product's Documentation or Specifications expressly state such third party's good is for use with the Product).



- d **Procedure for Warranty Coverage.** In order to be eligible for repair or replacement under this warranty Purchaser must (i) promptly contact Illumina's support department to report the non-conformance, (ii) cooperate with Illumina in confirming or diagnosing the non-conformance, and (iii) return this Product, transportation charges prepaid to Illumina following Illumina's instructions or, if agreed by Illumina and Purchaser, grant Illumina's authorized repair personnel access to this Product in order to confirm the non-conformance and make repairs.
  - e **Sole Remedy under Warranty.** Illumina will, at its option, repair or replace non-conforming Product that it confirms is covered by this warranty. Repaired or replaced Consumables come with a 30-day warranty. Hardware may be repaired or replaced with functionally equivalent, reconditioned, or new Hardware or components (if only a component of Hardware is non-conforming). If the Hardware is replaced in its entirety, the warranty period for the replacement is 90 days from the date of shipment or the remaining period on the original Hardware warranty, whichever is shorter. If only a component is being repaired or replaced, the warranty period for such component is 90 days from the date of shipment or the remaining period on the original Hardware warranty, whichever ends later. The preceding states Purchaser's sole remedy and Illumina's sole obligations under the warranty provided hereunder.
  - f **Third-Party Goods and Warranty.** Illumina has no warranty obligations with respect to any goods originating from a third party and supplied to Purchaser hereunder. Third-party goods are those that are labeled or branded with a third-party's name. The warranty for third-party goods, if any, is provided by the original manufacturer. Upon written request Illumina will attempt to pass through any such warranty to Purchaser.
- 9 **Indemnification.**
- a **Infringement Indemnification by Illumina.** Subject to these terms and conditions, including without limitation, the Exclusions to Illumina's Indemnification Obligations (Section 9(b) below), the Conditions to Indemnification Obligations (Section 9(d) below), Illumina shall (i) defend, indemnify and hold harmless Purchaser against any third-party claim or action alleging that this Product when used for research use purposes, in accordance with these terms and conditions, and in accordance with this Product's Documentation and Specifications infringes the valid and enforceable intellectual property rights of a third party, and (ii) pay all settlements entered into, and all final judgments and costs (including reasonable attorneys' fees) awarded against Purchaser in connection with such infringement claim. If this Product or any part thereof, becomes, or in Illumina's opinion may become, the subject of an infringement claim, Illumina shall have the right, at its option, to (A) procure for Purchaser the right to continue using this Product, (B) modify or replace this Product with a substantially equivalent non-infringing substitute, or (C) require the return of this Product and terminate the rights, license, and any other permissions provided to Purchaser with respect this Product and refund to Purchaser the depreciated value (as shown in Purchaser's official records) of the returned Product at the time of such return; provided that, no refund will be given for used-up or expired Consumables. This Section states the entire liability of Illumina for any infringement of third party intellectual property rights.
  - b **Exclusions to Illumina Indemnification Obligations.** Illumina has no obligation to defend, indemnify or hold harmless Purchaser for any Illumina Infringement Claim to the extent such infringement arises from: (i) the use of this Product in any manner or for any purpose outside the scope of research use purposes, (ii) the use of this Product in any manner not in accordance with its Specifications, its Documentation, the rights expressly granted to Purchaser hereunder, or any breach by Purchaser of these terms and conditions, (iii) the use of this Product in combination with any other products, materials, or services not supplied by Illumina, (iv) the use of this Product to perform any assay or other process not supplied by Illumina, or (v) Illumina's compliance with specifications or instructions for this Product furnished by, or on behalf of, Purchaser (each of (i) – (v), is referred to as an "Excluded Claim").
  - c **Indemnification by Purchaser.** Purchaser shall defend, indemnify and hold harmless Illumina, its affiliates, their non-affiliate collaborators and development partners that contributed to the development of this Product, and their respective officers, directors, representatives and employees against any claims, liabilities, damages, fines, penalties, causes of action, and losses of any and every kind, including without limitation, personal injury or death claims, and infringement of a third party's intellectual property rights, resulting from, relating to, or arising out of (i) Purchaser's breach of any of these terms and conditions, (ii) Purchaser's use of this Product outside of the scope of research use purposes, (iii) any use of this Product not in accordance with this Product's Specifications or Documentation, or (iv) any Excluded Claim.
  - d **Conditions to Indemnification Obligations.** The parties' indemnification obligations are conditioned upon the party seeking indemnification (i) promptly notifying the other party in writing of such claim or action, (ii) giving the other party exclusive control and authority over the defense and settlement of such claim or action, (iii) not admitting infringement of any intellectual property right without prior written consent of the other party, (iv) not entering into any settlement or compromise of any such claim or action without the other party's prior written consent, and (v) providing reasonable assistance to the other party in the defense of the claim or action; provided that, the party reimburses the indemnified party for its reasonable out-of-pocket expenses incurred in providing such assistance.
  - e **Third-Party Goods and Indemnification.** Illumina has no indemnification obligations with respect to any goods originating from a third party and supplied to Purchaser. Third-party goods are those that are labeled or branded with a third-party's name. Purchaser's indemnification rights, if any, with respect to third party goods shall be pursuant to the original manufacturer's or licensor's indemnity. Upon written request Illumina will attempt to pass through such indemnity, if any, to Purchaser.

