ATACseq_analyzer Manual

Mayur Divate and Edwin Cheung

Contact: mdivate@umac.mo

Table of Contents

How to install dependencies?

ATACseq_analyzer has dependencies which need to be installed on your system.

Java 1.8 or latest

Installing java on Mac OS

- To Download java go to https://java.com/en/download/
- Double-click the pkg file to launch it
- Double-click on the package icon to launch install Wizard
- The Install Wizard displays the Welcome to Java installation screen. Click Next
- Click the Next button to continue the installation.
- Click Close to finish the installation process.

For more details, please follow this link
https://www.java.com/en/download/help/mac_install.xml

Installing java on Linux OS

copy paste following command to the terminal

- sudo apt-get install oracle-java8-installer
- Or follow the link:
  https://java.com/en/download/help/linux_x64_install.xml

Bowtie version 1.1.2

- Download bowtie from here
  https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/
  Linux OS : bowtie-1.1.2-linux-x86_64.zip
  Mac OS : bowtie-1.1.2-macos-x86_64.zip
- Copy downloaded bowtie file or copy the bowtie*.zip file path
  To paste path on terminal use
  Mac => command + v
  Linux => ctrl + shift + v

- Open the terminal
- Use copy following commands to terminal and hit enter
  $ cp <bowtie file path> ~/
  $ cd ~/
  $ unzip bowtie-1.1.2*.zip
  $ cd bowtie-1.1.2/
    - Mac
  $ echo "export PATH=\$PATH:"`pwd` | cat - >> ~/.bash_profile
  $ source ~/.bash_profile
    - Linux

```
$ echo "export PATH=\$PATH:"`pwd` | cat - >> ~/.bashrc
$ ~/.bashrc
```

Python version 2.7

```
$ requires to run macs2
```

Mac OS

https://www.python.org/downloads/release/python-2711/

```
$ Download the Mac OS X 64-bit/32-bit installer (not the
  PPC installer) from the Python website.
$ Double-click the python-2.7.11-macosx10.6.pkg file in the
  Downloads folder.
$ If you have Gatekeeper enabled, the installation will be
  blocked. Open System Preferences > Security & Privacy and
  click Open Anyway.
$ Click Continue, Agree and Install buttons in the Install
  Python window.
```

Linux OS copy paste following command to terminal

```
$ sudo apt-get install python
```

MACS2

To install MACS2 on Mac or Linux OS use command below

```
$ pip install --user MACS2
```

SAMtools Version: 1.3.1

- Download samtools-1.3.1.tar.bz2 link: https://sourceforge.net/projects/samtools/files/samtools/1.3.1/
- Copy downloaded SAMtools file or copy the samtools-1.3.1.tar.bz2 file path
  To paste path on terminal use
  Mac => command + v
  Linux => ctrl + shift + v
- Open the terminal
- Use copy following commands to terminal and hit enter

```
$ cp <samtools file path> ~/
$ cd ~/
$ tar jxvf samtools-1.3.1.tar.bz2
$ cd samtools-1.3.1
$ make
```

- Mac

```
$ echo "export PATH=\$PATH:"`pwd` | cat - >>
  ~/.bash_profile
$ ~/.bash_profile
```

- Linux

```
$ echo "export PATH=\$PATH:"`pwd` | cat - >> ~/.bashrc
$ source ~/.bashrc
```

R Version: >= 3.3.0

Mac user can follow the link below for video tutorial to install R

- https://youtu.be/cX532N_XLIs?list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU

Linux user can use following link for installation

- https://cran.r-project.org

How to download and start ATACseq_analyzer?

