GeneLab Standard Operating Procedure:

Generate an iSeq QC report using HTStream

October 2020

Version 1.0

Document Revisions

Document Number	Revision Number	Date	Description of Changes
8.1	1	Oct. 2020	Original

Scope and Purpose

Introduction

- HTStream is a high throughput genomic data pre-processing tool, developed by Matt Settles from the UC Davis Bioinformatics Core. https://github.com/s4hts/HTStream.
- There are three HTStream components used to generate the iSeq quality control report: hts_SeqScreener, hts_Stats and hts_SuperDeduper.
- The iSeq quality control metrics for RNAseq studies are used to estimate the expected number of non-ribosomal reads per sample that will be generated in a subsequent NovaSeq run with the same library pool. Additionally, the iSeq QC metrics are also used to identify select bacterial ribosomal RNA (rRNA) contaminates and to estimate the % duplicates (not considering UMIs).
- rRNA reference fasta files retrieved from NCBI are used for human, mouse, and bacteria % rRNA calculations. These files are provided in the rRNA_fasta_files folder, which is inside the HTStream_scripts.zip folder. Note: rRNA reference fasta files from other organisms can be retrieved from NCBI and used to screen for rRNA contamination.
- Refer to SOP 7.1_sequencer_setup for Illumina iSeq and NovaSeq set-ups. Fastq files are generated from the iSeq run using the internal software, GenerateFASTQ, to convert bcl files to fastq files. However, the iSeq-generated fastq files may be slightly different than those generated by running bcl2fastq using the iSeq-generated bcl files as input. Since the purpose of the iSeq QC analysis is to predict the NovaSeq output, and only bcl files are generated from the NovaSeq run, we begin this protocol with the iSeq-generated bcl files.
- The following protocol is run using a command line interface (CLI) with the bash interpreter.
- All shell scripts indicated in the following instructions are provided in the HTStream_scripts.zip folder.
- The instructions below assume that the following software is installed on the machine you're using for the analysis and can be accessed through your $PATH
    - bcl2fastq (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)
    - HTStream (https://github.com/s4hts/HTStream)
    - bc (https://www.gnu.org/software/bc/)
    - For more information about what a $PATH is click here: https://astrobiomike.github.io/bash/modifying_your_path

The final report is used to:

- compare the sample reads to determine if the pool is balanced.
- analyze the % rRNA of organism of interest and determine if the sample's % rRNA meets the % cutoff (normally <10-20%).
- predict sample read depth on NovaSeq run from this iSeq read depth.

Procedure

Instructions to generate an iSEQ QC report for paired-end reads with HTStream

*Note: Instructions below are for paired-end reads generated from libraries with unique dual indices, in which Unique Molecular Identifiers (UMIs) are appended to the 3' end of the i7 index. Steps that apply to libraries without UMIs and single-end studies are indicated.

*** Steps 1 – 5 are for setting up your directory structure and copying your iSeq output files onto the machine you'll use to run the analysis. ***

1. Log in to the computer (or cluster, or server, or laptop, etc.) that you will use to run your analysis.

2. Create a new folder (aka directory) where you will store your input and output data.
   Note: The title of the folder cannot have any spaces or weird characters and it is good practice to include the study name, instrument used to generate the data, the date of the sequencing run, and your initials, see example below.
   Example directory name: RR6_iSeq_20190617_JD
   Study Name: RR6
   Instrument used to generate the data: iSeq
   Date of sequencing run: 20190617 (June 17th, 2019)
   Initials: JD (Jane Doe)

3. Copy the entire directory that's outputted from the iSeq run into the folder you created in step 2 then check that the copy completed successfully by checking directory/file sizes and the number of files.
   Note: The title of the iSeq directory cannot have any spaces or weird characters; it is good practice to keep the name generated by the iSeq run. The name of the iSeq directory in this example is:
   20190617_FS10000323_13_BPC29610-1014

4. Create a sub-directory in the folder you created in step 2, where you will store your bcl2fastq output.
   Note: The title of this sub-directory cannot have any spaces or weird characters; at GeneLab, we title this subdirectory as follows: out-name_of_iSeq_output_directory
   For example: out-20190617_FS10000323_13_BPC29610-1014.

5. At this point, you should have two sub-directories in the folder you created in step2 as follows:

RR6_iSeq_20190617_JD (directory made in step 2)
20190617_FS10000323_13_BPC29610-1014 (the directory that was outputted from the iSeq run)
out-20190617_FS10000323_13_BPC29610-1014 (the directory you made in step4, which is where the bcl2fastq output will be directed)

*** Steps 6 – 13 are only for iSeq runs with libraries with unique dual indices, in which UMIs are appended to the 3' end of the i7 index ***

*** If your iSeq run was with libraries that contain UMIs in a different location than that indicated, consult Illumina for the best way to extract UMI info during demultiplexing ***

*** If your iSeq run does NOT contain samples with UMIs, jump to step 14 ***

6. Open up a terminal window and change directories using the 'cd' command so that you're in the directory you made in step 2
    a. If you're new to the command line interface (CLI), some additional instructions are provided below (note that the '$' indicates a command line prompt, and what follows the '$' in the steps below are the commands that should be written in the CLI):
        i. In the terminal window, type the following command to see where you are in your computer's file system:
        $ pwd

        You should see an output similar to:
        /Users/username

        ii. Next type the following command to see what's in your current location:
        $ ls

        You should see a list of directories and maybe some files, including a directory titled 'Desktop'.

        iii. If you created the folder in step 2 on your desktop, go into your desktop by typing the following command:
        $ cd Desktop

        iv. Check what's on your desktop by typing the following command:
        $ ls

        You should see a list of the directories and files that are on your desktop, including the directory you made in step2 (in our example we titled the directory: RR6_iSeq_20190617_JD)

        v. Move into the directory you made in step 2 by typing the following command:
        $ cd RR6_iSeq_20190617_JD

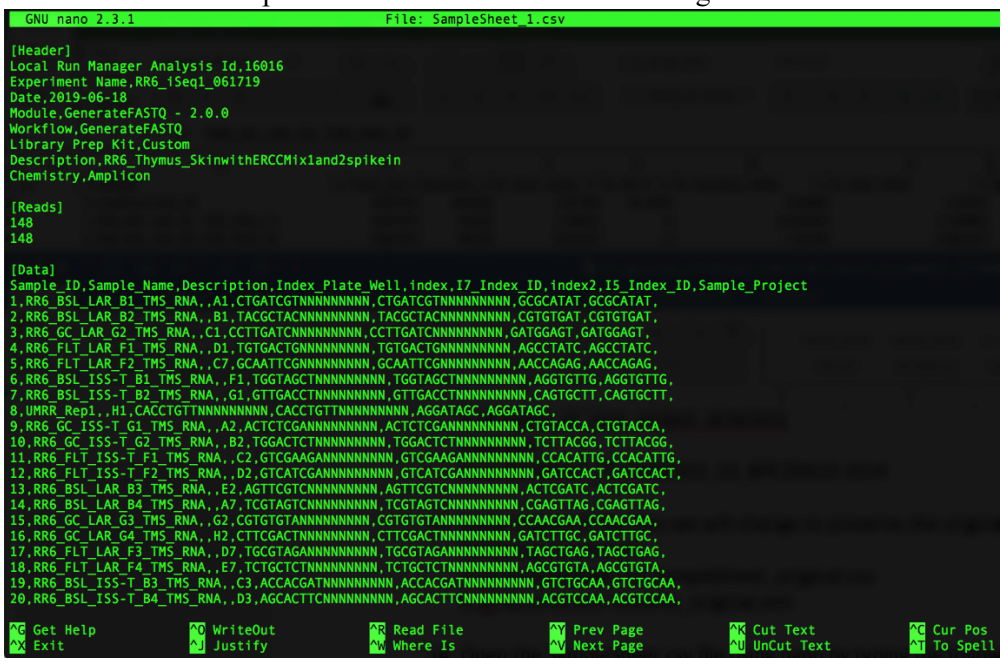
You should now be in the directory you made in step2, to check, type the following command to list the contents of the directory you're in:
$ ls

You should see the two subdirectories indicated in step5. You are now ready to move onto step7.