# Revision History

Part #	Revision	Date	Description of Change
15042295	C	February 2014	<p>Updated to changes introduced in MiSeq Reporter v2.4:</p> <ul style="list-style-type: none"> <li>• Added the alignment method to the description of the BAM file header.</li> <li>• Added the command line and annotation algorithm to the description of VCF file header.</li> <li>• Added information on configuring the FileCopyWaitFinishTimeInSeconds parameter.</li> </ul> <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>
15042295	B	August 2013	<p>Updated to changes introduced MiSeq Reporter v2.3:</p> <ul style="list-style-type: none"> <li>• Increased default for configuration setting MaximumHoursPerProcess from 1.5 to 72.</li> <li>• Changed letter designator for the TruSeq Amplicon workflow from C to TA.</li> <li>• Added description of genome VCF file, a file format optionally generated for the Enrichment, PCR Amplicon, and TruSeq Amplicon workflows.</li> </ul>
15042295	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 .....	v
Table of Contents .....	vii
<b>Chapter 1 Getting Started .....</b>	<b>1</b>
Introduction .....	2
Viewing MiSeq Reporter .....	3
MiSeq Reporter Concepts .....	4
MiSeq Reporter Interface .....	5
Requeue Analysis .....	10
Input File Requirements .....	11
Pre-Installed Databases and Genomes .....	12
<b>Chapter 2 Analysis Metrics and Procedures .....</b>	<b>13</b>
Introduction .....	14
Analysis Metrics .....	15
Demultiplexing .....	17
FASTQ File Generation .....	18
Alignment .....	19
Variant Calling .....	20
<b>Chapter 3 Folders, File Formats, and Settings .....</b>	<b>21</b>
MiSeqAnalysis Folder .....	22
Folder Structure .....	23
Analysis File Formats .....	24
MiSeq Reporter Configurable Settings .....	29
Restarting the Service .....	31
<b>Chapter 4 Installation and Troubleshooting .....</b>	<b>33</b>
MiSeq Reporter Off-Instrument Requirements .....	34
Installing MiSeq Reporter Off-Instrument .....	35
Using MiSeq Reporter Off-Instrument .....	37
Troubleshooting MiSeq Reporter .....	38
<b>Index .....</b>	<b>41</b>
<b>Technical Assistance .....</b>	<b>43</b>





# Getting Started

Introduction .....	2
Viewing MiSeq Reporter .....	3
MiSeq Reporter Concepts .....	4
MiSeq Reporter Interface .....	5
Requeue Analysis .....	10
Input File Requirements .....	11
Pre-Installed Databases and Genomes .....	12



## 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 secondary analysis when primary analysis 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, secondary analysis is stopped automatically.

To restart secondary analysis, 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 35.

## 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 29 and <i>Pre-Installed Databases and Genomes</i> on page 12.
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 RTA primary analysis software (MiSeqOutput folder) or the folder populated by MiSeq Reporter (MiSeqAnalysis). For more information, see <i>MiSeqAnalysis Folder</i> on page 22.
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.


