HW10

Tianqi Wu

4/20/2020

Problem 1

Since we have 42 barcodes, we have 42 subjects. If we are matching at the beginning of sequences with default arguments, the smallest and largest number of reads that you see belonging to any subject after demultiplexing is 0 and 2. There are 995 reads unmatched.

```
## cat hw_rna-seq.txt | bin/fastx_barcode_splitter.pl --bcfile barcodes.txt --bol --mismatches 2 --pref
```

Problem 2

Most of the reads are unmatched and it indicates there is a problem. After examine the FASTQ file, we find that all the read start with N and we may want to remove it. After trimming, the smallest and largest number of reads that you see belonging to any subject after demultiplexing is 3 and 58. There are 16 reads unmatched.

```
## fastx_trimmer -i hw_rna-seq.txt -o trimmed_rna.txt -f 2

## cat trimmed_rna.txt | bin/fastx_barcode_splitter.pl --bcfile barcodes.txt --bol --prefix tmp/bla_ --
```