

Отчет по финальному проекту
Алекберова Артема, группа 1

1. Сбор данных

Я взял гистоновую метку H3K27ac

Скачиваем данные:

```
wget
```

```
https://www.encodeproject.org/files/ENCFF491LQY/@@download/ENCFF491LQY.bed.gz
```

```
zcat ENCFF491LQY.bed.gz | cut -f1-5 > H3K27ac_MCF-7.ENCFF491LQY.hg38.bed
```

```
wget
```

```
https://www.encodeproject.org/files/ENCFF340KSH/@@download/ENCFF340KSH.bed.gz
```

```
zcat ENCFF340KSH.bed.gz | cut -f1-5 > H3K27ac_MCF-7.ENCFF340KSH.hg38.bed
```

Преобразовыве hg38 в hg19:

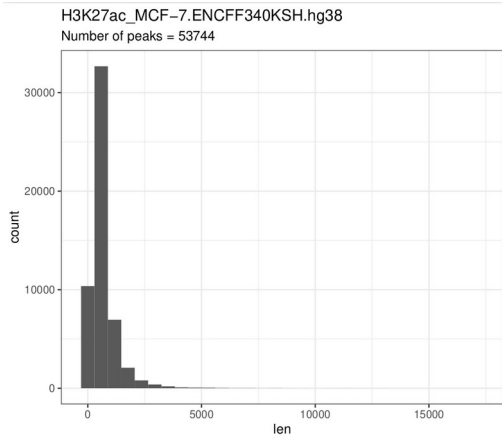
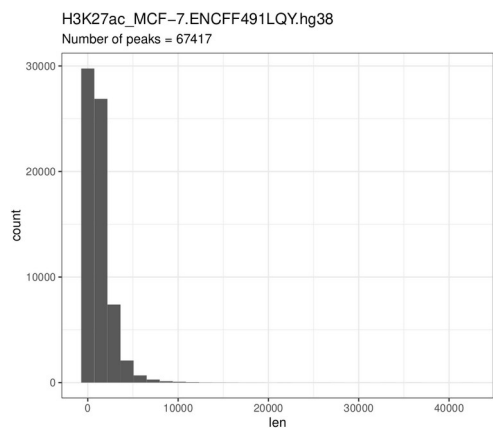
```
liftOver H3K27ac_MCF-7.ENCFF491LQY.hg38.bed hg38ToHg19.over.chain.gz  
H3K27ac_MCF-7.ENCFF491LQY.hg19.bed H3K27ac_MCF-7.ENCFF491LQY.unmapped.bed
```

```
liftOver H3K27ac_MCF-7.ENCFF340KSH.hg38.bed hg38ToHg19.over.chain.gz  
H3K27ac_MCF-7.ENCFF340KSH.hg19.bed H3K27ac_MCF-7.ENCFF340KSH.unmapped.bed
```

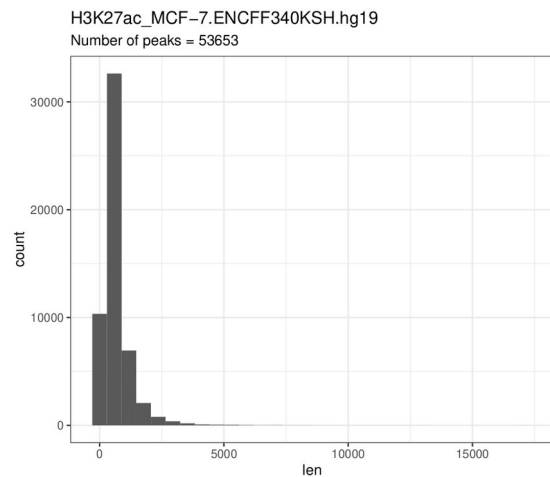
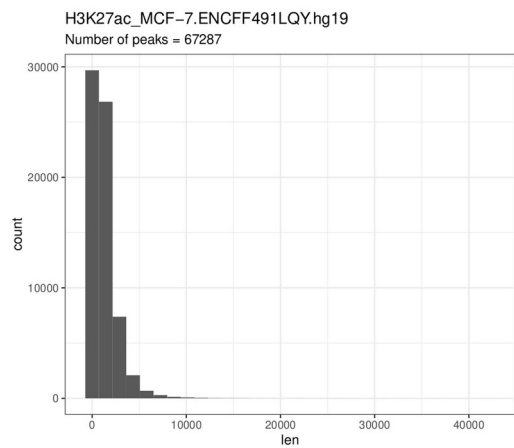
2. Гистограммы длин

