illumina®

Local Run Manager
Generate FASTQ Analysis Module

Workflow Guide

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

ILLUMINA PROPRIETARY

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

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

Overview

The Local Run Manager Generate FASTQ analysis module first demultiplexes indexed reads, if present, generates intermediate analysis files in the FASTQ file format, and then exits the workflow. No alignment or further analysis is performed. FASTQ files are required input for analysis with third-party analysis tools.

Compatible Library Types

The Generate FASTQ analysis module is compatible with specific library types represented by library kit categories on the Create Run screen. For a current list of compatible library kits, see the Local Run Manager support page on the Illumina website.

Input Requirements

The Generate FASTQ analysis module requires the base call files (*.bcl) and the run summary files generated during the sequencing run. Because the workflow ends after FASTQ file generation, no other input files are required.

About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis parameters for the Generate FASTQ analysis module. For information about the Local Run Manager dashboard and system settings, see the *Local Run Manager Software Guide (document # 1000000002702)*.

Set Parameters

1   Log in to Local Analysis Software.

2   Select **Create Run**, and then select **Generate FASTQ**.

3   Enter a run name that identifies the run from sequencing through analysis.
    The run name can contain alphanumeric characters, spaces, and the following special characters:
    `~!@#$%-_{}.

4   **[Optional]** Enter a run description to further identify the run.
    The run description can contain alphanumeric characters, spaces, and the following special characters:
    `~!@#$%-_{}.

Specify Run Settings

1   Select the library prep kit from the Library Prep Kit drop-down list.

2   Specify the number of index reads.
    ▶   0 for a run with no indexing
    ▶   1 for a single-indexed run
    ▶   2 for a dual-indexed run
    If your library prep kit supports only one option, the index read is automatically selected.

3   Select the read type for the run.
    If your library prep kit supports only one option, the read type is automatically selected.

4   Enter the number of cycles for the run.

5   **[Optional]** For Custom Primers, specify any custom primer information to be used for the run by selecting

the appropriate checkboxes.

Custom primer options vary based on your instrument or Local Run Manager implementation.

Specify Module-Specific Settings

1 Set the Adapter Trimming option.
   Adapter trimming is enabled by default.

2 Set the Reverse Complement option.
   Enabling this setting makes all reads reverse-complemented as they are written to FASTQ files. By default, the reverse complement setting is turned off.

The availability of these settings depend on your Library Prep kit, they are not available for every library prep kit option.

Specify Samples for the Run

Specify samples for the run using the following options:

▶ **Enter samples manually**—Use the blank table at the bottom of the Create Run screen.

▶ **Import sample sheet**—Navigate to an external file in a comma-separated values (*.csv) format.

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

Enter Samples Manually

1 Adjust the samples table to an appropriate number of rows.
   ▶ In the Rows field, use the up/down arrows or enter a number to specify the number of rows to add to the table. Select ➕ to add the rows to the table.
   ▶ Select ✖ to delete a row.
   ▶ Right-click on a row in the table and use the commands in the contextual menu.

2 Enter a unique sample ID in the Sample ID field.
   Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.

3 [Optional] Enter a sample description in the Description field.
   Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.

4 If you have a plated kit, select an index plate well from the Index well drop-down list and skip to step 7.

5 Specify an Index 1 sequence.
   Select **Show Index Sequence/Show Index Names** to toggle between showing the name of the index and the index sequence.

6 Specify an Index 2 sequence.

   > **NOTE**
   > During analysis, the iSeq™ 100, MiniSeq™, and NextSeq™ Systems automatically reverse complement the i5 indexes in custom library prep kits. Make sure that the i5 indexes are in the forward orientation.

7 Select a manifest file from the Manifest drop-down list.

8 [Optional] Enter a project name in the Sample Project field.
   Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.

9 [Optional] Select **Export Sample Sheet** to export the sample information in *.csv format.

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

The exported sample sheet can be used as a template or imported when creating new runs.

10 Select **Save Run**.

Import Sample Sheet

1 If you do not have a sample sheet to import, see *Enter Samples Manually* on page 4 for instructions on how to create and export a sample sheet. Edit the file as follows.

   a Open the sample sheet in a text editor.
   b Enter the sample information in the [Data] section of the file.
   c Save the file. Make sure that the sample IDs are unique.

