illumina®

MiSeq® Reporter Software Reference Guide for IVD Assays

FOR IN VITRO DIAGNOSTIC USE

ILLUMINA PROPRIETARY
Document # 15038356 Rev. 02
September 2017

Revision History

Document #	Date	Description of Change
Document # 15038356 v02	September 2017	Updated regulatory markings.
Document # 15038356 v01	December 2016	Corrected information in the Targets Table for Universal Kit 1.0 by changing insertions to deletions in the Deletions row. Removed Chrome from the list of supported browsers for MiSeq Reporter off-instrument use. Corrected formatting errors.
Part # 15038356 Rev. A	March 2014	Initial Release

Table of Contents

[This page intentionally left blank]

Overview

Chapter 1

Introduction

The MiSeqDx™ instrument includes three software applications that work in sequence to produce images of clusters on the flow cell, perform image analysis and base calling, and perform on-instrument secondary analysis.

- During the run, MiSeq Operating Software (MOS) captures images of clusters on the flow cell for image analysis, as well as operates the flow cell stage, gives commands to dispense reagents, and changes temperatures of the flow cell.
- The integrated primary analysis software, Real Time Analysis (RTA), performs image analysis and base calling, and assigns a quality score to each base for each cycle as the run progresses. The completion of primary analysis by RTA initiates MiSeq Reporter to begin secondary analysis.
- MiSeq Reporter performs on-instrument secondary analysis on base calls and quality scores generated by RTA during the sequencing run. MiSeq Reporter runs as a Windows service and is viewed through a web browser. Alternatively, it can be installed on an off-instrument computer. For more information, see *Installing MiSeq Reporter Off-Instrument* on page 35.

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 MiSeqDx instrument computing resources are dedicated to either sequencing or analysis. If a new sequencing run is started on the MiSeqDx before secondary analysis of an earlier run is complete, a confirmation dialog box appears. After confirming the sequencing run, secondary analysis stops.

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 the MiSeq Reporter

The MiSeq Reporter interface can only be viewed through a web browser. To view the MiSeq Reporter interface, open any web browser on a computer with access to the same network as the MiSeqDx instrument. 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 MiSeqDx followed by 8042

Network Name	HTTP Service Port	HTTP Address
MiSeqDx01, for example	8042	MiSeqDx01:8042

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
Manifest	The file that specifies a reference genome and targeted reference regions to be used in the alignment step. The manifest file used by the Cystic Fibrosis assays is pre-loaded on the MiSeqDx.
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).
Sample Sheet	A comma-separated values file (*.csv) which contains information required to set up and analyze a sequencing run, including a list of samples and their index sequences. This is created off-instrument using the Illumina Worklist Manager. The sample sheet must be provided during the run setup steps on the MiSeqDx. After the run begins, the sample sheet is automatically renamed to SampleSheet.csv and copied to the run folders: MiSeqOutput and MiSeqAnalysis.
Workflow	A secondary analysis procedure performed by MiSeq Reporter. The workflow for each run is specified in the sample sheet information.

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.

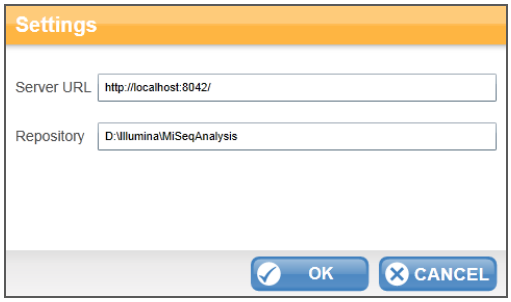
Figure 1 MiSeq Reporter Main Screen

Server URL or Repository Settings

Use the Settings feature to change 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



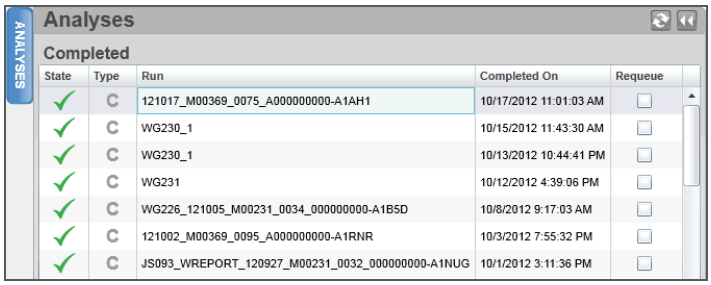
Typically, it is not necessary to change these settings unless MiSeq Reporter is running off-instrument. In this case, set the repository path to the network location of the MiSeqOutput folder. For more information, see *Using MiSeq Reporter Off-Instrument* on page 37.

Analyses Tab

The Analyses tab lists all the sequencing runs located in the specified repository. From this tab, results from any of the runs listed can be opened, or a selected run can be requeued for analysis.

Select the **Refresh Analysis List** icon in the upper-right corner to refresh the list at any time.

Figure 3 Analyses Tab Expanded



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 errors occurred and secondary analysis was not completed successfully.

- **Type**—Lists the analysis workflow associated with each run using a single letter designator. For the CF assays and the Universal Kit 1.0, the letter designator is **C**.
- **Run**—The name of the run folder in the MiSeqOutput and MiSeqAnalysis folders.
- **Completed On**—The date that secondary analysis completed.
- **Requeue**—Select the checkbox to requeue a specific job for analysis. The Requeue button appears. For more information, see *Requeue Analysis* on page 12.

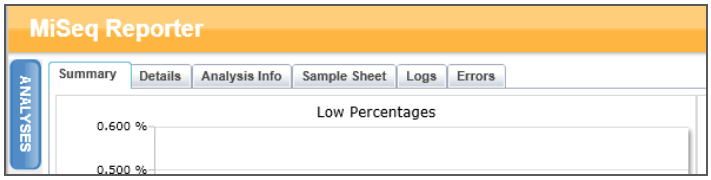
When analysis is queued, the run appears at the bottom of the Analyses tab and indicated as in-progress with the icon .

Figure 4 Queued Run in Analyses Tab

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: Summary, Details, Analysis Info, Sample Sheet, Logs, and Errors. Information on the Analysis Info and Sample Sheet tabs appear initially. All tabs are populated when analysis is complete.

