Illumina Data Analysis Guidelines

Workflow Guide

This guide provides instructions to set up run parameters for sequencing, and specifically details specifications of how to use the DNA Amplicon Module on Local Run Manager (LRM) for Paragon products. This guideline is written for the Illumina MiSeq, but concepts also translate for other Illumina sequencers with a LRM such as iSeq and NextSeq.

Get the latest user guide at: www.paragongenomics.com/product_documents/

UG7001-02
February 2024

Revision History

Document	Date	Description of Change
UG7001-01	October 2023	• Initial version
UG7001-02	February 2024	• Updated links to Illumina website in On-Instrument DNA Amplicon Analysis Workflow

Table of Content

Illumina On-Instrument DNA Amplicon Analysis Workflow Setup Recommendations

CleanPlex amplicon libraries sequenced on an Illumina platform can be analyzed with Illumina's Local Run Manager (LRM) DNA Amplicon Analysis Module. The LRM DNA Amplicon Analysis module v2.1.0 is available on iSeq (control software v1 or v2), MiSeq (MCS v3), NextSeq 500/550 (NCS 4.0), and NextSeq 500Dx (NCS 4.0). The LRM DNA Amplicon Analysis module v3.0.0.14 is available on MiSeq (MCS v4.0).

These instructions are a simple summary depicting how to set up analysis for an already completed run, or to set up a new run to include analysis. Samples must be grouped by genome for analysis: if there are different genomes in one sequencing run, the new run can only be set up to run analysis with one genome. However, these samples can be sequenced in the same run and later analyzed in batches by genome. If all samples in a sequencing run are from the same genome, users can utilize either method of analysis. This module aligns amplicon reads against the reference specified in the manifest file, and variants are called for the targeted regions. The workflow also yields a summary report of run quality and coverage information.

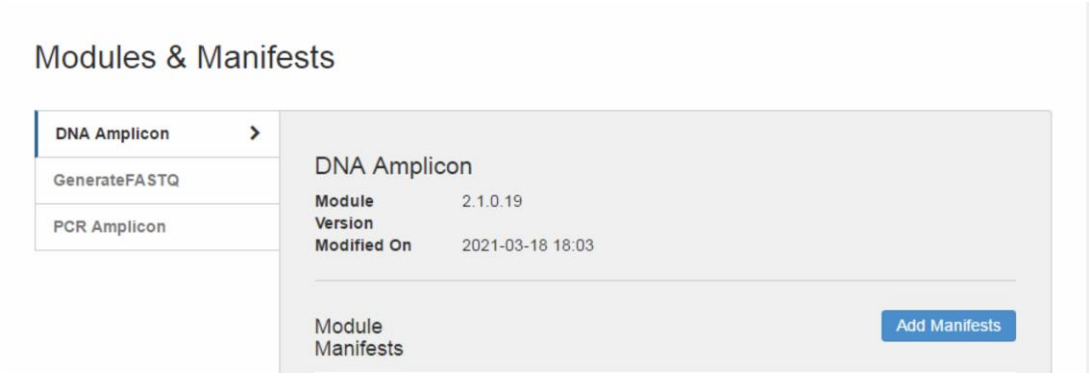
Helpful Tips and Resources:

- For additional details and troubleshooting, refer to Illumina's latest Local Run Manager DNA Amplicon Analysis Module Workflow Guide here: https://support.illumina.com/downloads/local-run-manager-dna-amplicon-analysis-module-workflow-guide.html
- If the program is not already installed on-instrument, please download the DNA Amplicon Analysis Module on your instrument from Illumina's support center here: https://support.illumina.com/downloads/local-run-manager-dna-amplicon-module.html
- To obtain the i5 and i7 Illumina index sequences by sequencer, please navigate to here: https://www.paragongenomics.com/customer-support/product_documents/ > Useful Tools > Illumina Index Sequence and Example Sample Sheets

Step 1: Import Critical Files: Reference, Manifest, and Genotype of Interests VCFs

The manifest file is available for download in the {panel_name}.DesignDetails.zip folder, under the design portal account. However, as the contents of a BED and manifest file are similar, a custom BED file can also be converted to a manifest file easily.

