

Day 9 Worksheet/Homework | QC multiple samples

Author: Rutendo F. Sigauke

1. Log into the AWS
2. Create a working directory for day9 in scratch

- a. `mkdir -p /scratch/Users/YourUsername/day9`
- b. `cd /scratch/Users/YourUsername/day9`
- c. `mkdir scripts qc fastq bam sam`
- d. `cd qc`
- e. `mkdir fastqc hisat_mapstats multiqc`

```
[rutendos@ip-172-31-38-192 ~]$ mkdir -p /scratch/Users/rutendos/day9
[rutendos@ip-172-31-38-192 ~]$ cd /scratch/Users/rutendos/day9/
[rutendos@ip-172-31-38-192 day9]$ mkdir scripts qc fastq bam sam
[rutendos@ip-172-31-38-192 day9]$ ls
bam fastq qc sam scripts
[rutendos@ip-172-31-38-192 day9]$ cd qc
[rutendos@ip-172-31-38-192 qc]$ mkdir fastqc hisat_mapstats
[rutendos@ip-172-31-38-192 qc]$ ls
fastqc hisat_mapstats
```

3. Moving scripts to scratch

- a. `cp /scratch/Shares/public/sread2021/scripts/day9/d9_fastqc.sbatch /scratch/Users/<YourUsername>/day9/scripts`
- b. `cp /scratch/Shares/public/sread2021/scripts/day9/d9_hisat2.sbatch /scratch/Users/<YourUsername>/day9/scripts`
- c. `cp /scratch/Shares/public/sread2021/scripts/day9/runmultiqc.sbatch /scratch/Users/<YourUsername>/day9/scripts`
 - i. Also add paths to qc folder
- d. `cp /scratch/Shares/public/sread2021/scripts/day9/d9_preseq.sbatch /scratch/Users/<YourUsername>/day9/scripts`
- e. `cp /scratch/Shares/public/sread2021/data_files/day9/qc_fastq/* /scratch/Users/<YourUsername>/day9/fastq`

```
[rutendos@ip-172-31-38-192 qc]$ cp /scratch/Shares/public/sread2021/scripts/day9/d9_fastqc.sbatch /scratch/Users/rutendos/day9/scripts
[rutendos@ip-172-31-38-192 qc]$ cp /scratch/Shares/public/sread2021/scripts/day9/d9_hisat2.sbatch /scratch/Users/rutendos/day9/scripts
[rutendos@ip-172-31-38-192 qc]$ cp /scratch/Shares/public/sread2021/scripts/day9/d9_multiqc.sbatch /scratch/Users/rutendos/day9/scripts
[rutendos@ip-172-31-38-192 qc]$ cp /scratch/Shares/public/sread2021/scripts/day9/d9_preseq.sbatch /scratch/Users/rutendos/day9/scripts
[rutendos@ip-172-31-38-192 qc]$ cp /scratch/Shares/public/sread2021/data_files/day9/qc_fastq/* /scratch/Users/rutendos/day9/fastq
```

4. Edit and run scripts fastqc and hisat2 scripts

- a. `cd /scratch/Users/<YourUsername>/day9/scripts/`
- b. Remove `#SBATCH --partition=short` from ALL scripts
- c. Replace path for indices for HISAT2 with
`/scratch/Shares/public/sread2021/data_files/genome/hg38/HISAT2/genome`

5. Run multiqc: (add this line to script)

- a. Move html file to local computer
`scp <username>@18.219.252.252:/scratch/Users/<username>/day9/scripts/multiqc_report.html local_computer`

6. You can now view the multiQC HTML report

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2021-07-28, 22:14 based on data in: /scratch/Users/rutendos/day9/qc

Welcome! Not sure where to start? [Watch a tutorial video](#) (6:06) [don't show again](#) ✕

General Statistics

[Copy table](#) [Configure Columns](#) [Plot](#) Showing $\frac{4}{4}$ rows and $\frac{4}{6}$ columns.

Sample Name	% Aligned	% Dups	% GC	M Seqs
chr22EliGABPAchip	96.3%	4.9%	46%	0.3
chr22ElizabethGABPAchip	95.6%	7.6%	47%	0.2
chr22EricGABPAchip	95.6%	3.7%	46%	0.3
chr22EthanGABPAchip	95.6%	4.6%	46%	0.2

- General Stats
- HISAT2
- FastQC
- Sequence Counts
- Sequence Quality Histograms
- Per Sequence Quality Scores
- Per Base Sequence Content
- Per Sequence GC Content
- Per Base N Content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Status Checks

Toolbox

- [Home](#)
- [Refresh](#)
- [Download](#)
- [Print](#)
- [Help](#)