ChIP-Seq data analysis workshop

Exercise 1. ChIP-Seq peak calling

1. Using Putty (Windows) or Terminal (Mac) to connect to your assigned computer.

   Create a directory /workdir/myUserID (replace myUserID with you BioHPC ID), copy the fastq.gz and bam files to the working directory, then de-compress the file.

```
mkdir /workdir/myUserID

cd /workdir/myUserID

cp /shared_data/ChIP_seq_workshop_2015/rep5_D12K4.txt_trim_uniq_sorted.bam* ./

cp /shared_data/ChIP_seq_workshop_2015/rep5_D12H3_rep4.txt_trim_uniq_sorted.bam* ./
```

2. Check sequencing quality from bam or fastq file, in this test sample, bam file used as input. The windows should have Xming installed and opened, otherwise firefox command doesn't work. You can copy rep5_D12K4.txt_trim_uniq_sorted.bam_ fastqc folder to local terminal by filezilla to look at results

```
fastqc  rep5_D12K4.txt_trim_uniq_sorted.bam
cd ./rep5_D12K4.txt_trim_uniq_sorted.bam_ fastqc
firefox fastqc_report.html
```

3. Cross-correlation analysis

```
wget http://compbio.med.harvard.edu/Supplements/ChIPseq/spp_1.10.tar.gz

R CMD INSTALL spp_1.10.tar.gz

R

library(spp)

chip.data <- read.bam.tags("rep5_D12K4.txt_trim_uniq_sorted.bam")

binding.characteristics <-
get.binding.characteristics(chip.data,srange=c(50,500),bin=5, accept.all.tags = T)

print(paste("binding peak separation distance
=",binding.characteristics$peak$x))

pdf(file="s_1.crosscorrelation.pdf",width=5,height=5)

par(mar = c(3.5,3.5,1.0,0.5), mgp = c(2,0.65,0), cex = 0.8)

plot(binding.characteristics$cross.correlation,type='l',xlab="strand
shift",ylab="cross-correlation")

abline(v=binding.characteristics$peak$x,lty=2,col=2)

dev.off()
```

Note based on https://sites.google.com/a/brown.edu/bioinformatics-in-biomed/spp-r-from-chip-seq

4. Peak calling using software MACS2

```
source /programs/bin/util/setup_macs2.sh

macs2 -t ./rep5_D12K4.txt_trim_uniq_sorted.bam -
c ./rep5_D12H3_rep4.txt_trim_uniq_sorted.bam -n MACS_output --nomodel --
shiftsize=70 --keep-dup=1 --bw=140 -g 'ce' -q 0.01 -m 2,10 -B
```