7. To extract UMI info during demultiplexing, we'll have to change some info in the files in the iSeq output directory, so change into that directory by typing the following command:
   $ cd name_of_iSeq_output_directory
   Example:
   $ cd 20190617_FS10000323_13_BPC29610-1014

8. Make copies of the files we will change to preserve the original file by typing the following commands:
   $ cp SampleSheet.csv SampleSheet_original.csv
   $ cp RunInfo.xml RunInfo_original.xml

9. Open the SampleSheet.csv file with a text editor (such as nano) by typing the following command (note: in the following steps we use the text editor, nano, but steps 9-XX can be done with any text editor, including vi, notepad, etc.):
   $ nano SampleSheet.csv
   You'll see a window open that looks similar to the following:



10. UMIs are expressed as Ns in the index sequence i7, example here.
    Remove the string of N's by doing the following:
    Press 'control' and '\ ' at the same time, you'll see the following:

```
  GNU nano 2.3.1                    File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGTNNNNNNNNN,CTGATCGTNNNNNNNNN,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTACNNNNNNNNN,TACGCTACNNNNNNNNN,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATCNNNNNNNNN,CCTTGATCNNNNNNNNN,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTGNNNNNNNNN,TGTGACTGNNNNNNNNN,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCGNNNNNNNNN,GCAATTCGNNNNNNNNN,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCTNNNNNNNNN,TGGTAGCTNNNNNNNNN,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCTNNNNNNNNN,GTTGACCTNNNNNNNNN,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTTNNNNNNNNN,CACCTGTTNNNNNNNNN,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGANNNNNNNNN,ACTCTCGANNNNNNNNN,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCTNNNNNNNNN,TGGACTCTNNNNNNNNN,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGANNNNNNNNN,GTCGAAGANNNNNNNNN,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGANNNNNNNNN,GTCATCGANNNNNNNNN,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTCNNNNNNNNN,AGTTCGTCNNNNNNNNN,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTCNNNNNNNNN,TCGTAGTCNNNNNNNNN,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTANNNNNNNNN,CGTGTGTANNNNNNNNN,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACTNNNNNNNNN,CTTCGACTNNNNNNNNN,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGANNNNNNNNN,TGCGTAGANNNNNNNNN,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCTNNNNNNNNN,TCTGCTCTNNNNNNNNN,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGATNNNNNNNNN,ACCACGATNNNNNNNNN,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTCNNNNNNNNN,AGCACTTCNNNNNNNNN,ACGTCCAA,ACGTCCAA,
Search (to replace):
^G Get Help      ^Y First Line    M-C Case Sens    M-R Regexp       ^N NextHstory
^C Cancel        ^V Last Line     M-B Backwards    ^P PrevHstory    ^R No Replace
```

Note that you'll get prompted for what you want to replace, type the following: NNNNNNNNN

It will look like this in your window:

```
  GNU nano 2.3.1                    File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGTNNNNNNNNN,CTGATCGTNNNNNNNNN,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTACNNNNNNNNN,TACGCTACNNNNNNNNN,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATCNNNNNNNNN,CCTTGATCNNNNNNNNN,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTGNNNNNNNNN,TGTGACTGNNNNNNNNN,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCGNNNNNNNNN,GCAATTCGNNNNNNNNN,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCTNNNNNNNNN,TGGTAGCTNNNNNNNNN,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCTNNNNNNNNN,GTTGACCTNNNNNNNNN,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTTNNNNNNNNN,CACCTGTTNNNNNNNNN,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGANNNNNNNNN,ACTCTCGANNNNNNNNN,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCTNNNNNNNNN,TGGACTCTNNNNNNNNN,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGANNNNNNNNN,GTCGAAGANNNNNNNNN,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGANNNNNNNNN,GTCATCGANNNNNNNNN,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTCNNNNNNNNN,AGTTCGTCNNNNNNNNN,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTCNNNNNNNNN,TCGTAGTCNNNNNNNNN,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTANNNNNNNNN,CGTGTGTANNNNNNNNN,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACTNNNNNNNNN,CTTCGACTNNNNNNNNN,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGANNNNNNNNN,TGCGTAGANNNNNNNNN,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCTNNNNNNNNN,TCTGCTCTNNNNNNNNN,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGATNNNNNNNNN,ACCACGATNNNNNNNNN,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTCNNNNNNNNN,AGCACTTCNNNNNNNNN,ACGTCCAA,ACGTCCAA,
Search (to replace): NNNNNNNNN
^G Get Help      ^Y First Line    M-C Case Sens    M-R Regexp       ^N NextHstory
^C Cancel        ^V Last Line     M-B Backwards    ^P PrevHstory    ^R No Replace
```

Then press ‘Enter’. You will now be prompted for what to replace the string of N’s with. It will look as follows:

```
  GNU nano 2.3.1                          File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGTNNNNNNNNN,CTGATCGTNNNNNNNNN,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTACNNNNNNNNN,TACGCTACNNNNNNNNN,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATCNNNNNNNNN,CCTTGATCNNNNNNNNN,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTGNNNNNNNNN,TGTGACTGNNNNNNNNN,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCGNNNNNNNNN,GCAATTCGNNNNNNNNN,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCTNNNNNNNNN,TGGTAGCTNNNNNNNNN,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCTNNNNNNNNN,GTTGACCTNNNNNNNNN,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTTNNNNNNNNN,CACCTGTTNNNNNNNNN,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGANNNNNNNNN,ACTCTCGANNNNNNNNN,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCTNNNNNNNNN,TGGACTCTNNNNNNNNN,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGANNNNNNNNN,GTCGAAGANNNNNNNNN,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGANNNNNNNNN,GTCATCGANNNNNNNNN,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTCNNNNNNNNN,AGTTCGTCNNNNNNNNN,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTCNNNNNNNNN,TCGTAGTCNNNNNNNNN,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTANNNNNNNNN,CGTGTGTANNNNNNNNN,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACTNNNNNNNNN,CTTCGACTNNNNNNNNN,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGANNNNNNNNN,TGCGTAGANNNNNNNNN,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCTNNNNNNNNN,TCTGCTCTNNNNNNNNN,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGATNNNNNNNNN,ACCACGATNNNNNNNNN,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTCNNNNNNNNN,AGCACTTCNNNNNNNNN,ACGTCCAA,ACGTCCAA,
Replace with:
^G Get Help        ^Y First Line        ^P PrevHstory
^C Cancel          ^V Last Line         ^N NextHstory
```

Since we want to remove the N’s (and not replace it with anything), just press ‘Enter’
Next, you’ll be asked which string of N’s you want to replace (or, in our case, remove), as follows:

```
  GNU nano 2.3.1                    File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGTNNNNNNNNN,CTGATCGTNNNNNNNNN,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTACNNNNNNNNN,TACGCTACNNNNNNNNN,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATCNNNNNNNNN,CCTTGATCNNNNNNNNN,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTGNNNNNNNNN,TGTGACTGNNNNNNNNN,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCGNNNNNNNNN,GCAATTCGNNNNNNNNN,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCTNNNNNNNNN,TGGTAGCTNNNNNNNNN,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCTNNNNNNNNN,GTTGACCTNNNNNNNNN,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTTNNNNNNNNN,CACCTGTTNNNNNNNNN,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGANNNNNNNNN,ACTCTCGANNNNNNNNN,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCTNNNNNNNNN,TGGACTCTNNNNNNNNN,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGANNNNNNNNN,GTCGAAGANNNNNNNNN,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGANNNNNNNNN,GTCATCGANNNNNNNNN,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTCNNNNNNNNN,AGTTCGTCNNNNNNNNN,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTCNNNNNNNNN,TCGTAGTCNNNNNNNNN,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTANNNNNNNNN,CGTGTGTANNNNNNNNN,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACTNNNNNNNNN,CTTCGACTNNNNNNNNN,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGANNNNNNNNN,TGCGTAGANNNNNNNNN,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCTNNNNNNNNN,TCTGCTCTNNNNNNNNN,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGATNNNNNNNNN,ACCACGATNNNNNNNNN,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTCNNNNNNNNN,AGCACTTCNNNNNNNNN,ACGTCCAA,ACGTCCAA,
Replace this instance?
 Y Yes          A All
 N No          ^C Cancel
```

Press ‘a’ to remove all of the strings of N’s. Note that all of the N’s will be removed and it will show how many strings of N’s were replaced as follows:

```
  GNU nano 2.3.1                    File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGT,CTGATCGT,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTAC,TACGCTAC,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATC,CCTTGATC,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTG,TGTGACTG,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCG,GCAATTCG,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCT,TGGTAGCT,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCT,GTTGACCT,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTT,CACCTGTT,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGA,ACTCTCGA,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCT,TGGACTCT,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGA,GTCGAAGA,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGA,GTCATCGA,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTC,AGTTCGTC,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTC,TCGTAGTC,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTA,CGTGTGTA,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACT,CTTCGACT,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGA,TGCGTAGA,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCT,TCTGCTCT,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGAT,ACCACGAT,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTC,AGCACTTC,ACGTCCAA,ACGTCCAA,
                                        [ Replaced 240 occurrences ]
^G Get Help    ^O WriteOut    ^R Read File   ^Y Prev Page   ^K Cut Text    ^C Cur Pos
^X Exit        ^J Justify     ^W Where Is    ^V Next Page   ^U UnCut Text  ^T To Spell
```

In our example, where the sequence of the i7 index + UMI is listed in both the 'index' and the 'I7_Index_ID' columns, the number of occurrences replaced should be two times the number of samples; i.e. in the example above there are 120 samples so we see that 240 occurrences were replaced.

