DUESSELPORE Webserver manual

This is the instruction of using Duesselpore webserver to process RNAseq data.

Data availability:

- Docker image on Dockerhub https://hub.docker.com/repository/docker/thachdt4/duesselpore:main
- Light weight testdata at https://iufduesseldorf-my.sharepoint.com/:u:/g/personal/thach_nguyen_iuf-duesseldorf_de/ES4BsdfJSKNHl-mDUR3BogcBEmdOawVTRy-eRXU3-XeG2A?e=Kq9O2e
- Full test data. https://iufduesseldorf-my.sharepoint.com/:u:/g/personal/thach_nguyen_iuf-duesseldorf_de/EWIk4CLauThHk61_5rItjEcBOP4CJstbyCN9yN3ty36A7g?e=zRUf1T

Install and configure webserver

System requirement

- CPU: 2.0 GHz (Intel architecture, acceleration support) 4 cores or higher
- System memory (RAM): 8 GB or higher
- Diskdrive: 100 GB free space or higher
- Host operating system Window 10, Linux (Ubuntu >=18.04 or Fedora) or MacOS (Intel)

Installation and run Dueselpore

Install Docker and setup webserver

Install Docker from https://www.docker.com/get-started. Open the terminal (on Linux, MacOS), commandline interface or WSL(on Window). You may have to run it as superuser.

```
docker run -it -p 8000:8000 thachdt4/duesselpore:running \
python3 /home/ag-rossi/projects/duesselpore/manage.py runserver 0.0.0.0:8000
```

Depend on your host Operating system and your IP address range, you may have to configure the Docker IP address (default is 172.17.0.2).

Using webserver

Access webserver

Now you can use your webserver within your Local Area Network (LAN) with a regular web browser (e.g., Firefox or Google Chrome HTTP port: 8000) http://localhost:8000/duesselpore.

Data preparation

Users can upload fastq files as ONE compressed zip file: each subfolder contains several replicas with one experimental condition. Our decompressor support

most common compression program in Window and Linux such as zip, gunzip, 7z etc. NOTE: files and folders' names must contain only alphabetic and numeric characters. Below is an example of data separated into two conditions, 'condition1' and 'condition2'. Please check the structure and the director name of your data carefully, all the name of analysis are generated by directory and file names.

```
fastq/(folder)
   condition1 (subfolder)
       condition1_replica1.fastq (single fastq file)
       condition1_replica2.fastq (single fastq file)
       condition1_replica3.fastq (single fastq file)
   condition2 (subfolder)
        condition2_replica1.fastq (single fastq file)
        condition2_replica2.fastq (single fastq file)
        condition2_replica3.fastq (single fastq file)
```

To merge multiple fastq files into a single file: On Linux terminal:

```
$ cat /path/to/fastq/files/*.fastq > /your/new/location/output.fastq
```

On Window command prompt (path symbol is different):

```
$ type \path\to\fastq\files\*.fastq> \your\new\location\output.fastq
```

Running parameter:

First, select one group among your groups as the reference group. Select the gene (or transcriptome) counting method, select the differential expression algorithm you want to analyze. To run the analysis, we have to set up other parameters of the analysis function. There are some optional parameters, e.g., ReadCountMinThreshold, Logfold change threshold, adjPValueThreshold. After submiting we have to keep the webbrowser open then wait for the result. Advanced users can customize the RNA.R code to develop a new workflow.

Collecting the results:

The run time depends on your data size and the system speed. For our standard 6 replicates dataset, approximate totals 16 million reads (around 15 Gb), the run time is around 6 hours. For lightweight test data, running time is one hours on normal laptop.
After the computation is completed, all the results are downloaded from the browser. The interactive HTML file is exported with different plots. Users can continue the offline analysis Docker directory at /home/agrossi/duesselpore/users_file/{your session id}. Experienced users will be able to further analyze their data by editing the R script. NGS data requires high volume space, therefore we recommend user erasing the data on the Docker regularly. The sample result is in the Support Information, or sample_result/report.html in sample_result.zip. The quality results are in QC

subfolder. Please note that while most of the analysis do not require an internet connection, except gene ontology and disease pathway will require an internet connection.