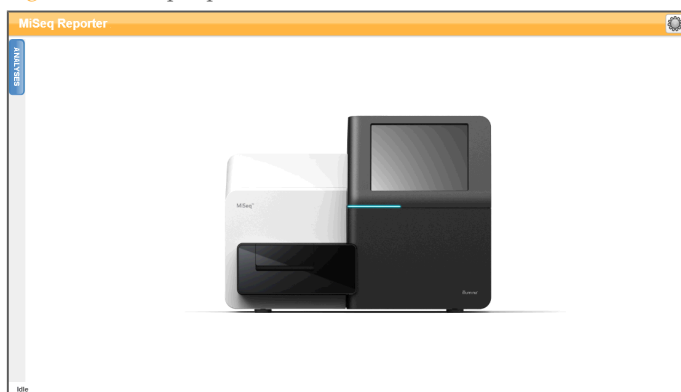

- ▶ **MiSeq Reporter Help**—Select the Help icon to open MiSeq Reporter documentation in the browser window.
- ▶ **Settings**—Select the Settings icon  to change the server URL and Repository path.
- ▶ **Analyses Tab**—Select Analyses to expand the tab. The Analyses tab shows a list of analysis runs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

Figure 1 MiSeq Reporter Main Screen

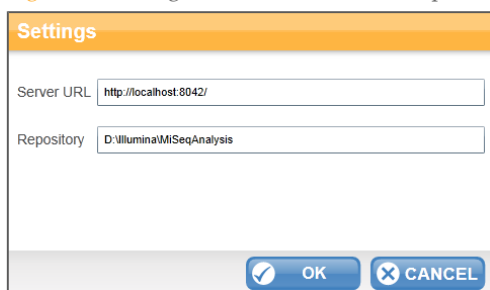


## Server URL or Repository Settings

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 37.

## 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 one of three status icons.


Table 1 State of Analysis Icons

Icon	Description
	Indicates that secondary analysis completed successfully.
	Indicates that secondary analysis is in progress.
	Indicates that secondary analysis 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
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 8.
- ▶ **Completed On**—The date that secondary 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, two eight-cycle index reads, and a final read of 151 cycles.

Row	Description
Start Time	The clock time that secondary analysis was started.
Completion Time	The clock time that secondary analysis was completed.
Data Folder	The root level of the output folder produced by RTA primary 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.

### 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.
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 (part # 15028392)*.

### 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.



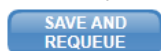
- ▶ To add a row to the sample sheet, click the row above the intended location of the new row and select **Add Row**.



- ▶ 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 5.

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 primary analysis files generated during the sequencing run to perform secondary analysis or to requeue analysis. Primary 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.

## Primary Analysis Files

MiSeq Reporter requires the following primary analysis files 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

## Pre-Installed Databases and Genomes

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

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

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 pre-installed 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](http://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 .....	14
Analysis Metrics .....	15
Demultiplexing .....	17
FASTQ File Generation .....	18
Alignment .....	19
Variant Calling .....	20



## 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 primary analysis metrics appear in secondary analysis reports:

- ▶ 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 one or two indices, 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 in-line 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

During primary analysis, filters and statistical estimates measure data quality and later include these metrics 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

During primary analysis, RTA filters 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 one 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](http://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

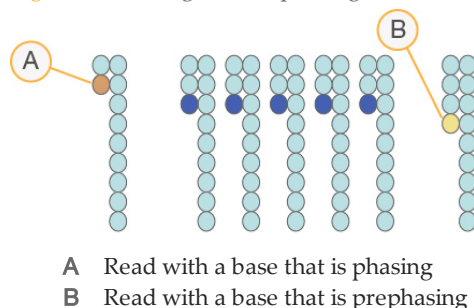
Table 4 ASCII Codes for Q-Scores 0–40

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 one 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



The number of cycles performed in a read is one more cycle than the number of cycles analyzed. For example, a paired-end 150-cycle run performs two 151-cycle reads ( $2 \times 151$ ) for a total of 302 cycles. At the end of the run,  $2 \times 150$  cycles are analyzed. The one 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 that 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.

When demultiplexing is complete, one demultiplexing file named `DemultiplexSummaryF1L1.txt` is written to the Alignment folder, and summarizes 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.
- ▶ Reports demultiplexing results in a table with one row per tile and one column per sample, including sample 0.
- ▶ Reports 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 24.

## 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 in-line 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 one sample, and the name of that sample is included in the FASTQ file name. For more information, see *FASTQ File Naming* on page 25.