To download ATACseq_analyzer use following link

Github link

Unzip

```
$ cd <ATACseq_analyzer folder>
$
```

To Start ATACseq_analyzer graphical user interface use following command

```
$ cd <ATACseq_analyzer folder>
$ java –jar ATACseq_analyzer.jar
```

ATACseq_analyzer graphical user interface

Start ATACseq_analyzer

```
$ java -jar ATACseq_analyzer.jar
```

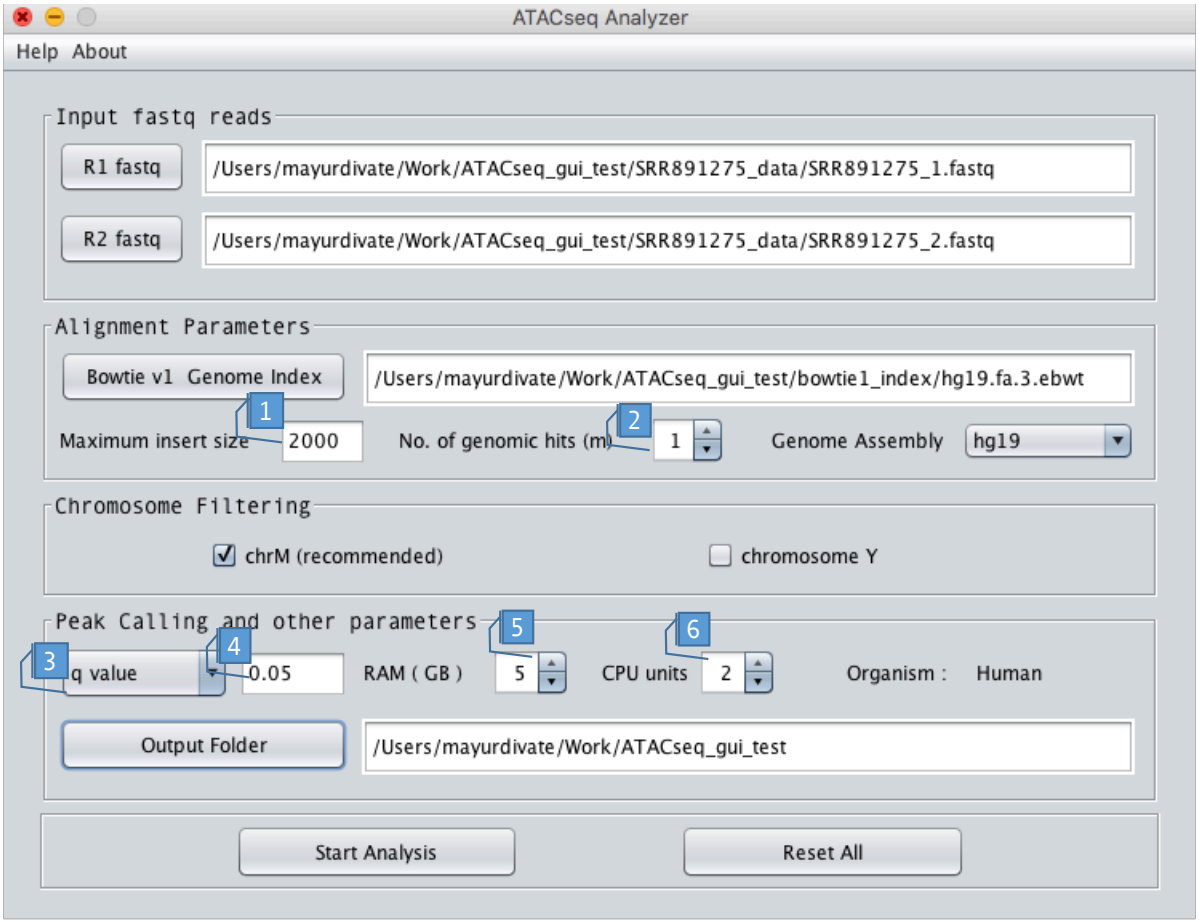


Figure 1: Graphical user interface of ATACseq_analyzer

User can select input FASTQ files and bowtie index of genome fasta file use R1 fastq, R2 fastq and Bowtie v1 Genome Index buttons respectively. Maximum distance for mapping of read mates can be set using box 1. To exclude reads pairs which are exceeding desired number of alignments (m) limits can be set using box 2. User can specify the assembly or version name using genome assembly drop down menu. If chromosome checkbox is selected, reads mapping to that chromosome will be discarded. Dropdown menu (box3) is provided to select p or q value and box 4 is for setting its cutoff to filter MACS2 peaks. RAM and threads can be set using box 5 and 6 respectively. Start analysis

button will start ATAC-seq data processing, if provided inputs values are valid else in valid values will be highlighted with red color.

