

Short Read Day 9: Peak-centric sequencing worksheets

- Author: Jessica Huynh-Westfall

Part 1: QC of your ChIP-seq data

As we discussed last week, ensuring your data's quality is a pertinent step before you move forward with your data analysis.

1. After you Log into the AWS, you will make a directory for day 9 in your scratch directory. Make the subdirectories including one for qc and its subdirectories. Create a working directory for day9 in scratch.

```
[<username>@<hostname> ~]$ cd /scratch/User/<username>
[<username>@<hostname> ~]$ mkdir day9
[<username>@<hostname> ~]$ cd day9
[<username>@<hostname> ~]$ mkdir scripts qc fastq bam sam
[<username>@<hostname> ~]$ cd qc
[<username>@<hostname> ~]$ mkdir fastqc hisat_mapstats multiqc
```

2. Copy scripts from github into your scripts directory a. d9_fastqc.sbatch b. d9_hisat2.sbatch c. d9_qc_chipseq.sbatch
3. Copy the fastq files over from the scratch directory to your fastq directory

```
[<username>@<hostname> ~]$ cp -r /scratch/Shares/public/sread2023/data_files/day9/fastq
/scratch/Users/<YourUsername>/day9/.
```

4. cd into your scripts directory. Edit and run scripts d9_fastqc.sbatch and d9_hisat2.sbatch. These two scripts will run FastQC and read alignment via HISAT2. ****note** If you did not make a directory in your scratch directory for your output and error files, make a directory.
5. Run the final QC script d9_qc_chipseq.sbatch which will run preseq and multiqc. Preseq predicting and estimating the complexity of a genomic sequencing library. MultiQC will summarize the output from numerous bioinformatic tools and compile a report for you to review the QC logs.

Part 2: MACS to call peaks

To study DNA enrichment assays such as ChIP-seq and ATAC-seq, we are introducing the analysis method, Model-based Analysis of ChIP-Seq (MACS). This method enables us to identify transcription factor binding sites and significant DNA read coverage through a combination of gene orientation and sequencing tag position.

1. In your day9 directory, make a subdirectory macs2.

```
[<username>@<hostname> ~]$ mkdir macs2
```

2. Copy the d9_macs.sbatch from the git repo and place it into your scripts directory. Edit and run script

Part 3: Bedtools

Annotation files are critical for downstream sequencing analysis, and text file manipulation and Bedtools are both very useful ways to work with them.

Follow the Day9_Bedtools worksheet now.