Figure 5 Information and Results Tabs

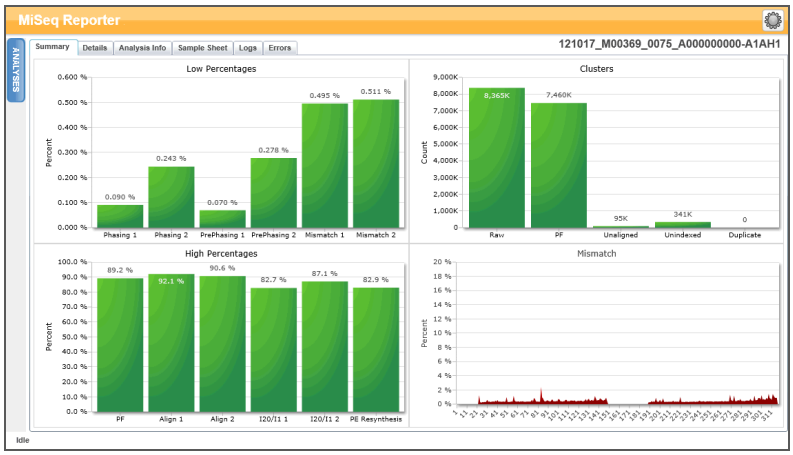


Summary Tab

The Summary tab contains a summary of analysis results. Four graphs appear on the Summary tab:

- **Low Percentages Graph**—Shows phasing, prephasing, and mismatches in percentages. Low percentages indicate good run statistics. For more information, see *Phasing and Prephasing* on page 13.
- **High Percentages Graph**—Shows clusters passing filter, alignment to a reference, and intensities in percentages. High percentages indicate good run statistics.
- **Clusters Graph**—Shows numbers of raw clusters, clusters passing filter, clusters that did not align, clusters not associated with an index, and duplicates.
- **Mismatch Graph**—Shows mismatches per cycle. A mismatch refers to any mismatch between the sequencing read and a reference genome after alignment.

Figure 6 Summary Tab



Details Tab

The Details tab contains details of analysis results. The following tables and graphs may appear on the Details tab depending on the assay or kit used:

- **Samples Table**—Summarizes the sequencing results for each sample.
- **Targets Table**—Shows statistics for the targeted regions of a selected sample. (Universal Kit 1.0 only)
- **Variants Table**—Shows differences between sample DNA and the reference.
- **Coverage Graph**—Shows how deeply the sample was sequenced by measuring the number of bases present in the sample sequence for each position of the reference.
- **Qscore Graph**—Shows the average quality score, which is the estimated probability of a base call error. For more information see *Qscore Graph* on page 29.
- **Variant Score Graph**—Shows the location of SNVs and indels.

Figure 7 Details Tab for CF 139-Variant Assay, Example

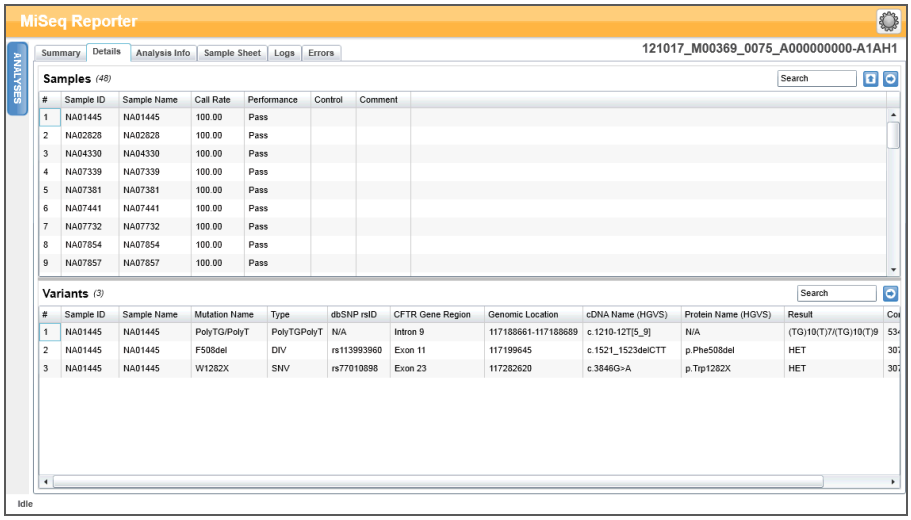
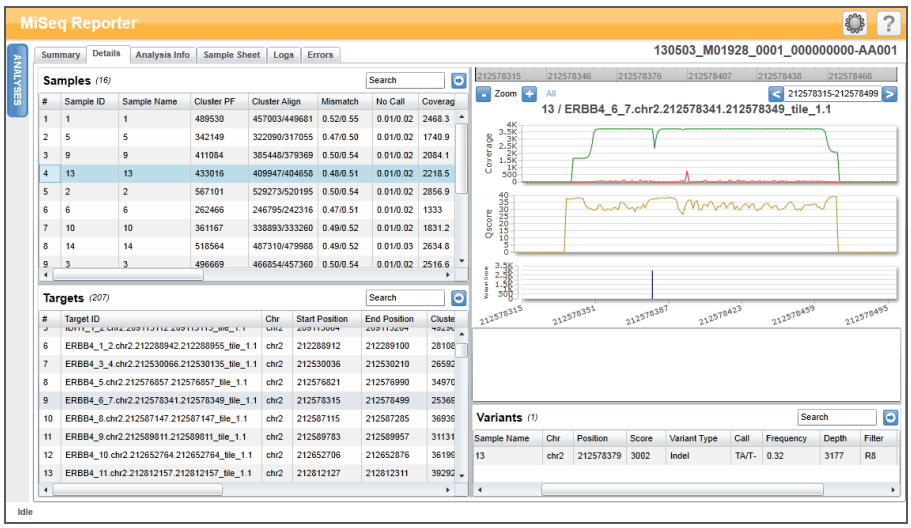


Figure 8 Details Tab for Universal Kit 1.0, Example



Results in the Samples, Targets, or Variants tables can be exported individually to a text file using the **Export table data to text file** icon. This export does not alter the analysis report file.

For CF assays, results can be exported to the CF analysis report file using the **Export data to CF report** icon. For more information, see *Analysis Output Files for the CF Assays* on page 31.

Analysis Info Tab

The Analysis Info tab contains logistical information about the run and analysis.

Figure 9 Analysis Info Tab

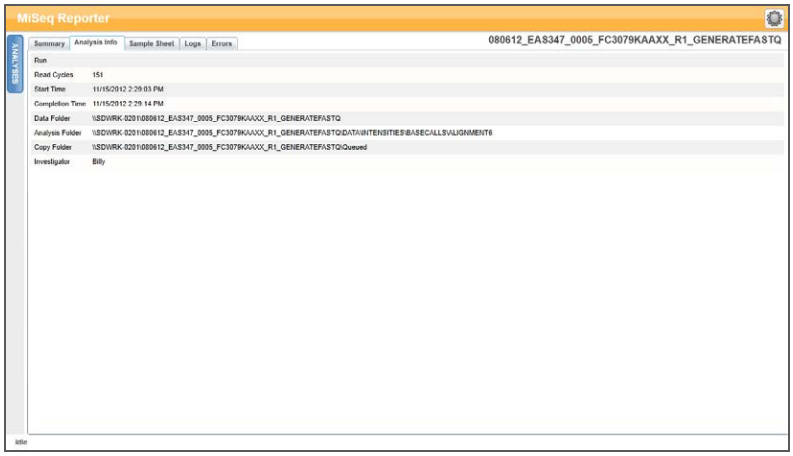


Table 2 Analysis Info Tab

Row	Description
Read Cycles	A representation of the number of cycles in each read, including notation for any index reads. For example, a run noted as 151, 8 (I), 8 (I), 151 indicates 151 cycles for the first read, 8 cycles for the first index read, 8 cycles for the second index read, and then a final read of 151 cycles.
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

The Sample Sheet tab contains run parameters specified in the sample sheet, and provides tools to edit the sample sheet and then requeue the run.

