Worksheet_9.3-Differential MD score generation with DAStk
Author: Joe Cardiello (joseph.cardiello@colorado.edu)

DAStk manual: https://github.com/Dowell-Lab/DAStk
DAStk paper: https://www.mdpi.com/1420-3049/23/5/1136/htm

Before you start:

1. Go to the directory with your ChIP or ATAC-seq scripts you created previously:

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

Running DAStk differential MD score analysis:

2. Copy and rename a template file to a new file for DAStk differential analysis:

```
rsync /scratch/Workshop/SR2019/scripts/template.sbatch
/scratch/Users/USERNAME/<ChIPorATAC-seq>/scripts/DAStkDiffMDAnalysis.sbatch
```

3. Use vim to view the new file::

```
vim /scratch/Users/<USERNAME>/ChIPorATAC-seq/scripts/ DAStkDiffMDAnalysis.sbatch
```

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

```
#SBATCH --job-name=DAStk_DiffMDAnalysis  # 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/DAStkDiffAnalysis.%j.ou
t

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

    d. For processing these files, these settings should work for the files we will analyze:

```
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --mem=8gb
#SBATCH --partition=compute
#SBATCH --time=00:10:00
```

5. Under “LOAD MODULES” enter the following command to load python3 where dastk is installed for this instance:

```
module load python/3.6.3
```

6. Under ‘Job Specific Variables’ section, enter the following variables:
    a. Provide the directory where the MACS2 derived bed files were saved:

```
BEDS='/scratch/Users/<USERNAME>/ChIP/macs2/'
```

    b. Specify the overall dastk outdirectory:

```
OUTDIR='/scratch/Users/joefranchesco/ChIP/dastk/'
```

    c. Specify the motif directory:

```
MOTIF='/scratch/Workshop/hg38/best_curated_Human_TFs_p1e-
6_grch38'
```

    d. Specify the appropriate genome

```
GENOME='hg38'
```

    e. Indicate the first dataset for which the called peaks will be compared

```
COND1=$1
```

    f. Indicate the second dataset for which the called peaks will be compared

```
COND2=$2
```

7. Write the function for differential MD score analysis:
    a. Specify the production of a new subdirectory to contain all results of the specific pairwise analysis being carried out:

```
mkdir $OUTDIR/${COND1}_vs_${COND2}
```

    b. Initiate the function for differential MD score analysis:

```
differential_md_score \
```

    c. Specify the two input files previously generated by using the DAStk processing:

```
-1 $OUTDIR/${COND1}_chr1_peaks_clean_md_scores.txt \
-2 $OUTDIR/${COND2}_chr1_peaks_clean_md_scores.txt \
```

    d. Specify a p value cutoff

```
-p .00001 \
```

    e. Specify the number of threads to be used (only 1 for this small run)

```
-t 1 \
```

    f. Specify that we will output all generated files into the new subdirectory

```
--output $OUTDIR/${COND1}_vs_${COND2} \
```

g. Add the b flag, indicating that we want to output barcode plots for both conditions for all TF's whose MD score is found to be significant, in addition to the MA plot to be generated.

```
-b
```

8. Save your file!

ESC then: :wq!

9. Run the analysis!!

```
sbatch DAStkDiffMDAnalysis.sbatch SRR5855055 SRR5855056
```

10. View the output!

Use X2go to navigate to the directory you've made and take a look at the MA plot, look which TFs show color coding for significant increases or decreases in nearby peaks between samples!

Also in X2go take a look at a few of the barcode plots for a TF between the two samples!