1. Open the Local Run Manager on the Illumina instrument.
2. Copy the manifest file {panel_name}.ampInsert.bed.manifest.txt, to the desired directory location.
3. Add the appropriate reference genome as a Fasta file (genome.fa) in the directory of the sequencer (C:\Illumina\Genomes).
    - The path to the genome location in the manifest file will need to be changed manually.
4. In the Local Run Manager, from the MENU, go to TOOLS tab > Modules & Manifests.
5. Click Add Manifests, navigate to the manifest file location, select the file, and click on Open.



6. Optionally if desired, add genotype of interest VCFs.
    1. The proper format for a VCF is a text file format. Take the known REF/ALT calls at a specific chr:position and configure the information to a compatible format.
    2. Please reference https://samtools.github.io/hts-specs/VCFv4.2.pdf for specifics.

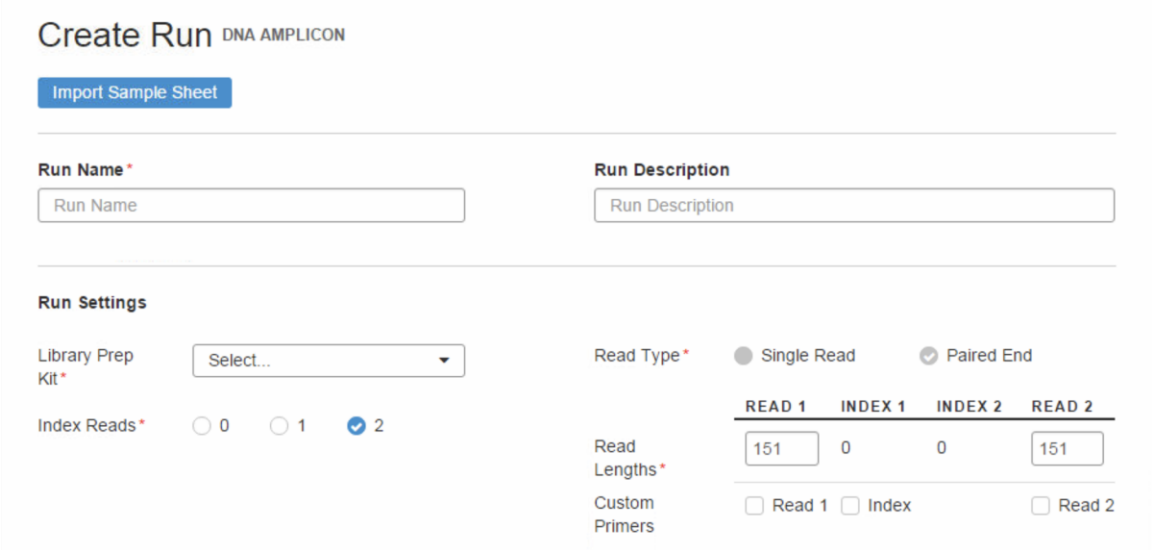
Step 2: Create a Run

The following instructions can be used to either configure the analysis as part of the sequencing run, or can be used after run completion, to reprocess and reconfigure the sequencing data. Please identify the proper use case and proceed accordingly.

2A. Configure analysis as part of the sequencing run using a Sample Sheet imported into the LRM.

1. Navigate to MENU > RUN DASHBOARD.
2. Select Create Run and click on the DNA Amplicon tab.
   Note: Paragon's libraries do not use custom sequence primers, but rather Illumina's standard indexing primers. Please refer to Illumina's guidelines on how to fill out the single index and dual index sample sheet at https://support.illumina.com/content/dam/illumina-

support/documents/documentation/system_documentation/miseq/miseq-sample-sheet-quick-ref-guide-15028392-j.pdf



Note: At this point you may either import a prefilled Sample Sheet OR manually enter the information into the LRM.

3. If importing a Sample Sheet, Select Import Sample Sheet and navigate to the SampleSheet.csv.

   Note: The steps below (3a - 3d) are all instructions on how to set up a Sample Sheet. Please utilize as necessary.
   Use the downloadable sample sheet template (as appropriate to your module version) located here: https://support.illumina.com/downloads/local-run-manager-dna-amplicon-module-v3.html.
   a. Make the following changes in the [Header] section

[Header]	
Experiment Name	**Change to desired run ID**
Date	**Change to date of run (mm/dd/yyyy)**
Module	DNA Amplicon - 3.0.0
Workflow	DNA Amplicon
Library Prep Kit	**Custom**
Index Kit	

   b. The [Reads] section will be based on the number of cycles to be performed in the run and is set at a default of 151. Configure and specify your inputas appropriate. For example, for 150 Paired ends, input 150 to both rows.
   c. Configure the setup [Settings] as appropriate:
      1. **Transcriptsource:** RefSeq supported for human genomes.
      2. **aligner**: Recommended BWA.
      3. **Variantcaller**: Select Somatic or Germline based on analysis requirements (refer to the appropriate Paragon User Guide for guidance based on the specific kit used).