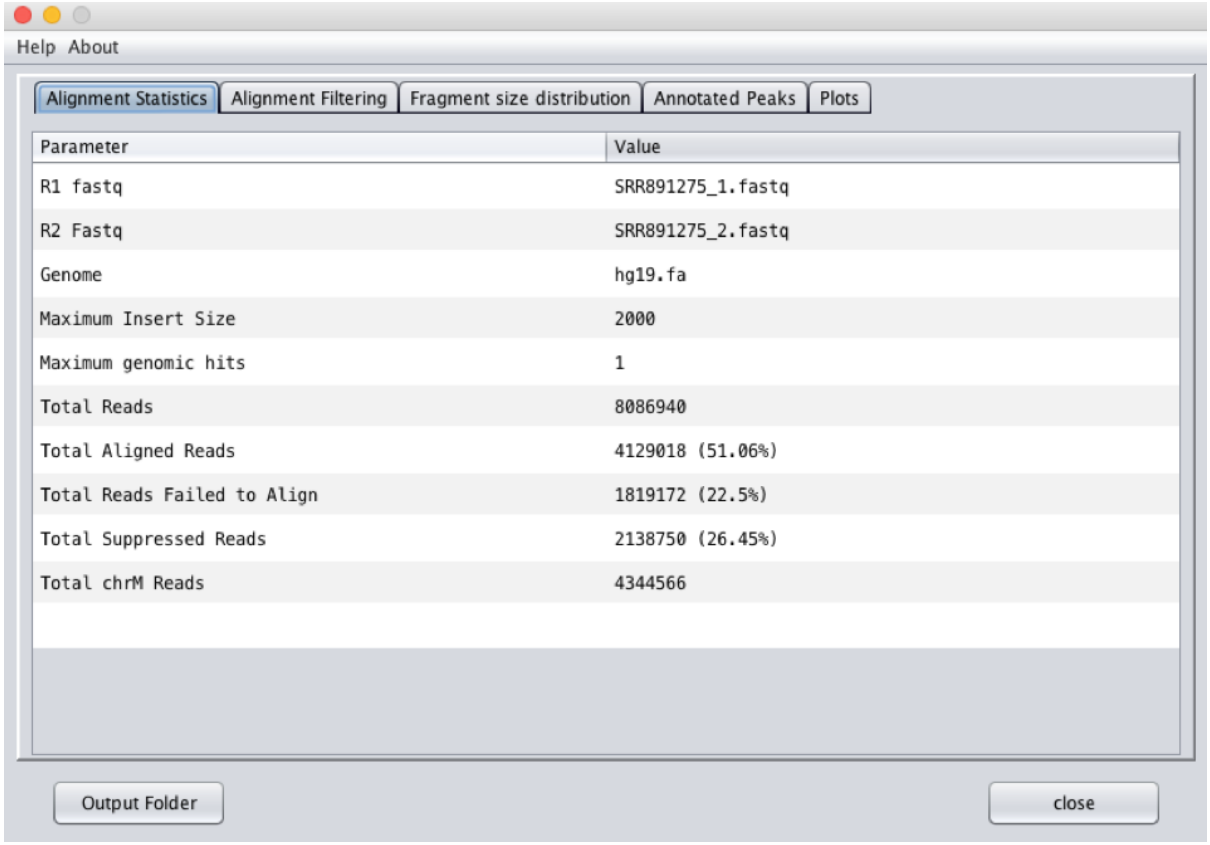


Figure 2: Result interface tab for alignment

This includes the input summary: names of FASTQ files and genome assembly used for analysis, maximum insert size and genomic hits. Total number of reads, aligned reads un aligned reads , fail to align due to m limit. Reads mapping to chr M and If chrY selected.