Figure 10 Sample Sheet Tab, Example Universal Kit 1.0

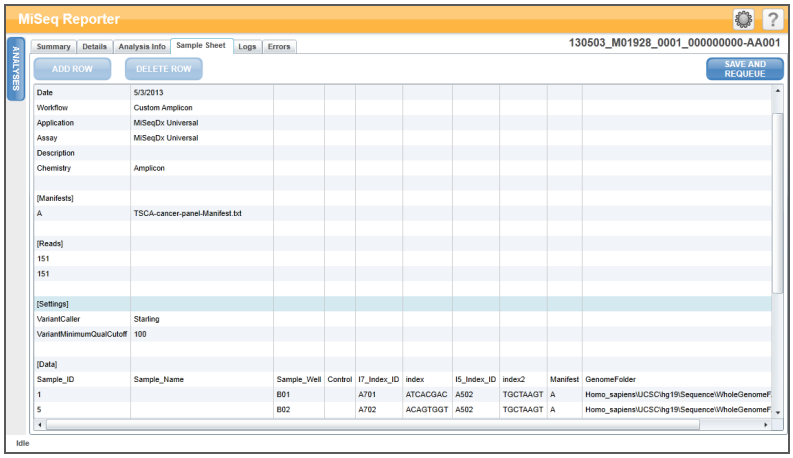


Table 3 Sample Sheet Tab Contents

Row	Description
Date	The date the sequencing run was performed.
Workflow	The analysis workflow for the run. For the CF Assays and the Universal Kit 1.0, the workflow name is Custom Amplicon.
Application	The application name. Used by the Illumina Worklist Manager software, this field indicates which assay or kit is used for the run.
Assay	The name of the assay or kit.
Chemistry	The chemistry name identifies recipe fragments used to build the run-specific recipe. For MiSeqDx runs, the chemistry name is amplicon.
Manifests	The name of the manifest file that specifies a reference genome and targeted reference regions to be used in the alignment step.
Reads	The number of cycles performed in Read 1 and Read 2. Index reads are not included in this section.
Settings	Optional run parameters.
Data	The sample ID, sample name, index sequences, and path to the genome folder. Requirements vary by workflow.

Logs Tab

The 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

The Errors tab lists any errors that occurred during analysis. A summary is written to AnalysisError.txt, which is an important file for troubleshooting purposes.

Editing the Sample Sheet in MiSeq Reporter

Sample sheet data can be edited 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.

CAUTION
Editing sample sheet information should be done with extreme care and caution. Sample tracking can be altered and potentially lead to incorrect results reporting.

- To edit a row in the sample sheet, click in the field and then make required changes.
- To add a row to the sample sheet, click in the row and select **Add Row**. The new row appears below the selected row.

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

- After changes to the sample sheet are complete, select **Save and Requeue**. This saves the changes and initiates secondary analysis using the edited sample sheet.

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

Saving Graphs as Images

MiSeq Reporter provides the option to save an image of the graphs generated for a run. Right-click over 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 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

It is possible to requeue analysis from the MiSeq Reporter web interface. Before proceeding, check that a sequencing run is not in progress.

Each time analysis is requeued, a new Alignment folder is created in the MiSeqAnalysis folder with a sequential number appended to the folder name. For example, Alignment, Alignment1, Alignment2.

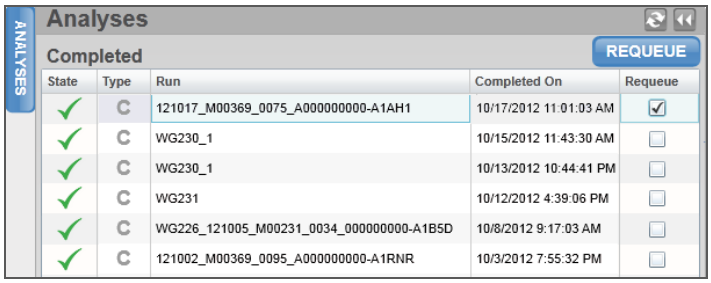
MiSeqAnalysis\<RunFolderName>\Data\Intensities\BaseCalls\Alignment2

1 From the MiSeq Reporter web interface, click **Analyses**.

2 Locate the run on the list of available runs and click the Requeue checkbox adjacent to the run name.

If the run is not listed, change the specified repository to the correct location. For more information, see *Server URL or Repository Settings* on page 5.

Figure 11 Requeue Analysis



3 Click **Requeue**. The State icon to the left of the run name changes to show that analysis is in progress .

NOTE
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.

Analysis Metrics

During the sequencing run, Real Time Analysis (RTA) generates data files that include analysis metrics used by MiSeq Reporter for secondary analysis. Metrics that appear in secondary analysis reports are clusters passing filter, base call quality scores, and phasing and prephasing values.

Clusters Passing Filter

Clusters passing filter is a measurement of cluster quality. This filter removes the least reliable data by filtering raw data to remove any reads that do not meet overall quality. Clusters passing filter are represented by PF in analysis reports.

Quality Scores

A quality score, or Q-score, is a prediction of the probability of an incorrect base call. During the sequencing run, base call quality scores are recorded for each cycle. During analysis, quality scores are recorded in FASTQ files in an ASCII encoded format.

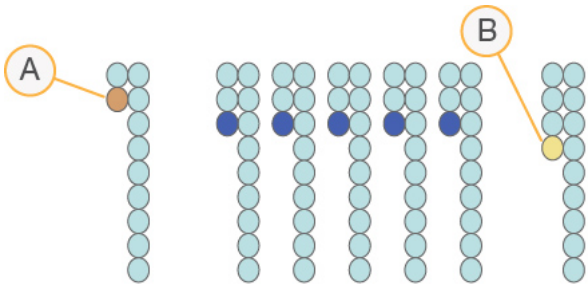
The following table shows the relationship between the quality score and error probability.

Quality Score	Error Probability
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)

Phasing and Prephasing

During the sequencing reaction, each DNA strand in a cluster extends by 1 base per cycle. A small portion of strands might become out of phase with the current incorporation cycle, either falling a base behind (phasing) or jumping a base ahead (prephasing). Phasing and prephasing rates indicate an estimate of the fraction of molecules that became phased or prephased in each cycle.

Figure 12 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 two 151-cycle reads (2 x 151) for a total of 302 cycles. At the end of the run, 2 x 150 cycles are analyzed. The one extra cycle for Read 1 and Read 2 is required for prephasing calculations.

