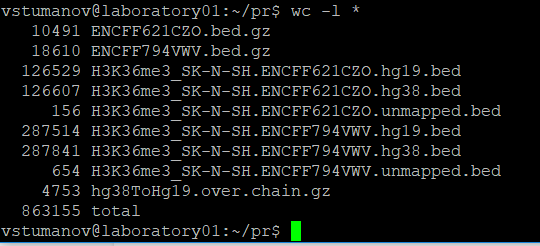
Отчёт по проекту по биоинформатике.

Туманов Вадим Сергеевич, группа 2.

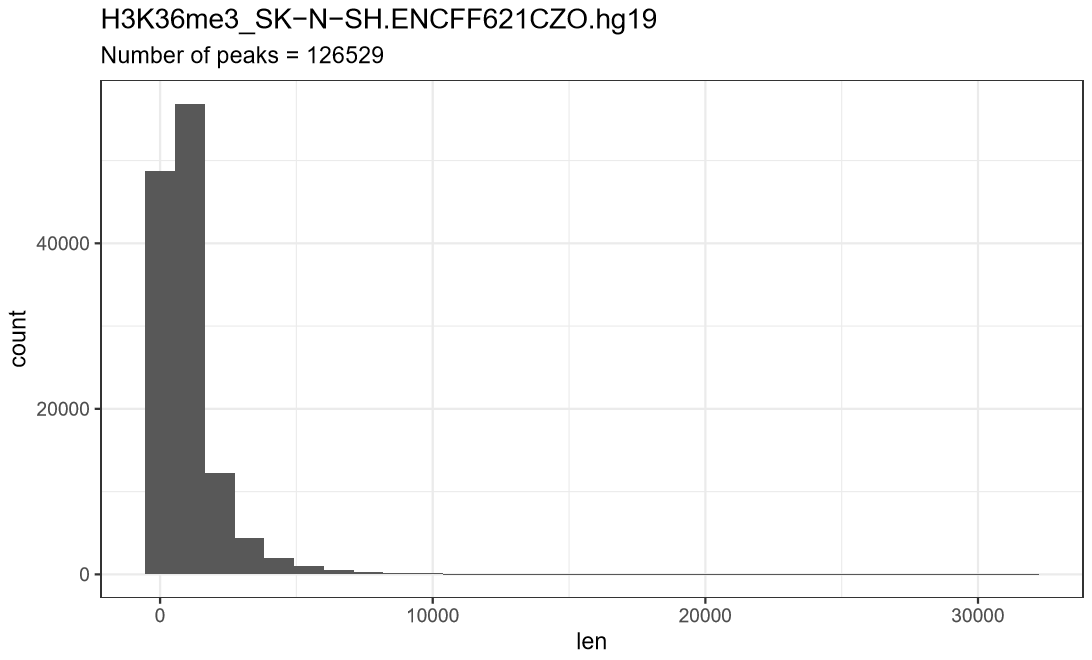
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <https://github.com/vadosi/hse21_H3K36me3_ZDNA_human> | | | | human(hg19) |
| ZDNA\_DeepZ | H3K36me3 | SK-N-SH | ENCFF794VWV | ENCFF621CZO | |

