

Demultiplexing Results

Helena Klein

October 16, 2018

Note: I had to re-run my script on talapas after learning more about the index quality filtering today in class, so my output is from my last run (my script takes ~15 hours, so it's not quite done yet). I will update this once I get my final results.

output:

```
Percentage of index swapping: 0.12%
Total Undetermined Reads: 4205183
Total Low-quality Reads: 5021
Total Reads: 363246735
```

```
Percentage of reads per index
Index    Percent of Total Sequencing Run
```

CGGTAATC	1.394343159065146
GCTACTCT	2.041740856941219
TATGGCAC	3.0789826644966265
CTCTGGAT	9.628823504772864
TACCGGAT	21.022585929093072
GATCAAGG	1.81339551475941
ACGATCAG	2.186627499900309
ATCGTGGT	1.8961194516999582
ATCATGCG	2.7770388631297678
AACAGCGA	2.4424263579409735
GATCTTGC	1.0023688168869571
CTAGCTCA	4.771422377684964
TAGCCATG	2.926284526686799
GTCCTAAG	2.430930590470414
GTAGCGTA	2.235186780137198
TCGACAAG	1.060807883104579
TCGAGAGT	3.2323888609762728
CGATCGAT	1.5430189620286607
TCTTCGAC	11.58829741442824
TCGGATTC	1.269481472421218
CACTTCAC	1.1538680450906187
AGAGTCCA	3.1154526413017862
TGTTCCGT	4.331217732762278
AGGATAGC	2.387682851437054

I ran my python script on the test file using the following command:

```
cd /projects/bgmp/hklein3/BI624/Demultiplexing
```

```
module load python3/3.6.5
```

```
python3 demultiplexing.py -R1 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R1_001.fastq.gz \
-R2 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R4_001.fastq.gz \
-I1 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R2_001.fastq.gz \
```

```
-I2 /projects/bgmp/shared/2017_sequencing/1294_S1_L008_R3_001.fastq.gz \
-key indexes_key.txt -T 20 > demultiplexing_stats_20.txt
```

Examples of fastq outputs from the R1 and R2 files with index “AACAGCGA” and the undetermined files.

R1 File

```
$ head -12 R1_AACAGCGA.fq
```

@K00337:83:HJKJNBXX:8:1101:1479:1701 1:N:0:1:AACAGCGA

CCTTTGCTTAATCTGCTCTTGTCTTTTCAGCACTAAGCTGTTCTTTTCTTCTTTTCATGGCTGAAATTTTGTGTTTCAGTTCTCTAACCTGGCGCTCAATAT
+

AAAAAJJFJJJJJJJJJJJJFJJJJFJJJJJFJJJJFFJJJJJJJJJJJJFJJJJJJAFJJJJJJJJJJJJFF<JJJJJJJJJJFJJJJJFJJJ

@K00337:83:HJKJNBXX:8:1101:1966:1701 1:N:0:1:AACAGCGA

ATGTTTCAGTTGTCCTCTGGACTTACAAGTTCCTCAGTACACTGCTCCAGAGCCTCCCTCTCTTCATCCATGTCCTGACTGGTCTCCTTGCTGGCTTCTGAGC
+

[illegible]

@K00337:83:HJKJNBXX:8:1101:2656:1701 1:N:0:1:AACAGCGA

CCGGTGCAGGAACCAGCAATCAATGCGGGTGAGCTCGTAGAGGCGCTCCACAGAGTAGCCAGCCCACAGAGCGGCAGCCACCACAAAGATCCGCTTATCCG
+

A-AAA--<AF7.IAA.IFA7<IA<FA<FA.I.I.FF7-<7-7FA.IA7A<F-7-7F.IF.I.I-F<F-A7AFFFA<.IFA-<A<777A7AA.I7A-F.I77AA.IAAAF.AFFF

```
# Associated R2 file
```

```
$ head -12 R2_AACAGCGA.fq
```

