

The second script summarizes data from module 1 that may be of interest at the moment of reporting parameters of data quality in papers, parameters like average sequence quality

It is located at

/bonn_data/Bonn_0_fastq/fastqcRawdata/summarizing_fastqc_1module_2data_mine-bis.sh

#It can be run by typing:

```
[vifehe@detritus fastqcRawdata]$ ./summarizing_fastqc_1module_2data_mine-bis.sh
```

[summarizing_fastqc_1module_2data_mine-bis.sh](#)

```
#!/bin/bash

##FastQC 0.10.1
#>>Basic Statistics pass
#Measure Value
#Filename SN7640211_14074_P1A01_MND1014_1_sequence.fq.gz
#File type Conventional base calls
#Encoding Sanger / Illumina 1.9
#Total Sequences 44012752
#Filtered Sequences 0
#Sequence length 101
#%GC 49
#>>END_MODULE
#>>Per base sequence quality pass
#Base Mean Median Lower Quartile Upper Quartile
10th Percentile 90th Percentile
#1 31.64506284451379 33.0 31.0 34.0 28.0 34.0
#2 31.880190722906853 34.0 31.0 34.0 28.0 34.0
#3 31.972653289210363 34.0 31.0 34.0 28.0 34.0
#4 35.39369340049448 37.0 35.0 37.0 32.0 37.0
#5 35.09201710449735 37.0 35.0 37.0 32.0 37.0
#6 35.08697933726116 37.0 35.0 37.0 32.0 37.0
#7 35.06162818448617 37.0 35.0 37.0 32.0 37.0
#8 35.052662714660514 37.0 35.0 37.0 32.0 37.0
#9 36.63084601026539 39.0 37.0 39.0 32.0 39.0
#10-14 36.86299192561283 39.2 37.2 39.4 32.0 39.4
#15-19 37.82707197677619 40.0 38.0 41.0 32.0 41.0
#20-24 37.68135608516368 40.0 38.0 41.0 31.0 41.0
#25-29 37.417561642134984 40.0 37.0 41.0 30.4 41.0
#30-34 37.13463264010394 40.0 36.8 41.0 30.0 41.0
#35-39 36.80116996546819 40.0 36.0 41.0 29.4 41.0
#40-44 36.6426807985104 40.0 35.8 41.0 28.8 41.0
#45-49 36.66400877636554 40.0 35.2 41.0 28.6 41.0
#50-54 36.30974728415073 39.2 35.0 41.0 28.0 41.0
#55-59 35.784558657000126 38.6 34.6 41.0 27.2 41.0
#60-64 35.131667385852175 37.2 34.0 40.0 26.0 41.0
#65-69 34.42042487140999 35.8 33.2 39.4 26.0 41.0
```

```
#70-74 33.645179519790084 35.0 33.0 38.0 25.4 40.4
#75-79 32.74637153341378 35.0 31.8 36.2 24.2 39.0
#80-84 32.329568739532576 35.0 32.0 35.6 24.0 37.2
#85-89 31.784759544233907 35.0 31.2 35.0 23.2 36.2
#90-94 31.437542837584893 34.8 31.0 35.0 23.0 35.6
#95-99 31.11940442170033 34.0 31.0 35.0 20.6 35.0
#100-101 30.526057039105396 34.0 29.5 35.0 19.0 35.0
```

```
idir=fastqcRawdata_P5
```

```
outfile=${idir}_modules_summary.stats
```

```
touch $outfile
```

```
printf "Sample\taverage_mean\taverage_median\n" > $outfile
```

```
for x in $idir/*.fq_fastqc/fastqc_data.txt
```

```
do
```

```
# echo $x
```

```
sample=(`echo $x | awk -F "/" {'print $2'} | awk -F "_" {'print $4'-'$5'}`) #this should output L5150-1
```

```
# echo $sample
```

```
basic_stats=${sample}_basic_stats.txt
```

```
touch $basic_stats
```

```
sed -n '14,41p' $x > $basic_stats
```

```
av_mean=0
```

```
av_median=0
```

```
count=0
```

```
while IFS=$"\t" read -r; do
```

```
mean=(`cat $basic_stats | awk -F"\t" {'print $2'}`)
```

```
median=(`cat $basic_stats | awk -F"\t" {'print $3'}`)
```

```
av_mean=`echo $av_mean+$mean | bc -l`
```

```
av_median=`echo $av_median+$median | bc -l`
```

```
# mean=$(echo "scale=2;${mean}+${line[0]}" | bc);
```

```
((count ++))
```

```
# echo "${line[1]}"
```

```
# median=$(echo "scale=2;${median}+${line[1]}" | bc);
```

```
# echo "$count"
```

```
done < $basic_stats
```

```
final_mean=`echo $av_mean/$count | bc -l`
```

```
# printf "%s\t%s" "$sample" "scale=2;${final_mean}";
```

```
# printf "final_mean="; echo "scale=2; $av_mean/$count" | bc;
```

```
final_median=`echo $av_median/$count | bc -l`
```

```
# printf "final_median="; echo "scale=2; $av_median/$count" | bc;
```

```
# printf "$sample\t"; echo "scale=2; $av_mean/$count" | bc; echo
```

```
"scale=2; $av_median/$count" | bc;
```

```
printf "%s\t%5.2f\t%5.2f\n" $sample $final_mean $final_median >>
```

```
$outfile
```

```
rm $basic_stats;
```

```
done
```

#and the output looks like this:

```
[vifehe@detritus fastqcRawdata]$ more fastqcRawdata_P1_modules_summary.stats
Sample average_mean average_median
MND1014-2 31.28 33.00
MND116-1 31.54 33.00
MND116-2 31.14 33.00
MND126-1 31.62 33.00
MND126-2 31.23 33.00
MND1405-1 31.62 33.00
MND1405-2 31.26 33.00
MND1493-1 31.56 33.00
MND1493-2 31.19 33.00
MND1725-1 31.58 33.00
MND1725-2 31.23 33.00
MND334-1 31.58 33.00
MND334-2 31.19 33.00
MND413-1 31.60 33.00
MND413-2 31.21 33.00
MND1021-1 31.45 33.00
MND1021-2 31.03 33.00
MND1165-1 31.54 33.00
MND1165-2 31.10 33.00
MND1263-1 31.58 33.00
MND1263-2 31.16 33.00
MND1408-1 31.59 33.00
MND1408-2 31.16 33.00
```

From:

<https://cortafuegos.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:

https://cortafuegos.fundacioace.com/wiki/doku.php?id=genetica:bioinf_process:fastqc:script2

Last update: **2020/08/04 10:58**