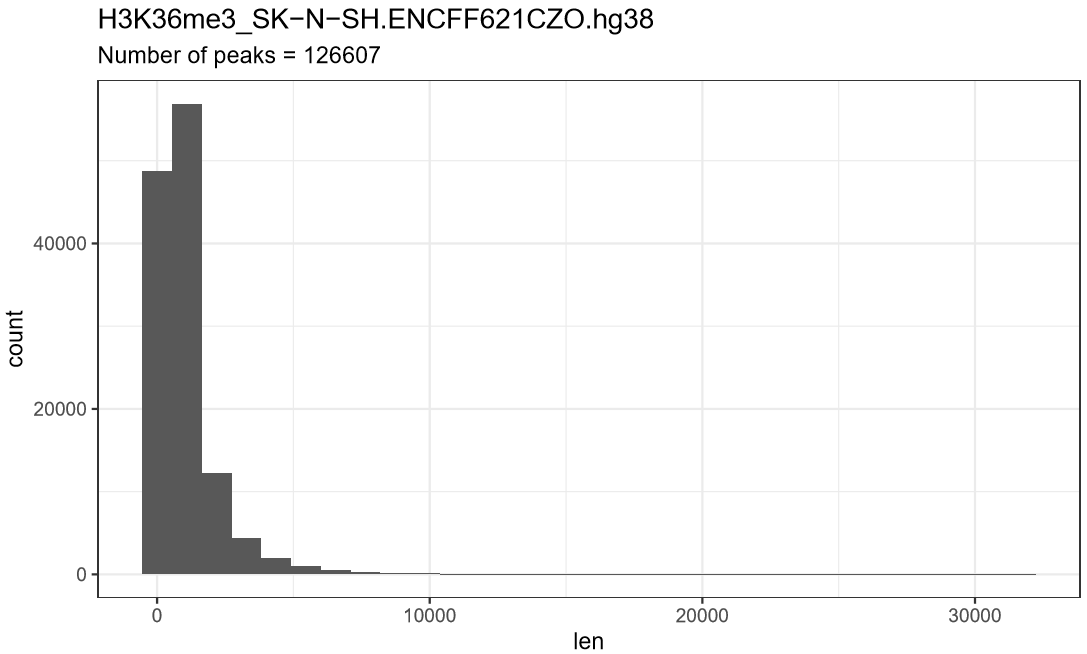
Были выбраны bed-файлы пиков из разных экспериментов.

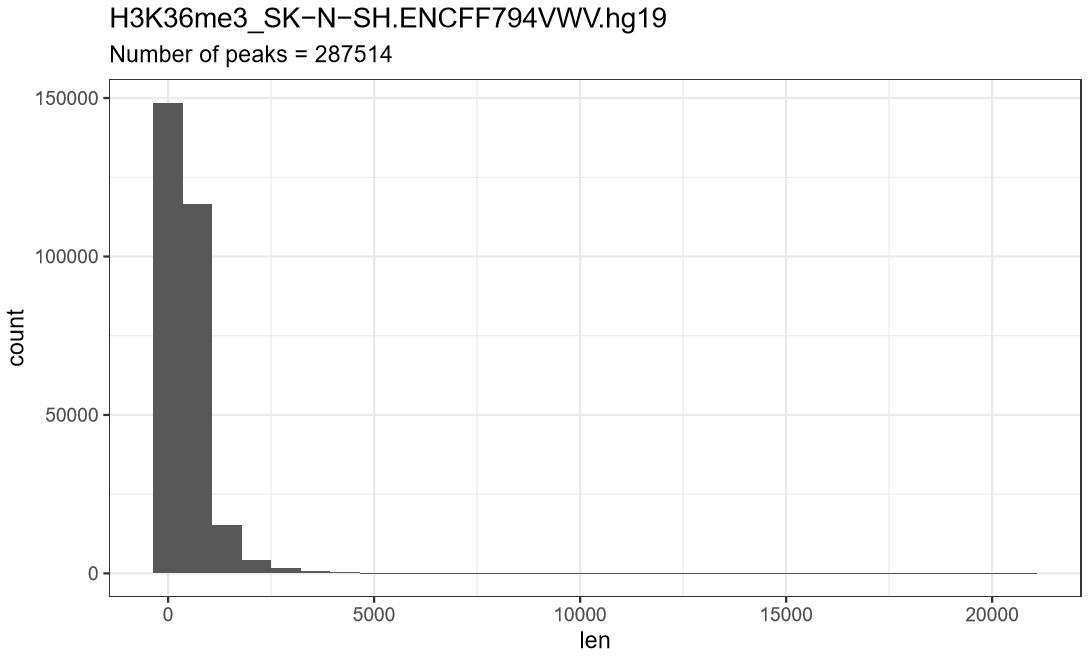
С помощью liftOver приводим hg38 к hg19.