## 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. You can configure FASTQ file generation to treat \*.bcl files that are missing or corrupt as no-calls (Ns) by changing the value to 1 (true).
- ▶ **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 29.

## Quality Trimming

FASTQ file generation optionally performs quality trimming of the 3' portion of non-index 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 (part # 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 25.

### 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 two 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

The Burrows-Wheeler Aligner (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 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](http://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 26.

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](http://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 three steps:

- ▶ Considers each position in the reference genome separately
- ▶ Counts bases at the given position for aligned reads that overlap the position
- ▶ Computes a variant score that measures the quality of the call. 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 two 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 .....	22
Folder Structure .....	23
Analysis File Formats .....	24
MiSeq Reporter Configurable Settings .....	29
Restarting the Service .....	31



## 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 primary analysis.
- ▶ Except for focus images and thumbnail images, RTA copies files to the MiSeqAnalysis folder in real time. When primary analysis is complete, RTA 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, one for each tile.
    - 📁 **RTA Logs**—Contains log files from primary 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 requeued, 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 one 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 four lines:

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

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

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



## FASTQ File Naming

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 on the SAM Tools website: [samtools.sourceforge.net](http://samtools.sourceforge.net).

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, and custom tags.

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

The read name includes the chromosome and start coordinate **chr22 17552189**, the alignment quality **8**, and the match descriptor **35M \* 0 0**.

