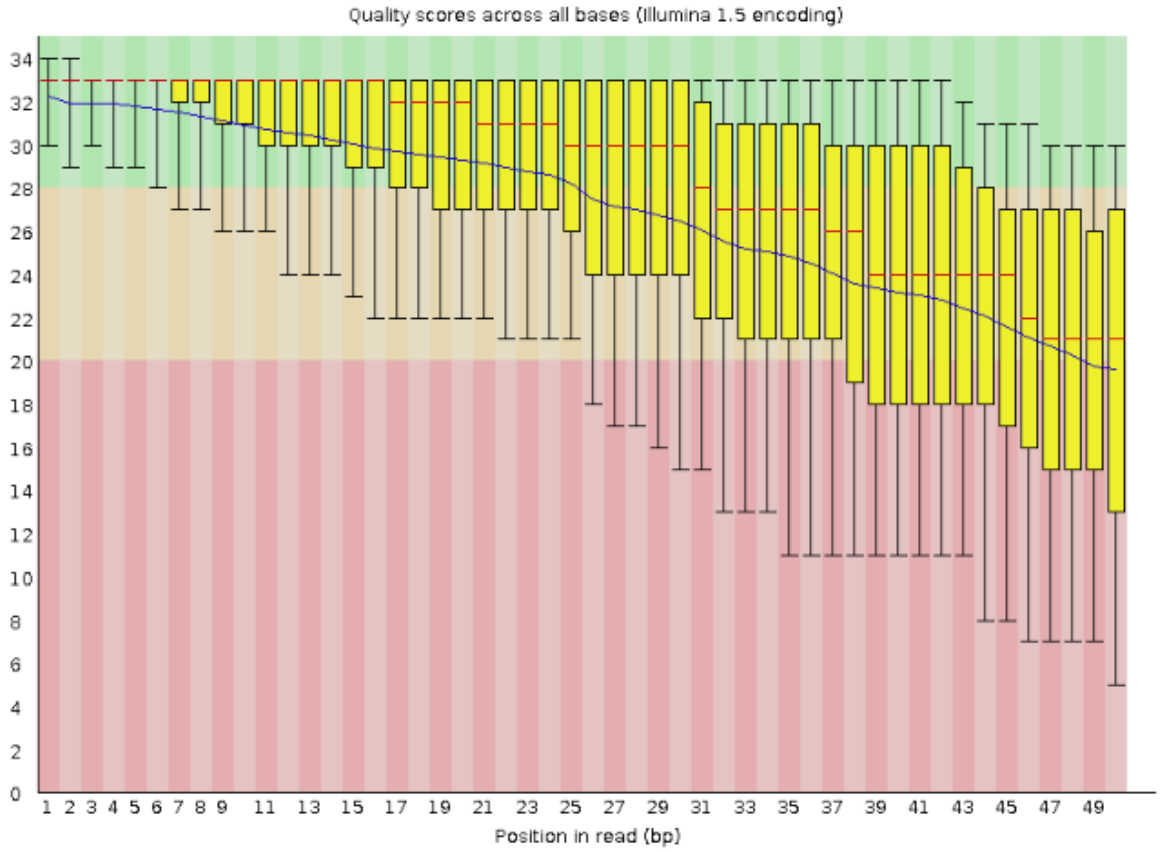
Khan Inan
Transcriptomics
Homework 3 Galaxy

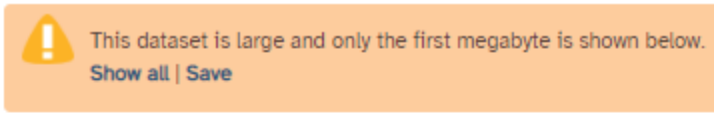
Basic Statistics

Measure	Value
Filename	NO3_2_fastq
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	5182417
Sequences flagged as poor quality	0
Sequence length	50
%GC	53

Per base sequence quality



This are some of the results I got after using the NO3_2.fastq file in FastQC tool in Galaxy. The best way I can interpret these results is that the quality of the sequence goes down the further along the sequence that we go. As we can see the blue line, which is essentially the mean line, it is trending downwards as the position in the read goes higher along the x-axis.




```
@HANNIBAL_4_FC308YYAAXX:8:1:5:840
ACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACCTGTAGG
+HANNIBAL_4_FC308YYAAXX:8:1:5:840
aaaaabbabaaaaaabaaaa\a[S^^VZZaaaaaaaaaUUU^_]_aaXS[
@HANNIBAL_4_FC308YYAAXX:8:1:5:1140
AGGTCACGGCGAGACGAGCCGTTTATCATTAC
+HANNIBAL_4_FC308YYAAXX:8:1:5:1140
aaa\aaa^V^[^[^^S[aaaaaaaaaaaaa][
@HANNIBAL_4_FC308YYAAXX:8:1:5:1083
ACCCGCTGAGTTTAAGCATATCAATAAGCGGAGGAAAAGAAACCTGTAGG
+HANNIBAL_4_FC308YYAAXX:8:1:5:1083
aabaaaaaaaaaaa[RXaaaaaaaaaZO[_`^[[^^[^X[[`ZU[^^XMU
@HANNIBAL_4_FC308YYAAXX:8:1:5:342
AACAAGTACCGCGAGGTAAAGATGAAAAGGACTTTGAAAAGAGAGTCAAA
+HANNIBAL_4_FC308YYAAXX:8:1:5:342
aaaaaaaaa_Z[a^V[^aa^[_`]_aa`]__aaaaaa_^^^^V[_[\Z^[
```

This was my result after putting the file into trimmomatic

QNAME
@HD VN:1.0 SO:coordinate
@SQ SN:Chr1 LN:30427671
@SQ SN:Chr2 LN:19698289
@SQ SN:Chr3 LN:23459830
@SQ SN:Chr4 LN:18585056
@SQ SN:Chr5 LN:26975502
@SQ SN:chloroplast LN:154478
@SQ SN:mitochondria LN:366924
@PG ID:hisat2 PN:hisat2 VN:2.2.1 CL:"/usr/local/bin/hisat2-align-s --wrapper basic-0 -p 8 -x genome --read-lengths 50,49,43,42,48,38,47,40,45,36,41,46,35,37,31,34,33,44,32,39,30,28,29,27,26,25,24,23,21,22,20,14,19,16,18,15,17,12,13,11,9,8,7,10,6,4,5,3,2 -U input_f.fastq"

This was my results after using HISAT2 to align my trimmomatic results with the Arabidopsis FASTA file

Category
__no_feature
__ambiguous
__too_low_aQual
__not_aligned
__alignment_not_unique

Afterwards I used HTSEQ-COUNT to create the readcount files and I arrived with a no feature file (shown above) and a regular file. I used the output from running the HISAT2 tool and used it with the Arabidopsis GTF file

GeneId	HISAT2 on data 26 and data 27: aligned reads (BAM)
AT1G01010	
AT1G01020	
AT1G01030	
AT1G01040	
AT1G01046	
AT1G01050	

This is my regular HTSEQ-COUNT file.


31 : htseq-count on data 28 and data 29 (no feature)
30 : htseq-count on data 28 and data 29
29 : HISAT2 on data 26 and data 27: aligned reads (BAM)
28 : Arabidopsis.gtf
27 : Arabidopsis.fa
26 : Trimmomatic on NO3_2.fastq
25 : FastQC on data 21: RawData
24 : FastQC on data 21: Webpage
23 : FastQC on data 20: RawData
22 : FastQC on data 20: Webpage
21 : NO3_2.fastq
20 : NO3_1.fastq
19 : KCL_2.fastq
18 : KCL_1.fastq