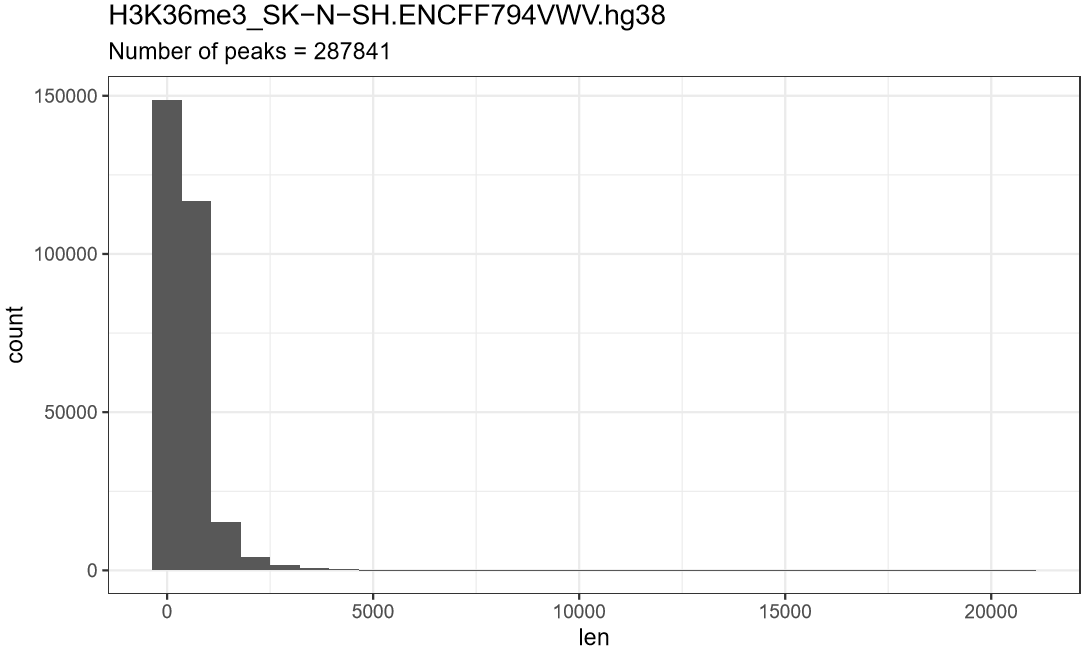


Гистограммы для hg19/hg38 (количество unmapped пиков минимально, это хорошо):



