

**Задание 0 (0,5 балла). Кратко опишите суть эксперимента.**

Суть эксперимента: чтобы получить информацию о расположении нуклеосом на ДНК, авторы провели ATAC-seq на данных для человека и для двух видов дрожжей: *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*. Для *Saccharomyces cerevisiae* ещё был проведен эксперимент с созданием осмотического стресса на клетках. После этого также провели ATAC-seq (замеры происходили после разных промежутков времени).

Собственно, с последними данными мы и работаем.

**Задание 1 (1 балл): какая команда используется для запуска анализа?**

`nextflow run nf-core/atacseq -profile test`

**Задание 2 (0,5 балла): приведите содержание вашей рабочей директории по окончании выполнения команды.**

```
(nf-core-atacseq-1.2.1) anastasia_rastvorova@hpc09:~/AtacPract$ ls
atacseq results work
(nf-core-atacseq-1.2.1) anastasia_rastvorova@hpc09:~/AtacPract$
```

**Задание 3 (0,5 балла): Приведите имя вашего запуска.**

`sad_meitner`

**Задание 4 (0,5 балла): Приведите добавочный параметр для поиска транскрипционных факторов.**

Нужно добавить `--narrow_peak`, чтобы уменьшить ширину пиков.

Собственно, команда:

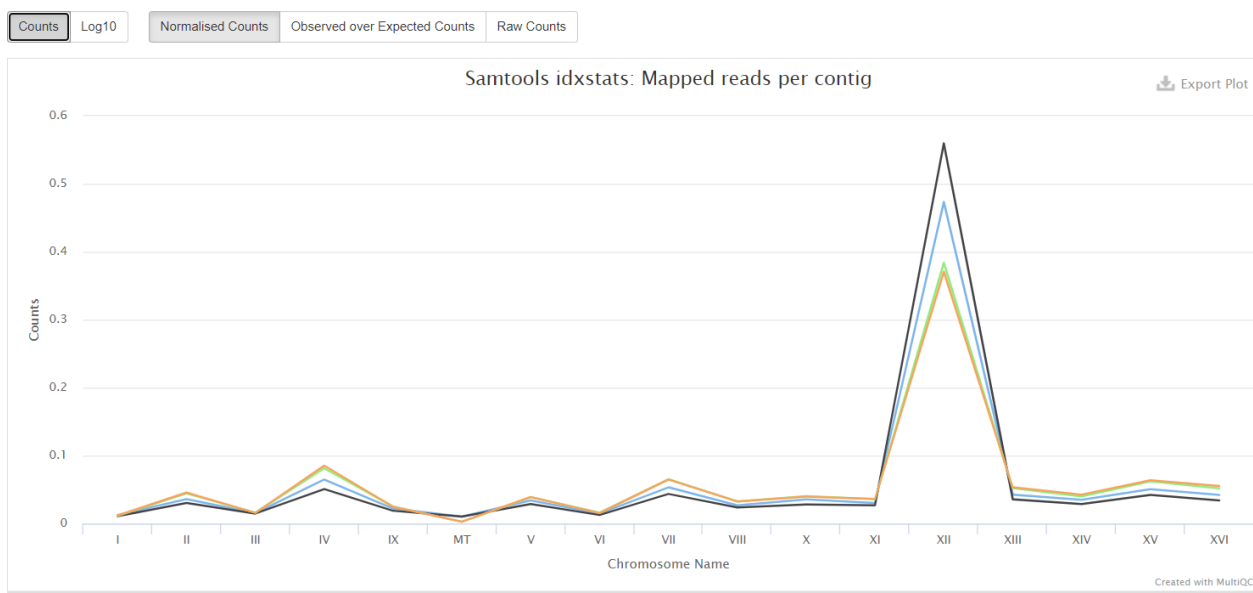
`nextflow run nf-core/atacseq -profile test -resume sad_meitner --narrow_peak`

Новое имя: `desperate_jang`

**Задание 5 (1 балл): Какова Fraction of Reads in Peaks для двух запусков? Почему? Удовлетворяют ли критерию? Связано ли с количеством пиков?**

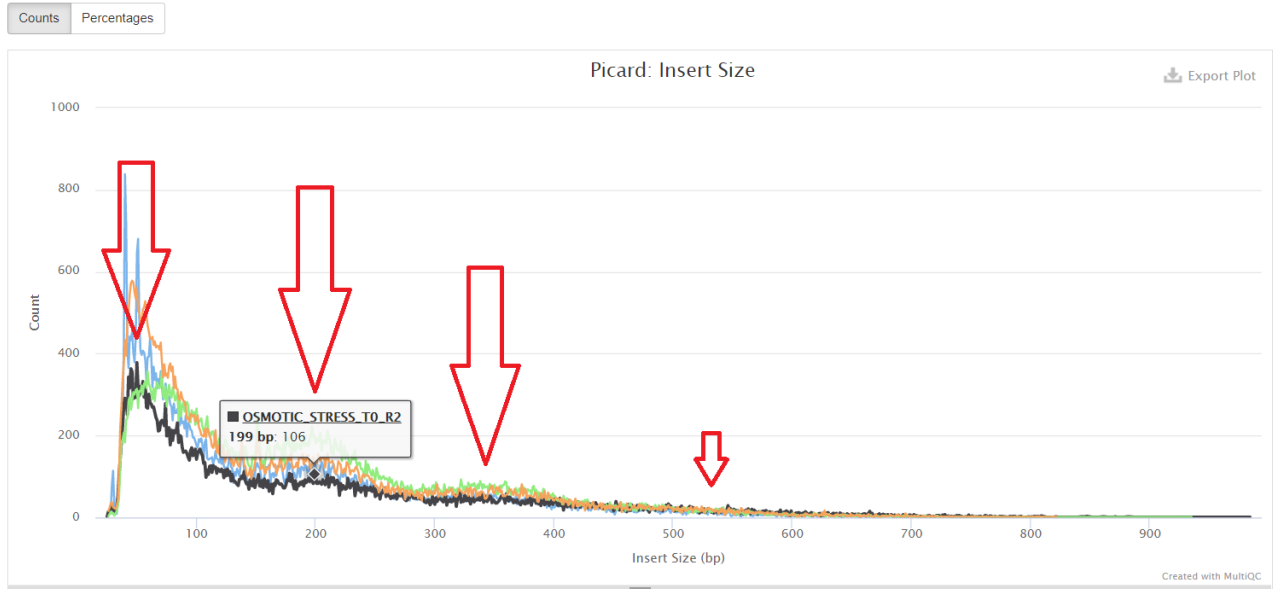
В случае с `broadPeaks` значение FRiP score для `OSMOTIC_STRESS_T0_R2` **0.36**, а в случае `narrowPeak` – **0.31**. И то, и другое значение меньше 0.4, что недостаточно. Чем шире пик, тем больше ридов в него попадает, тем выше значение FRiP score.