4. **Minimumcoveragedepth**: Refer to UG.
5. **Variantfrequencyemitcutoff:** (Only available for somatic calling). Refer to UG.
6. **Variantannotation**: None.
    - Recommended default is none but Refseq or Ensembl are viable options based on your requirements.
7. **Varientcallerrealignindel**s: 1 (on).

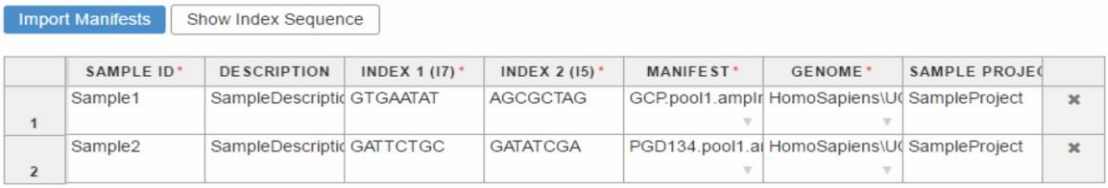
d. Make following changes to the [Manifests] section. Add a manifest file for each panel used in the run as manifest0, manifest1, manifest2, and so on.

[Manifests]	
Manifest0	Manifest1.txt
Manifest1	Manifest2.txt

- Add samples to the [Data] section and assign manifest file(s) to each sample and add location of the reference genome fasta file. For example, Homo_sapiens\UCSC\hg19\Sequence\WholeGenomeFasta.
- **Note**: The target genomes are located at C:\Illumina\Genomes.
- **Note**: For a reference genome not available on Illumina's default folder, please check the NCBI database or a database specific to your target organism.

After clicking open, the Create Run window should be automatically populated.

4. Input the desired Run Name.
    a. **Note**: Avoid using spaces in the run names to avoid errors.
5. Configure the Run settings as appropriate (if not already configured in the Sample Sheet):
    a. Change the Variant Caller under Module-Specific Settings tab to Somatic or Germline based on analysis requirements (refer to the appropriate Paragon User Guide for guidance based on the specific kit used).
    b. Select Variant Annotation: None.
        i. Recommended default is none but Refseq or Ensembl are viable options based on your requirements.
    c. Leave other tab settings as default.
    d. Recommended aligner is BWA.
    e. The genotype of interest VCF can be added if necessary, by selecting Add a Genotype VCF File.
    f. The setting for variant calling can be changed by selecting "Show advanced module settings…" and modifying default settings as needed.
    g. Indel realignment: On.
        i. Leave other tab settings as default.
6. If a sample sheet was imported, scroll down to the Sample Sheet section, navigate to the last column, and remove any samples that do not need data analysis by clicking the 'X' on the right.

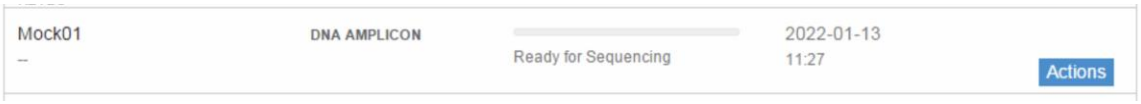


7. The manifest file is automatically populated for each sample as noted in the imported Sample Sheet. Refer to step 3 above as necessary.

