

Экстрагируйте риды и обрежьте их с помощью Trimmomatic.

Запустим команды для всех наборов данных. Я предполагаю, что UxuR_1 и UxuR_2 – парные риды, поэтому запущу их в Paired End Mode, в то время как остальные – в Single End Mode. Снизу под командами – output после выполнения.

```
> trimmomatic SE -threads 16 -phred33 /mnt/local/vse2020/shared/ChIP-
seq_data/control_ExuR.fastq control_ExuR_trim.fastq LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
> Input Reads: 19417311 Surviving: 19202163 (98.89%) Dropped: 215148 (1.11%)

> trimmomatic SE -threads 16 -phred33 /mnt/local/vse2020/shared/ChIP-
seq_data/ChIP_ExuR.fastq ChIP_ExuR_trim.fastq LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
> Input Reads: 21802259 Surviving: 21578386 (98.97%) Dropped: 223873 (1.03%)

> trimmomatic SE -threads 16 -phred33 /mnt/local/vse2020/shared/ChIP-
seq_data/control_UxuR.fastq control_UxuR_trim.fastq LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
> Input Reads: 22987727 Surviving: 22872857 (99.50%) Dropped: 114870 (0.50%)

> trimmomatic PE -threads 16 /mnt/local/vse2020/shared/ChIP-
seq_data/ChIP_UxuR_1.fastq /mnt/local/vse2020/shared/ChIP-
seq_data/ChIP_UxuR_2.fastq ChIP_UxuR_1_trim.fastq ChIP_UxuR_1_unpaired.fastq
ChIP_UxuR_2_trim.fastq ChIP_UxuR_2_unpaired.fastq LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
> Input Read Pairs: 23901482 Both Surviving: 23490359 (98.28%)
Forward Only Surviving: 135226 (0.57%)
Reverse Only Surviving: 274326 (1.15%)
Dropped: 1571 (0.01%)
```

Наложите на геном E. coli K-12 MG1655 genome с помощью Bowtie.

Скачаем файл с последовательностью генома кишечной палочки.

```
> wget "https://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?tool=portal&save=
file&log$=seqview&db=nucore&report=fasta&id=545778205&extrafeat=null&conwithfea
t=on&hide-cdd=on" -O U00096.3.fasta
```

Из файла с геномом бактерии получим файлы в формате .bw2, чтобы наложить образцы.

```
> bowtie2-build U00096.3.fasta U00096.3.bowtie2
```

Запускаем наложение для каждого файла с обрезанными ридами.

```
> bowtie2 -U ChIP_ExuR_trim.fastq -x U00096.3.bowtie2 -S ChIP_ExuR.sam --threads
16
> ... 94.62% overall alignment rate

> bowtie2 -U control_ExuR_trim.fastq -x U00096.3.bowtie2 -S control_ExuR.sam --
threads 16
> ... 90.37% overall alignment rate

> bowtie2 -U ChIP_UxuR_1_trim.fastq -x U00096.3.bowtie2 -S ChIP_UxuR_1.sam --
threads 16
> ... 65.44% overall alignment rate

> bowtie2 -U ChIP_UxuR_2_trim.fastq -x U00096.3.bowtie2 -S ChIP_UxuR_2.sam --
threads 16
> ... 84.80% overall alignment rate

> bowtie2 -U control_UxuR_trim.fastq -x U00096.3.bowtie2 -S control_UxuR.sam --
threads 16
> ... 66.96% overall alignment rate
```

Для файлов, которые я предположила парными, запустим общее наложение.

```
> bowtie2 -1 ChIP_UxuR_1_trim.fastq -2 ChIP_UxuR_2_trim.fastq -x U00096.3.bowtie2
-S ChIP_UxuR_all.sam --threads 16
> ... 75.07% overall alignment rate
```

Но согласованных выравниваний 0.0%.

Найдите пики с помощью MACS2.

Чтобы запустить поиск пиков, переведем получившиеся .sam файлы в .bam с помощью samtools.

```
> samtools view -bS -@ 16 ChIP_ExuR.sam > ChIP_ExuR.bam
> samtools view -bS -@ 16 control_ExuR.sam > control_ExuR.bam
> samtools view -bS -@ 16 ChIP_UxuR_1.sam > ChIP_UxuR_1.bam
> samtools view -bS -@ 16 ChIP_UxuR_2.sam > ChIP_UxuR_2.bam
> samtools view -bS -@ 16 control_UxuR.sam > control_UxuR.bam
```