Analysis Procedures

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

Demultiplexing

Demultiplexing is the first step in analysis if the sample sheet lists multiple samples and the run has index reads.

Demultiplexing separates data from pooled samples based on short index sequences that tag samples from different libraries. Each index read sequence is compared to the index sequences specified in the sample sheet. No quality values are considered in this step.

FASTQ File Generation

After demultiplexing, this procedure generates intermediate files in the FASTQ file format, which is a text format used to represent sequences. FASTQ files are the primary input for the alignment step. FASTQ files contain the reads for each sample and the quality scores, excluding reads from any clusters that did not pass filter.

Alignment

Alignment compares sequences against the reference to identify a relationship between the sequences and assigns a score based on regions of similarity. Aligned reads are written to files in BAM format.

For data generated on the MiSeq Reporter, MiSeqDx uses a banded Smith-Waterman algorithm, which 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.

Variant Calling

Variant calling records single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and other structural variants.

For data generated on the MiSeqDx instrument, variant calling is performed by the Starling variant caller in MiSeq Reporter. Starling calls SNPs and small indels, and summarizes depth and probabilities of error for every site in the genome. For each SNP or indel call, the probability of an error is provided as a variant quality score.

Upon completion, Starling produces html-formatted reports of SNPs and indels, and tab-delimited text files containing variants in the Variant Call Format (VCF). For more information, *VCF File Format* on page 42.

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.

[This page intentionally left blank]

Data Visualization

Chapter 2

Introduction

MiSeq Reporter performs secondary analysis and generates various types of information specific to the assay upon completion of analysis. Results appear on the MiSeq Reporter web interface in the form of graphs and tables for each run. MiSeqDx products include those listed in the table below:

Product	Description
Cystic Fibrosis 139-Variant Assay	Detects 139 clinically relevant variants in the CFTR gene from a maximum of 48 samples.
Cystic Fibrosis Clinical Sequencing Assay	Detects mutations in the protein coding regions including intron/exon boundaries, two large deletions, and two-deep intronic mutations in the CFTR gene from a maximum of 8 samples.
Universal Kit 1.0	Set of reagents and consumables used along with a user-supplied custom oligo to perform targeted resequencing of specific genomic regions of interest.

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.

Pre-Installed Databases and Genomes

The MiSeqDx includes pre-installed databases and genomes.

Pre-Installed	Description
Databases	dbSNP for human, version 131 refGene for human
Genomes	human (*Homo sapiens*) build hg19

Custom Amplicon Workflow

The Custom Amplicon workflow used for the CF assays and Universal Kit 1.0 evaluates short regions of amplified DNA, or amplicons, for variants. Focused sequencing of amplicons enables high coverage of particular regions across a large number of samples.

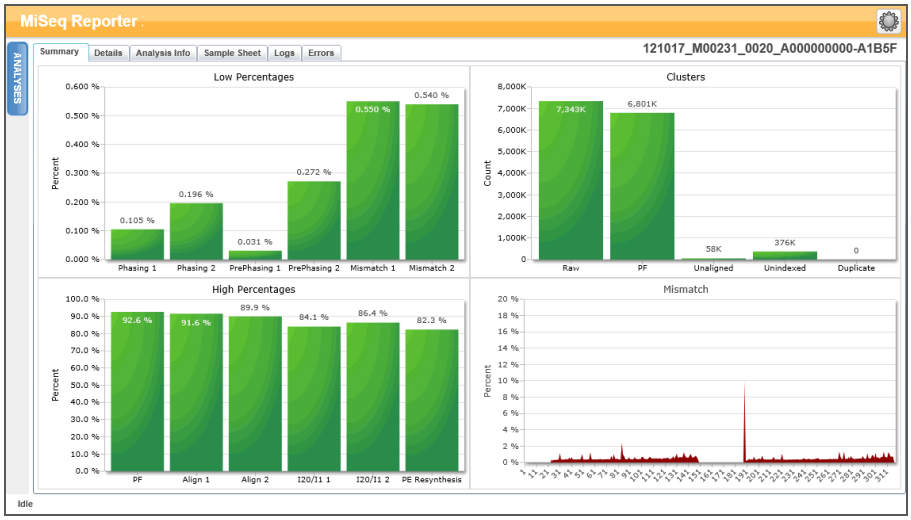
After demultiplexing and FASTQ file generation, the workflow performs the following steps:

- **Alignment**—Clusters from each sample are aligned against amplicon sequences specified in the manifest file.
  - For paired-end data, each read is initially evaluated in terms of its alignment to the relevant probe sequences for that read. Read 1 is evaluated against the reverse compliment of the downstream locus specific oligos (DLSO) and Read 2 is evaluated against the upstream locus specific oligos (ULSO). If the start of a read sequence matches a probe sequence with no more than one mismatch, the full length of the read is then aligned against the amplicon target sequence for that probe sequence. This alignment is performed along the length of the amplicon target sequences using a banded Smith-Waterman alignment.
  - Indels within the DLSO and ULSO are not observed given the assay chemistry.
- **Paired-end evaluation**—For paired-end runs, the top-scoring alignment for each read is considered. If either read did not align or aligned to different chromosomes, the reads are flagged as an unresolved pair. Additionally, if the two alignments come from different amplicons (i.e., different rows in the Targets section of the manifest), the reads are flagged as an unresolved pair.
- **Bin/Sort**—Reads are grouped by sample and chromosome, and then sorted by chromosome position. Results are written to one BAM file per sample.
- **Variant calling**—Mutations are identified by the variant caller. For more information, see *Variant Calling* on page 14.
- **Variant analysis and annotation**—Using a pre-installed SNP database (dbsnp.txt), any known mutations are flagged in the analysis report file.
- **Statistics reporting**—Statistics are summarized and reported.

Summary Tab

The information that appears on the Summary tab includes a low percentages graph, high percentages graph, clusters graph, and mismatch graph.

Figure 13 Example Summary Tab



Low Percentages Graph

Y Axis	X Axis	Description
Percent	Phasing 1	The percentage of molecules in a cluster that fall behind the current cycle within Read 1.
	Phasing 2	The percentage of molecules in a cluster that fall behind the current cycle within Read 2.
	PrePhasing 1	The percentage of molecules in a cluster that run ahead of the current cycle within Read 1.
	PrePhasing 2	The percentage of molecules in a cluster that run ahead of the current cycle within Read 2.
	Mismatch 1	The average percentage of mismatches for Read 1 over all cycles.
	Mismatch 2	The average percentage of mismatches for Read 2 over all cycles.

High Percentages Graph

Y Axis	X Axis	Description
Percent	PF	The percentage of clusters passing filters.
	Align 1	The percentage of clusters that aligned to the reference in Read 1.
	Align 2	The percentage of clusters that aligned to the reference in Read 2.
	I20 / I1 1	The ratio of intensities at cycle 20 to the intensities at cycle 1 for Read 1.
	I20 / I1 2	The ratio of intensities at cycle 20 to the intensities at cycle 1 for Read 2.
	PE Resynthesis	The ratio of first cycle intensities for Read 1 to first cycle intensities for Read 2.