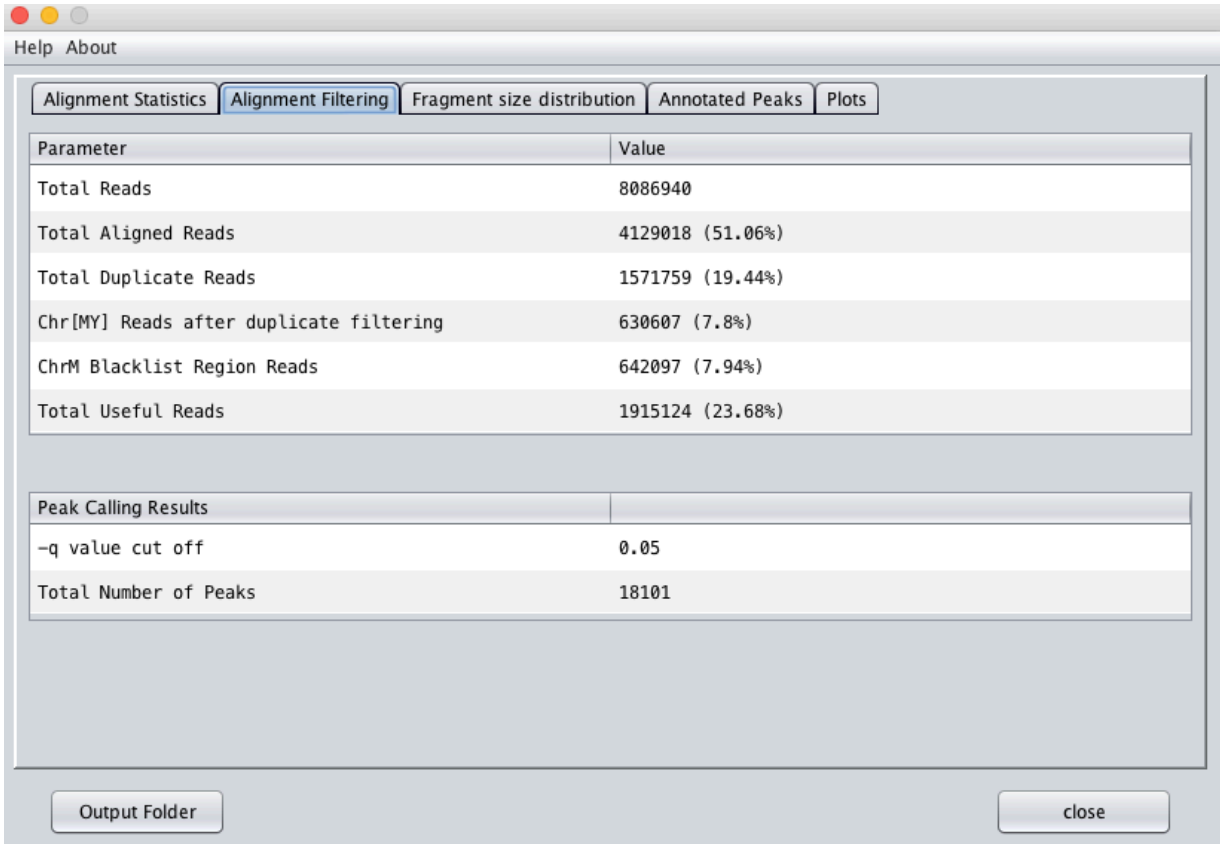


Figure 2: Alignment filtering and peak calling summary

First table provides number and percentage reads relative to total input reads i) aligned reads ii) duplicate reads iii) reads aligned to chromosome M and Y (based on selection) iv) reads mapped in blacklisted regions v) Useful reads represents the actual reads that qualifies for further processing. Second table gives you total number of peaks called and p/q value cutoff as per input

