"prepare ATAC-seq data from Luca/Pique-Regi lab for analysis"

Heejung Shim

November 1, 2016

Contents

1 Copy new ATAC-seq data from Luca/Pique-Regi lab to the rice cluster

Create directories in the rice cluster to save ATAC-seq data:

```
cd /home/hjshim/d/shared_data/internal_restricted/
mkdir atacseq_Luca_PiqueRegi
mkdir atacseq_Luca_PiqueRegi/rawbamfiles
mkdir atacseq_Luca_PiqueRegi/info
```

Copy infomation on each file:

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
scp gb9571@grid.wayne.edu:/nfs/rprscratch/atac.lcl/deseq/cv.txt info/
```

Copy bam files:

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
scp gb9571@grid.wayne.edu:/nfs/rprscratch/atac.lcl/merge/*.bam rawbamfiles/
```

Copy merged bam files:

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
scp gb9571@grid.wayne.edu:/nfs/rprscratch/atac.lcl/mergeReps/*.bam rawbamfiles/
```

2 Convert bam file to hdf5 format

2.1 Prepare directory and install genome library, samtools, and python

2.1.1 Create directory

Create directory to save hdf5 files:

```
cd /home/hjshim/d/shared_data/internal/
mkdir hdf5
cd hdf5
mkdir hg19
```

2.1.2 Install genome library

Clone Graham's genome repo and follow his instruction.

```
cd /home/hjshim/d/hjshim/software/genome
```

2.1.3 Install samtools

Download samtools on rice cluster:

```
cd /home/hjshim/software/source/
wget https://github.com/samtools/samtools/releases/download/1.3.1/samtools
    -1.3.1.tar.bz2
tar -vxjf samtools-1.3.1.tar.bz2
cd samtools-1.3.1
make
```

Install it on rice cluster:

```
make prefix=~/software/install/samtools-1.3.1 install
```

Make a symbolic link:

```
cd /home/hjshim/bin
ln -s ~/software/install/samtools-1.3.1/bin/* .
ls -al
```

2.1.4 Install python

```
module add python
pip install pysam --user
```

2.2 Copy chromosome.h5 file to rice cluster

I will copy the in PPS cluster to rice cluster here:

```
cd /home/hjshim/d/shared_data/internal/hdf5/hg19/
```

Now login into PPS cluster and copy chromosome.h5 file:

```
cd /data/internal/genome_db/hg19/
scp chromosome.h5 hjshim@rice.rcac.purdue.edu:/home/hjshim/d/shared_data/
    internal/hdf5/hg19/
```

2.3 Filter reads with low alignment quality 10

2.3.1 Create directory to save preprocessed bam files

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
mkdir processedbamfiles
```

2.3.2 Filter reads with low alignment quality 10

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    rawbamfiles/

samtools view -b -q 10 AT2-N701-N501.bam > ../processedbamfiles/AT2-N701-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N701-N502.bam > ../processedbamfiles/AT2-N701-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N701-N503.bam > ../processedbamfiles/AT2-N701-N503.
    qfiltered10.bam

samtools view -b -q 10 AT2-N702-N501.bam > ../processedbamfiles/AT2-N702-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N702-N502.bam > ../processedbamfiles/AT2-N702-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N702-N503.bam > ../processedbamfiles/AT2-N702-N503.
    qfiltered10.bam

samtools view -b -q 10 AT2-N703-N501.bam > ../processedbamfiles/AT2-N703-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N703-N502.bam > ../processedbamfiles/AT2-N703-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N703-N503.bam > ../processedbamfiles/AT2-N703-N503.
    qfiltered10.bam

samtools view -b -q 10 AT2-N704-N501.bam > ../processedbamfiles/AT2-N704-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N704-N502.bam > ../processedbamfiles/AT2-N704-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N704-N503.bam > ../processedbamfiles/AT2-N704-N503.
    qfiltered10.bam
```

```
samtools view -b -q 10 AT2-N705-N501.bam > ../processedbamfiles/AT2-N705-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N705-N502.bam > ../processedbamfiles/AT2-N705-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N705-N503.bam > ../processedbamfiles/AT2-N705-N503.
    qfiltered10.bam

samtools view -b -q 10 AT2-N706-N501.bam > ../processedbamfiles/AT2-N706-N501.
    qfiltered10.bam
samtools view -b -q 10 AT2-N706-N502.bam > ../processedbamfiles/AT2-N706-N502.
    qfiltered10.bam
samtools view -b -q 10 AT2-N706-N503.bam > ../processedbamfiles/AT2-N706-N503.
    qfiltered10.bam
```

2.4   Index filtered bam files