You can now save the changes you made by pressing 'control' and 'x' at the same time, which will then ask you if you want to save your changes as shown below:

```
  GNU nano 2.3.1                          File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGT,CTGATCGT,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTAC,TACGCTAC,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATC,CCTTGATC,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTG,TGTGACTG,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCG,GCAATTCG,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCT,TGGTAGCT,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCT,GTTGACCT,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTT,CACCTGTT,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGA,ACTCTCGA,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCT,TGGACTCT,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGA,GTCGAAGA,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGA,GTCATCGA,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTC,AGTTCGTC,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTC,TCGTAGTC,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTA,CGTGTGTA,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACT,CTTCGACT,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGA,TGCGTAGA,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCT,TCTGCTCT,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGAT,ACCACGAT,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTC,AGCACTTC,ACGTCCAA,ACGTCCAA,
Save modified buffer (ANSWERING "No" WILL DESTROY CHANGES) ?
 Y Yes
 N No           ^C Cancel
```

Press 'y' to save your changes and it will prompt you with the name of the file. Make sure it says SampleSheet.csv (as shown below) then press 'Enter'

```
  GNU nano 2.3.1                          File: SampleSheet_1.csv

[Header]
Local Run Manager Analysis Id,16016
Experiment Name,RR6_iSeq1_061719
Date,2019-06-18
Module,GenerateFASTQ - 2.0.0
Workflow,GenerateFASTQ
Library Prep Kit,Custom
Description,RR6_Thymus_SkinwithERCCMix1and2spikein
Chemistry,Amplicon

[Reads]
148
148

[Data]
Sample_ID,Sample_Name,Description,Index_Plate_Well,index,I7_Index_ID,index2,I5_Index_ID,Sample_Project
1,RR6_BSL_LAR_B1_TMS_RNA,,A1,CTGATCGT,CTGATCGT,GCGCATAT,GCGCATAT,
2,RR6_BSL_LAR_B2_TMS_RNA,,B1,TACGCTAC,TACGCTAC,CGTGTGAT,CGTGTGAT,
3,RR6_GC_LAR_G2_TMS_RNA,,C1,CCTTGATC,CCTTGATC,GATGGAGT,GATGGAGT,
4,RR6_FLT_LAR_F1_TMS_RNA,,D1,TGTGACTG,TGTGACTG,AGCCTATC,AGCCTATC,
5,RR6_FLT_LAR_F2_TMS_RNA,,C7,GCAATTCG,GCAATTCG,AACCAGAG,AACCAGAG,
6,RR6_BSL_ISS-T_B1_TMS_RNA,,F1,TGGTAGCT,TGGTAGCT,AGGTGTTG,AGGTGTTG,
7,RR6_BSL_ISS-T_B2_TMS_RNA,,G1,GTTGACCT,GTTGACCT,CAGTGCTT,CAGTGCTT,
8,UMRR_Rep1,,H1,CACCTGTT,CACCTGTT,AGGATAGC,AGGATAGC,
9,RR6_GC_ISS-T_G1_TMS_RNA,,A2,ACTCTCGA,ACTCTCGA,CTGTACCA,CTGTACCA,
10,RR6_GC_ISS-T_G2_TMS_RNA,,B2,TGGACTCT,TGGACTCT,TCTTACGG,TCTTACGG,
11,RR6_FLT_ISS-T_F1_TMS_RNA,,C2,GTCGAAGA,GTCGAAGA,CCACATTG,CCACATTG,
12,RR6_FLT_ISS-T_F2_TMS_RNA,,D2,GTCATCGA,GTCATCGA,GATCCACT,GATCCACT,
13,RR6_BSL_LAR_B3_TMS_RNA,,E2,AGTTCGTC,AGTTCGTC,ACTCGATC,ACTCGATC,
14,RR6_BSL_LAR_B4_TMS_RNA,,A7,TCGTAGTC,TCGTAGTC,CGAGTTAG,CGAGTTAG,
15,RR6_GC_LAR_G3_TMS_RNA,,G2,CGTGTGTA,CGTGTGTA,CCAACGAA,CCAACGAA,
16,RR6_GC_LAR_G4_TMS_RNA,,H2,CTTCGACT,CTTCGACT,GATCTTGC,GATCTTGC,
17,RR6_FLT_LAR_F3_TMS_RNA,,D7,TGCGTAGA,TGCGTAGA,TAGCTGAG,TAGCTGAG,
18,RR6_FLT_LAR_F4_TMS_RNA,,E7,TCTGCTCT,TCTGCTCT,AGCGTGTA,AGCGTGTA,
19,RR6_BSL_ISS-T_B3_TMS_RNA,,C3,ACCACGAT,ACCACGAT,GTCTGCAA,GTCTGCAA,
20,RR6_BSL_ISS-T_B4_TMS_RNA,,D3,AGCACTTC,AGCACTTC,ACGTCCAA,ACGTCCAA,
File Name to Write [DOS Format]: SampleSheet.csv
^G Get Help        M-D DOS Format     M-A Append         M-B Backup File
^C Cancel          M-M Mac Format     M-P Prepend
```

11. Open the RunInfo.xml file using nano (or a text editor of your choosing) by typing the following command:
    $ nano RunInfo.xml
    You'll see a window open that looks similar to the following:

```
  GNU nano 2.3.1                          File: RunInfo.xml

<?xml version="1.0"?>
<RunInfo xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Version="5">
  <Run Id="20190617_FS10000323_13_BPC29610-1014" Number="13">
    <Flowcell>BPC29610-1014</Flowcell>
    <Instrument>FS10000323</Instrument>
    <Date>190617</Date>
    <Reads>
      <Read Number="1" NumCycles="148" IsIndexedRead="N" />
      <Read Number="2" NumCycles="17" IsIndexedRead="Y" />
      <Read Number="3" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="4" NumCycles="148" IsIndexedRead="N" />
    </Reads>
    <FlowcellLayout LaneCount="1" SurfaceCount="1" SwathCount="1" TileCount="16" SectionPerLane="1" LanePerSection="1">
      <TileSet TileNamingConvention="FourDigit">
        <Tiles>
          <Tile>1_1101</Tile>
          <Tile>1_1102</Tile>
          <Tile>1_1103</Tile>
          <Tile>1_1104</Tile>
          <Tile>1_1105</Tile>
          <Tile>1_1106</Tile>
          <Tile>1_1107</Tile>
          <Tile>1_1108</Tile>
          <Tile>1_1109</Tile>
          <Tile>1_1110</Tile>
          <Tile>1_1111</Tile>

^G Get Help     ^O WriteOut     ^R Read File    ^Y Prev Page    ^K Cut Text     ^C Cur Pos
^X Exit         ^J Justify      ^W Where Is     ^V Next Page    ^U UnCut Text   ^T To Spell
```

12. Modify the Read Numbers to account for the UMIs attached to the index by doing the following:
    Add another Read Number line between Read Number="2" and Read Number="3" (note: you can do this by copying the Read Number="2" line and pasting the copy below that line as follows):

```
  GNU nano 2.3.1                          File: RunInfo.xml                                    Modif

<?xml version="1.0"?>
<RunInfo xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Version="5">
  <Run Id="20190617_FS10000323_13_BPC29610-1014" Number="13">
    <Flowcell>BPC29610-1014</Flowcell>
    <Instrument>FS10000323</Instrument>
    <Date>190617</Date>
    <Reads>
      <Read Number="1" NumCycles="148" IsIndexedRead="N" />
      <Read Number="2" NumCycles="17" IsIndexedRead="Y" />
      <Read Number="2" NumCycles="17" IsIndexedRead="Y" />
      <Read Number="3" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="4" NumCycles="148" IsIndexedRead="N" />
    </Reads>
    <FlowcellLayout LaneCount="1" SurfaceCount="1" SwathCount="1" TileCount="16" SectionPerLane="1" LanePerSection="1">
      <TileSet TileNamingConvention="FourDigit">
        <Tiles>
          <Tile>1_1101</Tile>
          <Tile>1_1102</Tile>
          <Tile>1_1103</Tile>
          <Tile>1_1104</Tile>
          <Tile>1_1105</Tile>
          <Tile>1_1106</Tile>
          <Tile>1_1107</Tile>
          <Tile>1_1108</Tile>
          <Tile>1_1109</Tile>
          <Tile>1_1110</Tile>

^G Get Help     ^O WriteOut     ^R Read File    ^Y Prev Page    ^K Cut Text     ^C Cur Pos
^X Exit         ^J Justify      ^W Where Is     ^V Next Page    ^U UnCut Text   ^T To Spell
```

Next, you need to modify the info for each Read Number (note that the index and UMI read lengths could be larger or shorter depending on the length of the indexes and UMIs that you used when preparing your libraries): They should be labeled 1-5.
Read Number="1" is your forward read, has a read length (aka NumCycles) of 148 and is not an indexed read.
Read Number="2" is your i7 index, has a read length (aka NumCycles) of 8 and is an indexed read.
Read Number="3" is your UMI read, has a read length (aka NumCycles) of 9 and is not an indexed read.
Read Number="4" is your i5 index, has a read length (aka NumCycles) of 8 and is an indexed read.
Read Number="5" is your reverse read, has a read length (aka NumCycles) of 148 and is not an indexed read.