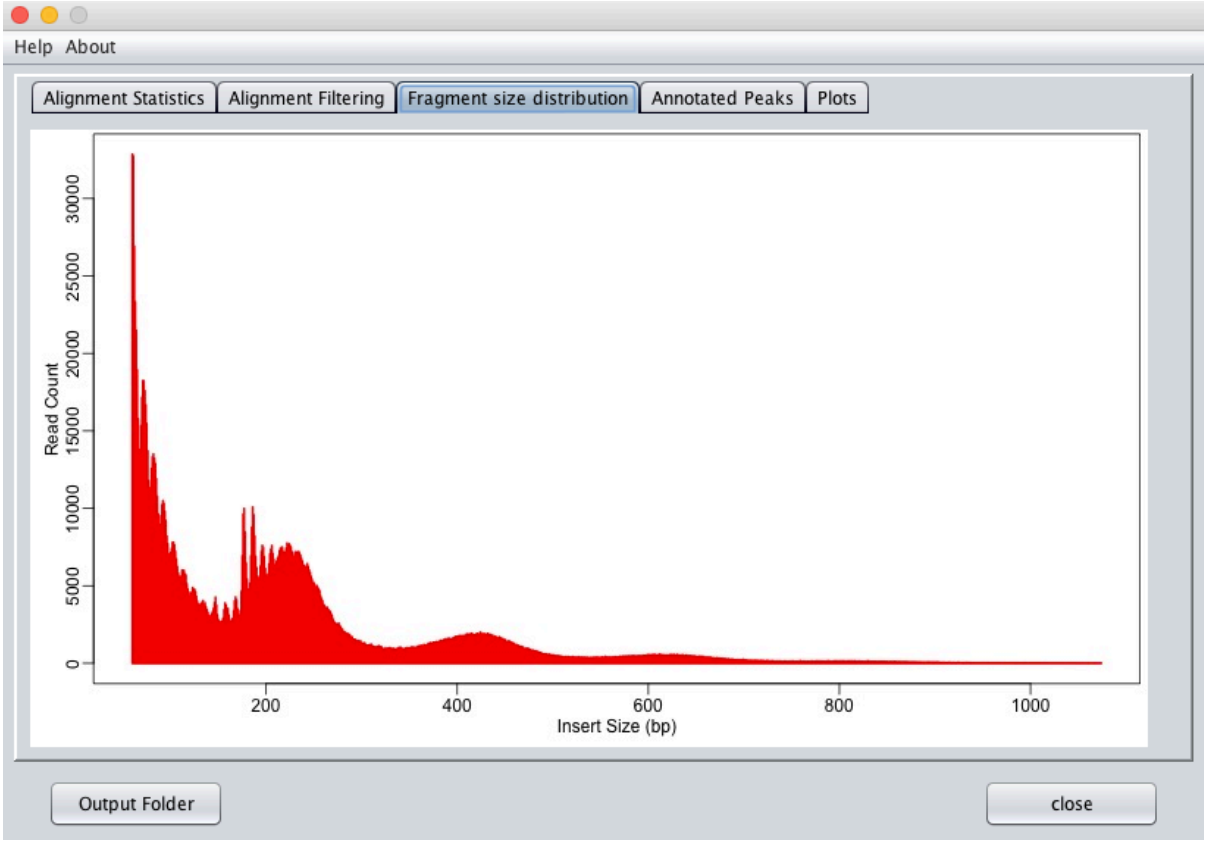
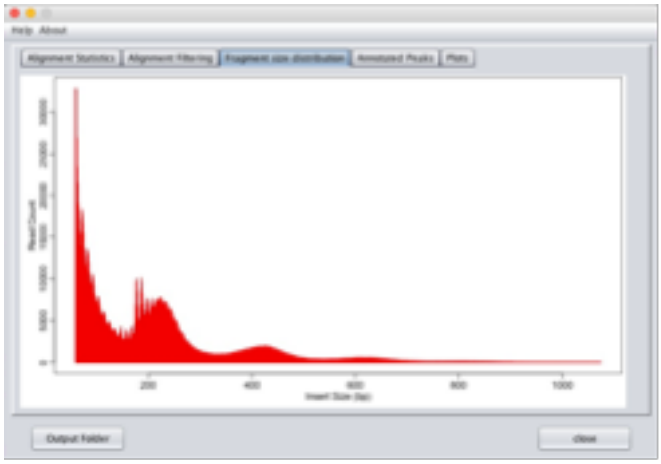
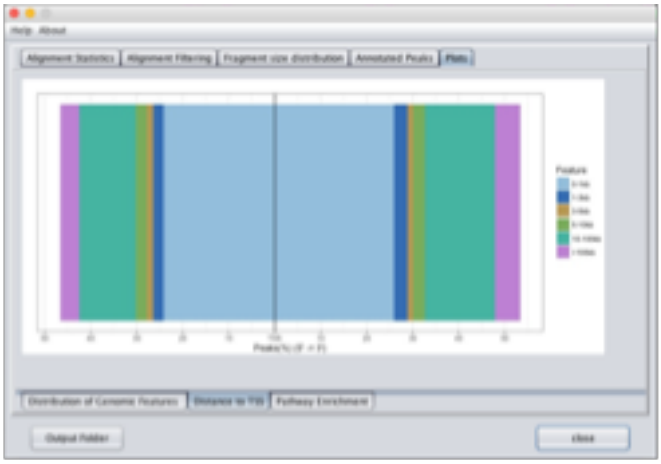
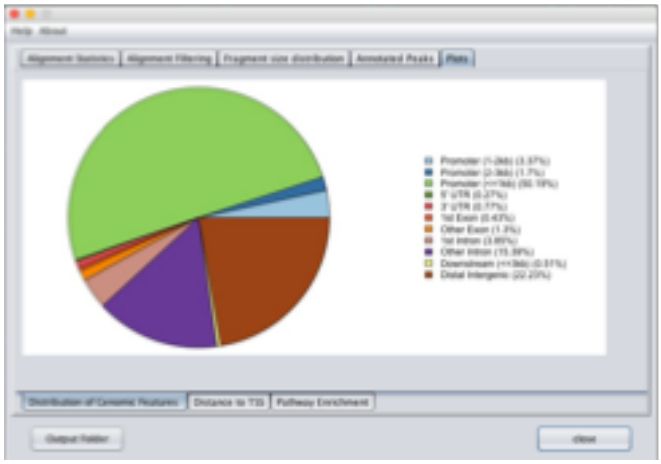


