RSeQC

Background: RSeQC is a tool that can be used to generate QC reports for RNA-seq. For more information, please check: Tool Homepage

Objectives: In this section, we will try to generate a QC report for a data set downloaded from RSeQC website.

Files needed:

- Aligned bam file.
- Index file for the aligned bam.
- A RefSeq bed file.

Copy RSeQC Data

set your working directory and copy the necessary files

```
mkdir -p ~/workspace/rnaseq/
cp -r ~/CourseData/RNA_data/RSeQC/RSeQC.zip ~/workspace/rnaseq/
cd ~/workspace/rnaseq/
```

Unzip the RSeQC file

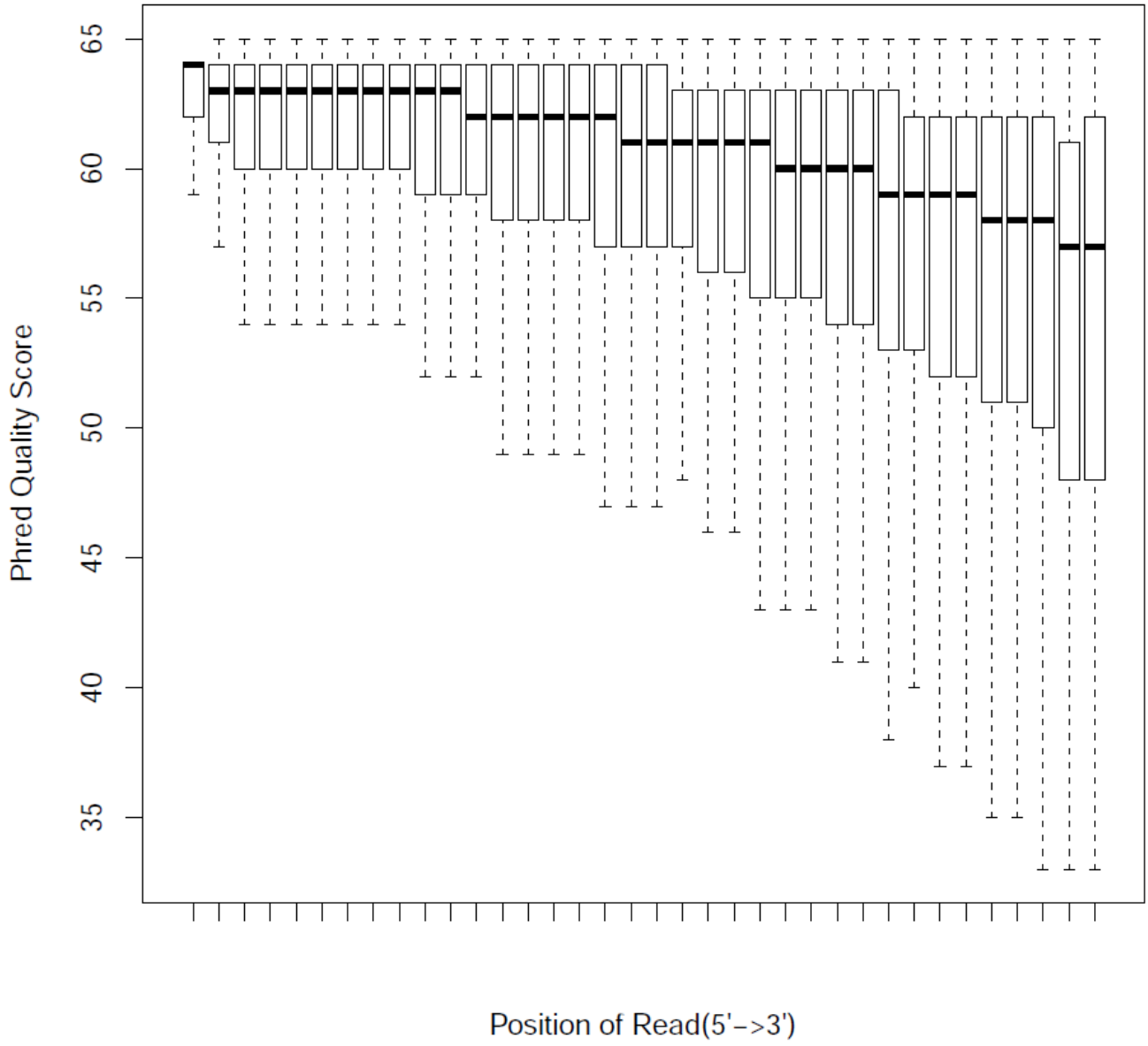
```
unzip RSeQC.zip
cd RSeQC/
```

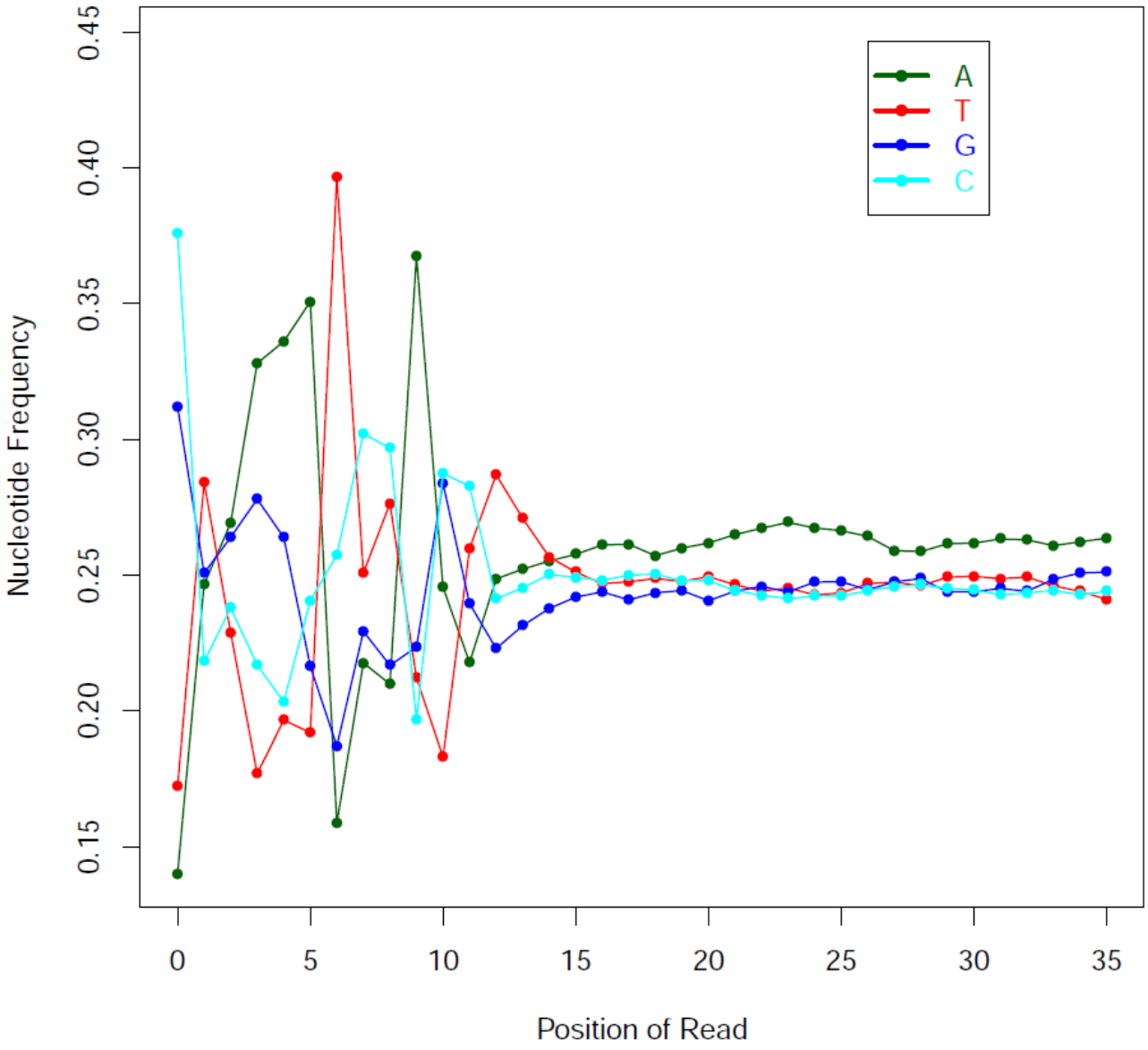
Note: You should now see the bam, index, and RefSeq bed files listed