Строим гистограмму длин участков для каждого эксперимента до и после конвертации к нужной версии генома. Количество пиков можно увидеть в подписи на графике

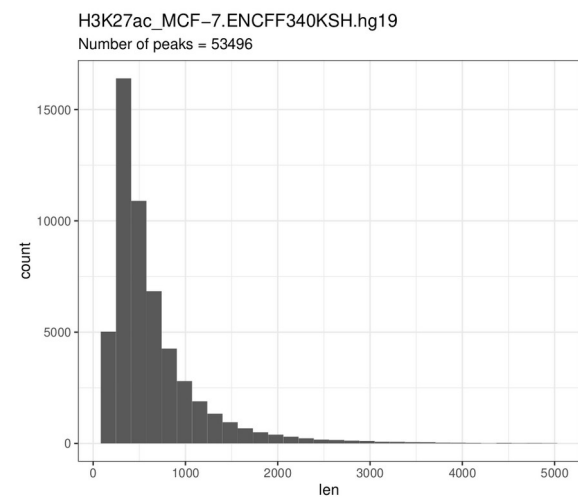
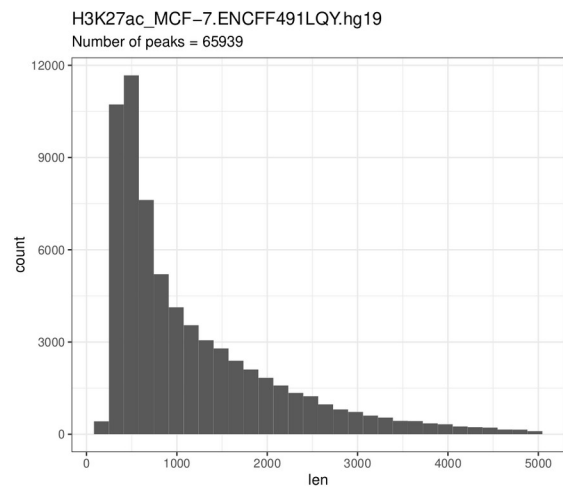
hg38:



hg19:



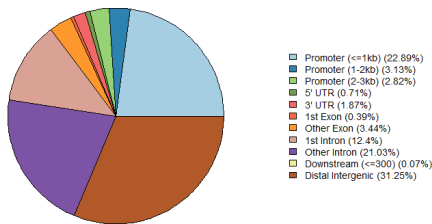
filtered hg19 (в качестве порога взята длина 5.000):



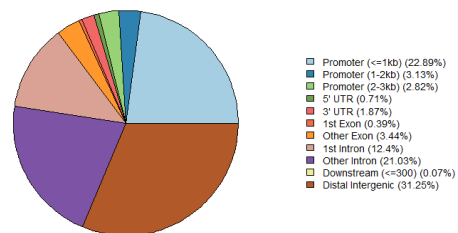
3. Пай-чарты

Смотрим, где располагаются пики гистоновой метки относительно аннотированных генов:

ENCFF340KSH



ENCFF491LQY



Видим, что где-то четверть пиков приходится на промютеры

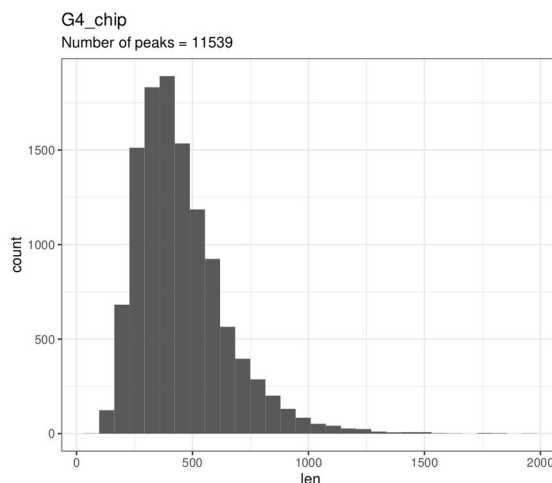
Объединяем пики с помощью команды:

```
cat *.filtered.bed | sort -k1,1 -k2,2n | bedtools merge > H3K27ac_MCF-7.merged.bed
```

4. Анализ участков вторичной стр-ры ДНК

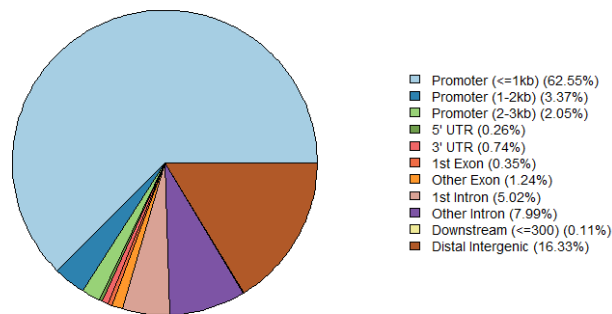
Я выбрал G4_ChIP

а) Гистограмма и количество пиков



b) Пай-чарт:

Видим, что большая часть пиков приходится на промоторы

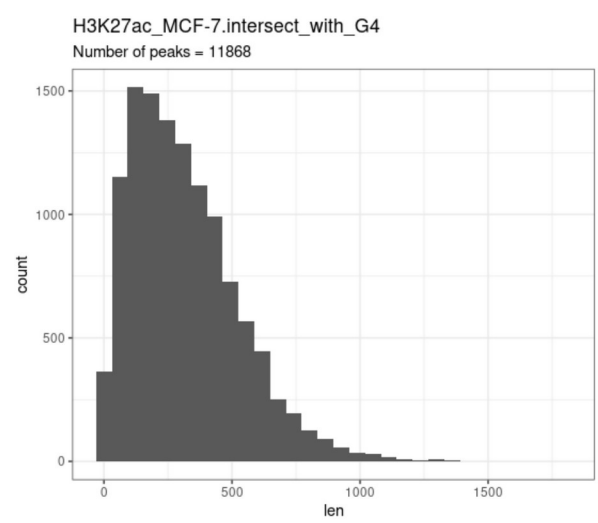


5. Анализ пересечений гистоновой метки и стр-ры ДНК

Находим пересечения гистоновой меткой и стр-ры ДНК с помощью следующей команды:

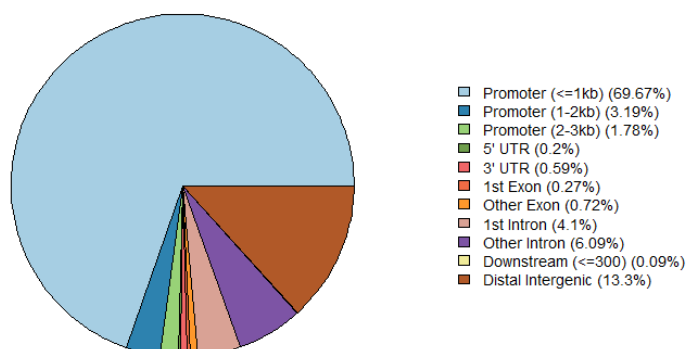
```
bedtools intersect -a G4_chip.bed -b H3K27ac_MCF-7.merged.bed >  
H3K27ac_MCF-7.intersect_with_G4.bed
```

а) Гистограмма длин и количество пиков:



b) Пай-чарт:

Видим, что большая часть пиков также приходится на промоутеры



6. Визуализация в геномном браузере

Сессия: http://genome.ucsc.edu/s/art591/hse21_H3K27ac_MCF-7_human

Визуализация выполнена следующими командами:

- добавляем ENCFF491LQY:

```
track visibility=dense name="ENCFF491LQY"
```

```
description="H3K27ac_MCF-7.ENCFF491LQY.hg19.filtered.bed"
```

```
https://raw.githubusercontent.com/art591/hse21_H3K27ac_G4_human/main/data/H3K27ac_MCF-7.ENCFF491LQY.hg19.filtered.bed
```

- добавляем ENCFF340KSH

```
track visibility=dense name="ENCFF340KSH"
```

```
description="H3K27ac_MCF-7.ENCFF340KSH.hg19.filtered.bed"
```

```
https://raw.githubusercontent.com/art591/hse21_H3K27ac_G4_human/main/data/H3K27ac_MCF-7.ENCFF340KSH.hg19.filtered.bed
```

- добавляем их объединение:

```
track visibility=dense name="ChIP_merge" color=50,50,200
```

```
description="H3K27ac_MCF-7.merge.hg19.bed"
```

https://raw.githubusercontent.com/art591/hse21_H3K27ac_G4_human/main/data/H3K27ac_MCF-7.merged.bed

- добавляем G4:

`track visibility=dense name="G4_chip" color=0,200,0`
`description="G4_chip"`

https://raw.githubusercontent.com/art591/hse21_H3K27ac_G4_human/main/data/G4_chip.bed

