DOI

Note

The demo files are mainly for Arabidopsis and data analysis of NAD tagSeq protocol

Table of content

Softwares and code

(1) MinKNOW 19.6.8, with base-caller of Guppy embedded, from Oxford Nanopore Technology
(2) Ubuntu 18.04.3 LTS, Linux-based operating system (https://ubuntu.com/download)
***The following code packages should be installed on Ubuntu**
(3) Python 2.7 and 3.7 (http://www.python.org/downloads/)
(4) Miniconda3 (Miniconda3-latest-Linux-x86_64.sh)(https://dos.conda.io/projects/conda/en/latest/user-guide/install/linux.html) for pycoQC uses;
(5) PycoQC (https://github.com/a-slide/pycoQC) to analyze the basecalling results;
(6) Homemade python script to sort out tagged and untagged RNA (https://github.com/rocketjishao/NAD-tagSeq/blob/master/main.py)
(7) Minimap2 (https://github.com/lh3/minimap2) to align the sequenced RNA to genome or transcriptome databases for interpretation of the RNA identities;
(8) featureCounts 1.6.0 (http://bioinf.wehi.edu.au/featureCounts/) to map and count the reads of tagged RNA to genes in different samples.
(9) Samtools 1.7 (http://samtools.sourceforge.net/) to translate the sam file to bam file and obtain its bam.bai file;
(10) Integrative Genomics Viewer 2.7.2 (https://software.broadinstitute.org/software/igv/) to visualize the RNA structures;

Software installation and initiation

python2.7 and python3.6

```
$ sudo apt-get install python2
      # Then type in password
$ sudo apt-get install python-pip
      # or try $ python get-pip.py

$ sudo add-apt-repository ppa:jonathonf/python-3.6
      # Then type in password, or try $ sudo apt-get install python3
```

Miniconda3

(https://conda.io/projects/conda/en/latest/user-guide/install/linux.html):
a. Download the installer: Miniconda installer for Linux.(https://docs.conda.io/en/latest/miniconda.html#linux-installers)
b. Verify your installer hashes, in a terminal window enter:

```
$ sha256sum Downloads/Miniconda3-latest-Linux-x86_64.sh
```

c.In your terminal window, run Miniconda:

```
$ bash Downloads/Miniconda3-latest-Linux-x86_64.sh
```

d.Follow the prompts on the installer screens.
e.If you are unsure about any setting, accept the defaults. You can change them later.
f.To make the changes take effect, type in the command below, or close and then re-open your terminal window.

```
$ source ~/.bashrc
   # Then you can see "(base)" in the front of the terminal command line.
```

g.Test your installation. In your terminal window, run the command:

```
$ conda list
```

h. A list of installed packages appears if it has been installed correctly.
i. To change the automatic conda activation because *auto_activate_base* is set to True. You can check this using the following command

```
$ conda config --show | grep auto_activate_base
```

pycoQC

(https://a-slide.github.io/pycoQC/installation/)
a. Create a clean virtual environment (only needed for the 1st run):

```
$ conda create -n pycoQC python=3.6
# Note: python 2 is not supported by pycoQC
```

b. Install pycoQC with miniconda3:(only needed for the 1st run)

```
$ conda install -c aleg pycoqc
```

c. Run pycoQC by the command:

```
$ pycoQC -f sequencing_summary.txt -o pycoQC_output.html
```

d. Quit conda

```
$ conda deactivate
```

e. To enter and exit conda for 2nd, 3rd,... run
To change the automatic conda activation because *auto_activate_base* is set to True. You can check this using the following command

```
$ conda config --show | grep auto_activate_base
```

To set it false

```
$ conda config --set auto_activate_base False
$ source ~/.bashrc
```

To reactivate set it to True

```
$ conda config --set auto_activate_base True
$ source ~/.bashrc
```

Minimap2

(https://github.com/lh3/minimap2) :

```
$ git clone https://github.com/lh3/minimap2
$ cd minimap2 && make
```

featureCounts

(http://subread.sourceforge.net/):

```
$ sudo apt-get install subread
    # then type in password
```

Samtools

(https://gist.github.com/adefelicibus/f6fd06df1b4bb104ceeaccdd7325b856)
(http://www.htslib.org/download/)

```
$ sudo apt-get install -y samtools
    # then type in password
```

IGV for Linux OS

Download the IGV file: https://data.broadinstitute.org/igv/projects/downloads/2.8/IGV_Linux_2.8.0.zip;
Unzip the package;
In the terminal window, start IGV by the command line:

```
$ java --module-path=lib -Xmx4g @igv.args --module=org.igv/org.broad.igv.ui.Main
```

