Run Setup and Bioinformatic Analysis

Accel-NGS® 2S MID Indexing Kits

Sequencing MID Libraries

For MiSeq®, HiSeq®, and NextSeq® instruments:

- Modify the config file to create a fastq for index reads
- Using the Illumina® Experiment Manager software, specify 2 index reads for the run.
- In the CSV file, specify the MID index and the sample index sequences.
  - Samples will be demultiplexed based only on their sample index.
- Using a custom script, join MIDs to their respective fastq read headers, align these fastq, and analyze the reads with a common genomic coordinate.
- **The following slides contain instructions for the MiSeq instrument, but these can also apply to HiSeq and NextSeq instruments as well.**

Swift BIOSCIENCES™

The Swift MID Adapter Sequence

Below is the sequence of the standard LT P7 adapter:

GATCGGAAGAGCACACGTCTGAACTCCAGTCAC**[i7]**ATCTCGTATGCCGTCTTCTGCTTG

Below is the sequence of the Swift MID P5 adapter, where 9 random N bases (instead of the standard Index 2, 8 bp D501-508 index sequence) is inserted:

AATGATACGGCGACCACCGAGATCTACAC***NNNNNNNNN***ACACTCTTTCCCTACACGACG CTCTTCCGATCT

Swift BIOSCIENCES™

Pre-Run Instrument Set-Up

Modify the config file to allow generation of an index fastq file during data analysis:

1. Stop the MiSeq® Reporter process (if it is running).
2. Locate the “MiSeq Reporter.exe.config” file located in C:/Illumina/MiSeq Reporter.
3. Open config file and search for a line that reads:

   “<add key=”CreateFastqForIndexReads” value=”0”/>”.

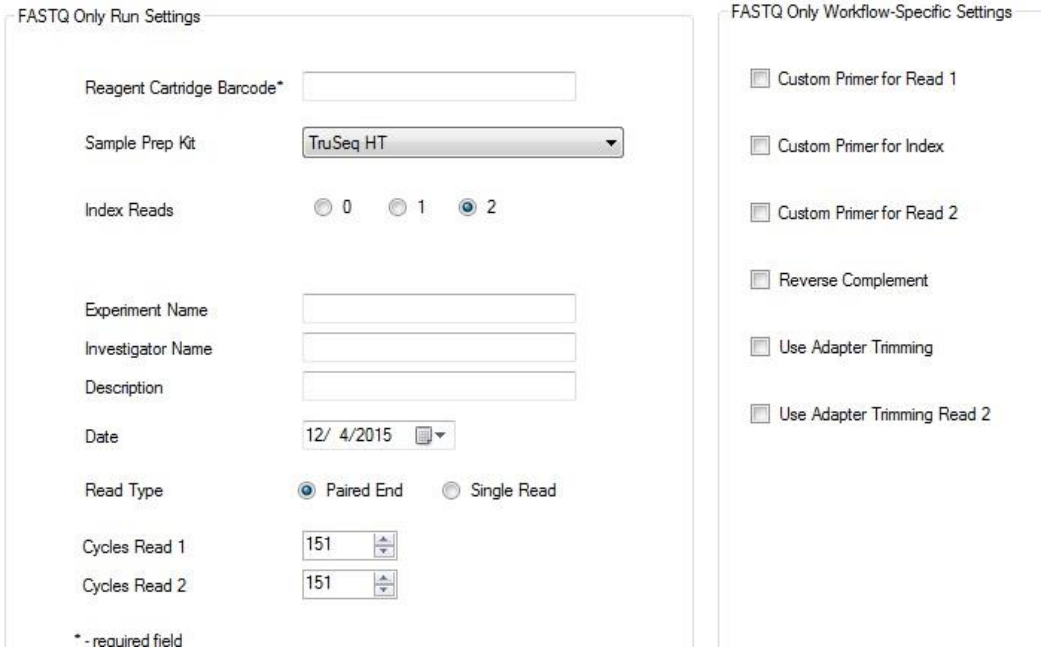
   - If this line is present, change the value from “0” to “1”.
   - If this line is not present, add the line to the config file using the add keys function under the app Settings tab.