Clusters Graph

Y Axis	X Axis	Description
Clusters	Raw	The total number of clusters detected in the run.
	PF	The total number of clusters passing filter in the run.
	Unaligned	The total number of clusters passing filter that did not align to the reference genome, if applicable. Clusters that are unindexed are not included in the unaligned count.
	Unindexed	The total number of clusters passing filter that were not associated with any index sequence in the run.
	Duplicate	This value is not applicable to CF assays or the Universal Kit 1.0, so it will always be zero.

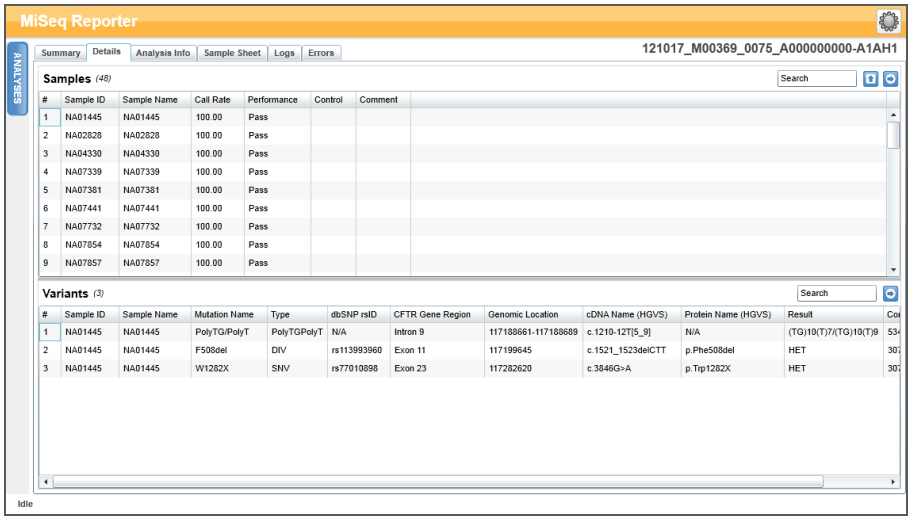
Mismatch Graph

Y Axis	X Axis	Description
Percent	Cycle	Plots the percentage of mismatches for all clusters in a run by cycle.

Details Tab for CF 139-Variant Assay

The information that appears on the Details tab for the CF 139-Variant Assay includes a samples table and a variants table.

Figure 14 Details Tab for CF 139-Variant Assay, Example



Samples Table for CF 139-Variant Assay

Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.

Column	Description
Call Rate	The number of mutational positions that meet a predefined confidence value threshold divided by the total mutational positions interrogated. Call rate is described on a per-sample basis and reported as percentage that is calculated as 1 minus [number of positions with incomplete calls divided by the total number of positions sequenced].
Performance	Pass or Fail rating based on the call rate. For a positive control sample: • PASS—with a call rate ≥ 99% • FAIL—with a call rate < 99% For a negative control sample: • PASS—with a call rate ≤ 10% • FAIL—with a call rate > 10% For a sample not labeled as a positive or negative control: • PASS—with a call rate ≥ 99% • FAIL—with a call rate < 99%
Control	The type of control as listed in the sample sheet. Values are positive or negative. A blank field indicates sample only.
Comment	An optional text field for comments. Comments entered in this field are saved in the analysis report file, MiSeqDxCF139VariantAssay.txt. If analysis is requeued, a new report file is written. Comments from a previous analysis run do not carry over to the next analysis run.

Variants Table for CF 139-Variant Assay

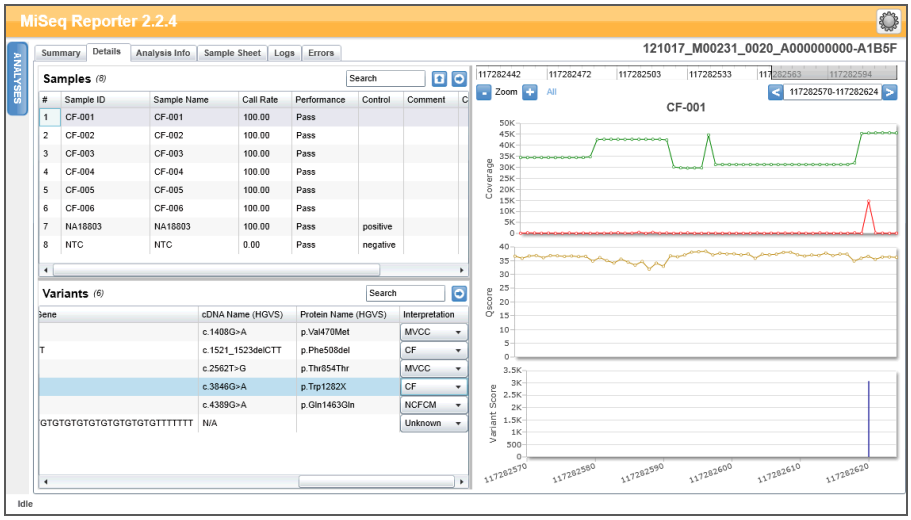
Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Mutations (Common Name)	Common name of the Cystic Fibrosis variant, as described in the CFTR2 database.
Mutation Type	The type of variant. • SNV—Single Nucleotide Variant • DIV—Deletion Insertion Variant • DEL—Large deletion • PolyTGPolyT—PolyTG/PolyT genotype in CF gene
dbSNP rsID	The dbSNP rsID of the variant, if applicable.
CFTR Gene Region	CFTR gene region (exon # or intron #) where variant is present.
Genomic Location	Genomic location of the variant.

Column	Description
cDNA Name (HGVS)	Description of a variant at the DNA-level using the coding DNA (cDNA) sequence nomenclature as recommended by the Human Genome Variation Society (HGVS).
Protein Name (HGVS)	Description of a variant at the protein-level using the protein sequence nomenclature as recommended by the Human Genome Variation Society (HGVS).
Result	Variant genotype. For SNVs, DIVs and DELs: • HET—Heterozygous • HOM—Homozygous For PolyTGPolyT variant, the actual genotype is reported. **NOTE**: PolyTGPolyT is reported only when the R117H variant is detected.

Details Tab for CF Clinical Sequencing Assay

The information that appears on the Details tab for the CF Clinical Sequencing Assay includes a samples table, variants table, coverage graph, Qscore graph, and variant score graph.

Figure 15 Details Tab for CF Clinical Sequencing Assay, Example



Samples Table for CF Clinical Sequencing Assay

Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.