Download genome file from IGV for A. thaliana, human, mouse, or E.coli:
Genome > Load Genome from Server > Select the genome file

Demo files

Step	sofware	input_files	output_files	demo files
1	pycoQC	sequencing_summary.txt	pycoQC.html (raw data)	no
2	Windows OS CMS	fastq files (ADPRC+_1.fastq,ADPRC+_2.fastq,ADPRC+_3.fastq; ADPRC-.fastq	ADPRC+.fastq, ADPRC-.fastq	demo
3	main.py	ADPRC+.fastq; ADPRC-.fastq	ADPRC+_tagged.fastq; ADPRC+_untagged.fastq; ADPRC-_tagged.fastq; ADPRC-_untagged.fastq	deomo
4	minimap2	ADPRC+_tagged.fastq, ADPRC+_untagged.fastq; ADPRC-_tagged.fastq; ADPRC-_untagged.fastq; reference_file (A. thaliana TAIR10.fas)	ADPRC+_tagged.sam, ADPRC+_untagged.sam; ADPRC-_tagged.sam; ADPRC-_untagged.sam	demo
5	featureCounts	ADPRC+_tagged.sam; ADPRC+_untagged.sam; ADPRC-_tagged.sam; ADPRC-_untagged.sam; annotation file (TAIR10.gff)	all; all.summary	demo
6	samtools	ADPRC+_tagged.sam	ADPRC+_tagged.bam; ADPRC+_tagged_sort.bam; [ADPRC+_tagged_sort.bam.bai	demo
7	IGV	ADPRC+_tagged_sort.bam; ADPRC+_tagged_sort.bam.bai; ADPRC+_untagged_sort.bam; ADPRC+_untagged_sort.bam.bai; genome files (mm10.genome)	IGV figure	no

NAD-tagSeq data analysis procedure

1. Run pycoQC in MiniConda3 active virtual environment

To visualize the summary file generated from the sequencing and do the quality control analysis of the basecalling results: Type in the command below. Open the html file with web browser to visualize the results.

```
$ pycoQC –f sequencing_summary.txt –o pycoQC.html
```

2. Combine fastq files to one fastq file

In Windows OS CMD:

```
$ copy ADPRC+_*.fastq ADPRC+.fastq
```

In Linux OS:

```
$ cat ADPRC+_*.fastq > ADPRC+.fastq
```

3. Sort out the RNA with and without tag in the first 40 nt

Download main.py from our Git-Hub repository: https://github.com/rocketjishao/NAD-tagSeq/blob/master/main.py
Change directory to the file pathway of main.py; Sort out the RNAs with and without tag RNA sequence by typing in:

```
$ python main.py ADPRC+.fastq ADPRC+_tagged.fastq ADPRC+_untagged.fastq
   # result files: ADPRC+_tagged.fastq and ADPRC+_untagged.fastq
```

4. Minimap2 to align the reads to reference sequence

Run Minimap2 for analyzing the Nanopore direct RNA sequencing data by typing in the command:

```
$ ./minimap2 -ax splice -uf -k14 reference.fa ADPRC+_tagged.fastq > ADPRC+_tagged.sam
   # reference file like TAIR10.fa, result file is ADPRC+_tagged.sam
```

5. Use featureCounts to count the aligned reads to genes

Use simultaneously the tagged and untagged counterparts (or map each gene to the tagged RNA in ADPRC- and ADPRC+ samples.)
And download gene annotation files in gtf format from Ensembl or GenBank (https://www.ncbi.nlm.nih.gov/genbank/), avoid UCSC
Run the command below:

```
$ featureCounts -L -a annotation -o all ADPRC+_tagged.sam ADPRC+_untagged.sam ADPRC-_tagged.sam
   # annotation file like TAIR10.gff, result files are all and all.summary
```

6. Samtools to translate the sam file to bam file and obtain its bam.bai file

Run Samtools by typing in (one by one):

```
$ samtools view -bS ADPRC+_tagged.sam > ADPRC+_tagged.bam
$ samtools sort -O BAM -o ADPRC+_tagged_sort.bam ADPRC+_tagged.bam
$ samtools index ADPRC+_tagged_sort.bam
   # result files: ADPRC+_tagged.bam, ADPRC+_tagged_sort.bam, ADPRC+_tagged_sort.bam.bai
$ samtools stats ADPRC+_tagged.bam | grep '^SN' | cut -f 2-
   # use this to visualize the # mismatches / bases mapped (cigar), which should be smaller than
```

7. IGV to visualize the RNA structure

Import the bam and bam.bai to IGV by:

File > Load from File > Select the ADPRC+_tagged_sort.bam file