1-Counting from raw data

BAI Qiang*

2021-09-22 13:05:12 +0200

Contents

*University Liege, mail qiang.bai@uliege.be

1 Description

Total RNA was extracted and concentrated from the different samples with the RNA Clean & Concentrator 5 (Zymo Research). Possible DNA contaminant were removed with DNAse I. RNA quality and quantity were evaluated using a 2100 bioanalyzer (Agilent) and the Quant-iT™ RiboGreen™ RNA Assay Kit (ThermoFisher). The RNA Integrity Number (RIN) was greater than 7 for all samples. In order to generate the libraries using the Truseq stranded mRNA kit (Illumina), 100 ng of RNA was used. These libraries were sequenced on an Illumina Novaseq sequencer on a SP flow cell in single read 100 bp. Sequence alignment with the human genome (version GRCh37), sequence counting and quality control were performed using the nf-core/rnaseq pipeline [1].

2 Counting from fastq data using nf-core/rnaseq pipeline

The following codes were used to do the mapping and counting.

```
nextflow run /home/mass/opt/griddata/Pipelines/nf-core/rnaseq-master -c
/script/nextflow.config --star_index
/references/Homo_sapiens/Ensembl/GRCh37/Sequence/STARIndex --gtf
/references/Homo_sapiens/Ensembl/GRCh37/Sequence/STARIndex/genes.gtf --
   singleEnd -profile conda --reads
'/fastq/*_R1.fastq.gz'
```

3 Session information

```
Workflow profile: conda
Workflow container: nfcore/rnaseq:1.3
Nextflow version: version 19.07.0, build 5106 (27-07-2019 13:22 UTC)
```

References

1. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* 2020;