Worksheet_9.1-Peak calling ChIP-seq/ATAC-seq data using MACS2
Author: Joe Cardiello (joseph.cardiello@colorado.edu)

MACS2 manual: https://github.com/taoliu/MACS

Before you start:

1. Make a directory to store your peak calls under /scratch/:

```
mkdir '/scratch/Users/<USERNAME>/ChIPorATAC-seq'

mkdir '/scratch/Users/<USERNAME>/ChIPorATAC-seq/macs2/'
```

2. Create the following directory to store the scripts you'll create:

```
mkdir /scratch/Users/USERNAME/ChIPorATACseq/scripts
```

3. Go to the directory you created:

```
cd /scratch/Users/USERNAME/ChIPorATACseq/scripts
```

Running MACS2 to peak call from your ChIPseq data

1. Rename the template file to a new file for peak calling:

```
rsync /scratch/Workshop/SR2019/scripts/template.sbatch
/scratch/Users/USERNAME/ChIPorATAC-seq/scripts/MACS2_BroadPeaks.sbatch
```

2. Use vim to view the new peak calling file:

```
vim /scratch/Users/USERNAME/ChIPorATAC-seq/scripts/MACS2_BroadPeaks.sbatch
```

3. Edit the sbatch parameters:
    a. Write a useful job name

```
#SBATCH --job-name=macs2broad  # Job name
```

    b. Enter your email

```
#SBATCH --mail-type=ALL # Mail events
 #SBATCH --mail-user=<YOUR EMAIL>
```

    c. Edit the error and output directory to match your appropriate directory

```
#SBATCH --
output=/scratch/Users/<USERNAME>/e_and_o/macs2broad.%j.out

#SBATCH --
error=/scratch/Users/<USERNAME>/e_and_o/macs2broad.%j.err
```

d. For peak calling, these settings should work for the file we will analyze:

```
#SBATCH --nodes=1
#SBATCH --ntasks=1 # Number of CPU (processer cores i.e. tasks)
#SBATCH --time=00:30:00 # Time limit hrs:min:sec
#SBATCH --mem=1gb # Memory limit
#SBATCH --partition=compute
```

4. Under 'LOAD MODULES', enter the following command to load Macs2

```
module load python/2.7.14/MACS/2.1.1
```

5. We will also include a command to remove blacklist regions, so also include the following command to load bedtools:

```
module load bedtools/2.25.0
```

6. Under 'Job Specific Variables' section, enter the following variables:
   a. Conveniently, for this class we have placed all of the relevant ChIP-seq BAM files into one directory, so specify the following directory as our BAM directory:

```
BAMS_DIRECTORY='/scratch/Workshop/SR2019/ChIP-seq/mapped/bams/'
```

   b. Control files are important for calibrating MACS's peak calling to background noise. In this case we will use the SRR5855054 file for the input ChIP BAM file:

```
ControlBAM='/scratch/Workshop/SR2019/ChIP-seq/mapped/bams/SRRFILE_chr1.sorted.bam'
```

   c. Use the $1 variable to indicate that the first variable given after the script name will be plugged in as the file to be peak called on, and this will also be used as the rootname of the file:

```
ROOTNAME=$1
```

   d. Specify the directory your made as your output directory:

```
OUT='/scratch/Users/<USERNAME>/ChIP-seq/macs2'
```

   e. We will also be removing regions that overlap with a 'Blacklist'- regions across the genome that are almost always enriched in ChIP-seq data for reasons such as containing repetitive regions:

```
BLACKLIST='/scratch/Workshop/hg38/wgEncodeDacMapabilityConsensusExcludable_grch38.bed'
```

7. Finally for running the MACS broad peak calling on this broad H3K27Ac ChIP-seq data:
    a. Call the macs2 function callpeak:

```
macs2 callpeak \
```

    b. Add the control bam file and ChIPseq file of interest to the command:

```
-c ${ControlBAM} \
-t $BAMS_DIRECTORY/${ROOTNAME}.sorted.bam \
```

    c. Specify the outdirectory

```
--outdir ${OUT} \
```

    d. Add the Rootname as the sample name

```
-n ${ROOTNAME} \
```

    e. Specify the human genome

```
-g hs \
```

    f. The SPMR tag specifies to scale the color coding of the ouput bed file

```
--SPMR \
```

    g. Add the broad tag to the command, and a very small q (adjusted p value)

```
--broad \
-q 0.00001
```

8. Finally, we will use bedtools to remove blacklist regions from the output peak file:
    a. Start the bedtools intersect command:

```
bedtools intersect \
```

    b. Specify the the file to be effectively filtered is your broadpeak output file, and the blacklist regions will be compared against this file

```
-a $OUT/${ROOTNAME}_peaks.broadPeak \
-b ${BLACKLIST} \
```

    c. Specify the -v option, which says to throw out any regions in your peak file that have ANY overlap with blacklist regions in b.

```
-v \
```

    d. Specify the output directory, and the final file name:

```
$OUT/${ROOTNAME}_peaks_clean.broadPeak
```

9. Save the program!

ESC then: :wq!

10. Run the program, initially using the H3K27ac DMSO sample for peak calling!

```
sbatch MACS2_BroadPeaks.sbatch SRR5855055
```

11. Also run it for the largazole treatment sample, so we can use DAStk later!

```
sbatch MACS2_BroadPeaks.sbatch SRR5855056
```