**Задание 6 (1 балл): Достаточно ли глубина секвенирования? У какой хромосомы наибольшее покрытие?**



Не уверена, что это нужный график. Но глубина секвенирования недостаточна. Наибольшее покрытие у хромосомы XII.

Задание 7 (2 балл): Какой размер нуклеосомы исходя из вашего графика. Отметьте это на графике, добавьте в отчет и поясните.



Размер нуклеосомы примерно 150 bp (расстояние между пиками, хотя на моем самом низком графике они почти не заметны).

Задание 8 (1 балл): Найдите нуклеосомные димеры, тримеры, тетрамеры. Опишите результаты в отчете и поясните почему.

Для графика OSMOTIC\_STRESS\_T0\_R2 довольно сложно определить димеры, тримеры и тем более тетрамеры, потому что пики очень и очень низкие. Но судя по тому, что в принципе остальные графики имеют похожую тенденцию, димер находится в районе 350 bp, тример должен быть в районе 500 bp, а третрамер – в районе 650 bp.

Задание 9 (1 балла): какие нашлись мотивы? Приложите лого лучшего из них. Является ли эта находка статистически значимой?

Не уверена, что я правильно нашла narrowPeak файл. Команда для получения fasta из narrowPeak: `bedtools getfasta -fi /anastasia_rastvorova/AtacPract/results/genome/genome.fa -bed /anastasia_rastvorova/AtacPract/results/bwa/mergedLibrary/macsnarrowPeak/OSMOTIC_STRESS_T0_R2.mLb.cIN_peaks.narrowPeak -fo narrowPeak_test.fa`

	Logo	E-value ?	Sites ?	Width ?
1.		1.2e-137	368	20
2.		3.9e-025	115	14
3.		7.2e-008	276	11
4.		8.4e-007	14	20
5.		1.0e-006	19	20

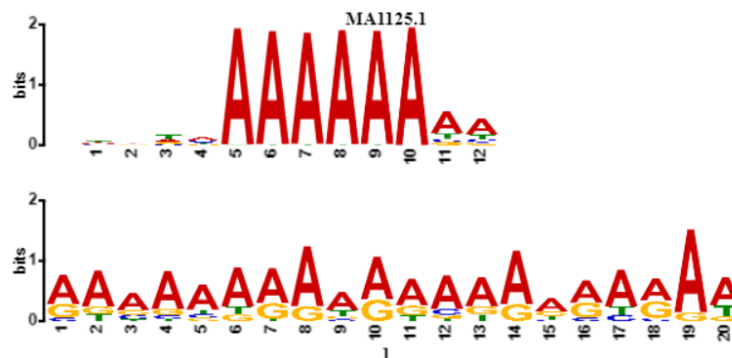
Прилагаю сразу все лого. Первое лого является статистически значимым, собственно, как и остальные (чем меньше E-value, тем лучше). Хотя лого 4 и 5 имеют большее информационное содержание, поэтому кажутся мне более интересными.

Задание 10 (1 балла): запустите поиск похожих мотивов с помощью Tomtom. Приложите лого наиболее похожего. Является ли эта находка статистически значимой? Опишите функцию данного транскрипционного фактора.

#### Summary

<b>Name</b>	<a href="#">MA1125.1 (ZNF384)</a>
<b>Database</b>	JASPAR2018_CORE vertebrates non-redundant
<b>p-value</b>	8.02e-05
<b>E-value</b>	1.45e-01
<b>q-value</b>	2.90e-01
<b>Overlap</b>	12
<b>Offset</b>	-1
<b>Orientation</b>	Normal
	<a href="#">Show logo download options</a>

#### Optimal Alignment



Tomtom решил, что наиболее похожий мотив для первого, это человеческий (?) мотив ZNF384 транскрипционного фактора из группы «цинковых пальцев». Они связывают белки, регулирующие транскрипцию.

Находка статистически значимая, но не так чтобы очень сильно.

Информация из Uniprot: Фактор транскрипции, связывающий консенсусную последовательность ДНК [GC] AAAAAA. Видимо, что они связывают и регулируют промоторы MMP1, MMP3, MMP7 и COL1A1.