

fastQ

Lecture 9
Sept 19, 2016

ANNOUNCEMENTS

- Codes??
- No class on Wednesday
- Practice launching AWS instance!!

fastQ

```
@HSQ-7001360:67:H88RHADXX:1:1101:1448:2158 1:N:0:CAGATC
ATCTATCTGAGACTGATACGCCTTCGGCTTAATTTATACAAG
+
BBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```

Quality Scores https://en.wikipedia.org/wiki/FASTQ_format

BBBFFFFFFFFFIIFFFFFFFFFFF

```

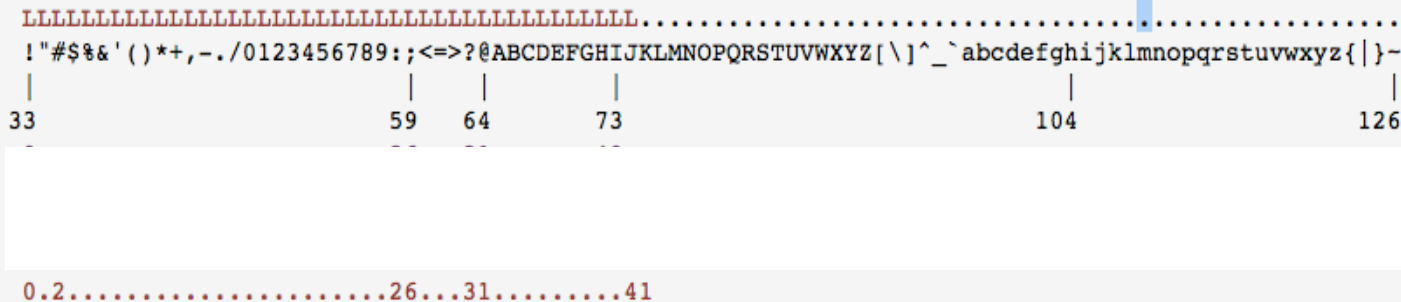
SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....
.....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
.....IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII..
.....JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ...
LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....
!"#$%&'()*+,-./0123456789;<=>?@ABCDEFGHIJKLMN
OPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
|                                     |
33                               59   64   73               104           126
0.....26...31.....40
          -5...0.....9.....40
              0.....9.....40
                  3.....9.....40
0.2.....26...31.....41

```

S - Sanger Phred+33, raw reads typically (0, 40)
X - Solexa Solexa+64, raw reads typically (-5, 40)
I - Illumina 1.3+ Phred+64, raw reads typically (0, 40)
J - Illumina 1.5+ Phred+64, raw reads typically (3, 40)
with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)
(Note: See discussion above).
L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)

Quality Scores

BBBFFFFFFFFFFFFF|||||||||||||||||||||||||||||||||||||||||FFFFFFFFFFFFFFFFFFFF

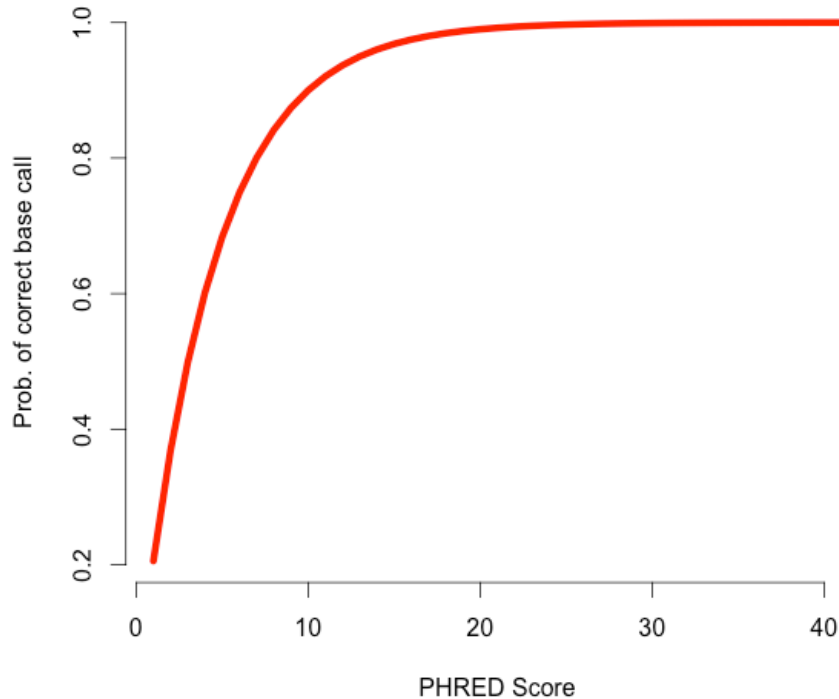


L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)

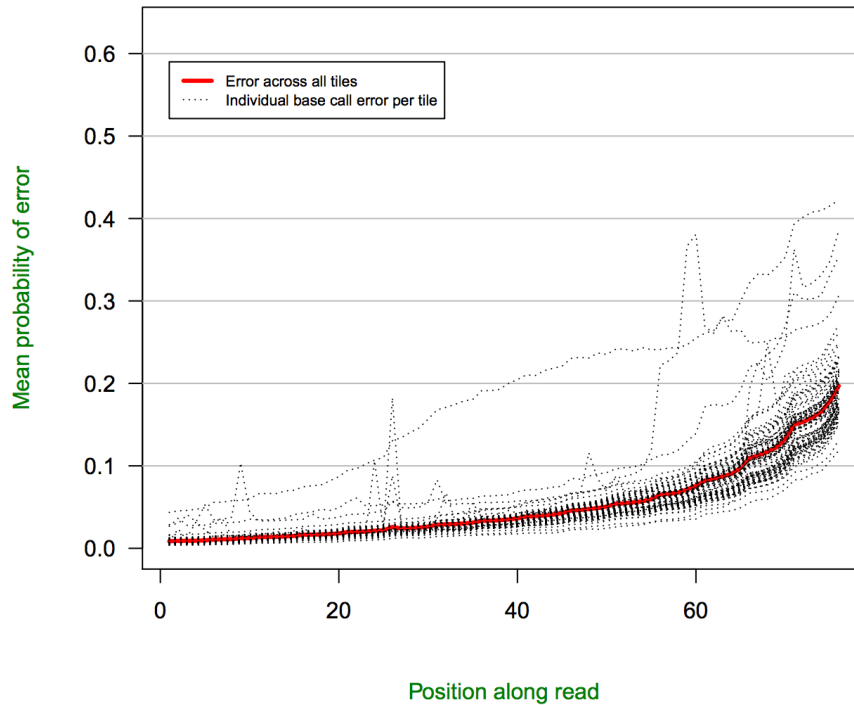
Quality Scores

$$p_{correct} = 1 - [10^{-\left(\frac{Q}{10}\right)}]$$

Q=Phred score

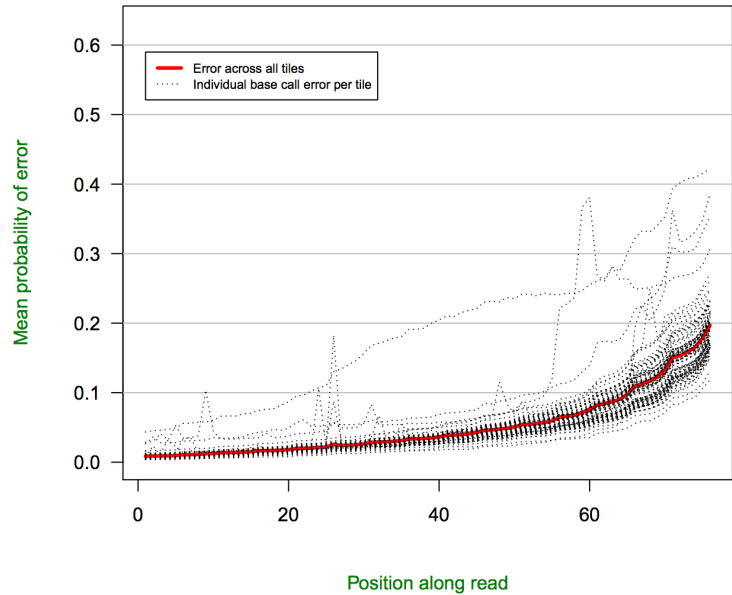


Quality Scores

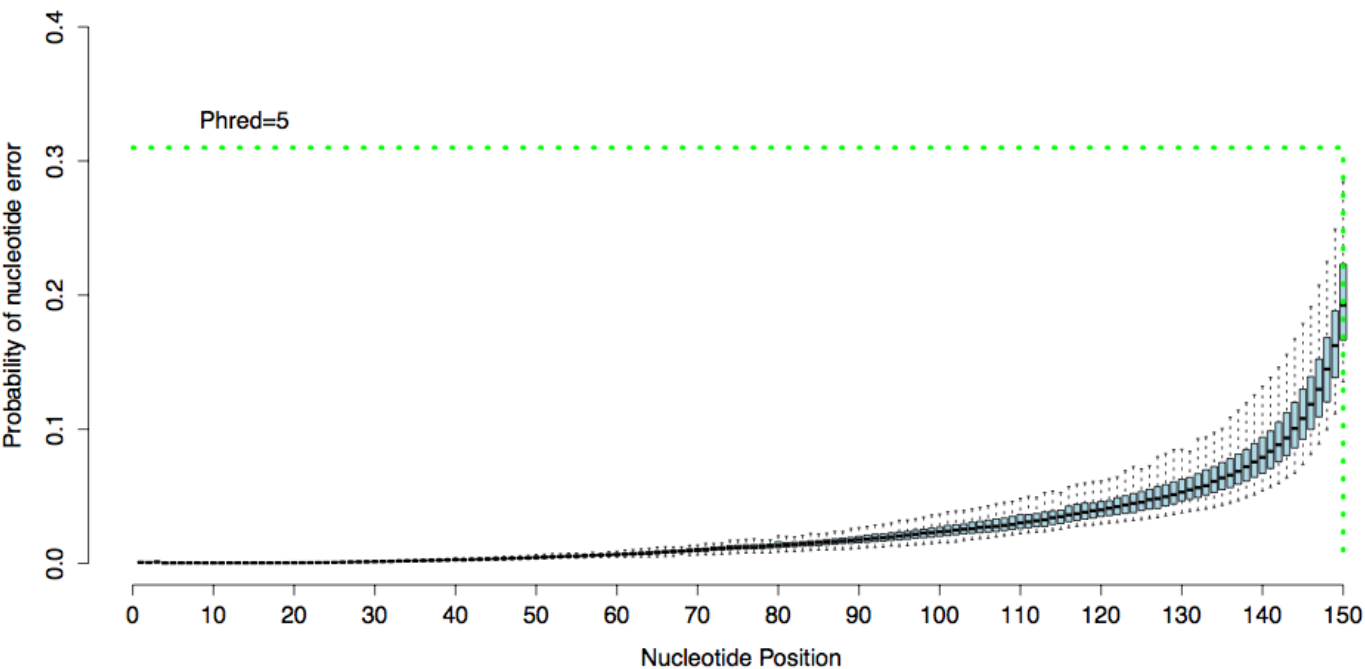


Quality Scores

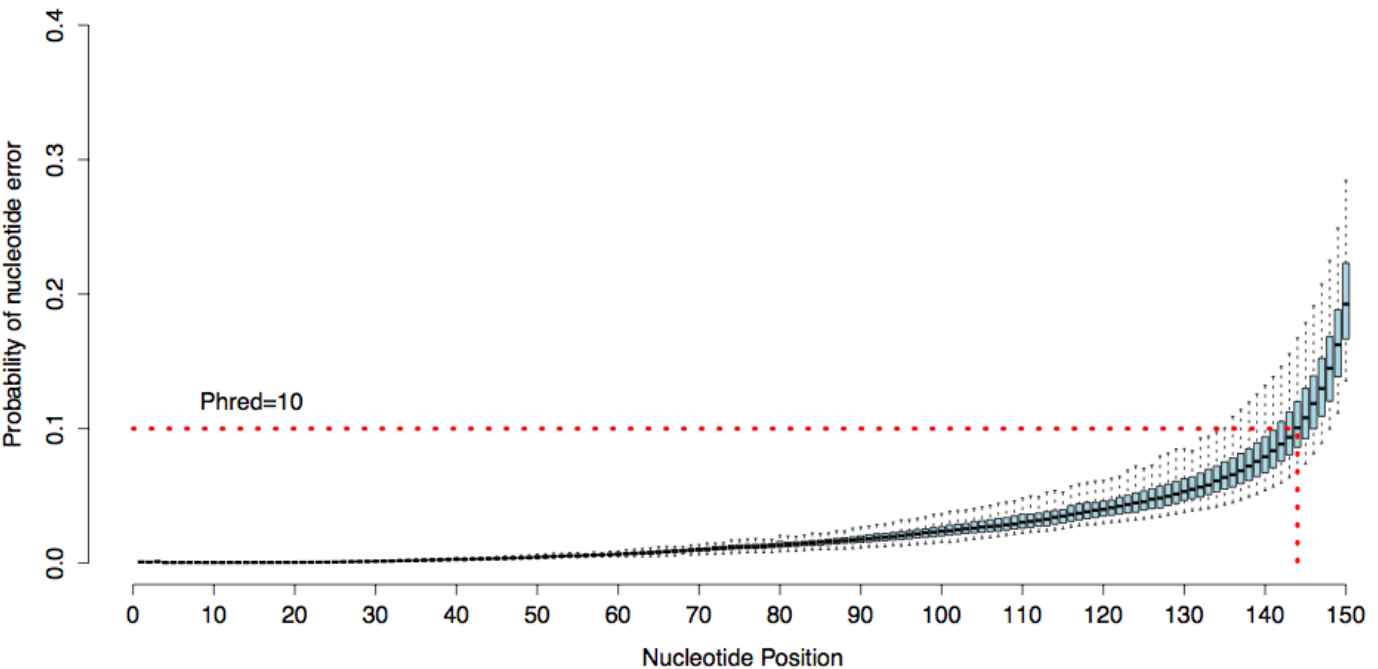
What should we do about this?



What to do? Trimming?



What to do? Trimming?



What to do? Trimming?