When modification is complete, it should look similar to the following:

```
  GNU nano 2.3.1                    File: RunInfo.xml                                   Mo

<?xml version="1.0"?>
<RunInfo xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Version="5">
  <Run Id="20190617_FS10000323_13_BPC29610-1014" Number="13">
    <Flowcell>BPC29610-1014</Flowcell>
    <Instrument>FS10000323</Instrument>
    <Date>190617</Date>
    <Reads>
      <Read Number="1" NumCycles="148" IsIndexedRead="N" />
      <Read Number="2" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="3" NumCycles="9" IsIndexedRead="N" />
      <Read Number="4" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="5" NumCycles="148" IsIndexedRead="N" />
    </Reads>
    <FlowcellLayout LaneCount="1" SurfaceCount="1" SwathCount="1" TileCount="16" SectionPerLane="1" LanePerSection="1">
      <TileSet TileNamingConvention="FourDigit">
        <Tiles>
          <Tile>1_1101</Tile>
          <Tile>1_1102</Tile>
          <Tile>1_1103</Tile>
          <Tile>1_1104</Tile>
          <Tile>1_1105</Tile>
          <Tile>1_1106</Tile>
          <Tile>1_1107</Tile>
          <Tile>1_1108</Tile>
          <Tile>1_1109</Tile>
          <Tile>1_1110</Tile>

^G Get Help     ^O WriteOut     ^R Read File    ^Y Prev Page    ^K Cut Text     ^C Cur Pos
^X Exit         ^J Justify      ^W Where Is     ^V Next Page    ^U UnCut Text   ^T To Spell
```

You can now save the changes you made by pressing ‘control’ and ‘x’ at the same time, which will then ask you if you want to save your changes as shown below:

```
  GNU nano 2.3.1                    File: RunInfo.xml                                   Modi

<?xml version="1.0"?>
<RunInfo xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Version="5">
  <Run Id="20190617_FS10000323_13_BPC29610-1014" Number="13">
    <Flowcell>BPC29610-1014</Flowcell>
    <Instrument>FS10000323</Instrument>
    <Date>190617</Date>
    <Reads>
      <Read Number="1" NumCycles="148" IsIndexedRead="N" />
      <Read Number="2" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="3" NumCycles="9" IsIndexedRead="N" />
      <Read Number="4" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="5" NumCycles="148" IsIndexedRead="N" />
    </Reads>
    <FlowcellLayout LaneCount="1" SurfaceCount="1" SwathCount="1" TileCount="16" SectionPerLane="1" LanePerSection="1">
      <TileSet TileNamingConvention="FourDigit">
        <Tiles>
          <Tile>1_1101</Tile>
          <Tile>1_1102</Tile>
          <Tile>1_1103</Tile>
          <Tile>1_1104</Tile>
          <Tile>1_1105</Tile>
          <Tile>1_1106</Tile>
          <Tile>1_1107</Tile>
          <Tile>1_1108</Tile>
          <Tile>1_1109</Tile>
          <Tile>1_1110</Tile>
Save modified buffer (ANSWERING "No" WILL DESTROY CHANGES) ?
 Y Yes
 N No           ^C Cancel
```

Press ‘y’ to save your changes and it will prompt you with the name of the file. Make sure it says RunInfo.xml (as shown below) then press ‘Enter’

```
  GNU nano 2.3.1                    File: RunInfo.xml

<?xml version="1.0"?>
<RunInfo xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Version="5">
  <Run Id="20190617_FS10000323_13_BPC29610-1014" Number="13">
    <Flowcell>BPC29610-1014</Flowcell>
    <Instrument>FS10000323</Instrument>
    <Date>190617</Date>
    <Reads>
      <Read Number="1" NumCycles="148" IsIndexedRead="N" />
      <Read Number="2" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="3" NumCycles="9" IsIndexedRead="N" />
      <Read Number="4" NumCycles="8" IsIndexedRead="Y" />
      <Read Number="5" NumCycles="148" IsIndexedRead="N" />
    </Reads>
    <FlowcellLayout LaneCount="1" SurfaceCount="1" SwathCount="1" TileCount="16" SectionPerLane="1" LanePerSection="1">
      <TileSet TileNamingConvention="FourDigit">
        <Tiles>
          <Tile>1_1101</Tile>
          <Tile>1_1102</Tile>
          <Tile>1_1103</Tile>
          <Tile>1_1104</Tile>
          <Tile>1_1105</Tile>
          <Tile>1_1106</Tile>
          <Tile>1_1107</Tile>
          <Tile>1_1108</Tile>
          <Tile>1_1109</Tile>
          <Tile>1_1110</Tile>
File Name to Write [DOS Format]: RunInfo.xml
^G Get Help        M-D DOS Format     M-A Append         M-B Backup File
^C Cancel          M-M Mac Format     M-P Prepend
```

13. Move up one directory to return to the directory you made in step 2 by typing the following command:
    $ cd ..

14. Set up your bcl2fastq script to convert the bcl files (which are located in the iSeq output directory) to fastq files, while assigning each sequence read to the appropriate sample (a process known as demultiplexing). We provide an example bcl2fastq script (0_bcl2fastq.sh), that you can use as a template.
    Copy the 0_bcl2fastq.sh script (which can be found in the HTStream_scripts.zip folder) into your current directory (the directory you made in step 2).

15. Open the 0_bcl2fastq.sh script using nano (or a text editor of your choosing) by typing the following command:
    $ nano 0_bcl2fastq.sh
    You'll see a window open that looks as follows:

```
  GNU nano 2.3.1                    File: 0_bcl2fastq.sh

start=$(date +%s)
echo "start time: $start"

input_dir=./iSeq_output_directory
output_dir=./out-iSeq_output_directory

bcl2fastq --runfolder-dir $input_dir --output-dir $output_dir --loading-threads 8 --writing-threads 8 --processing-threads 15 --barcode-mis$

end=$(date +%s)
echo "end time: $end"
runtime_s=$(echo $(( end - start )))
echo "total run time(s): $runtime_s"
sec_per_min=60
sec_per_hr=3600
runtime_m=$(echo "scale=2; $runtime_s / $sec_per_min;" | bc)
                                        [ Read 27 lines ]
^G Get Help    ^O WriteOut    ^R Read File   ^Y Prev Page   ^K Cut Text    ^C Cur Pos
^X Exit        ^J Justify     ^W Where Is    ^V Next Page   ^U UnCut Text  ^T To Spell
```

16. Replace 'iSeq_output_directory' with the title of your iSeq output directory – in this example, it is: 20190617_FS10000323_13_BPC29610-1014
    When the modification is complete, it will look as follows:

```
  GNU nano 2.3.1                    File: 0_bcl2fastq.sh                                    Modified

start=$(date +%s)
echo "start time: $start"

input_dir=./20190617_FS10000323_13_BPC29610-1014
output_dir=./out-20190617_FS10000323_13_BPC29610-1014

bcl2fastq --runfolder-dir $input_dir --output-dir $output_dir --loading-threads 8 --writing-threads 8 --processing-threads 15 --barcode-mis$

end=$(date +%s)
echo "end time: $end"
runtime_s=$(echo $(( end - start )))
echo "total run time(s): $runtime_s"
sec_per_min=60
sec_per_hr=3600
runtime_m=$(echo "scale=2; $runtime_s / $sec_per_min;" | bc)

^G Get Help     ^O WriteOut     ^R Read File    ^Y Prev Page    ^K Cut Text     ^C Cur Pos
^X Exit         ^J Justify      ^W Where Is     ^V Next Page    ^U UnCut Text   ^T To Spell
```