Column	Description
Call Rate	The number of bases that meet a quality score threshold divided by the total bases interrogated. Call rate is described on a per-sample basis and reported as percentage that is calculated as 1 minus [number of positions with incomplete calls divided by the total number of bases/positions sequenced].
Performance	Pass or Fail rating based on the call rate. For a positive control sample: • PASS—with a call rate ≥ 99% • FAIL—with a call rate < 99% For a negative control sample: • PASS—with a call rate ≤ 10% • FAIL—with a call rate > 10% For a sample not labeled as a positive or negative control: • PASS—with a call rate ≥ 99% • FAIL—with a call rate < 99%
Control	The type of control as listed in the sample sheet. Values are positive or negative. A blank field indicates sample only.
Comment	An optional text field for comments. Comments entered in this field are saved in the analysis report file, MiSeqDxCFClinicalSequencing.txt. If analysis is requeued, a new report file is written. Comments from a previous analysis run do not carry over to the next analysis run.
Coordinates Not Called	Genome coordinates within the targeted region where a call was not reported due to low confidence values.

Variants Table for CF Clinical Sequencing Assay

Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Chr	The reference target or chromosome name.
Position	The position at which the variant was found.
Variant Type	The type of variant. • SNV—Single Nucleotide Variant • DIV—Deletion Insertion Variant • DEL—Large deletion • PolyTGPolyT—PolyTG/PolyT genotype in CF gene
Call	A string representing how the base or bases changed at this location in the reference.

Column	Description
Frequency	The fraction of reads for the sample that includes the variant. For example, if the reference base at a particular position is A and sample 1 has 60 A reads and 40 T reads, then the SNV has a variant frequency of 0.4.
Depth	The number of reads for a sample covering a particular position.
Filter	The criteria for a filtered variant.
dbSNP ID	The dbSNP name of the variant.
RefGene	The gene according to RefGene in which this variant appears.
cDNA Name (HGVS)	Description of a variant at the DNA-level using the coding DNA (cDNA) sequence nomenclature as recommended by the Human Genome Variation Society (HGVS).
Protein Name (HGVS)	Description of a variant at the protein-level using the protein sequence nomenclature as recommended by the Human Genome Variation Society (HGVS).
Interpretation	This field enables the Medical Geneticist to provide clinical interpretation of the mutation for each sample. The following options are included in the drop-down list for each sample: • CF—CF causing • MVCC—Mutation of Varying Clinical Consequence • MOUS—Mutation of Unknown Significance • NCFCM—Non CF Causing Mutation • Unknown A new report can be generated using the icon.

Variants Table Interpretation Column

The Interpretation column provides selections that enable the Medical Geneticist to provide a clinical interpretation of the mutation for each sample. The following options are included in the Interpretation drop-down list:

- **CF**—CF causing
- **MVCC**—Mutation of Varying Clinical Consequence
- **MOUS**—Mutation of Unknown Significance
- **NCFCM**—Non CF Causing Mutation
- **Unknown**

Figure 16 Interpretation Column

Results in the Variants tables can be exported individually to a text file using the **Export table data to text file** icon. This export does not alter the analysis report file.

After the Medical Geneticist has completed determining variant significance, interpretation settings can be saved to the analysis report. The file name of the original analysis report will automatically be appended with a time/date stamp.

Coverage Graph for CF Clinical Sequencing Assay

Y Axis	X Axis	Description
Coverage	Position	The green curve is the number of aligned reads that cover each position in the reference. The red curve is the number of aligned reads that have a miscall at this position in the reference. SNVs and other variants show up as spikes in the red curve.

Qscore Graph

Y Axis	X Axis	Description
Qscore	Position	The average quality score of bases at the given position of the reference.

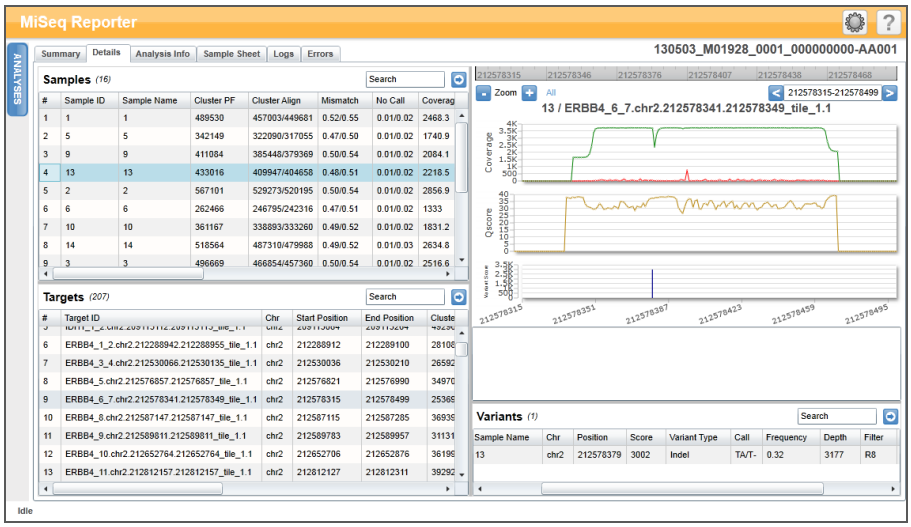
Variant Score Graph for CF Clinical Sequencing Assay

Y Axis	X Axis	Description
Score	Position	Graphically depicts quality score and the position of SNVs and indels.

Details Tab for Universal Kit 1.0

The information that appears on the Details tab for the Universal Kit 1.0 includes a samples table, targets table, coverage graph, Q-score graph, variant score graph, and variants table.

Figure 17 Details Tab for Universal Kit 1.0, Example



Samples Table for Universal Kit 1.0

Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Cluster PF	The number of clusters passing filter for the sample.
Cluster Align	The total count of PF clusters aligning for the sample (Read 1/Read 2).
Mismatch	The percentage mismatch to reference averaged over cycles per read (Read 1/Read 2).
No Call	The percentage of bases that could not be called (no-call) for the sample averaged over cycles per read (Read 1/Read 2).
Coverage	Median coverage (number of bases aligned to a given reference position) averaged over all positions.
Het SNPs	The number of heterozygous SNPs detected for the sample.
Hom SNPs	The number of homozygous SNPs detected for the sample.
Insertions	The number of insertions detected for the sample.
Deletions	The number of deletions detected for the sample.
Manifest	The file that specifies a reference genome and targeted reference regions to be used in the alignment step.
Genome	The name of the reference genome.

Targets Table for Universal Kit 1.0

Column	Description
#	An ordinal identification number in the table.
Target ID	The name of the target in the manifest.
Chr	The reference target or chromosome name.
Start Position	The start position of the target region.
End Position	The end position of the target region.
Cluster PF	Number of clusters passing filter for the target displayed per read (Read 1/Read 2).
Mismatch	The percentage of mismatched bases to target averaged over all cycles, displayed per read. Mismatch = [mean(errors count in cycles) / cluster PF] * 100.