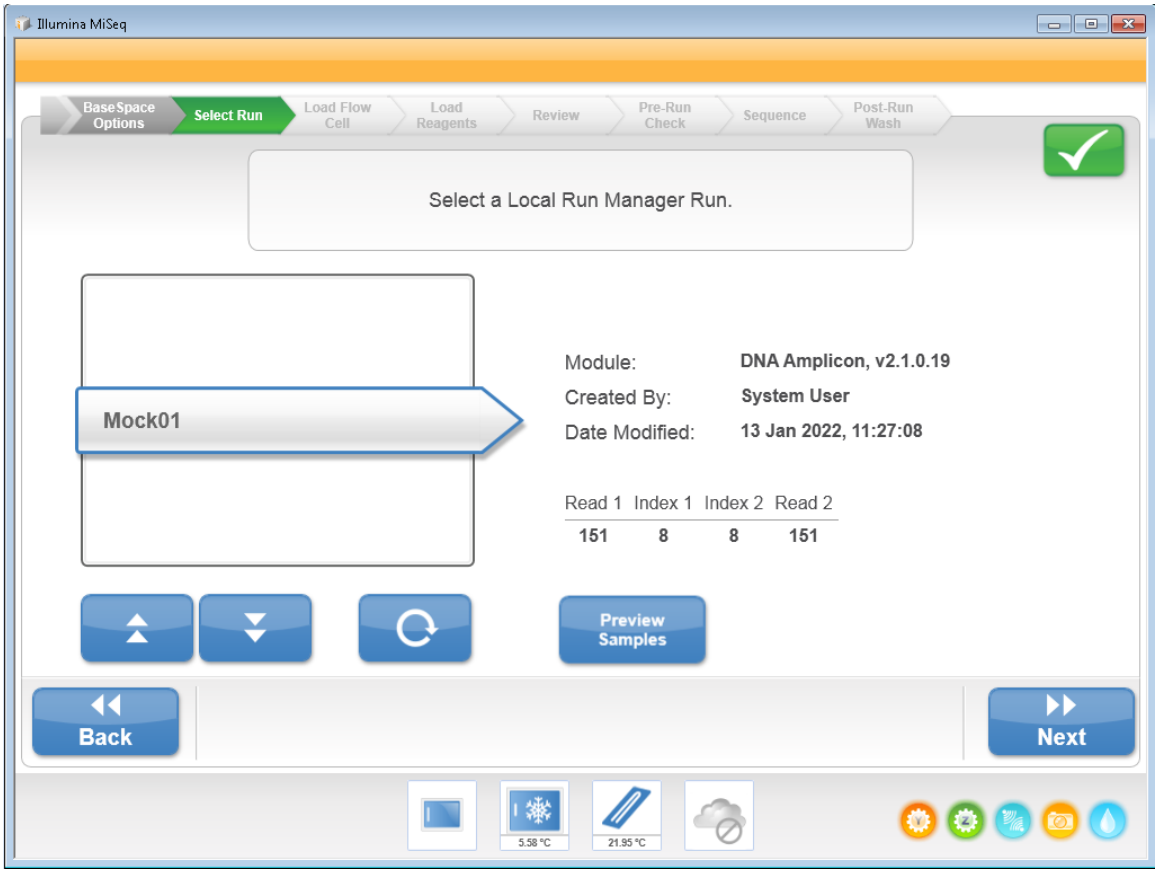
8. The reference genome will be populated automatically as specified in the Sample Sheet. Note that only one unique genome can be analyzed per analysis. If multiple genomes are to be sequenced in one run, multiple analyses will need to be performed successively).

	SAMPLE ID *	DESCRIPTION	INDEX 1 (I7) *	INDEX 2 (I5) *	MANIFEST *	GENOME *	SAMPLE PROJE(	
1	Sample1	SampleDescriptic	GTGAATAT	AGCGCTAG	GCP.pool1.amplr	HomoSapiens\U(	SampleProject	✖
2	Sample2	SampleDescriptic	GATTCTGC	GATATCGA	PGD134.pool1.a	HomoSapiens\U(	SampleProject	✖

9. After the mandatory fields (marked with asterisks *) are populated, save the run by selecting > Save Run. The saved run appears on the RUN DASHBOARD with its STATUS as Ready for Sequencing.



10. Once the run is populated on LRM, go to the MiSeq or appropriate Illumina platform's Control Software, click on Local Run Manager, and select Next.
11. The scheduled run appears on the LRM, verify the run information, and click on Next to start the run.



2B. Configure analysis after an already completed run.