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 27.

```
##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 semi-colon 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.
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.
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, <b>PASS</b> is written in the filter column.</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b>—Applied to sites with depth of coverage below a cutoff. Configure cutoff using the <b>MinimumCoverageDepth</b> sample sheet setting.</li> <li>• <b>LowGQ</b>—The genotyping quality (GQ) is below a cutoff. Configure cutoff using the <b>VariationMinimumGQCutoff</b> sample sheet setting.</li> <li>• <b>LowQual</b>—The variant quality (QUAL) is below a cutoff. Configure using the <b>VariationMinimumQualCutoff</b> sample sheet setting.</li> <li>• <b>LowVariantFreq</b>—The variant frequency is less than the given threshold. Configure using the <b>VariationFrequencyFilterCutoff</b> sample sheet setting.</li> <li>• <b>R8</b>—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 <b>IndelRepeatFilterCutoff</b> setting in the config file or the sample sheet.</li> <li>• <b>SB</b>—The strand bias is more than the given threshold. This filter is configurable using the <b>StrandBiasFilter</b> sample sheet setting; available only for somatic variant caller and GATK.</li> </ul> <p>For more information about sample sheet settings, see <i>MiSeq Sample Sheet Quick Reference Guide (part # 15028392)</i>.</p>

Heading	Description
INFO	<p>Possible entries in the INFO column include:</p> <ul style="list-style-type: none"> <li>• <b>AC</b>—Allele count in genotypes for each ALT allele, in the same order as listed.</li> <li>• <b>AF</b>—Allele Frequency for each ALT allele, in the same order as listed.</li> <li>• <b>AN</b>—The total number of alleles in called genotypes.</li> <li>• <b>CD</b>—A flag indicating that the SNP occurs within the coding region of at least one RefGene entry.</li> <li>• <b>DP</b>—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.</li> <li>• <b>Exon</b>—A comma-separated list of exon regions read from RefGene.</li> <li>• <b>FC</b>—Functional Consequence.</li> <li>• <b>GI</b>—A comma-separated list of gene IDs read from RefGene.</li> <li>• <b>QD</b>—Variant Confidence/Quality by Depth.</li> <li>• <b>TI</b>—A comma-separated list of transcript IDs read from RefGene.</li> </ul>
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"> <li>• <b>AD</b>—Entry of the form X,Y, where X is the number of reference calls, and Y is the number of alternate calls.</li> <li>• <b>DP</b>—Approximate read depth; reads with MQ=255 or with bad mates are filtered.</li> <li>• <b>GQ</b>—Genotype quality.</li> <li>• <b>GQX</b>—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.</li> <li>• <b>GT</b>—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.</li> <li>• <b>NL</b>—Noise level; an estimate of base calling noise at this position.</li> <li>• <b>PL</b>—Normalized, Phred-scaled likelihoods for genotypes.</li> <li>• <b>SB</b>—Strand bias at this position. Larger negative values indicate less bias; values near zero indicate more bias.</li> <li>• <b>VF</b>—Variant frequency; the percentage of reads supporting the alternate allele.</li> </ul>
SAMPLE	<p>The sample column gives the values specified in the FORMAT column.</p>

## 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 <OutputDirectory> specified in the RTAConfiguration.xml file, which is located in Data\Intensities.
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"> <li>• Set to a higher value to retain more reads.</li> <li>• Set to 0 to disable down-sampling. <i>Disabling down-sampling can lead to increased runtime and memory use on high-coverage runs.</i></li> </ul>
IndelRepeatFilterCutoff	8 (default) By default, indels are flagged as filtered if the reference has a 1- or 2-base motif repeated eight 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 or ELAND (deprecated in v2.2).
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 .....	34
Installing MiSeq Reporter Off-Instrument .....	35
Using MiSeq Reporter Off-Instrument .....	37
Troubleshooting MiSeq Reporter .....	38



## 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 35.

### Computing Requirements

MiSeq Reporter software requires the following computing components:

- ▶ 64-bit Windows OS (Vista, Windows 7, Windows Server 2008 64-bit)
- ▶ ≥ 8 GB RAM minimum; ≥ 16 GB RAM recommended
- ▶ ≥ 1 TB disk space
- ▶ Quad core processor (2.8 GHz or higher)
- ▶ Microsoft .NET 4

### Supported Browsers

MiSeq Reporter can be viewed with the following web browsers:

- ▶ Safari 5.1.7, or later
- ▶ Chrome 20.0, or later
- ▶ Firefox 13.0.1, or later
- ▶ Internet Explorer 8, 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

To install MiSeq Reporter on an off-instrument Windows computer, first set up **Log on as a service** permission, and then run the installation wizard. Then, configure the software to point to the appropriate Repository and GenomePath.

### Uninstall Previous Versions of MiSeq Reporter

If MiSeq Reporter v1.0.27, or earlier, is installed on the computer, first uninstall it before running the installation wizard.



#### NOTE

If a later version is installed, skip to *Set Up User or Group Accounts on Windows 7*.

- 1 [Optional] Save a copy of the folder where the FASTA files for the reference genomes are stored.
- 2 From the Windows **Start** menu, select **Control Panel**, and then click **Programs**.
- 3 Click **Programs and Features**.
- 4 Right-click **MiSeq Reporter**, and then click **Uninstall**.
- 5 Click **OK** through any prompts.