Column	Description
No Call	The percentage of no-call bases for the target averaged over cycles, displayed per read.
Het SNPs	The number of heterozygous SNPs detected for the target across all samples.
Hom SNPs	The number of homozygous SNPs detected for the target across all samples.
Insertions	The number of insertions detected for the target across all samples.
Deletions	The number of deletions detected for the target across all samples.
Manifest	The file that specifies a reference genome and targeted reference regions to be used in the alignment step.

Coverage Graph for Universal Kit 1.0

Y Axis	X Axis	Description
Coverage	Position	The green curve is the number of aligned reads that cover each position in the reference. The red curve is the number of aligned reads that have a miscall at this position in the reference. SNPs and other variants show up as spikes in the red curve.

Qscore Graph

Y Axis	X Axis	Description
Qscore	Position	The average quality score of bases at the given position of the reference.

Variant Score Graph for Universal Kit 1.0

Y Axis	X Axis	Description
Score	Position	Graphically depicts variant quality score and the position of SNPs and indels.

Variants Table for Universal Kit 1.0

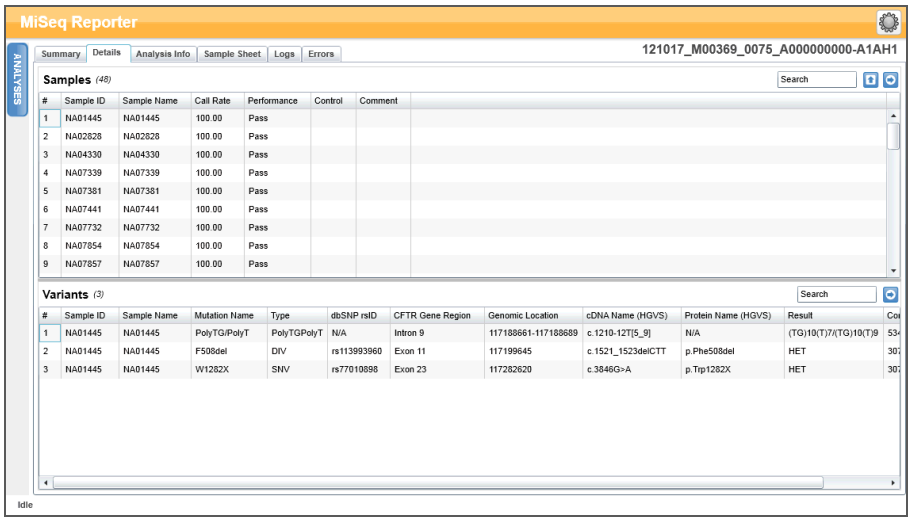
Column	Description
#	An ordinal identification number in the table.
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Chr	The reference target or chromosome name.
Position	The position at which the variant was found.
Score	The variant quality score for this variant.

Column	Description
Variant Type	The variant type, which can be either SNP or indel.
Call	A representation of how the base or bases changed at this location in the reference. • SNPs are listed in the format Reference > AlleleA/AlleleB. • Insertions are listed in the format Reference/Insertion. G-/GA shows the insertion of A. • Deletions are listed in the format Reference/Deletion. AGG/A-- shows the deletion of GG.
Frequency	The fraction of reads for the sample that includes the variant. For example, if the reference base is A, and sample 1 has 60 A reads and 40 T reads, then the SNP has a variant frequency of 0.4.
Depth	The number of reads for a sample covering a particular position.
Filter	The criteria for a filtered variant. If all filters are passed, PASS is written in the filter column. For more information, see *VCF File Headings and Annotations* on page 43.
dbSNP	The dbSNP name of the variant, if applicable.
RefGene	The gene according to RefGene in which this variant appears.

Analysis Output Files for the CF Assays

Analysis results for the CF assays appear on the Details tab.

Figure 18 Details Tab for CF 139-Variant Assay, Example



Results in the Variants tables can be exported individually to a text file using the **Export table data to text file** icon. This export does not alter the analysis report file.

After the Medical Geneticist has completed determining variant significance, interpretation settings can be saved to the analysis report. The file name of the original analysis report will automatically be appended with a time/date stamp.

Output files for the CF assays are also summarized in one tab-delimited text file named after the assay used for the run. These results are identical to what is found on the Details tab.

- For the CF 139-Variant Assay, the file is named MiSeqDxCF139VariantAssay.txt.
- For the CF Clinical Sequencing Assay, the file is named MiSeqDxCFClinicalSequencingAssay.txt.

When analysis is complete, the output file is written to the Alignment folder for the run. For example:

MiSeqAnalysis\<RunFolderName>\Data\Intensities\BaseCalls\Alignment

If analysis has been repeated or requeued, a new report file is written to the Alignment for that analysis run. For more information, see *Requeue Analysis* on page 12.

The output file contains a header that includes the following information about the run:

Header	Description
Test	This describes the test that was performed.
Run ID	This is the run ID that was generated by MOS at the beginning of the sequencing run.
Run Date	This is the date (DDMMYY) that the sequencing run was started in MOS.
Analysis Version	This is the version of MiSeq Reporter that was used for analysis.

Figure 19 Header for the CF 139-Variant Assay Output File, Example

```
Test CF 139-Variant Assay
For In Vitro Diagnostic Use.
Run ID 140212_M01018_0071_000000000-A2618
Run Date 140212
Analysis Version 2.2.31.1
```

Following the header is a summary section for each sample ID that contains columns for each reported value. For column descriptions, see *Details Tab for CF 139-Variant Assay* on page 22 and *Details Tab for CF Clinical Sequencing Assay* on page 24.

NOTE

The analysis pipeline that generates output files is not identical between the CF assays and the Universal Kit 1.0. The output files generated for the Universal Kit 1.0 are *.bam files, *.vcf files, and AmpliconCoverage_M#.tsv files. For more information on output files for the Universal Kit 1.0, see Appendix A Universal Kit 1.0 Analysis Output Files.

Installation and Troubleshooting

Chapter 3

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 MiSeqDx 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
- 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.

Set Up User or Group Accounts on Windows 7

Administrator rights are required to configure user or group accounts to enable **Log on as a service** permission. If needed, contact the local facility administrator for assistance.

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 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.
   Example:
   ```
   <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.
Example:
```
<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, folders containing run data and reference genomes must be accessible.

1. Unless a network location for sequencing data and reference genomes is used, copy the following folders to the local computer:
   - Copy run data from the MiSeqDx computer in D:\MiSeqOutput\<RunFolder>.
   - Copy reference genomes from the MiSeqDx computer in C:\Illumina\MiSeq Reporter\Genomes.
2. Open a web browser to http://localhost:8042, which opens the MiSeq Reporter web interface.
3. Change the path to the Repository using the **Settings** icon in the top-right corner of the web interface.