2 Select **Import Sample Sheet** at the top of the Create Run screen and browse to the location of the sample sheet.
Make sure that the information in the sample sheet is correct. Incorrect information can impact the sequencing run.

> **NOTE**
>
> During analysis, the iSeq™ 100, MiniSeq™, and NextSeq™ Systems automatically reverse complement the i5 indexes in custom library prep kits. If you are importing a sample sheet for a custom library prep kit, make sure that the i5 indexes are in the forward orientation.

3 When finished, select **Save Run**.

Analysis Methods

The Generate FASTQ analysis module performs the following analysis steps and then writes analysis output files to the folder.

▶ Demultiplexes index reads
▶ Generates FASTQ files

Demultiplexing

Demultiplexing compares each Index Read sequence to the index sequences specified for the run. No quality values are considered in this step.

Index reads are identified using the following steps:

▶ Samples are numbered starting from 1 based on the order they are listed for the run.
▶ Sample number 0 is reserved for clusters that were not assigned to a sample.
▶ Clusters are assigned to a sample when the index sequence matches exactly.

FASTQ File Generation

After demultiplexing, the software generates intermediate analysis files in the FASTQ format, which is a text format used to represent sequences. FASTQ files contain reads for each sample and the associated quality scores. Any controls used for the run and clusters that did not pass filter are excluded.

Each FASTQ file contains reads for only one sample, and the name of that sample is included in the FASTQ file name. FASTQ files are the primary input for alignment.

View Analysis Results

1 From the Local Analysis Software dashboard, select the run name.

2 From the Run Overview tab, review the sequencing run metrics.

3 To change the analysis data file location for future requeues of the selected run, select the Edit ☑ icon, and edit the output run folder file path.
   The file path leading up to the output run folder is editable. The output run folder name cannot be changed.

4 [Optional] Select the Copy to Clipboard ▣ icon to copy the output run folder file path.

5 Select the Sequencing Information tab to review run parameters and consumables information.

6 Select the Samples & Results tab to view the analysis report.
   ▶ If analysis was requeued, select the appropriate analysis from the Select Analysis drop-down list.
   ▶ From the left navigation bar, select a sample ID to view the report for another sample.

7 [Optional] Select the Copy to Clipboard ▣ icon to copy the Analysis Folder file path.

Analysis Report

Analysis results are provided on the Samples and Results tab.

Indexing

Table 1 Indexing Table

Column Heading	Description
Index Number	An assigned ID based on the order that samples are listed in the sample table.
Sample ID	The sample ID provided when the run was created.
Sample Description	The sample description provided when the run was created.
Index 1 (i7)	The Index 1 adapter used with the sample.
Index 2 (i5)	The Index 2 adapter used with the sample.

Analysis Output Files

The following analysis output files are generated for the Generate FASTQ analysis module.

File Name	Description
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.

Demultiplexing File Format

The process of demultiplexing reads the index sequence attached to each cluster to determine from which sample the cluster originated. The mapping between clusters and sample number is written to a demultiplexing (*.demux) file for each tile of the flow cell.

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

Demultiplexing files start with a header:

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

▶ Version (4-byte integer), currently 1

▶ Cluster count (4-byte integer)

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

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

▶ In the file name, **F1** represents the flow cell number.

▶ In the file name, **L1** represents the lane number.

▶ Demultiplexing results in a table with one row per tile and one column per sample, including sample 0.

▶ The most commonly occurring sequences in index reads.

FASTQ File Format

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

▶ The identifier

▶ The sequence

▶ A plus sign (+)

▶ The Phred quality scores in an ASCII + 33 encoded format

The identifier is formatted as:

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

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

FASTQ File Names

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

…\samplename_S1_L001_R1_001.fastq.gz

▶ **samplename**—The sample name listed for the sample. 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 for the run starting with 1. In this example, S1 indicates that this sample is the first sample listed for the run.

> **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.

▶ **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. When generated, the Index Reads are I1 or I2.

▶ **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).

Supplementary Output Files

The following output files provide supplementary information, or summarize run results and analysis errors. Although these files are not required for assessing analysis results, they can be used for troubleshooting purposes. All files are located in the Alignment folder unless otherwise specified.