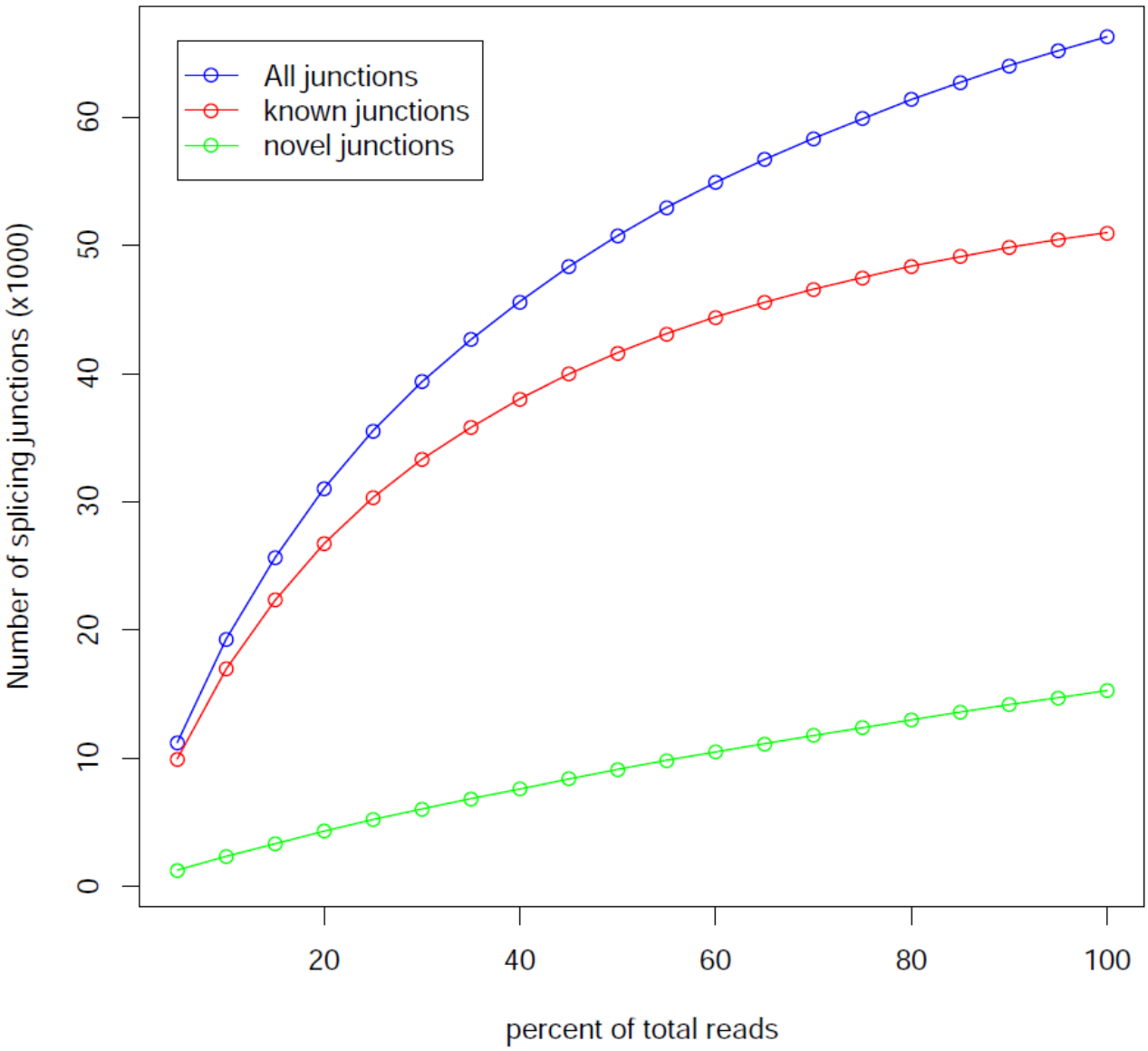
Run RSeQC commands:

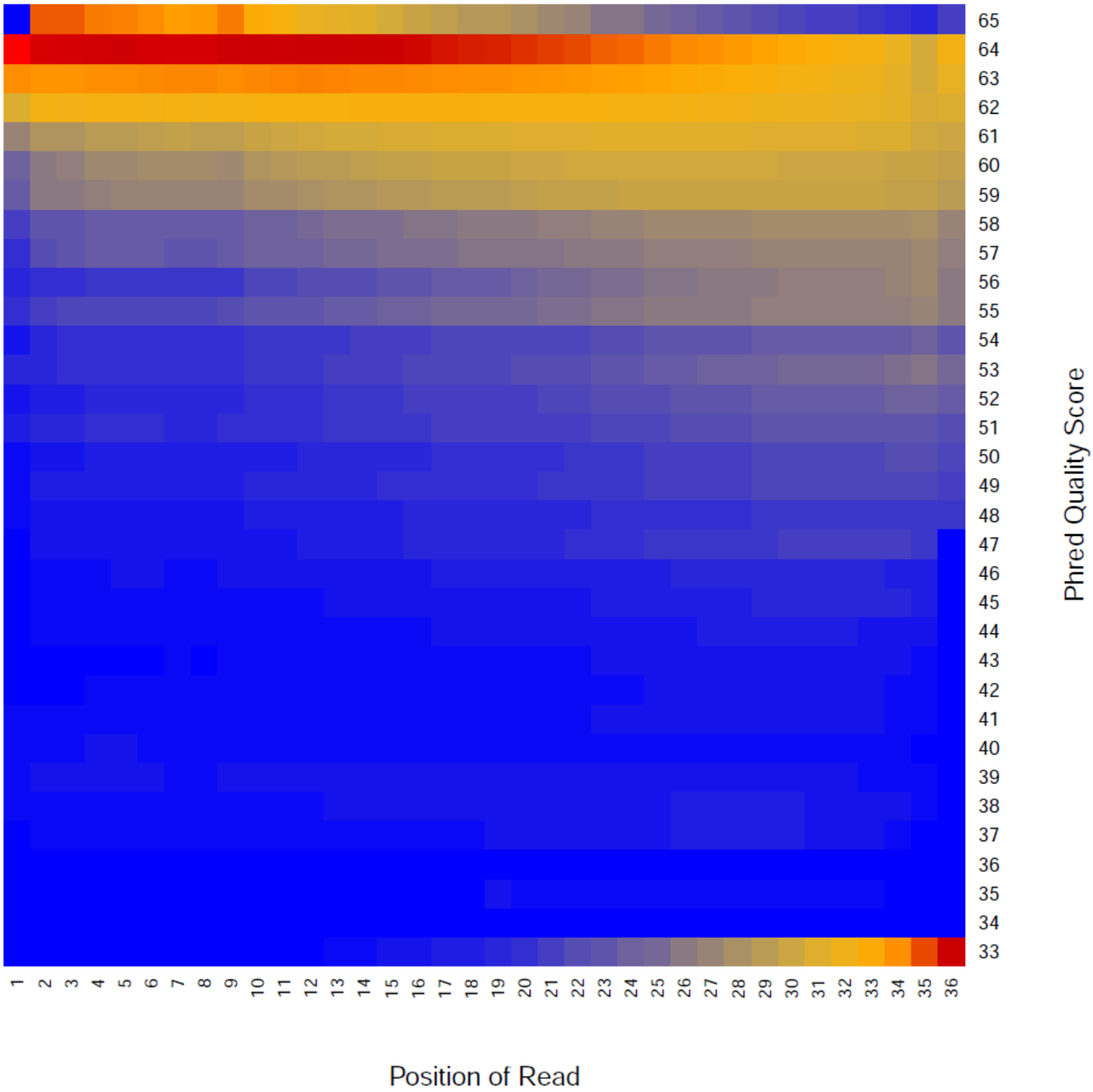
```
 bam_stat.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam
 clipping_profile.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 geneBody_coverage.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 infer_experiment.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam
 inner_distance.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 junction_annotation.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 junction_saturation.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 read_distribution.py -r hg19_RefSeq.bed -i
Pairend_nonStrandSpecific_36mer_Human_hg19.bam
 read_duplication.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 read_GC.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
 read_NVC.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
```