You can now save the changes you made by pressing ‘control’ and ‘x’ at the same time, which will then ask you if you want to save your changes as shown below:

```
  GNU nano 2.3.1                    File: 0_bcl2fastq.sh                                    Modified

start=$(date +%s)
echo "start time: $start"

input_dir=./20190617_FS10000323_13_BPC29610-1014
output_dir=./out-20190617_FS10000323_13_BPC29610-1014

bcl2fastq --runfolder-dir $input_dir --output-dir $output_dir --loading-threads 8 --writing-threads 8 --processing-threads 15 --barcode-mis$

end=$(date +%s)
echo "end time: $end"
runtime_s=$(echo $(( end - start )))
echo "total run time(s): $runtime_s"
sec_per_min=60
sec_per_hr=3600
runtime_m=$(echo "scale=2; $runtime_s / $sec_per_min;" | bc)
Save modified buffer (ANSWERING "No" WILL DESTROY CHANGES) ?
 Y Yes
 N No           ^C Cancel
```

Press ‘y’ to save your changes and it will prompt you with the name of the file. Make sure it says 0_bcl2fastq.sh then press ‘Enter’

17. Change permissions on the 0_bcl2fastq.sh file, so the computer knows it’s a file containing commands that can be executed, by typing the following command:
    $ chmod a+x 0_bcl2fastq.sh

18. You’re now ready to run bcl2fastq by typing the following command:
    $ nohup ./0_bcl2fastq.sh &
    Note: Running this command with the ‘nohup’ and ‘&’ at the beginning and end, respectively, will allow the job to continue running as long as your computer is on (i.e. it won’t stop if you accidently ‘x’ out of the terminal window, or if your computer goes to sleep, etc.). This will also put the standard output of the bcl2fastq run into a file called ‘nohup.out’. If you prefer to let the job run so that your standard output is printed to the terminal window as it goes, remove the ‘nohup’ and ‘&’ at the beginning and end of the command, but be sure that your terminal window stays open and your computer does not go to sleep.

19. After your job is complete, execute the ls command to see what’s in your directory.
    $ ls

You should now see a file titled:
'nohup.out'

20. Look at that file by typing the following command:
    $ cat nohup.out
    The end of that file should look similar to the following:

```
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep1/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep1/unknown"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep11"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep11/AGGCATAG+A
AGAGGCA"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep11/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep11/unknown"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep2"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep2/AGAACGAG+AT
TAGCCG"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep2/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/UMRR_Rep2/unknown"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/Undetermined"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/Undetermined/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/Undetermined/unknown"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/all/all"
2019-06-20 11:52:26 [2967880] INFO: creating directory "./out-20190617_FS10000323_13_BPC29610-1014_take2/Reports/html/BPC29610-1014/default/all/unknown"
2019-06-20 11:52:29 [2967880] Processing completed with 0 errors and 0 warnings.
end time: 1561056749
total run time(s): 101
total run time(m): 1.68
total run time(h): .02
```

    Note: You want to see that the "Processing completed with 0 errors and 0 warnings."

21. Change directory into your bcl2fastq output directory by typing the following command:
    $ cd out-iSeq_output_directory
    In this example, the command looks as follows:
    $ cd out-20190617_FS10000323_13_BPC29610-1014

22. Execute the ls command to view the bcl2fastq files and directories.
    $ ls
    You should see two directories titled 'Reports' and 'Stats'
    <u>For samples with UMIs (set up as indicated in this example)</u>: You should also see one R1, one R2, and one R3 fastq.gz file for each sample
    <u>For samples without UMIs</u>: You should also see one R1 and one R2 fastq.gz file for each sample

23. Make sure all samples were converted to fastq files by typing (or copy and pasting) the following command:
    $ ls -1 *_R1_* | wc -l
    The number that is outputted should be the number of samples plus 1 (the additional sample is undetermined). If you get less than that, there may be a mistake in the indices in the sample sheet for the sample(s) that do not have fastq.gz files.

24. Make a 'Fastq' directory in your current directory (which should be the bcl2fastq output directory, out-20190617_FS10000323_13_BPC29610-1014 in this example) by typing the following command:
    $ mkdir Fastq

25. Move all of your fastq.gz files into the Fastq directory you made in step 24 by typing the following command:
    $ mv *fastq.gz Fastq

26. Execute the ls command to view what's in your bcl2fastq output directory.

$ ls
You should only see the following 3 directories:
'Fastq', 'Reports', and 'Stats'

27. Follow the instructions below to make a text file that contains a list of the samples you ran on the iSeq and save it as 'samples.txt'

Go into the Fastq folder you generated in step 24 by typing the following command:

$ cd Fastq

Once there, type the following command to get a single column list of all your samples.

Command:

$ ls -1 *_R1_*

Copy that list (highlight, ctrl+c)!

Next, move up one folder so that you're back in your bcl2fastq output folder (for this example, it would be out-20190617_FS10000323_13_BPC29610-1014) using the following command:

$ cd ..

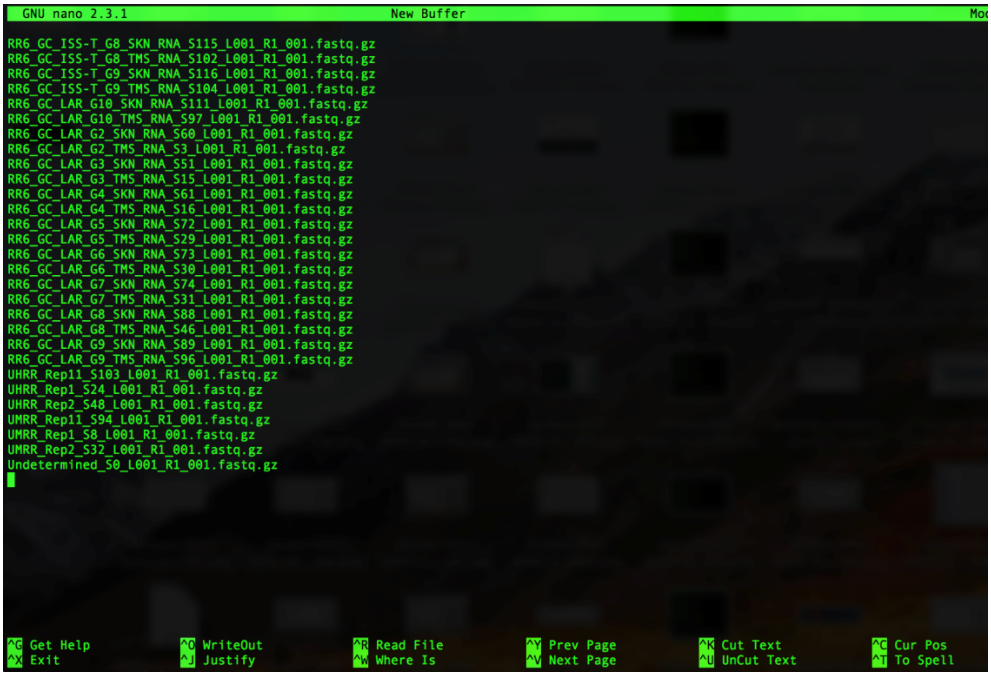
You will now use the nano text editor (or a text editor of your choosing) to make a samples.txt file by following the instructions below.