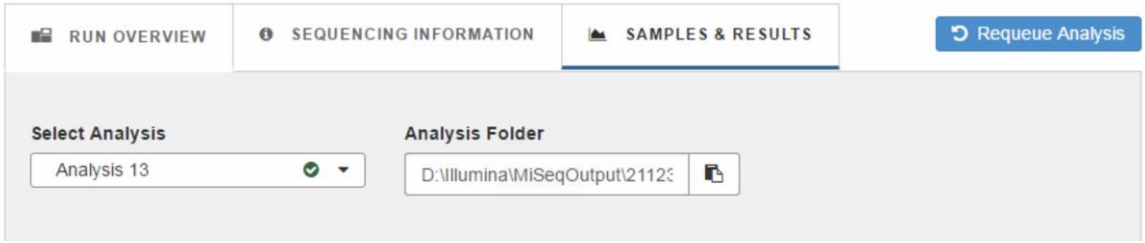
For a run that has already been completed and just needs to be analyzed, please locate the Sample Sheet that was automatically generated from the sequencing run.

Note: Different genomes from the same sequencing run cannot be analyzed together but will need to be processed separately.

9. After the mandatory fields (marked with asterisks *) are populated, save the run by selecting > Save Run. The saved run appears on the RUN DASHBOARD with its STATUS as “Ready for Sequencing”.
    a. Even if these samples have already been sequenced and just need to be reprocessed and analyzed, the status will still display “Ready for Sequencing”.
10. Select Actions > Import > Select the run folder by copying the path for the run directory located in (D:\Illumina\MiSeqAnalysis).
11. Create a run folder with the name of your choice (for example, D:\Illumina\MiSeqOutput\MiSeq001DNAAmpliconAnalysis) and copy its path as the Base Output Folder tab. Once the input and output directories are set, click on Import Run.
12. Once the analysis is successful, the STATUS of the RUN NAME will change to Analysis Completed.
    a. Note: If the analysis fails, a message will be displayed as ‘Analysis Failed’

Step 3: Review the QC Metrics Summary

1. Results are stored in the output folder present in the directory D:\Illumina\MiSeqOutput for MiSeq and D:\Analysis for iSeq.
    a. A specific output directory can be created if desired and RUN ID can be different from the output folder name.
    b. Note: for other Illumina instruments, please change the path as appropriate to the relevant instrument name.
2. To see the results, navigate to the RUN DASHBOARD, click on the RUN NAME and click on SAMPLES & RESULTS tab.



3. An Aggregate Report (available in PDF file format) which displays all sequencing metrics is generated in the Analysis folder.
4. Individual sample results can be viewed by navigating to the Sample name tab, displayed below the Aggregate Report tab.

Illumina Off-Instrument DNA Amplicon Workflow Auxiliary File Generation Instructions

Overview

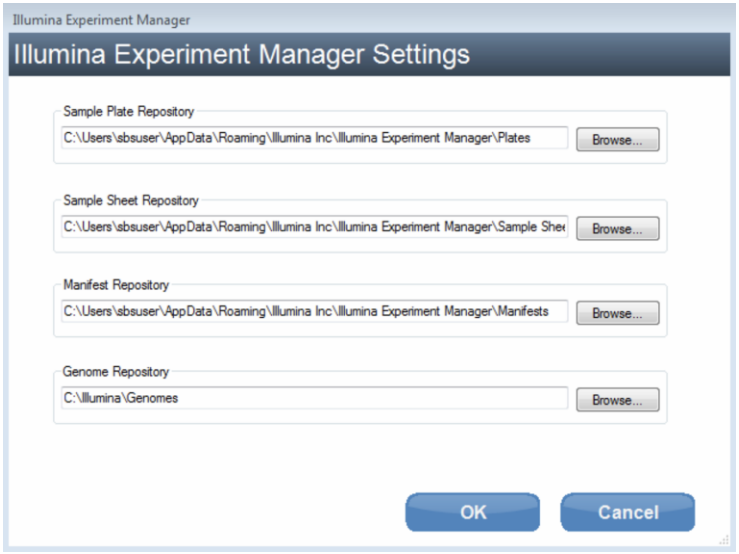
The following section contains recommendations to generate auxiliary files on the Illumina Experiment Manager (IEM) necessary for running DNA Amplicon Module. .The following guidelines are written using the MiSeq platform, and are to be adjusted accordingly for other instruments.