### 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.

For more information, see [technet.microsoft.com/en-us/library/cc739424\(WS.10\).aspx](http://technet.microsoft.com/en-us/library/cc739424(WS.10).aspx) on the Microsoft website.

### Run the MiSeq Reporter Installation Wizard

- 1 Download and unzip the MiSeq Reporter installation package from the Illumina website.
- 2 Double-click the setup.exe file.
- 3 Click **Next** through the prompts in the installation wizard.

- 4 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.
- 5 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\<RunFolder>`.
  - 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 8.

## 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 35.

For more information, see [msdn.microsoft.com/en-us/library/ms189964.aspx](http://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 the Transfer of MiSeq Reporter Analysis Files to a Shared Linux Network Storage (part # 770-2014-004)* for assistance. The technote is on the Documentation and Literature page of [support.illumina.com](http://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 check 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 29
- 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 31.



### 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 25
- \*.bam.bai 25
- \*.bcl files 11
- \*.demux 24
- \*.fastq.gz 25
- \*.filter files 11
- \*.locs.files 11
- \*.vcf 26

## A

- AdapterTrimmingStringency 29
- alignment
  - BWA 19
  - scores 19
  - Smith-Waterman 19
- analysis
  - during sequencing 2
- analysis folder 7, 22
- analysis tab 7
- AnalysisError.txt 39
- AnalysisLog.txt 39
- ASCII codes 15

## B

- BAM files
  - file format 25
  - in alignment 19
- BAM index files 25
- base call files 11
- bcl files 11
- BWA 19

## C

- CD coding region 27
- clusters passing filter 15
- computing requirements 34
- configurable settings 29
- ConvertMissingBclsToNoCalls 18, 29
- copy folder 7
- CopyToRTAOutputPath 29
- CreateFastqForIndexReads 18, 29
- customer support 43

## D

- data folder 7
- databases, pre-installed 12
- dbSNP database 12
- demultiplexing 17, 24
- DemultiplexSummaryFIL1.txt 17
- details tab 7
- documentation 43
- DP depth 27

## E

- editing the sample sheet 8

- EnableHTTPService 29
- error probability 15
- errors tab 7

## F

- FASTQ files
  - config settings 18
  - file format 24
  - file naming 25
  - generation 18
  - quality trimming 18
- FASTQ files for index reads 29
- files fail to copy 38-39
- filter files 11
- FilterNonPFReads 18, 29

## G

- GATK 20
- GATKDownsampleDepth 29
- genome path 29, 36
- GI gene ID 27
- GNU zip format 25
- GT genotype 27

## H

- help, technical 43

## I

- icons, state of analysis 5
- iGenomes 12
- IndelRepeatFilterCutoff 27, 29
- input files 11
- installation, off-instrument 35
- IP address, MiSeq Reporter 3

## L

- license (EULA) 34
- Linux 38
- local security policy 35
- Local System account 36
- localhost 3
- locs files 11
- log files 39
- log on as a service 35
- logs tab 7
- LowDP 27
- LowGQ 27
- LowVariantFreq 27

## M

- manifest file
  - definition 4
  - in sample sheet 8
- MaxGigabytesPerProcess 29
- MaxHoursPerProcess 29

MaxMegabasesAssembly 29  
 MinimumAlignReadLength 29  
 MinimumCoverageDepth 27  
 miRbase database 12  
 MiSeq Reporter.exe.config 29  
 MiSeqAnalysis folder 22  
 MiSeqOutput folder 22

## N

NL noise level 27  
 NMaskShortAdapterReads 29

## P

passing filter (PF) 15  
 phasing 16  
 Phred scale 15  
 prephasing 16

## Q

Q-scores 15  
 q20 27  
 quality score 20  
 quality scores 15  
 QualityScoreTrim 18

## R

r8s 27  
 read cycles 7  
 reference genome  
   file format 4  
 reference genomes  
   custom genomes 12  
   file format 12  
   pre-installed 12  
 refGene database 12  
 repository path 5, 29, 36  
 requeue analysis 5, 8, 10  
 RetainTempFiles 29  
 RTAComplete.txt 11  
 run folder  
   definition 4  
   relationship 22  
 RunInfo.xml 11

## S

SAM tools 25  
 sample number 0 17, 25  
 sample sheet  
   definition 4  
   editing 8  
 sample sheet tab 7  
 SampleSheet.csv 11  
 SB strand bias 27  
 sb0.5 27  
 server URL 5  
 service fails to start 38  
 Smith-Waterman 19  
 SNPs 20  
 somatic variant caller 20  
 Starling 20  
 StrandBiasFilter 27  
 summary tab 7

## T

technical assistance 43  
 TI transcript ID 27  
 timeout error 38-39  
 troubleshooting  
   files fail to copy 38-39  
   log files 39  
   service fails to start 38

## V

variant caller  
   GATK 20  
   somatic variant caller 20  
   Starling 20  
 VariantFilterQualityCutoff 27, 29  
 VariantFrequencyFilterCutoff 27  
 VariantMinimumGQCutoff 27  
 VCF files  
   annotations 27  
   file format 26  
   filter annotations 27  
   in variant calling 20  
 VF variant frequency 27  
 viewing MiSeq Reporter 3

## W

Windows service  
   about 2  
   Log on as service 38  
   restart the service 31  
 workflows  
   letter designators 5

## Technical Assistance

For technical assistance, contact Illumina Technical Support.

**Table 5** Illumina General Contact Information

<b>Illumina Website</b>	www.illumina.com
<b>Email</b>	techsupport@illumina.com

**Table 6** Illumina Customer Support Telephone Numbers

<b>Region</b>	<b>Contact Number</b>	<b>Region</b>	<b>Contact Number</b>
North America	1.800.809.4566	Italy	800.874909
Austria	0800.296575	Netherlands	0800.0223859
Belgium	0800.81102	Norway	800.16836
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000

### Safety Data Sheets

Safety data sheets (SDSs) are available on the Illumina website at [www.illumina.com/msds](http://www.illumina.com/msds).

### Product Documentation

Product documentation in PDF is available for download from the Illumina website. Go to [www.illumina.com/support](http://www.illumina.com/support), select a product, then click **Documentation & Literature**.