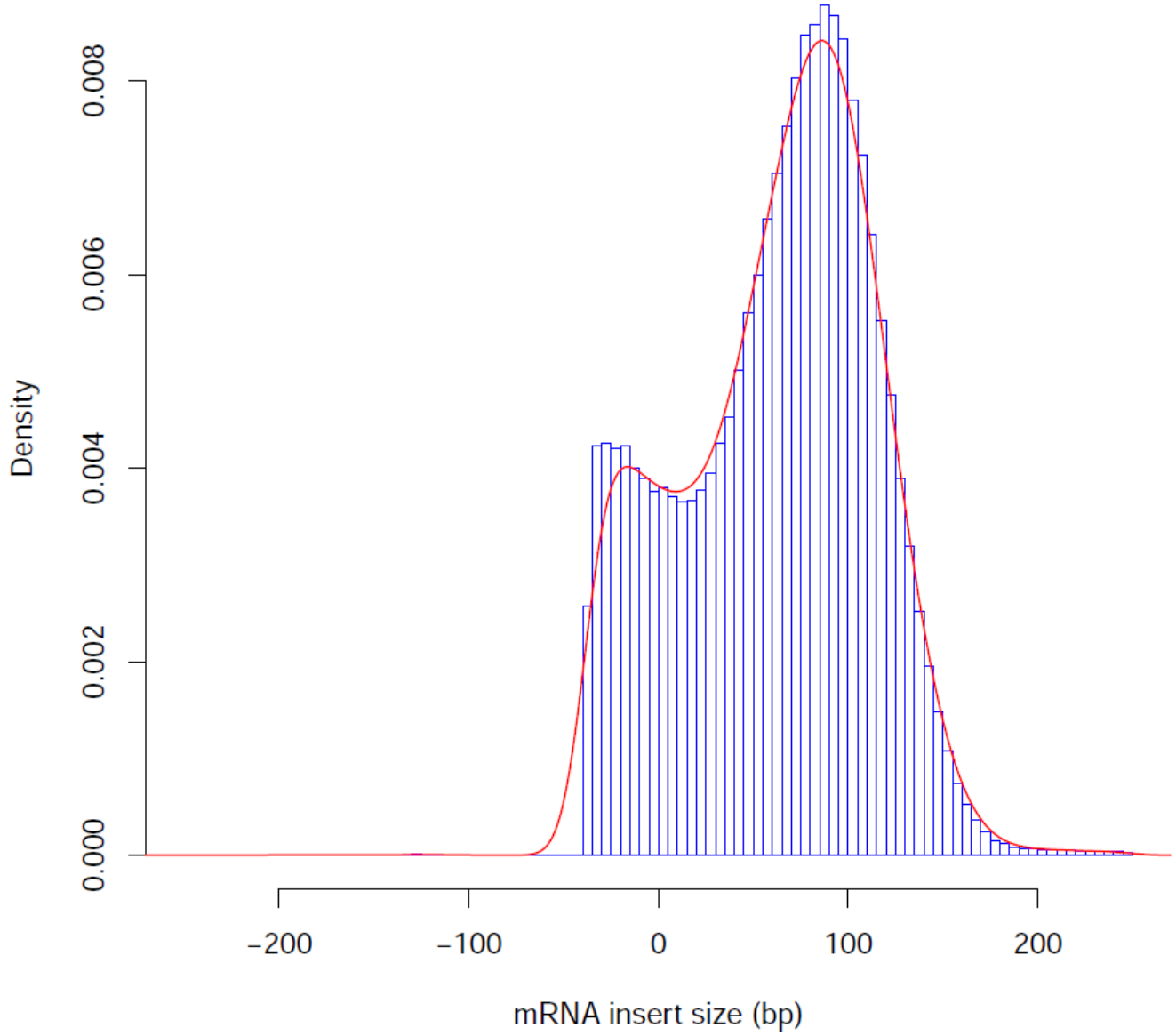
```
read_quality.py -i Pairend_nonStrandSpecific_36mer_Human_hg19.bam -o tutorial
```