Запускаем MACS2 с получившимися файлами.

```
> macs2 callpeak -t ChIP_ExuR.bam -c control_ExuR.bam -f BAM -g 4.7e+6 --extsize
147 --nomodel --keep-dup=auto --outdir macs2/ExuR -n ExuR 2>
macs2/ChIP_ExuR.log

> macs2 callpeak -t ChIP_UxuR_1.bam -c control_UxuR.bam -f BAM -g 4.7e+6 --
extsize 147 --nomodel --keep-dup=auto --outdir macs2/UxuR_1 -n UxuR_1 2>
macs2/ChIP_UxuR_1.log

> macs2 callpeak -t ChIP_UxuR_2.bam -c control_UxuR.bam -f BAM -g 4.7e+6 --
extsize 147 --nomodel --keep-dup=auto --outdir macs2/UxuR_2 -n UxuR_2 2>
macs2/ChIP_UxuR_2.log
```

Посчитаем количество получившихся пиков.

```
> wc -l macs2/ExuR/ExuR_peaks.narrowPeak
> 189

> wc -l macs2/UxuR_1/UxuR_1_peaks.narrowPeak
> 1765

> wc -l macs2/UxuR_2/UxuR_2_peaks.narrowPeak
> 218
```

Гомологичны ли белки ExuR и UxuR?

Да, вероятно, ExuR и UxuR являются гомологами. Оба являются DNA-binding transcriptional repressor, по данным из NCBI, и имеют приблизительно одинаковую длину.

В каком эксперименте есть проблемы, с чем они могут быть связаны и как их решить?

В части с macs2 встречается ошибка, что не хватает парных пиков. Программа предлагает исправить это с помощью флагов `-extsize 147 -nomodel`. Ещё пришлось добавить флаг `--keep-dup=auto`. Я сначала предположила, что UxuR_1 и UxuR_2 – это парные риды, поэтому в trimmomatic использовала Paired End Mode. Но, видимо, это не так, и это просто

разные версии экспериментов, так как количество пиков и процент выравнивания сильно отличаются, а также потому что bowtie2 для парных файлов не нашел согласованных выравниваний.

Предположите, чем могут отличаться эксперименты UxuR1 и UxuR2 с биологической точки зрения.

Вероятно, в первом эксперименте белок UxiR экспрессировался, а во втором UxiR не экспрессировался, поэтому у UxiR_1 больше пиков и меньше покрытие. Либо наоборот.

Используя ChIPMunk, найдите мотив для одного из этих белков.

С помощью bedtools получим .fasta файл с участками пиков.

```
> bedtools getfasta -fi U00096.3.fasta -bed macs2/UxuR_2/UxuR_2_peaks.narrowPeak  
-fo Motif UxuR 2.fasta
```

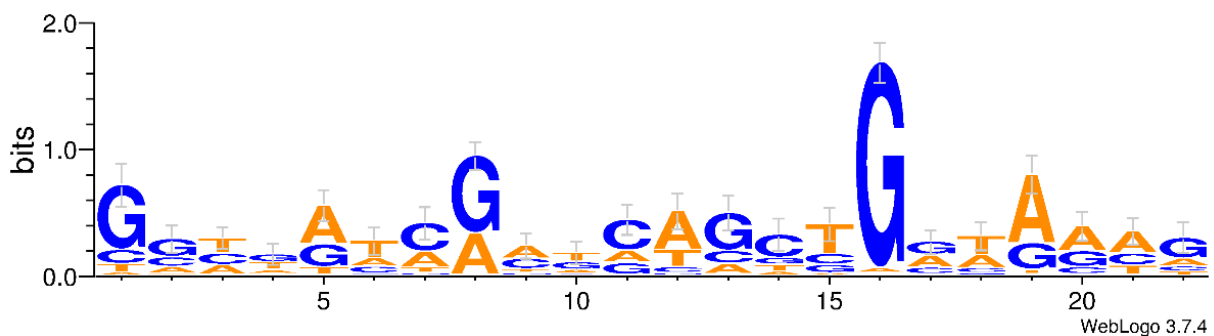
Запустим поиск мотивов в ChIPMunk.

```
> java -cp chipmunk.jar ru.autosome.ChIPMunk s:Mot UxuR 2.fasta > results.txt
```

Из полученного файла с результатами я с помощью скрипта на python вытащила информацию о последовательностях, представляющих мотив, и получила из них .fasta файл, который подала в WebLogo 3.

```
> weblogo -f output.txt -D fasta -o motif -F png
```

В итоге получился следующий мотив.



Поскольку вам рассказывали про MEME: сравните мотивы, полученные с помощью ChIPMunk и MEME. Какой мотив лучше и почему?

Последовательности, полученные с помощью `bedtools`, в предыдущем пункте я подавала в MEME со параметрами по умолчанию.

DISCOVERED MOTIFS

	Logo	E-value	Sites	Width
1.		1.1e-098	12	43
2.		2.7e-087	7	50
3.		2.7e-087	7	50
Stopped because requested number of motifs (3) found.				

Мотивы, полученные с помощью MEME имеют очень хороший e-value и высокую консервативность, особенно 2 и 3. Мотив из ChIPMunk имеет p-value $4.8840E^{-8}$, что неплохо, но хуже результатов MEME, плюс консервативность очень низкая.