Type the following command to open the nano text editor:

$ nano

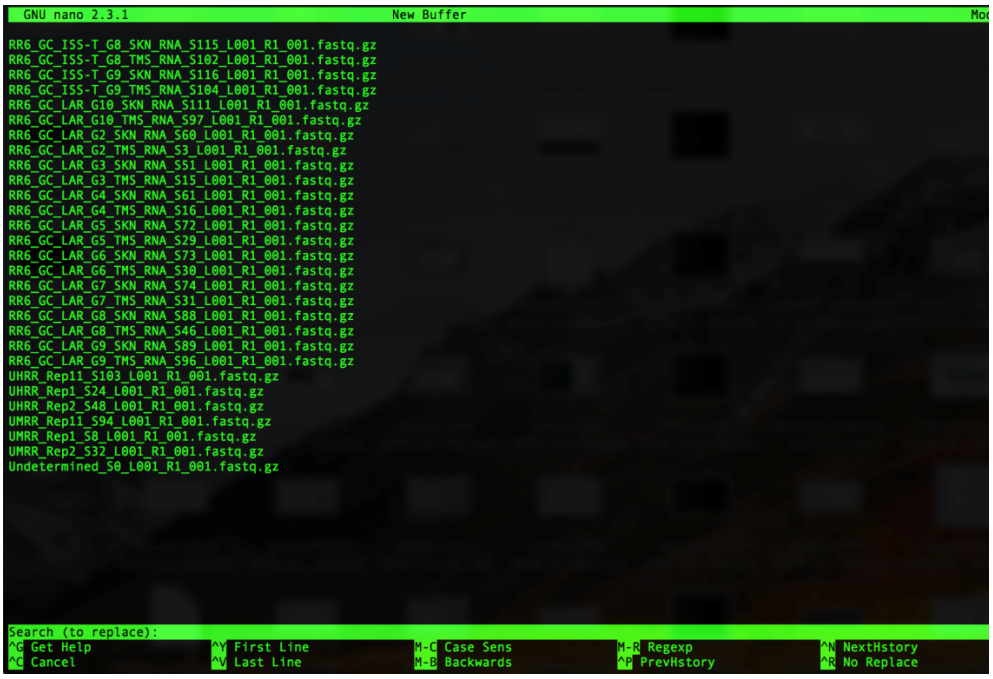
The following will appear:

Paste (ctrl+v) the fastq file names that you copied above. It should look similar to the following:



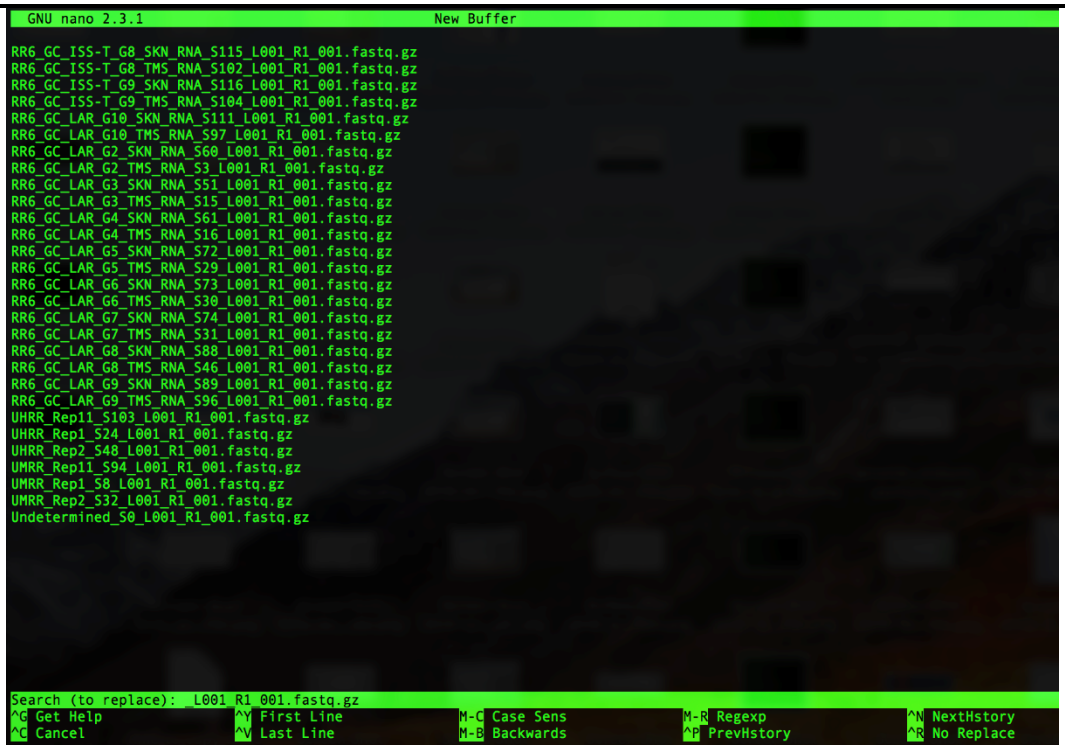
Remove everything that comes after the S# as follows:

Press ‘control’ and ‘\ ‘ at the same time, you’ll see the following:

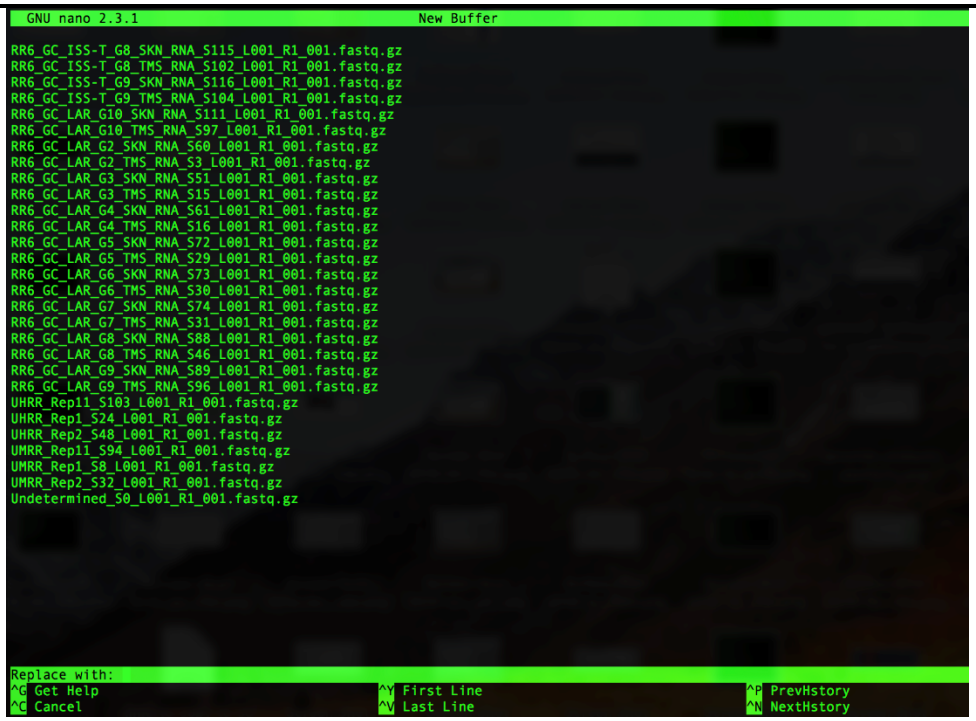


Note that you’ll get prompted for what you want to replace, type the following:
_L001_R1_001.fastq.gz

It will look like this in your window:



Then press ‘Enter’. You will now be prompted for what to replace _L001_R1_001.fastq.gz with. It will look as follows:



Since we want to remove _L001_R1_001.fastq.gz (and not replace it with anything), just press ‘Enter’

Next, you’ll be asked which instance of _L001_R1_001.fastq.gz you want to replace (or, in our case, remove), as follows:

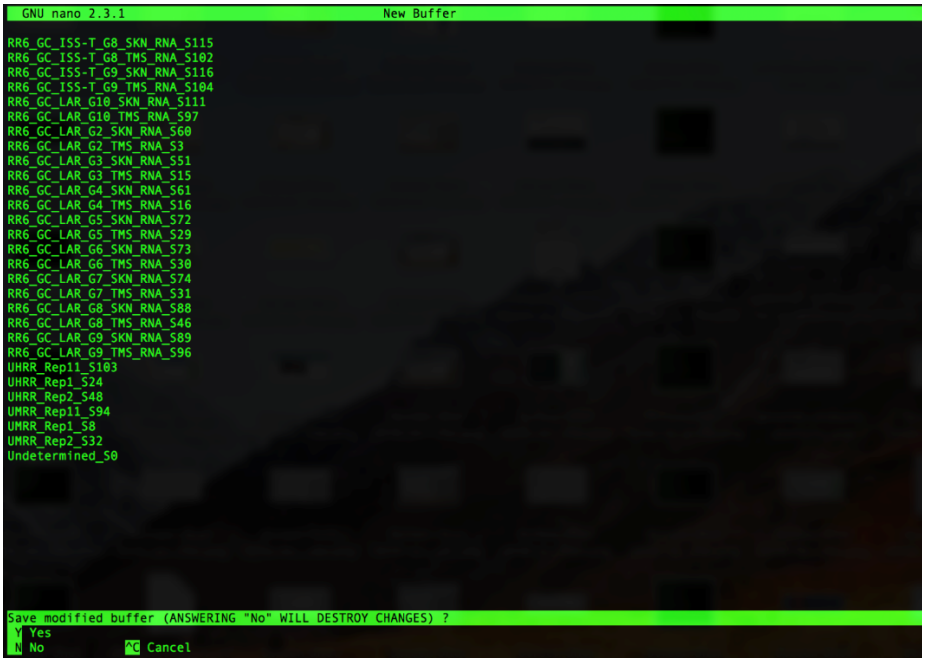
```
  GNU nano 2.3.1                                New Buffer

RR6_BSL_ISS-T_B1_TMS_RNA_S6_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B2_SKN_RNA_S54_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B2_TMS_RNA_S7_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B3_SKN_RNA_S64_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B3_TMS_RNA_S19_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B4_SKN_RNA_S65_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B4_TMS_RNA_S20_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B5_SKN_RNA_S79_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B5_TMS_RNA_S36_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B6_SKN_RNA_S80_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B6_TMS_RNA_S37_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B7_SKN_RNA_S92_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B7_TMS_RNA_S38_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B8_SKN_RNA_S78_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B8_TMS_RNA_S100_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B9_SKN_RNA_S112_L001_R1_001.fastq.gz
RR6_BSL_ISS-T_B9_TMS_RNA_S101_L001_R1_001.fastq.gz
RR6_BSL_LAR_B10_SKN_RNA_S110_L001_R1_001.fastq.gz
RR6_BSL_LAR_B10_TMS_RNA_S95_L001_R1_001.fastq.gz
RR6_BSL_LAR_B1_SKN_RNA_S49_L001_R1_001.fastq.gz
RR6_BSL_LAR_B1_TMS_RNA_S1_L001_R1_001.fastq.gz
RR6_BSL_LAR_B2_SKN_RNA_S50_L001_R1_001.fastq.gz
RR6_BSL_LAR_B2_TMS_RNA_S2_L001_R1_001.fastq.gz
RR6_BSL_LAR_B3_SKN_RNA_S58_L001_R1_001.fastq.gz
RR6_BSL_LAR_B3_TMS_RNA_S13_L001_R1_001.fastq.gz
RR6_BSL_LAR_B4_SKN_RNA_S59_L001_R1_001.fastq.gz
RR6_BSL_LAR_B4_TMS_RNA_S14_L001_R1_001.fastq.gz
RR6_BSL_LAR_B5_SKN_RNA_S86_L001_R1_001.fastq.gz
RR6_BSL_LAR_B5_TMS_RNA_S26_L001_R1_001.fastq.gz
RR6_BSL_LAR_B6_SKN_RNA_S70_L001_R1_001.fastq.gz
RR6_BSL_LAR_B6_TMS_RNA_S27_L001_R1_001.fastq.gz
RR6_BSL_LAR_B7_SKN_RNA_S71_L001_R1_001.fastq.gz
RR6_BSL_LAR_B7_TMS_RNA_S28_L001_R1_001.fastq.gz
RR6_BSL_LAR_B8_SKN_RNA_S87_L001_R1_001.fastq.gz
RR6_BSL_LAR_B8_TMS_RNA_S45_L001_R1_001.fastq.gz
RR6_BSL_LAR_B9_SKN_RNA_S109_L001_R1_001.fastq.gz
RR6_BSL_LAR_B9_TMS_RNA_S93_L001_R1_001.fastq.gz
RR6_FLT_ISS-T_F10_SKN_RNA_S120_L001_R1_001.fastq.gz
RR6_FLT_ISS-T_F10_TMS_RNA_S108_L001_R1_001.fastq.gz
Replace this instance?
 Y Yes          A All
 N No          ^C Cancel
```

Press ‘a’ to remove all of the instances of _L001_R1_001.fastq.gz. Note that all of the instances of _L001_R1_001.fastq.gz will be removed and it will show how many instances of _L001_R1_001.fastq.gz were replaced as follows:

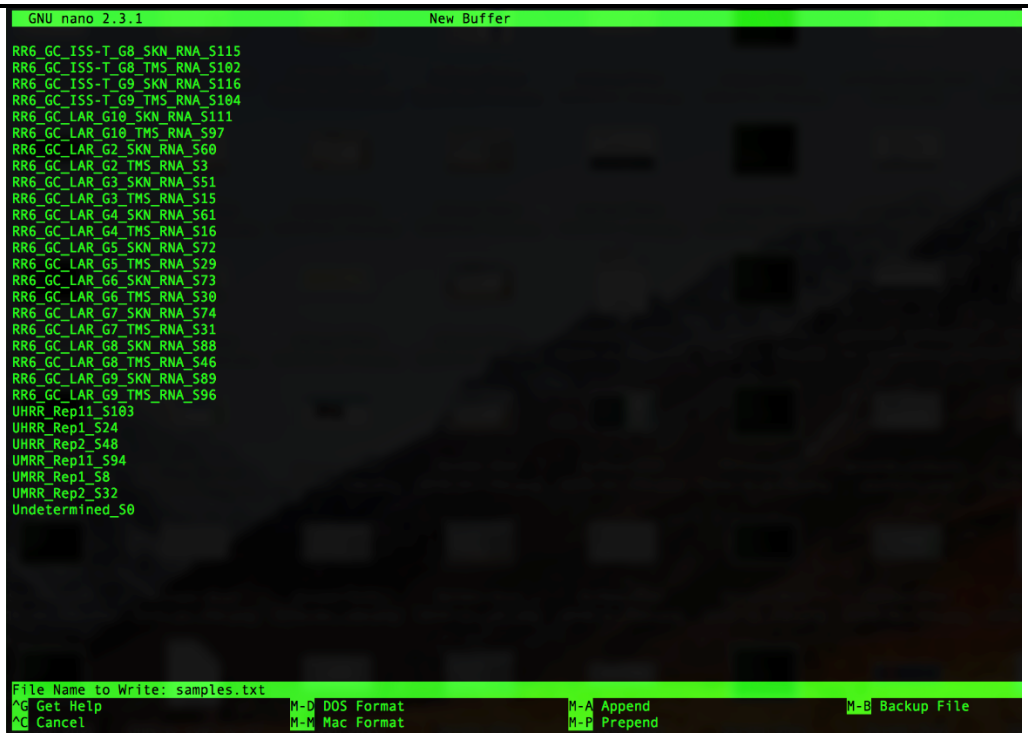
```
  GNU nano 2.3.1                                New Buffer

RR6_GC_ISS-T_G8_SKN_RNA_S115
RR6_GC_ISS-T_G8_TMS_RNA_S102
RR6_GC_ISS-T_G9_SKN_RNA_S116
RR6_GC_ISS-T_G9_TMS_RNA_S104
RR6_GC_LAR_G10_SKN_RNA_S111
RR6_GC_LAR_G10_TMS_RNA_S97
RR6_GC_LAR_G2_SKN_RNA_S60
RR6_GC_LAR_G2_TMS_RNA_S3
RR6_GC_LAR_G3_SKN_RNA_S51
RR6_GC_LAR_G3_TMS_RNA_S15
RR6_GC_LAR_G4_SKN_RNA_S61
RR6_GC_LAR_G4_TMS_RNA_S16
RR6_GC_LAR_G5_SKN_RNA_S72
RR6_GC_LAR_G5_TMS_RNA_S29
RR6_GC_LAR_G6_SKN_RNA_S73
RR6_GC_LAR_G6_TMS_RNA_S30
RR6_GC_LAR_G7_SKN_RNA_S74
RR6_GC_LAR_G7_TMS_RNA_S31
RR6_GC_LAR_G8_SKN_RNA_S88
RR6_GC_LAR_G8_TMS_RNA_S46
RR6_GC_LAR_G9_SKN_RNA_S89
RR6_GC_LAR_G9_TMS_RNA_S96
UHRR_Rep11_S103
UHRR_Rep1_S24
UHRR_Rep2_S48
UMRR_Rep11_S94
UMRR_Rep1_S8
UMRR_Rep2_S32
Undetermined_S0

                          [ Replaced 121 occurrences ]
^G Get Help    ^O WriteOut    ^R Read File   ^Y Prev Page   ^K Cut Text    ^C Cur Pos
^X Exit        ^J Justify     ^W Where Is    ^V Next Page   ^U UnCut Text  ^T To Spell
```

The number of occurrences replaced should be the number of samples plus one (for undetermined); i.e. in the example above there are 120 samples so we see that 121 occurrences were replaced.

You can now save the changes you made by pressing ‘control’ and ‘x’ at the same time, which will then ask you if you want to save your changes as shown below:



Press ‘y’ to save your changes and it will prompt you to name the file. Type ‘samples.txt’ as shown below then press ‘Enter’



28. Type the following command to check what's in your folder.
    Command:
    $ ls

    You should see the following in your bcl2fastq output folder:
    'Fastq', 'Reports', and 'Stats' directories and the 'samples.txt' file you created in step 27.

