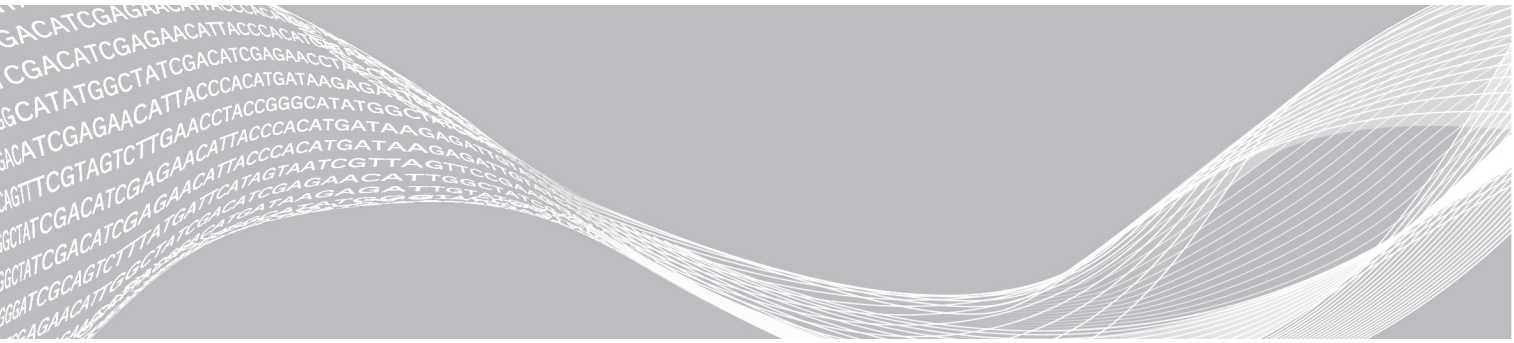
illumina®

Local Run Manager Assembly Workflow Module

Workflow Guide



Document # 1000000057983 v00

ILLUMINA PROPRIETARY

June 2018

Overview

The Assembly workflow assembles small genomes (< 20 Mb) either with or without the use of a reference genome. It is best suited for the assembly of bacterial genomes, such as *E. coli*.

The Assembly workflow uses the Velvet software and writes assembly results in the FASTA file format.

Compatible Library Types

The Assembly 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 Assembly analysis module only requires the sequencing data files generated during a sequencing run.

About This Guide

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

Set Parameters

1 If needed, log in to Local Run Manager.

2 Select **Create Run**, and select **Assembly**.

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

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

Specify Run Settings

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

2 Specify the number of index reads.
- **0** for a run with no indexing
- **1** for a single-indexed run
- **2** for a dual-indexed run

3 Select the read type for the run, if a change is possible.

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 Enter a K-Mer Size.
This setting overrides the k-mer size used by Velvet. The default size is 31 and the minimum value is 3. Odd-numbered values up to 255 are supported.

2 Select the **On/Off** toggle to enable or disable the following setting.
- **Reverse Complement**—(Only available with Nextera Mate Pair library prep kits) Off by default. When enabled, all reads are reverse-complemented as they are written to FASTQ files.

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 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 ✖.

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.

3 Enter a sample name in the Sample Name field.
Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.

4 **[Optional]** Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, underscores, or spaces.

5 If you have a plated kit, select an index plate well from the Index well drop-down list and skip to step 8.

6 If applicable, 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.

7 If applicable, 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.

8 If a reference is required, select a reference genome from the Genome Folder drop-down list.

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

10 **[Optional]** Select **Export Sample Sheet** to export the sample information in *.csv format.
The exported sample sheet can be used as a template or imported when creating new runs.

11 Select **Save Run**.

Import Sample Sheet

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

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

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

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

Sample Sheet Fields

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

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

Analysis Methods

The Assembly workflow uses a *de Bruijn* graph methodology to assemble reads into contigs, which are consensus DNA sequences representing overlapping sets of reads. The resulting contigs are written to a FASTA file named contigs.fa in a subfolder of the Alignment folder named AssemblyN, where N is the sample number.

Reads are randomly subsampled from the total data output to produce SN_L00#_Rx.fastq.gz files, where SN refers to the sample number , # refers to the lane number, and x refers to the read number. These files contain the reads used in the assembly process. The selection process is random but not stochastic, meaning the same subset of reads are selected each time that the Assembly workflow is run. The subsampling of reads prevents overloading of the RAM built into MiSeq instrument computer.

If a reference genome is specified, the workflow performs the following steps:

- Compares contigs against the reference genome.
- Reorders contigs to match the order of the reference genome, as closely as possible.
- Generates the samples graph (dot-plot), which summarizes the match between contigs and the reference genome. For more information, see *Samples Graph* on page 1.

NOTE

A reference genome is optional. Reference genomes can use either a *.fasta or *.fa file extension.

The assembly process uses the Velvet software. For a description of Velvet, see *Velvet: algorithms for de novo short read assembly using de Bruijn graphs*, Zerbino and Birney, Genome Research 2008.

View Analysis Results

1. From the Local Run Manager dashboard, select the run name.
2. From the Run Overview tab, review the sequencing run metrics.
3. To change the analysis data file location for future requeues of the selected run, select the **Edit** icon, and edit the output run folder file path.
   The file path leading up to the output run folder is editable. The output run folder name cannot be changed.
4. **[Optional]** Select the **Copy to Clipboard** icon to copy the output run folder file path.
5. Select the Sequencing Information tab to review run parameters and consumables information.
6. Select the Samples & Results tab to view the analysis report.
   - If analysis was requeued, select the appropriate analysis from the Select Analysis drop-down list.
7. **[Optional]** Select the **Copy to Clipboard** icon to copy the Analysis Folder file path.

Analysis Report

Analysis results are summarized on the Samples & Results tab. The report is also available in a PDF file format for each sample in the Analysis folder.

Sample Information

Column	Description
Sample Name	The sample name from the sample sheet.
Number of Contigs	The total number of contigs identified in the sample.
N50	The length for which the collection of all contigs of that length, or longer, accumulates to half of the total bases in the sample.
Minimum Contig Length	The number of bases in the shortest contig in the sample.
Median Contig Length	The median number of bases averaged over all contigs in the sample.
Mean Contig Length	The mean number of bases averaged over all contigs in the sample.

Column	Description
Maximum Contig Length	The number of bases in the longest contig in the sample.
Base Count	The total number of bases in the sample.
Contig Plot	Available when a reference genome is specified. Summarizes the match between contigs and the reference genome.

Analysis Output Files

The following analysis output files are generated for the Assembly analysis module and provide analysis results. Analysis output files are located in the Alignment folder.

File Name	Description
Contigs.fa	Contains the contigs for each assembly.
DotPlot.png	File for the Contig Plot. Summarizes the match between contigs and the reference genome.

Supplementary Output Files

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

File Name	Description
AdapterTrimming.txt	Lists the number of trimmed bases and percentage of bases for each tile. This file is present only if adapter trimming was specified for the run.
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages. Located in the root level of the run folder.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file will be empty if no errors occurred. Located in the root level of the run folder.
AssemblyRunStatistics.xml	Contains summary statistics specific to the run. Located in the root level 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.
Summary.xml	Contains summary statistics specific to the run. Contains a summary of mismatch rates and other base calling results.
Summary.htm	Contains a summary web page generated from Summary.xml.

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**.

Document # 1000000057983 v00

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.

illumina®