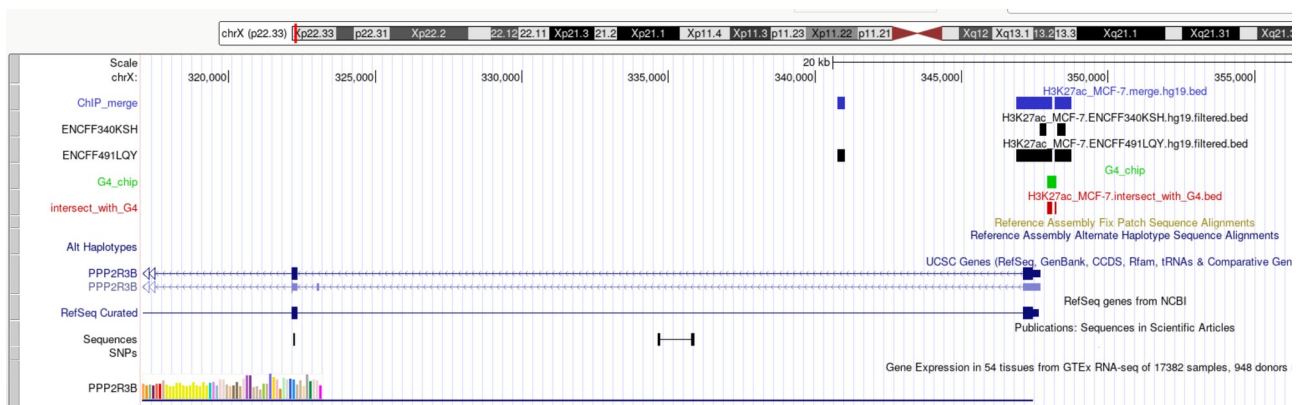
- добавляем пересечение с G4:

`track visibility=dense name="intersect_with_G4" color=200,0,0`
`description="H3K27ac_MCF-7.intersect_with_G4.bed"`

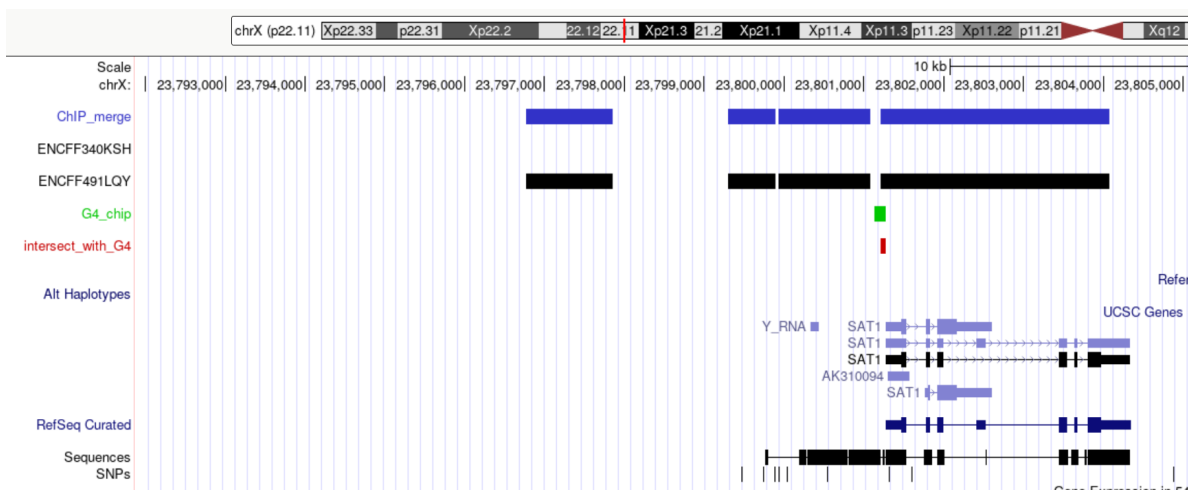
https://raw.githubusercontent.com/art591/hse21_H3K27ac_G4_human/main/data/H3K27ac_MCF-7.intersect_with_G4.bed

Скриншоты:

chrX:317,036-384,136



chrX:23,791,891-23,822,256:



7. Ассоциируем полученные пересечения с ближайшими генами с помощью ChIPpeakAnno

Кол-во пиков, которые удалось проассоциировать с генами: 8567

Общее кол-во уникальных генов: 6528

GO-анализ для полученных уникальных генов:

Наиболее значимые категории:

GO biological process complete	expected	Fold Enrichment	+/-	raw P value	▼ FDR
mitotic nuclear membrane reassembly	15.20	2.04	+	2.32E-03	4.98E-02
positive regulation of telomere maintenance	15.20	2.04	+	2.32E-03	4.97E-02
mitotic nuclear membrane organization	15.20	2.04	+	2.32E-03	4.97E-02
neutrophil migration	25.93	.39	-	2.25E-03	4.85E-02