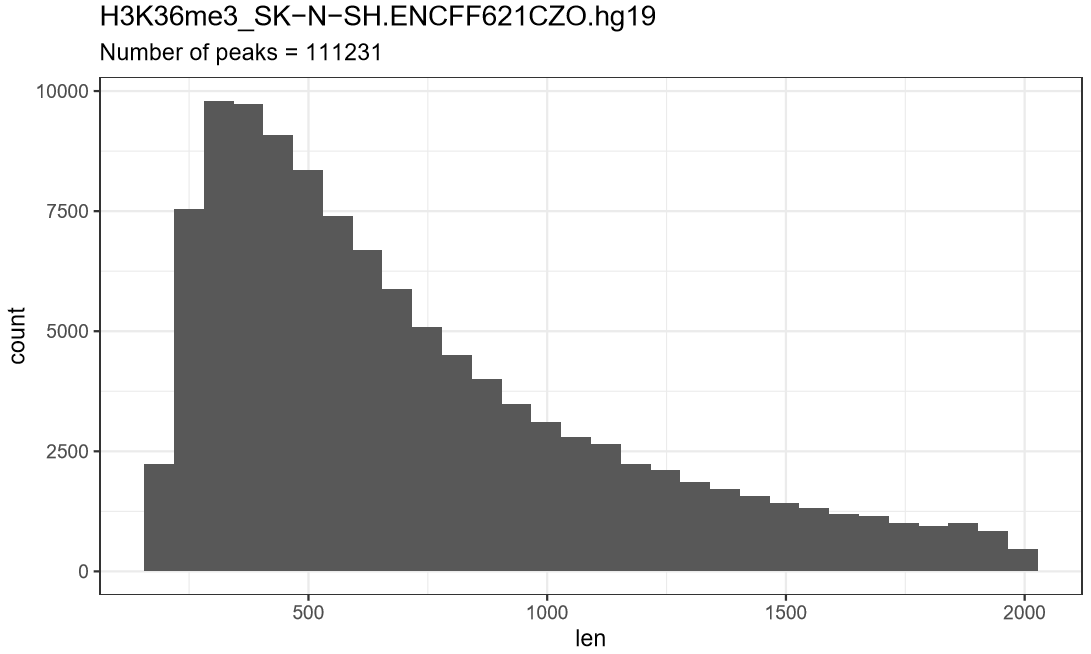




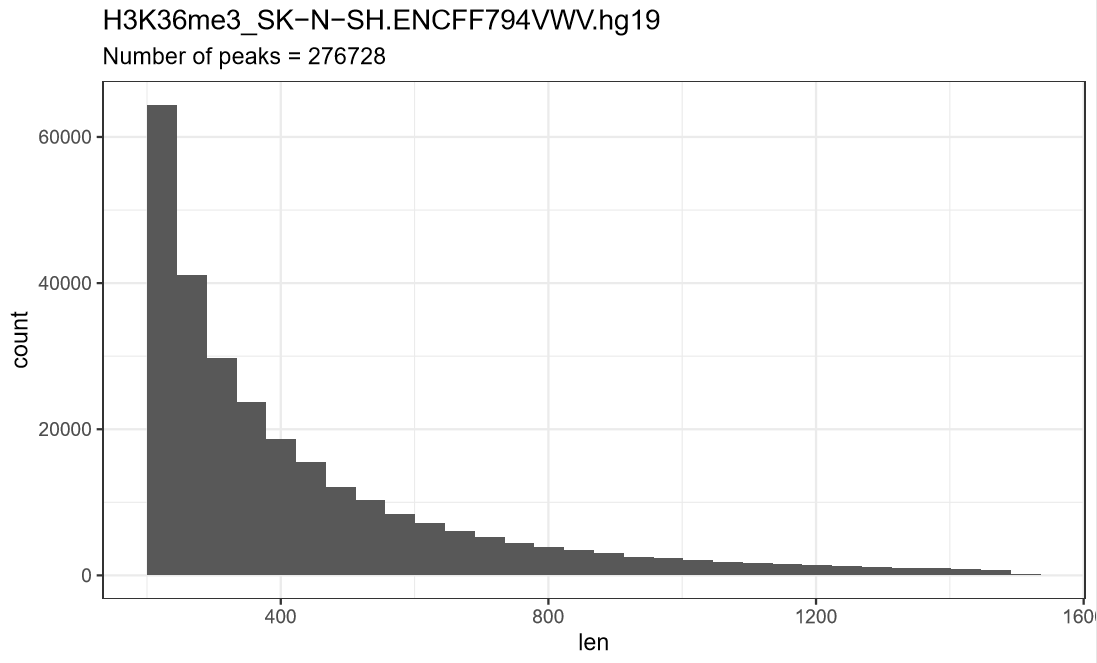


Отфильтрованные пики (количество указано на гистограмме):

* Для одного файла оптимальный порог длины пика – 2000:

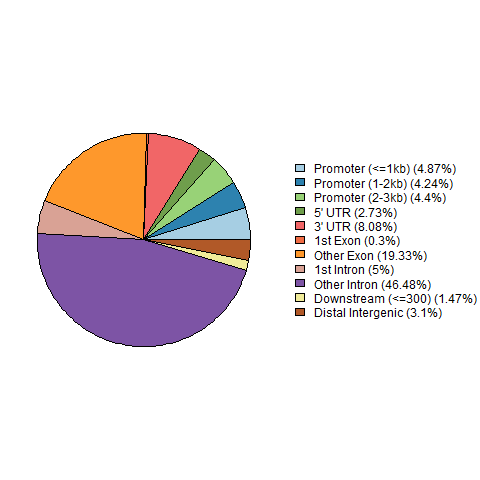


* Для другого – 1600:

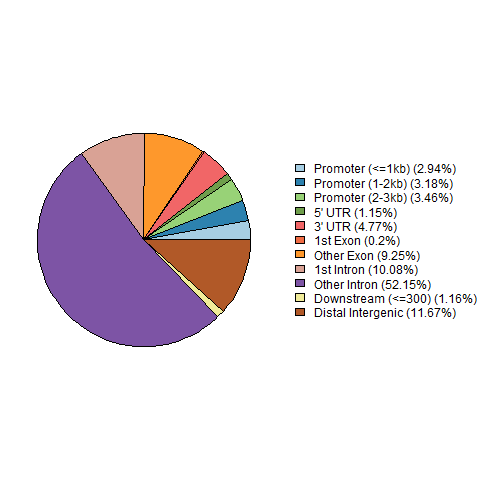


Пайчарты для:

* H3K36me3\_SK-N-SH.ENCFF621CZO.hg19.filtered



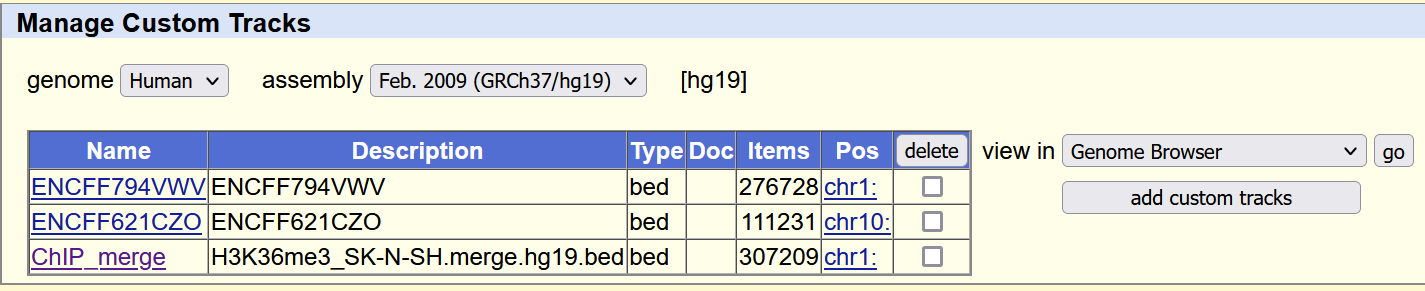
* H3K36me3\_SK-N-SH.ENCFF794VWV.hg19.filtered

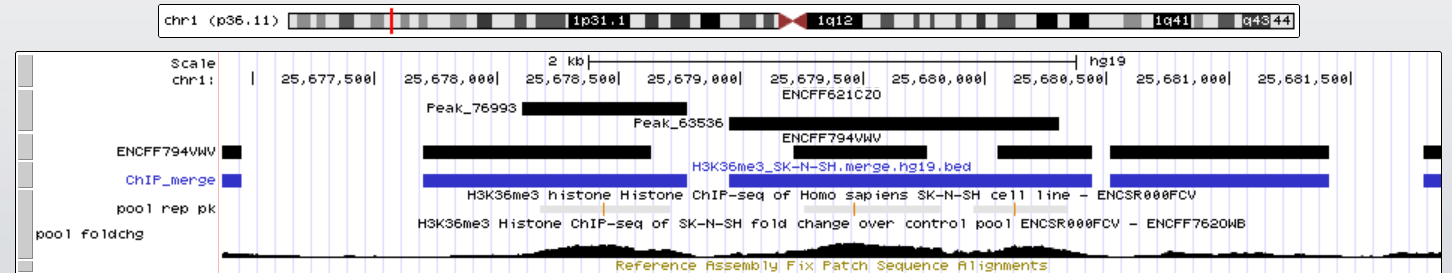
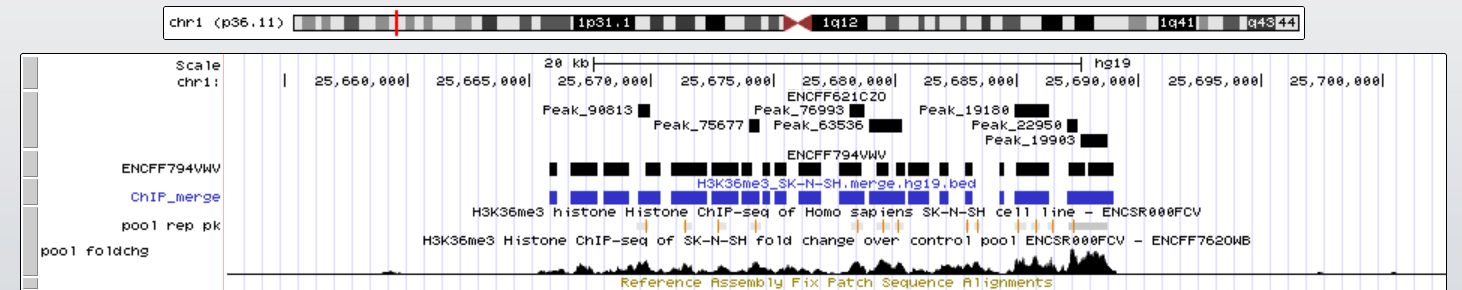