```
cd /home/hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/

samtools index AT2-N701-N501.qfiltered10.bam
samtools index AT2-N701-N502.qfiltered10.bam
samtools index AT2-N701-N503.qfiltered10.bam

samtools index AT2-N702-N501.qfiltered10.bam
samtools index AT2-N702-N502.qfiltered10.bam
samtools index AT2-N702-N503.qfiltered10.bam

samtools index AT2-N703-N501.qfiltered10.bam
samtools index AT2-N703-N502.qfiltered10.bam
samtools index AT2-N703-N503.qfiltered10.bam

samtools index AT2-N704-N501.qfiltered10.bam
samtools index AT2-N704-N502.qfiltered10.bam
samtools index AT2-N704-N503.qfiltered10.bam

samtools index AT2-N705-N501.qfiltered10.bam
samtools index AT2-N705-N502.qfiltered10.bam
```

```
samtools index AT2-N705-N503.qfiltered10.bam

samtools index AT2-N706-N501.qfiltered10.bam
samtools index AT2-N706-N502.qfiltered10.bam
samtools index AT2-N706-N503.qfiltered10.bam
```

2.5   Convert filtered bam files into hdf5

```
cd /home/hjshim/d/hjshim/software/genome/python/script/db/

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N701N501.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N701-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N701N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N701-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N701N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N701-N503.qfiltered10.bam

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N702N501.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N702-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N702N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N702-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N702N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N702-N503.qfiltered10.bam

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N703N501.qfiltered10.rev /home/
```

```
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N703-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N703N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N703-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N703N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N703-N503.qfiltered10.bam

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N704N501.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N704-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N704N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N704-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N704N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N704-N503.qfiltered10.bam

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N705N501.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N705-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N705N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N705-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N705N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N705-N503.qfiltered10.bam

python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N501
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N706N501.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
```

```
    processedbamfiles/AT2-N706-N501.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N502
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N706N502.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N706-N502.qfiltered10.bam
python load_bam_5prime_ends.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N503
    .qfiltered10.fwd /atacseq_Luca_PiqueRegi/N706N503.qfiltered10.rev /home/
    hjshim/d/shared_data/internal_restricted/atacseq_Luca_PiqueRegi/
    processedbamfiles/AT2-N706-N503.qfiltered10.bam
```

2.6 Obtain library read depth

First set track stats using set_track_stats.py and get library read depth using get_track_-
stats.py. One downside might be that it computes library read depth over all chromosomes
including sex chromosomes.

```
cd /home/hjshim/d/hjshim/software/genome/python/script/db/

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N503.
    qfiltered10.fwd

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N503.
    qfiltered10.rev

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N503.
    qfiltered10.fwd
```

```
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N503.
    qfiltered10.rev

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N503.
    qfiltered10.fwd

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N503.
    qfiltered10.rev

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N503.
    qfiltered10.fwd

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N503.
    qfiltered10.rev

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N503.
```

```
    qfiltered10.fwd

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N503.
    qfiltered10.rev

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N501.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N502.
    qfiltered10.fwd
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N503.
    qfiltered10.fwd

python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N501.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N502.
    qfiltered10.rev
python set_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N503.
    qfiltered10.rev
```

```
cd /home/hjshim/d/hjshim/software/genome/python/script/db/

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N502.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N701N503.
    qfiltered10.rev
```

```
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N502.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N702N503.
    qfiltered10.rev

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N502.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N703N503.
    qfiltered10.rev

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N502.
    qfiltered10.rev
```

```
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N704N503.
    qfiltered10.rev

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N502.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N705N503.
    qfiltered10.rev

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N501.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N502.
    qfiltered10.fwd
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N503.
    qfiltered10.fwd

python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N501.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N502.
    qfiltered10.rev
python get_track_stats.py --assembly hg19 /atacseq_Luca_PiqueRegi/N706N503.
    qfiltered10.rev
```

Library read depths have been saved in:

```
/home/hjshim/d/shared_data/internal/hdf5/hg19/atacseq_Luca_PiqueRegi/library.
    depth.txt
```