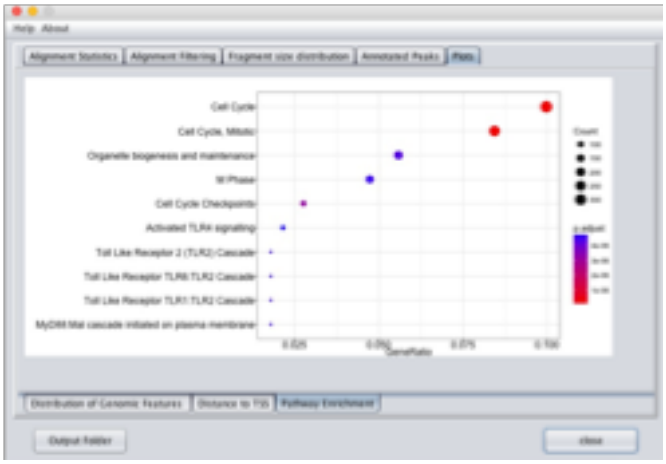
Figure 2: fragment size distribution graph

This graph









ATACseq_analyzer command-line interface

To print help message

```
$ java –jar ATACseq_analyzer.jar –h
```

Usage and option summary

Usage: $ java -jar ATAC_GUI.jar [options]*

Options	Description
R1	Path to the FASTQ file containing upstream mates
R2	Path to the FASTQ file containing upstream mates
g	Path to bowtie index of genome fasta file
a	Genome assembly version [hg18,hg18,hg38,mm9,mm10]
value	p | q value for MACS2 peak filtering default: q
c | cutoff	Cutoff for p/q value e.g. 0.05, 5E-2 default: 0.05
X	Maximum distance from each other at which read mates can map to the genome default: 2000
m	Report alignment for pair, if maximum number of reportable alignments for pair is less or equal to m default: 1
O | outdir	Path to the output directory default: current directory
ram	RAM memory to use in GBs default: 1
cpu	Number of threads to use default: 1
chrM	Remove(T) or keep(F) reads mapping to mitochondrial chromosome default: T
chrY	Remove(T) or keep(F) reads mapping to chromosome Y default: F
H | help	Print help mesage

Table 1: Usage and options for

Compulsory options are shown in blue color

How to create bowtie index of genome fasta file?

If you already have genome fasta file, please follow the command below to create bowtie genome index. Bowtie uses this index to speed up alignment process.

```
$ cd <path to genome fasta file>
$ bowtie-build <genome.fasta> <genome>
```

Note: It is time consuming step

Optionally you can create fasta index of genome

```
$ samtools faidx <genome.fasta>
```