   NOTE
   Specifying the Repository path in Settings is temporary. The next time the computer is restarted, 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 analysis of a run can be requeued using an off-instrument installation of MiSeq Reporter, the path to the GenomeFolder must be updated in the sample sheet, which can be done from the Sample Sheet tab. After updating the GenomeFolder path, click **Save and Requeue**.
   For more information, see *Editing the Sample Sheet in MiSeq Reporter* on page 11.

Troubleshooting MiSeq Reporter

MiSeq Reporter runs as Windows service application. User accounts must be configured to enable **Log on as a service** permissions 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.

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 Fail 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:
   - For off-instrument installations, check the repository location using Settings on the MiSeq Reporter web interface.
   - For on-instrument installations, check the MiSeqOutput folder location on the MOS Run Options screen, Folder Settings tab.

   The full UNC path (e.g., \\server1\Runs) must be used. Because MiSeq Reporter runs as a Windows service, it does not recognize user-mapped drives (e.g., Z:\Runs).
2. Confirm write-access to the output folder location. For assistance, contact the local facility administrator.
3. Check that copying is not disabled in the MiSeq Reporter.exe.config. This setting is located in the <appSettings> section and the value should be set to **1**.

```
<add key="CopyToRTAOutputPath" value="1"/>
```

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.

Universal Kit 1.0 Analysis Output Files

Appendix A

Analysis Output File Types

The following table describes output files generated for the Universal Kit 1.0, which provide analysis results for alignment, variant calling, and coverage.

File Name	Description
*.bam files	Contains aligned reads for a given sample. Located in Data\Intensities\BaseCalls\Alignment.
*.vcf files	Contains information about variants found at specific positions in a reference genome. Located in Data\Intensities\BaseCalls\Alignment.
AmpliconCoverage_M#.tsv	Contains details about the resulting coverage per amplicon per sample. M# represents the manifest number. Located in Data\Intensities\BaseCalls\Alignment.

NOTE
The analysis pipeline that generates these output files is not identical between the CF assays and the Universal Kit 1.0. This section describes the analysis output files for Universal Kit 1.0 only.

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.

BAM files are written to the alignment folder in Data\Intensities\BaseCalls\Alignment in 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 and sample length. Alignments in the alignments section are associated with specific information in the header section.
- **Alignments**—Contains read name, read sequence, read quality, and custom tags.

  Figure 20 Example, BAM File Alignment Section

  ```
  GA23_40:8:1:10271:11781 64 chr22 17552189 8 35M * 0 0
  TACAGACATCCACCACCACACCCAGCTAATTTTTG
  IIIII>FA?C::B=:GGGB>GGGEGIIIHI3EEE#
  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.

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. The last line in the header is column headings for the data lines. For more information, see *VCF File Headings and Annotations* on page 43.

  Figure 21 Example, VCF File Header

```
##fileformat=VCFv4.1
##FORMAT=<ID=GQX,Number=1,Type=Integer,Description="Minimum of
   {Genotype quality assuming variant position,Genotype quality
   assuming non-variant position}">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Float,Description="Genotype
   Quality">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths
   for the ref and alt alleles in the order listed">
##FORMAT=<ID=VF,Number=1,Type=Float,Description="Variant
   Frequency, the ratio of the sum of the called variant depth to
   the total depth">
##INFO=<ID=TI,Number=.,Type=String,Description="Transcript ID">
##INFO=<ID=GI,Number=.,Type=String,Description="Gene ID">
##INFO=<ID=EXON,Number=0,Type=Flag,Description="Exon Region">
##INFO=<ID=FC,Number=.,Type=String,Description="Functional
   Consequence">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in
   genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency,
   for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of
   alleles in called genotypes">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read
   depth; some reads may have been filtered">
##FILTER=<ID=LowVariantFreq,Description="Low variant frequency <
   0.20">
##FILTER=<ID=LowGQ,Description="GQ below < 20.00">
##FILTER=<ID=LowQual,Description="QUAL below < 100.00">
##FILTER=<ID=R8,Description="IndelRepeatLength is greater than
   8">
##fileDate=20130506
##source=Starling 0.3
##phasing=none
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT
```

- **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 and Annotations

The VCF file format is flexible and extensible. The following tables describe the VCF file headings and annotations generated by MiSeq Reporter.

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For SNPs, this position is the reference base with the variant; for indels or deletions, this position is the reference base immediately before the variant.
ID	The rs number for the SNP obtained from dbSNP.txt, if applicable. If there are multiple rs numbers at this location, the list is 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	If all filters are passed, PASS is written in the filter column. • **LowDP**—Applied to sites with depth of coverage below the cutoff. • **LowGQ**—The genotyping quality (GQ) is below the cutoff. • **LowQual**—The variant quality (QUAL) is below the cutoff. • **LowVariantFreq**—The variant frequency is less than the threshold. • **R8**—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8.
INFO	• **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. • **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. • **TI**—A comma-separated list of transcript IDs read from RefGene.

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

Amplicon Coverage File

One amplicon coverage file is generated for each manifest. The M# in the file name represents the manifest number as it is listed in the sample sheet.

Each file begins with a header row that contains the sample IDs associated with the manifest. The first column contains the Target IDs. Each additional column lists coverage depth for the associated sample ID.

Supplementary Output Files

The following output files provide supplementary information, or summarize run results and analysis errors. Although, these files are not required for assessing analysis results, they can be used for troubleshooting purposes.

File Name	Description
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages. Located in the root level of the run folder.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file is present only if errors occurred. Located in the root level of the run folder.
AmpliconRunStatistics.xml	Contains summary statistics specific to the run. Located in the root level of the run folder.
CompletedJobInfo.xml	Written after analysis is complete, contains information about the run, such as date, flow cell ID, software version, and other parameters. Located in the root level of the run folder.
DemultiplexSummaryF1L1.txt	Reports demultiplexing results in a table with one row per tile and one column per sample. Located in Data\Intensities\BaseCalls\Alignment.
ErrorsAndNoCallsByLaneTile ReadCycle.csv	A comma-separated values file that contains the percentage of errors and no-calls for each tile, read, and cycle. Located in Data\Intensities\BaseCalls\Alignment.
Mismatch.htm	Contains histograms of mismatches per cycle and no-calls per cycle for each tile. Located in Data\Intensities\BaseCalls\Alignment.
Summary.xml	Contains a summary of mismatch rates and other base calling results. Located in Data\Intensities\BaseCalls\Alignment.
Summary.htm	Contains a summary web page generated from Summary.xml. Located in Data\Intensities\BaseCalls\Alignment.

Index

S

T

U

V

W

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 4 Illumina General Contact Information

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

Table 5 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Italy	800.874909
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
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 support.illumina.com/sds.html.

Product Documentation

Product documentation in PDF is available for download from the Illumina website. Go to support.illumina.com, select a product, then click **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

EC REP
Ilumina Cambridge Limited
Chesterford Research Park, Little Chesterford
Saffron Walden, CB10 1XL
UNITED KINGDOM