Step 1: Set a Reference Genome File

1. Add the genome.fa file for your genome of interest to the MiSeq instrument folder C:\Illumina\Genomes using your Institution's recommended directory structure path.
    a. **Note:** We recommend the following directory/sub-directory structure: (C:\Illumina\Genomes\{species_name}\InstitutionName\{reference_file_name}\Sequence\WholeGenomeFasta).

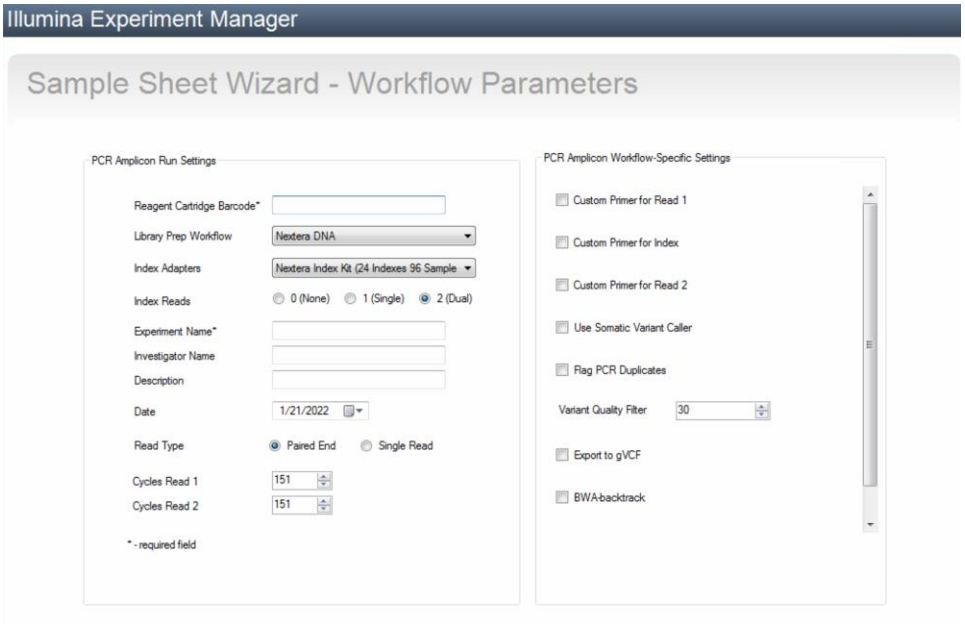
Step 2: Generate a GenomeSize.xml file

1. Open Illumina Experiment Manager on the MiSeq Desktop and select Settings (see image below).
2. Select the Genome Repository with your custom genome.fa file of interest located at C:\Illumina\Genomes\. Once set, click Okay.