4. Restart the MiSeq reporter process.
5. Re-queue the run for data analysis (if required).

Swift BIOSCIENCES™

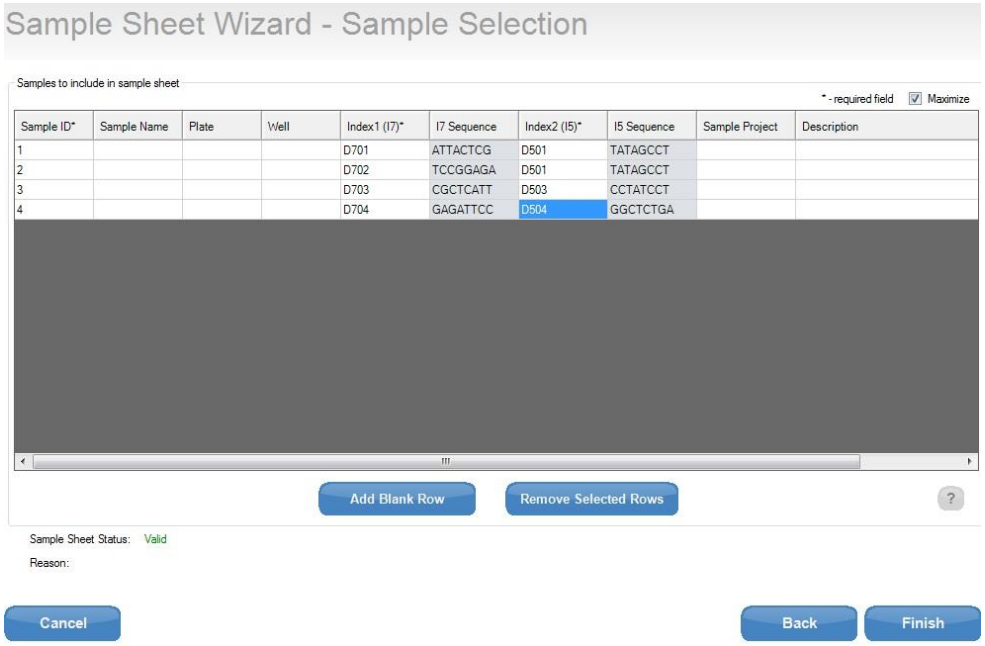
Instrument Set-Up and Sample Sheet Preparation

- Experimental set-up on Illumina® Experiment Manager software
  - Select MiSeq
  - On the MiSeq Application Selection page, select Other --> FASTQ Only
  - Enter Reagent Cartridge Barcode, choose TruSeq HT, choose “2" for Index Reads, ... Read Type: Paired End, enter desired number of cycles for each read, and uncheck all the boxes on the right (including adapter trimming)
    - **By selecting HT, we can alter the index sequences read in the sample sheet**
    - **2 Index reads ensures P7 index (Index 1) is read, as well as the P5 MID located in place of the P5 index (Index 2)**



Instrument Set-Up and Sample Sheet Preparation

- On the next step, enter your index sequence if using high throughput indices
- If using low throughput indices, select a random index sequence until sample sheet status is valid
  - We will specify the low throughput sequence in the sample sheet
- For I5 Sequence, enter a random index number on the pull down menu
  - We will specify the MID (NNNNNNNNN) in the sample sheet
  - Click Finish to generate the CSV file



Instrument Set-Up and Sample Sheet Preparation

- Alter your sample sheet to represent the real index sequences
  - When using a low throughput index, include the index and the next 2 bases
    - » For example with I2: Enter **CGATGT**AT. Bold represents the index whereas the underlined bases represent the next two bases on the adapter
  - For Index 2, enter the mID sequence: NNNNNNNNN
  - During the MiSeq run, the samples will be separated based only on their Index 1 indexes, since any Index 2 reads will be valid