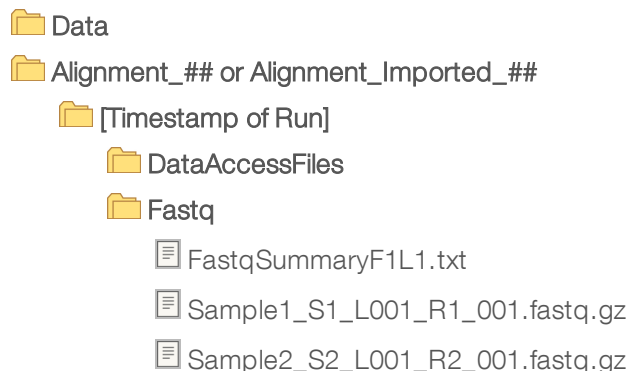| File Name	Description
AdapterCounts.txt	Contains a summary of the number of reads that had adapter trimming performed per sample.
AdapterTrimming.txt	Lists the number of trimmed bases and percentage of bases for each tile. This file is present only if adapter trimming was specified for the run.
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file will be empty if no errors occurred.
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.
GenerateFASTQRunStatistics.xml	Contains summary statistics specific to the run.

Analysis Folder

The analysis folder holds the files generated by the Local Run Manager software.

The relationship between the output folder and analysis folder is summarized as follows:

▶ During sequencing, Real-Time Analysis (RTA) populates the output folder with files generated during image analysis, base calling, and quality scoring.

▶ RTA copies files to the analysis folder in real time. After RTA assigns a quality score to each base for each cycle, the software writes the file RTAComplete.txt to both folders.

▶ When the file RTAComplete.txt is present, analysis begins.

▶ As analysis continues, Local Run Manager writes output files to the analysis folder, and then copies the files back to the output folder.

Folder Structure

📁 Data

📁 Alignment_## or Alignment_Imported_##

   📁 [Timestamp of Run]

      📁 DataAccessFiles

      📁 Fastq

         📄 FastqSummaryF1L1.txt

         📄 Sample1_S1_L001_R1_001.fastq.gz

         📄 Sample2_S2_L001_R2_001.fastq.gz

📄 Undetermined_S0_L001_R1_001.fastq.gz

📄 Undetermined_S0_L001_R2_001.fastq.gz

📁 Logging

📄 BuildFastq0.stdout.txt

📄 BuildFastq1.stdout.txt

📄 commands.txt

📁 Plots

📄 AdapterCounts.txt

📄 AdapterTrimming.txt

📄 AnalysisError.txt

📄 AnalysisLog.txt

📄 Checkpoint.txt

📄 CompletedJobInfo.xml

📄 DemultiplexSummaryF1L1.txt

📄 GenerateFASTQRunStatistics.xml

📄 SampleSheetUsed.csv

Custom Analysis Settings

Custom analysis settings are intended for technically advanced users. If settings are applied incorrectly, serious problems can occur.

Add a Custom Analysis Setting

1 From the Module-Specific Settings section of the Create Run screen, select **Show advanced module settings**.

2 Select **+ Add custom setting**.

3 In the custom setting field, enter the setting name as listed in the Available Analysis Settings section.

4 In the setting value field, enter the setting value.

5 To remove a setting, select ✖.

Available Analysis Settings

▶ **Adapter Trimming**—By default, adapter trimming is enabled in the Generate FASTQ analysis module. To specify a different adapter, use the Adapter setting. The same adapter sequence is trimmed for Read 1 and Read 2.

    ▶ To specify 2 adapter sequences, separate the sequences with a plus (+) sign.

    ▶ To specify a different adapter sequence for Read 2, use the AdapterRead2 setting.

Setting Name	Setting Value
Adapter	Enter the sequence of the adapter to be trimmed.
AdapterRead2	Enter the sequence of the adapter to be trimmed.

Revision History

Document	Date	Description of Change
Document # 100000003344 v02	April 2018	Added i5 index sequence instruction when creating custom prep kit sample sheets for use on the iSeq, MiniSeq, and NextSeq systems.
Document # 100000003344 v01	January 2018	Updated software descriptions for Generate FASTQ v2.0, which includes updates to: • Sample sheet import • Output folder structure • Indexing table • Removed Most Popular Index Sequences table • Supplementary output files
Document # 100000003344 v00	January 2016	Initial release.

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

Technical Assistance

For technical assistance, contact Illumina Technical Support.

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

Illumina Customer Support Telephone Numbers

Region	Toll Free	Regional
North America	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgium	+32 80077160	+32 34002973
China	400.066.5835	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.

Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

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

illumına®