@K00337:83:HJKJNBXX:8:1101:1479:1701 4:N:0:1:TCGCTGTT

CAACGAGACGCGCGCTAAGCTCGACGAGCTTTCTGCTAAGCGAGAAACGAGTGAGAGAAATCCAGACAACCTAAGGGATGCCAGCAGGATGCAAGAGACA
+

[illegible]

@K00337:83:HJKJNBXX:8:1101:1966:1701 4:N:0:1:TCGCTGTT

CTTGTGAGAACTGAACCACGAGAGCAGCTCCTTGATCAGCTGCAAAAGAAACAACCACCAATGATGCTAAACAGCTCAGAAGCCAGCAAGGAGACCAGTCA
+

AAFFFIILIA.IIIIIIIIIIIIIIIIIIIIIIIIE.IIE.II.IE.II

@K00337:83:HJKJNBXX:8:1101:2656:1701 4:N:0:1:TCGCTGTT

GCTGGAGACTCCGACGGATAAGCGGATCTTTGTGGTGGCTGCCGCTCTGTGGGCTGGCTACTCTGTGGAGCGCCTCTACGAGCTACCCGCATTGATTGCT
+

```
<AF<FAIFIIIIIIIFIIIIIIIA<AIIIIIIIIIIIIIFIIIIIIIIIIIIIFIIIIIIIFIIIFIIII<IFEEIIIF
```

A

Undetermined File R1

```
$ head -12 R1_undetermined.fq
```

@K00337:83:HJKJNBXX:8:1101:1265:1191 1:N:0:1:NCTTCGAC

GNCTGGCATTCCCAGAGACATCAGTACCCAGTTGGTTCAGACAGTTCCTCTATTGGTTGACAAGGTCTTCATTTCTAGTGATATCAACACGGTGTCTACAA
+

A#A-<F.I.I.I.<J.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.F.I.I.I.I.F.F.I.I.F.I.I.I.A.I.I.I.I.-A.I.I.I.I.I.I.I.F.F.I.I.I.I.I.I.F.F.A-7<A.I.I.I.F.F.A.I.I.I.I.I.F<F--.I.I.I.I.I.I.F.-A-F7.I.I.I.I

@K00337:83:HJKJNBXX:8:1101:1286:1191 1:N:0:1:NACAGCGA

CNACCTGTCCCCAGCTCACAGGACAGCACACCAAAGGCGGCAACCCACACCCAGTTTTACAGCCACACAGTGCCTTGTTTTACTTGAGGACCCCCCACTCC

A#AAEIIIIIIIEIIIIIIIIIIIIIIIIIIIIIEIIIIIIIIIIIIAIIIIIIIIIIIEIIIIIEEEEIIIIIIIIIIIIIIIIII77E

@K00337:83:HHKINBBYX:8:1101:1347:1191 1:N:0:1:NTCCTAAG

GNGGTCTTCTACCTTTCTCTCTTTTTGGAGGAGTAGAATGTTGAGAGTCAGCAGTAGCCTCATCATCACTAGATGGCATTCTTCTGAGCAAAACAGGT

A#AFFFIEIEIIIEIIIIIIAIEIIIEIEI7EAEIEIEIEIEIEIAAIIIEIIIIIIIIIIIIIIAIIIEAIIIIIEIIAIIIE-

A#A1 1 1 31 331 33331 33333333A331 333331 31 31 / 1 A1 331 31 331 31 3331 3AA3331 3333333333333333A3331 A333331 1 333A333 / 1

Undetermined File R2

```
$ head -12 R2_undetermined.fq
```

@K00337:83:HJKJNBXX:8:1101:1265:1191 4:N:0:1:NTCGAAGA

NTTTTGATTACCTTTCAGCCAATGAGAAGGCCGTTTCATGCAGACTTTTTTAATGATTTTGAAGACCTTTTGTATGATGATGATGTCCAGTGAGGCCTCCC

[illegible]