	A	B	C	D	E	F	G	H	I	J
1	[Header]									
2	IEMFileVersion	4								
3	Investigator Name	Dr. Lenhart								
4	Experiment Name	mID								
5	Date	12/2/2015								
6	Workflow	GenerateFASTQ								
7	Application	FASTQ Only								
8	Assay	TruSeq HT								
9	Description									
10	Chemistry	Amplicon								
11										
12	[Reads]									
13	76									
14	76									
15										
16	[Settings]									
17	ReverseComplement	0								
18										
19	[Data]									
20	Sample_ID	Sample_Name	Sample_Plate	Sample_Well	I7_Index_ID	index	I5_Index_ID	index2	Sample_Project	Description
21	1	Sample #1			D701	CGATGTAT	D501	NNNNNNNNN		
22	2	Sample #2			D702	CAGATCAT	D502	NNNNNNNNN		
23	3	Sample #3			D703	TGACCAAT	D503	NNNNNNNNN		
24	4	Sample #4			D704	GCCAATAT	D504	NNNNNNNNN		

Swift BIOSCIENCES™

Instrument Set-Up and Sample Sheet Preparation

- After the run, all reads will be separated based on Index 1 reads

Analysis | Imaging | Summary | Tile Status | TruSeq Controls | Indexing

Reads mapped to Index Id

Total Reads	PF Reads	% Reads Identified (PF)
30804952	27330618	98.0736

Index Number	Sample Id	Project	Index 1 (I7)	Index 2 (I5)
1	1	NA	CGATGTAT	NNNNNNNNN
2	2	NA	CAGATCAT	NNNNNNNNN
3	3	NA	TGACCAAT	NNNNNNNNN
4	4	NA	GCCAATAT	NNNNNNNNN

Non-Identified (Undetermined) reads are due to either poor quality or reads containing absent index sequences.

Swift BIOSCIENCES™

Options for Retrieving Index Sequence Files

Instrument	bcl2fastq	MiSeq reporter	BaseSpace
MiniSeq	✓		
MiSeq	✓	✓	
NextSeq 500	✓		
HiSeq 2500	✓		
HiSeq 4000	✓		

If for any reason the data is not extracted correctly by the Sample Sheet setup, bcl2fastq Conversion Software can be used to correctly extract the data.

Logic for Bioinformatic Analysis of ChIP-Seq and cfDNA

- After the sequencing run, all reads will be separated based on Index 1 only. Non-identified (or undetermined) reads are either due to poor quality or reads missing index sequences.
- Using a custom script, join the MIDs to the respective fastq read headers.
- Align these fastq using BWA.
- Using another custom script, analyze the reads with a common genomic coordinate.
  - If the reads have unique MID sequences, they represent fragment duplicates and should be retained as unique reads
  - If the reads have identical MID sequences, they represent PCR duplicates. Mark all reads except one as duplicates based on mapping and base quality.

Swift BIOSCIENCES™

Logic for Bioinformatic Analysis of Low Frequency Variants

- After the sequencing run, all reads will be separated based on Index 1 only. Non-identified (or undetermined) reads are either due to poor quality or reads missing index sequences.
- Using a custom script, join the MIDs to the respective Fastq read headers.
- Align these Fastq using BWA.
- Use Picard Tools or Samtools to collect metrics like % duplication, reads on target, etc.
- Using another custom script, group fragments with at least 3 PCR duplicate reads per MID for variant calling (also considering Flag and Cigar values). Within the group, determine a consensus sequence for each fragment which eliminates sequencing and PCR errors present at less than 50%.
  - The following publication contains some details concerning this kind of analysis, and may be a useful reference:
    - S.R. Kennedy, et al. Nat Protoc. 2014 November; 9(11):2586-2606

Swift BIOSCIENCES™

Fulcrum Genomics Analysis Tools

- To perform this analysis, we recommend using UMI tools from the fgbio package (https://github.com/fulcrumgenomics/fgbio)
  - It enables you to tag your bam file with the MIDs from your Index 2 fastq (AnnotateBamWithUmis)
  - Use the Picard MarkDuplicate with “BARCODE_TAG=RX” option to get duplicate rate based on MIDs.
  - Use “GroupReadsByUmi” to group reads with the same MID.
  - Use “CallMolecularConsensusReads” to enrich for MID reads for variant calling.

Swift BIOSCIENCES™

THANK YOU

www.swiftbiosci.com

Swift BIOSCIENCES™