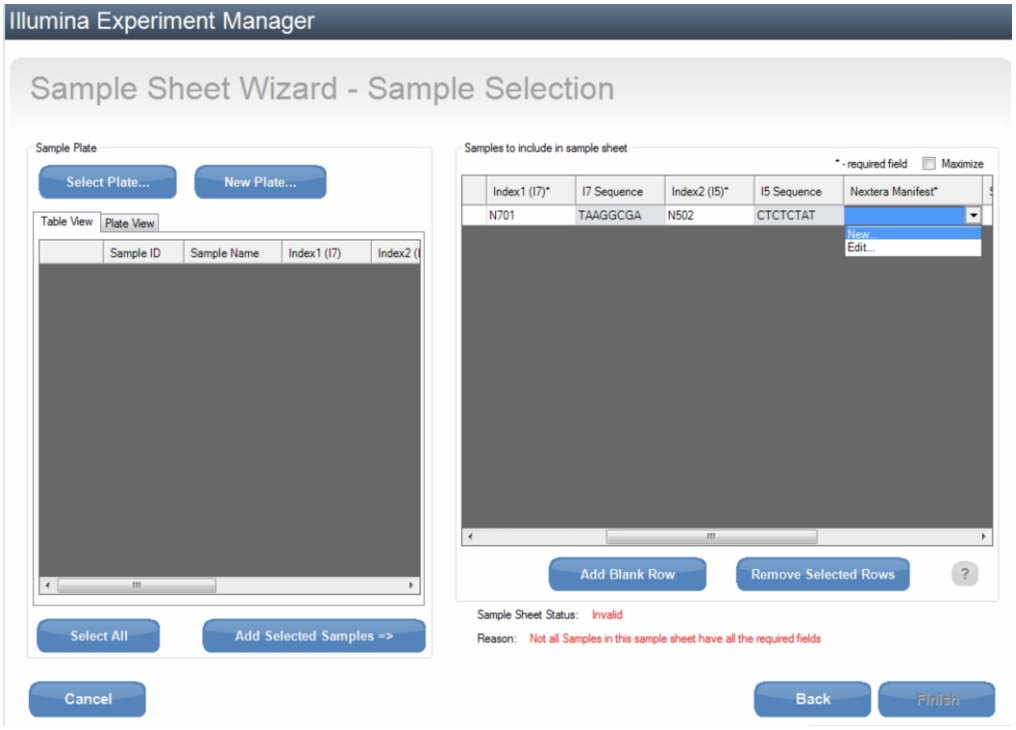


3. Select Create Sample Sheet, and then MiSeq, Category: Targeted Resequencing, Application: PCR Amplicon (see image below).

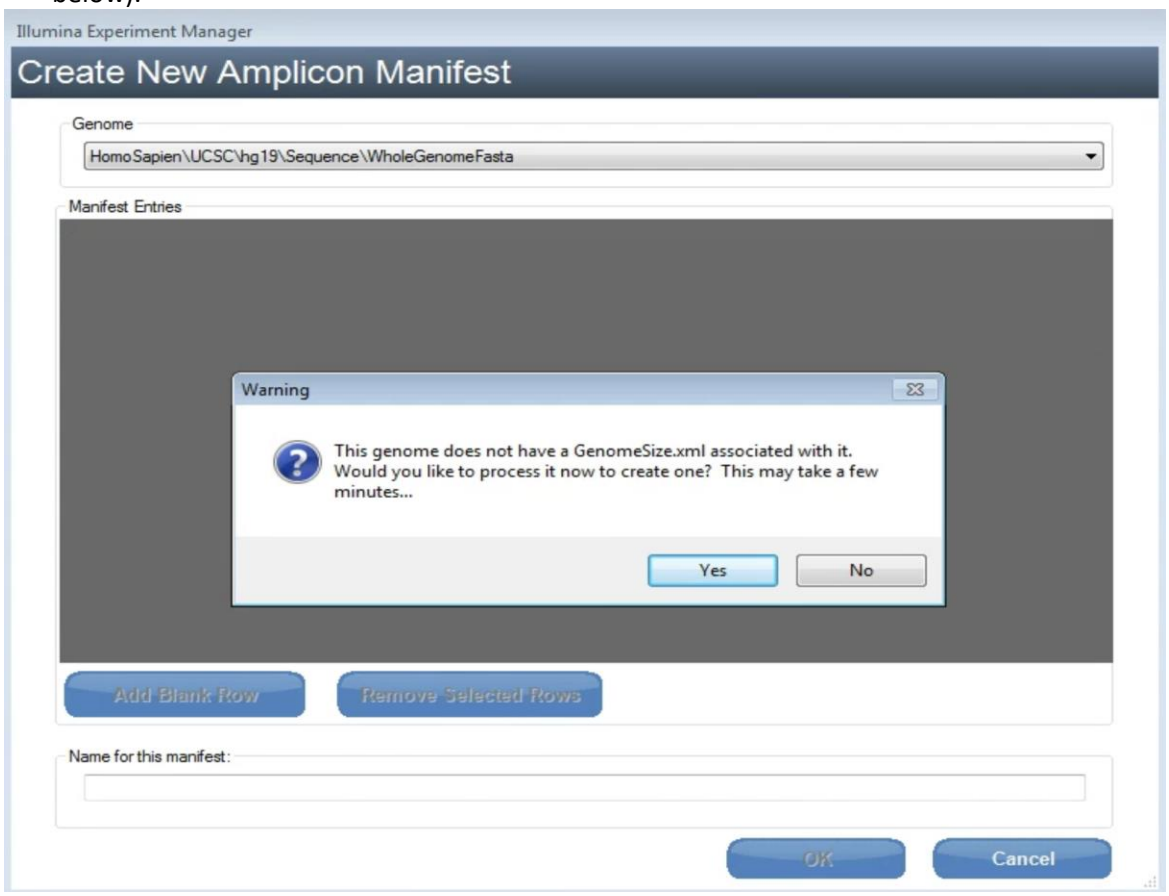
4. Fill out the required fields (Reagent Cartridge Barcode* and Experiment Name*), and others as necessary.