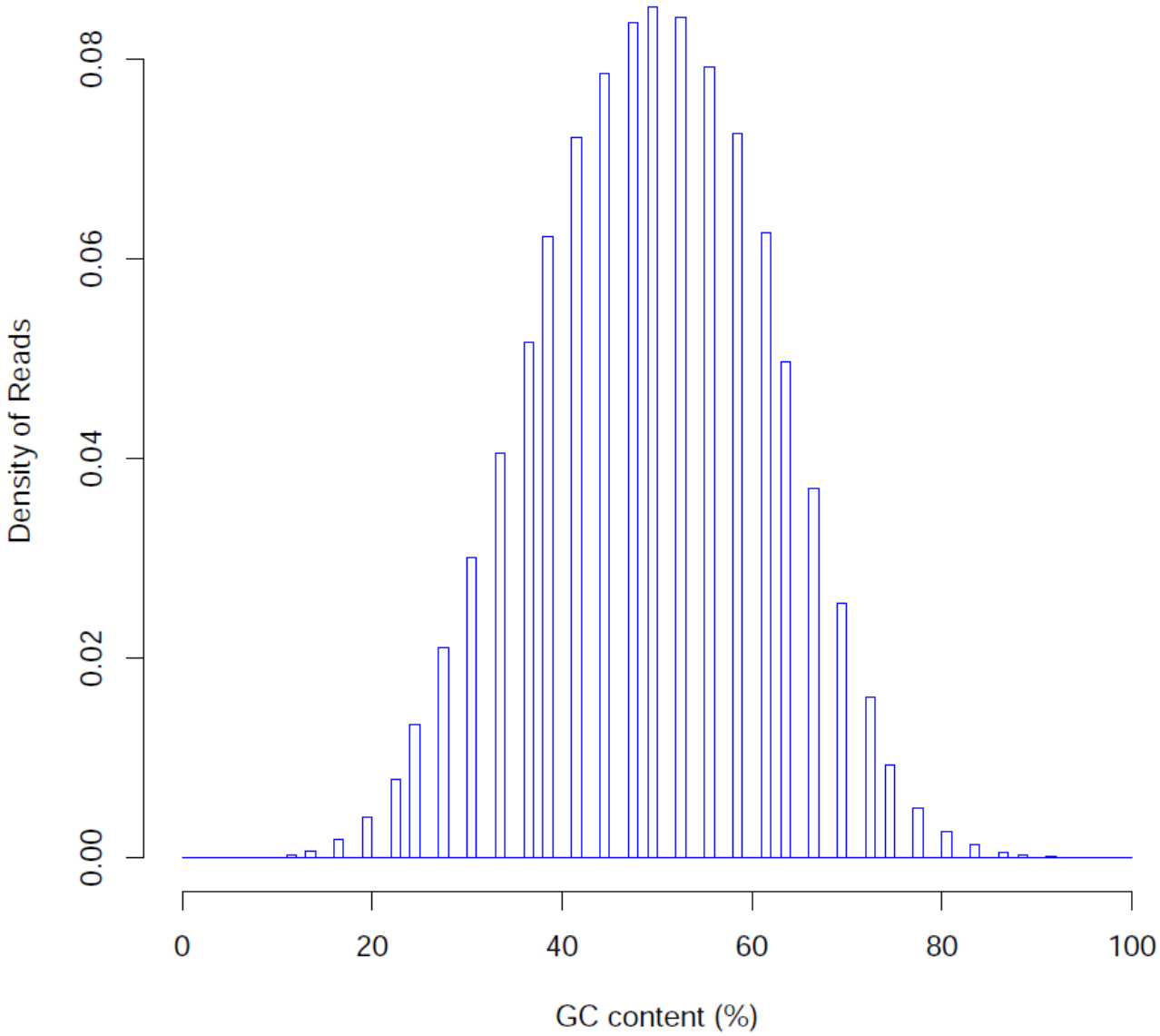


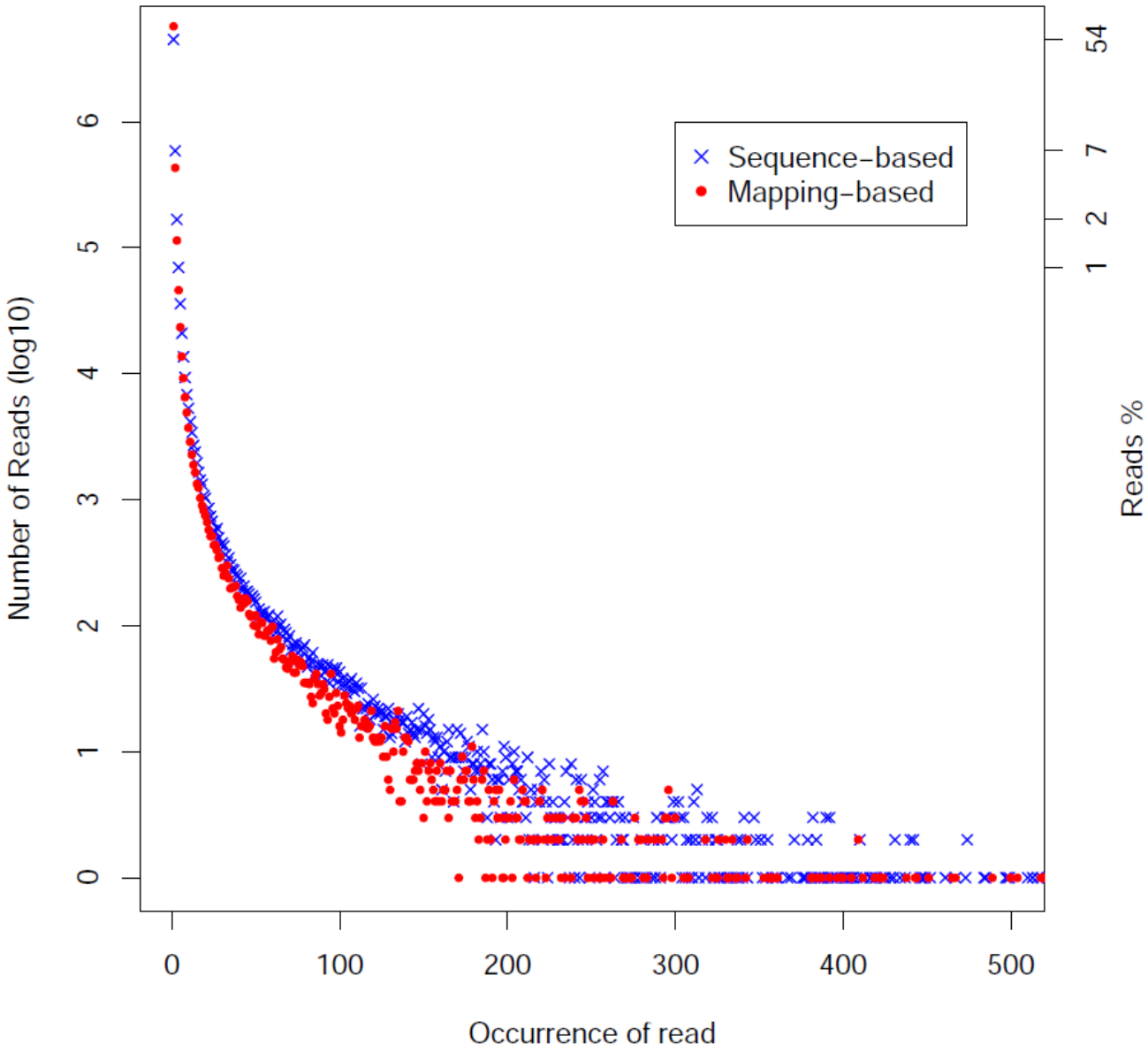

A
T
G
C
Nucleotide Frequency
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0
5
10
15
20
25
30
35
Position of Read


All junctions
known junctions
novel junctions
Number of splicing junctions (x1000)
0
10
20
30
40
50
60
20
40
60
80
100
percent of total reads


65
64
63
62
61
60
59
58
57
56
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
Phred Quality Score
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
Position of Read


Mean=61.7058439571647;SD=50.3152604799405
Density
0.000
0.002
0.004
0.006
0.008
−200
−100
0
100
200
mRNA insert size (bp)


Density of Reads
0.00
0.02
0.04
0.06
0.08
0
20
40
60
80
100
GC content (%)


Sequence-based
Mapping-based
Number of Reads (log10)
Reads %
Occurrence of read
0
1
2
3
4
5
6
1
2
7
54
0
100
200
300
400
500