AS-Quant User Manual

Naima Ahmed Fahmi, Jeongsik Yong, Wei Zhang
May, 2019

1. Download

AS-Quant tool can be downloaded directly from github. User need to have python installed in their machine. It can work on Windows, Linux and Mac platform.

2. Required software

a. Python (version 3.0 or higher)
b. Samtools 0.1.8

3. Required python packages

a. Matplotlib
b. Panda
c. csv

4. Run AS-Quant

AS-Quant consists of five python files, which are:

i) Samtools_view.py: processes the input aligned bam files into chromosome wise separated text files
ii) Initial.py: initializes the variables and classes
iii) ARC.py: the main function which the user need to run
iv) Methods.py: all the functions(methods) to run the AS-Quant
v) count_pvalue.py: computes the association among two samples and generate the final table
vi) makePlots.py: generates figures for visual representation of data

AS-Quant comprises of three major phases: 1) Pre-processing, 2) Generate and Compare and 3) Visualization. The process of running the tool is described here step by step with functionality of each phase. User **must** follow each step sequentially in order to run AS-Quant properly.

a) Pre-processing:
   → Open the **ARC.py** file. Change **current_path** variable according to your current path [change in code].
   → Add two sample names in the **samplenames** list [change in code].

→ Make a folder inside of current path named **input**, and make two subfolders inside of that **input** folder with name same as the two sample names**.**

→ Provide the bam files for two input samples aligned with the reference genome, named as **accepted_hits.bam** and copy these files inside of each **sample** folder**.**

→ For mouse, the chromosome lists will be **chr1,ch2,...,chr19,chrX,chrY** and for human it will be **chr1,chr2,...,chr22,chrX,chrY.** Change **chromosome** list accordingly [change in code].

→ The target exon list for all five types of alternative splicing are **SE.csv, RI.csv, A3SS.csv, A5SS.csv** and **MXE.csv.** Copy then into the parent directory.

→ Copy **annotation.csv** in the parent directory. The annotation files are used from UCSC Genome Browser and for mouse it is mm10, and for human, it is hg19.

→ Then, **run samtools_view.py** in the command prompt and it will divide the total aligned bam file into chunks according to each chromosome.
$python3 samtools_view.py

b) Generate and Compare: **Run ARC.py**
$python3 ARC.py
This will generate the

c) Visualization:

→ Run **makePlots.py** and provide input in a specific pattern: chromosome:geneName:start-end

→ $python3 makePlots.py

→ Sample input format: chr1:Plcd4:74561841-74561937

d) bbbb