5. On the next page, select Add Blank Row, and click under the Nextera Manifest* column to display a dropdown arrow from which to select New… (see image below).



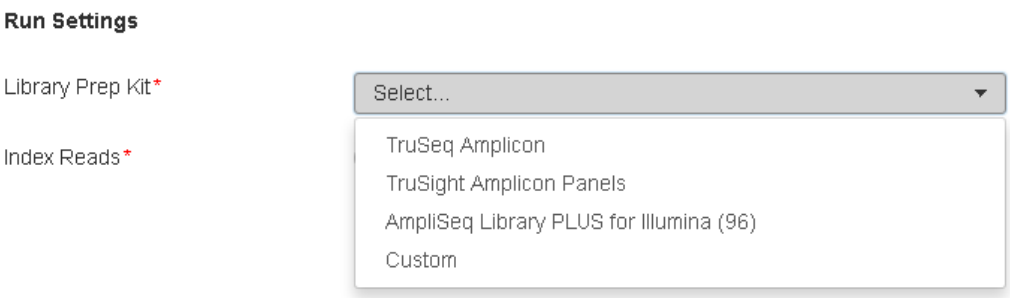
6. Select the genome from the dropdown arrow and auxiliary files are generated (see image below).



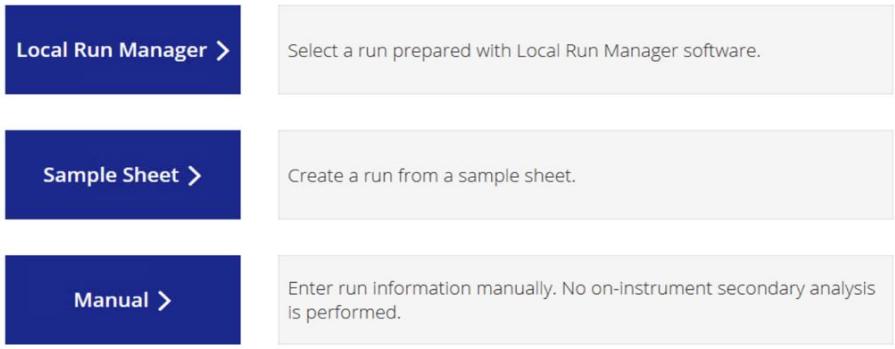
Frequently Asked Questions (FAQs)

1. How do I enter my Paragon library as a “custom” library?

In the DNA Amplicon module Run settings, there is a dropdown option where you can select custom library.



2. How do I generate a sample sheet?



Local Run Manager: A sample sheet can be generated for using the DNA amplicon module on LRM using a template provided here: https://support.illumina.com/downloads/local-run-manager-dna-amplicon-module-v3.html

Manually generate a CSV file with relevant headers and information: https://support.illumina.com/downloads/sample-sheet-v2-template.html

Use IEM to fill out the sample information and export a sample sheet that then gets imported in the beginning of a run: https://www.youtube.com/watch?v=bfPpbJtj3LE

3. What indexes should I use on the sample sheet?

Please refer to: https://www.paragongenomics.com/customer-support/product_documents/ for exact sequences. For different sequencers, the i5 sequence may be different. For further questions, please refer to Illumina’s documentation on indexed sequencing technology located here: https://support.illumina.com/downloads/indexed-sequencing-overview-15057455.html

4. How can we select from Illumina's dropdown menu for the index sequence?

Illumina's Truseq index sequence overlaps with Paragon's Set A index primers and can be selected from the dropdown menu as a replacement. However, the rest of the index sets (B, C, and D) will need to be input manually. Please reference Paragon's index spreadsheet available at: https://www.paragongenomics.com/customer-support/product_documents/ and copy and paste the sequences into the appropriate samples.

5. How do I upload the sample sheet? Can I use either LRM or IEM?

LRM and IEM are used to generate sample sheets, which can then be uploaded into the MiSeq Control Software to start a sequencing run. The LRM is also used to create a sequencing run that can then be initiated from the MiSeq Control Software. So either the LRM can be used to input sample information, which will act as a sample sheet, OR the IEM can be used to generate a sample sheet that can be imported to LRM when creating a run.