29. Type the following command to check your 'samples.txt' file.
    Command:
    $ cat samples.txt

    Example:

```
RR6_GC_ISS-T_G1_SKN_RNA_S57
RR6_GC_ISS-T_G1_TMS_RNA_S9
RR6_GC_ISS-T_G2_SKN_RNA_S113
RR6_GC_ISS-T_G2_TMS_RNA_S10
RR6_GC_ISS-T_G3_SKN_RNA_S68
RR6_GC_ISS-T_G3_TMS_RNA_S21
RR6_GC_ISS-T_G4_SKN_RNA_S69
RR6_GC_ISS-T_G4_TMS_RNA_S22
RR6_GC_ISS-T_G5_SKN_RNA_S84
RR6_GC_ISS-T_G5_TMS_RNA_S39
RR6_GC_ISS-T_G6_SKN_RNA_S85
RR6_GC_ISS-T_G6_TMS_RNA_S40
RR6_GC_ISS-T_G7_SKN_RNA_S114
RR6_GC_ISS-T_G7_TMS_RNA_S41
RR6_GC_ISS-T_G8_SKN_RNA_S115
RR6_GC_ISS-T_G8_TMS_RNA_S102
RR6_GC_ISS-T_G9_SKN_RNA_S116
RR6_GC_ISS-T_G9_TMS_RNA_S104
RR6_GC_LAR_G10_SKN_RNA_S111
RR6_GC_LAR_G10_TMS_RNA_S97
RR6_GC_LAR_G2_SKN_RNA_S60
RR6_GC_LAR_G2_TMS_RNA_S3
RR6_GC_LAR_G3_SKN_RNA_S51
RR6_GC_LAR_G3_TMS_RNA_S15
RR6_GC_LAR_G4_SKN_RNA_S61
RR6_GC_LAR_G4_TMS_RNA_S16
RR6_GC_LAR_G5_SKN_RNA_S72
RR6_GC_LAR_G5_TMS_RNA_S29
RR6_GC_LAR_G6_SKN_RNA_S73
RR6_GC_LAR_G6_TMS_RNA_S30
RR6_GC_LAR_G7_SKN_RNA_S74
RR6_GC_LAR_G7_TMS_RNA_S31
RR6_GC_LAR_G8_SKN_RNA_S88
RR6_GC_LAR_G8_TMS_RNA_S46
RR6_GC_LAR_G9_SKN_RNA_S89
RR6_GC_LAR_G9_TMS_RNA_S96
UHRR_Rep11_S103
UHRR_Rep1_S24
UHRR_Rep2_S48
UMRR_Rep11_S94
UMRR_Rep1_S8
UMRR_Rep2_S32
Undetermined_S0
```

Note that there are no spaces or empty lines above or below the list.

30. Set up the directory structure that you will need for the HTStream analysis by copying the make_directories.sh script (which can be found in the HTStream_scripts.zip folder) into your current directory, which is the bcl2fastq output folder (for this example, it would be out-20190617_FS10000323_13_BPC29610-1014).

31. Change permissions of the make_directories.sh file, so the computer knows it's a file containing commands that can be executed, by typing the following command:
    $ chmod a+x make_directories.sh

32. You're now ready to run bcl2fastq by typing the following command:
    $ ./make_directories.sh

33. Execute the ls command to see what changed. You should now see a directory titled 'HTStream_output' in addition to what's listed in step 28.

34. Copy the following HTStream scripts, header.txt file, and rRNA_fasta_files directory (all of which can be found in the HTStream_scripts.zip folder) into your current directory, which is the bcl2fastq output folder (for this example, it would be out-20190617_FS10000323_13_BPC29610-1014)
    1_HTStream_rRNA_SD_PE_wUMI_loop.sh

    1_HTStream_rRNA_SD_PE_noUMI_loop.sh

    1_HTStream_rRNA_SD_SE_loop.sh

2_calc_totalreads_loop.sh

3_make_rRNA_reports_PE_loop.sh

3_make_rRNA_reports_SE_loop.sh

4_make_complete_rRNA_report.sh

header.txt

/rRNA_fasta_files

35. Change permissions of the HTStream script files, so the computer knows these files contain commands that can be executed, by typing the following command:
    $ chmod a+x *.sh

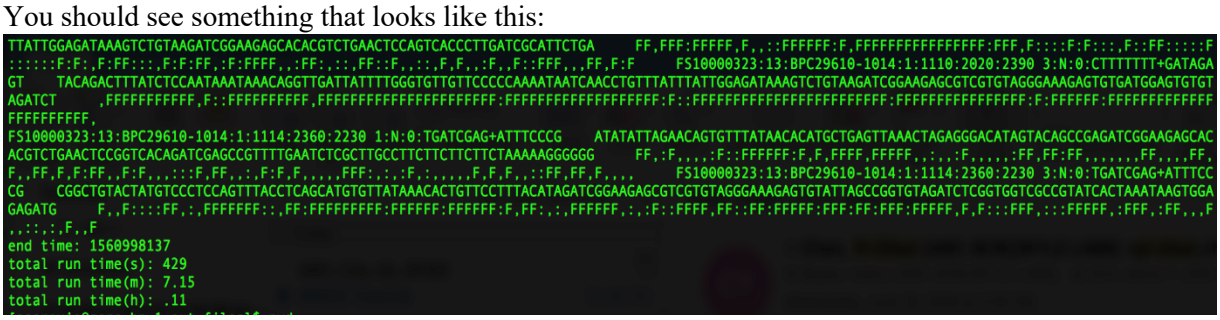
36. You are now ready to run the HTStream program. To do so, type one of the following commands (depending on how your libraries were prepared and if you used paired-end (PE), or single-end (SE) sequencing).
    Command (for PE samples with UMIs):
    $ nohup ./1_HTStream_rRNA_SD_loop.sh &

    Command (for PE samples without UMIs):
    $ nohup ./1_HTStream_rRNA_SD_noUMI_loop.sh &

    Command (for SE samples):
    $ nohup ./1_HTStream_rRNA_SD_SE_loop.sh &

    Note: See note about 'nohup' and '&' in step 18 above

37. Once complete, look at the end of the standard output by typing the following command:
    $ tail ./nohup.out
    You should see something that looks like this:



Note: The hts_SuperDeduper function of HTStream removes reads that are identified as duplicates (there is no option to only report these reads). Since we only want to report the number of duplicate reads identified by hts_SuperDeduper (but not remove them), rather than modify the input fastq files, we instead print the modified fastq files to standard out (leaving the original fastq files intact). This means that your nohup.out file will be very large (roughly the size of all the original input fastq files combined). To save space on your machine, once you are confident the script ran successfully, you may want to delete the nohup.out file.

38. Once complete, there should be a .log file generated for each sample. To view the files that have been generated, use the following command.
    Command:
    $ ls ./HTStream_output/1_log_files

39. Type the following command to calculate the total number of reads generated by the iSeq run.
    Command:
    $ ./2_calc_totalreads_loop.sh

40. To view the files generated, type the command below.

    Command:
    $ ls HTStream_output/2_totreads

    *Note: You should see a .txt file for each sample and a 'totalreads.txt' file

41. Use the totalreads.txt and .log files to generate a rRNA report for each sample by typing the command below.
    Command (for PE samples with or without UMIs):
    $ ./3_make_rRNA_reports_PE_loop.sh

    Command (for SE samples):
    $ ./3_make_rRNA_reports_SE_loop.sh

    Note: This command may take a few minutes to run and will generate several .txt files

42. To view the files generated, type the command below.
    Command:
    $ ls HTStream_output/3_rRNA_reports

    *Note: You should see the following two txt files for each sample (the other two txt files per sample that are generated are in HTStream_output/3_tmp_files):
    <sample_name>_rRNA-report_tmp.txt
    <sample_name>_rRNA-report.txt

Files in HTStream_output/3_tmp_files are as follows:
<sample_name>_tmp.txt
<sample_name>_.txt

43. Use the rRNA reports from each sample to generate a complete rRNA report containing all the info from all the samples by typing the command below.
Command:
$ ./4_make_complete_rRNA_report.sh

Note: This command will run very quickly (within seconds) and will generate one additional .txt file

44. Use the following command to see the output files that were generated from step 43:
Command:
$ ls HTStream_output

*Note: You should see the following txt file (in addition to directories that contain outputs from the previous 4 commands):
complete_rRNA-report.txt

45. Open the complete_rRNA-report.txt in excel and save it as an excel file (in .xls or .xlsx format). The following steps are done in excel.
Follow SOP#8.2 "Instructions to use iSeq HTStream output to calculate library pooling values" to calculate re-pooling volumes or to calculate the projected number of non-ribosomal reads to expect from the pool on a given NovaSeq flow cell.