После использования betlools merge был получен следующий файл с 307209 пиками:

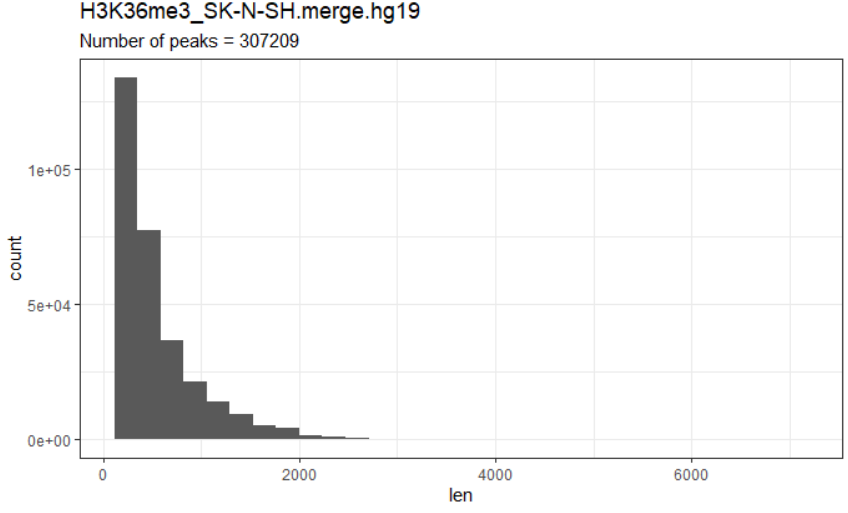
H3K36me3\_SK-N-SH.merge.hg19.bed

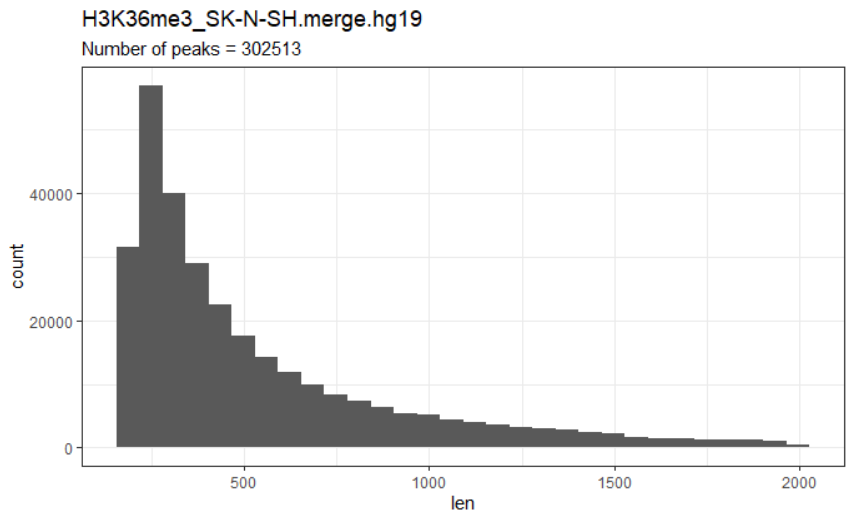
Визуализация в UCSC Genome Browser:

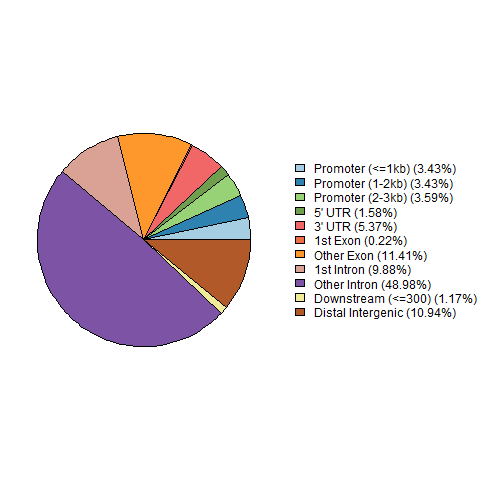




До фильтрации:



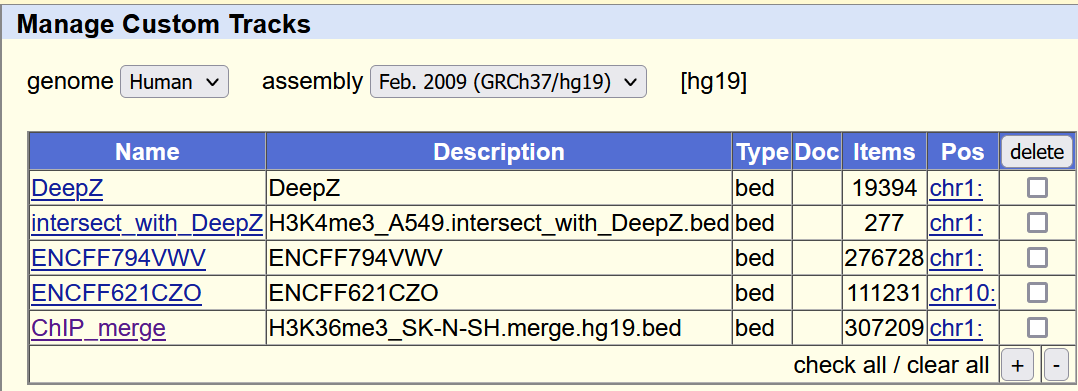
После: 

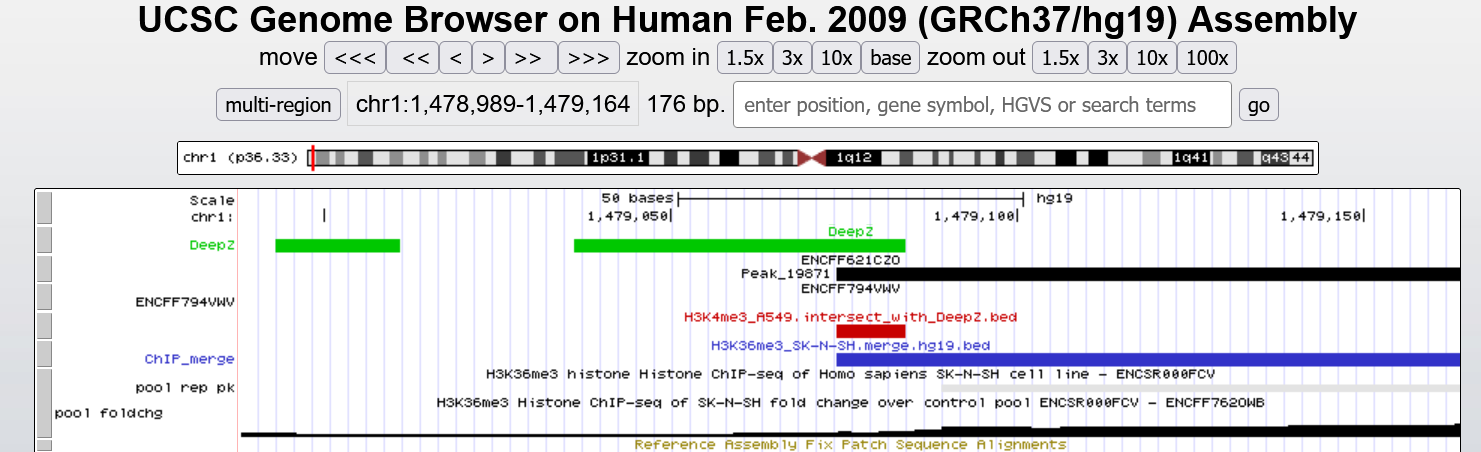


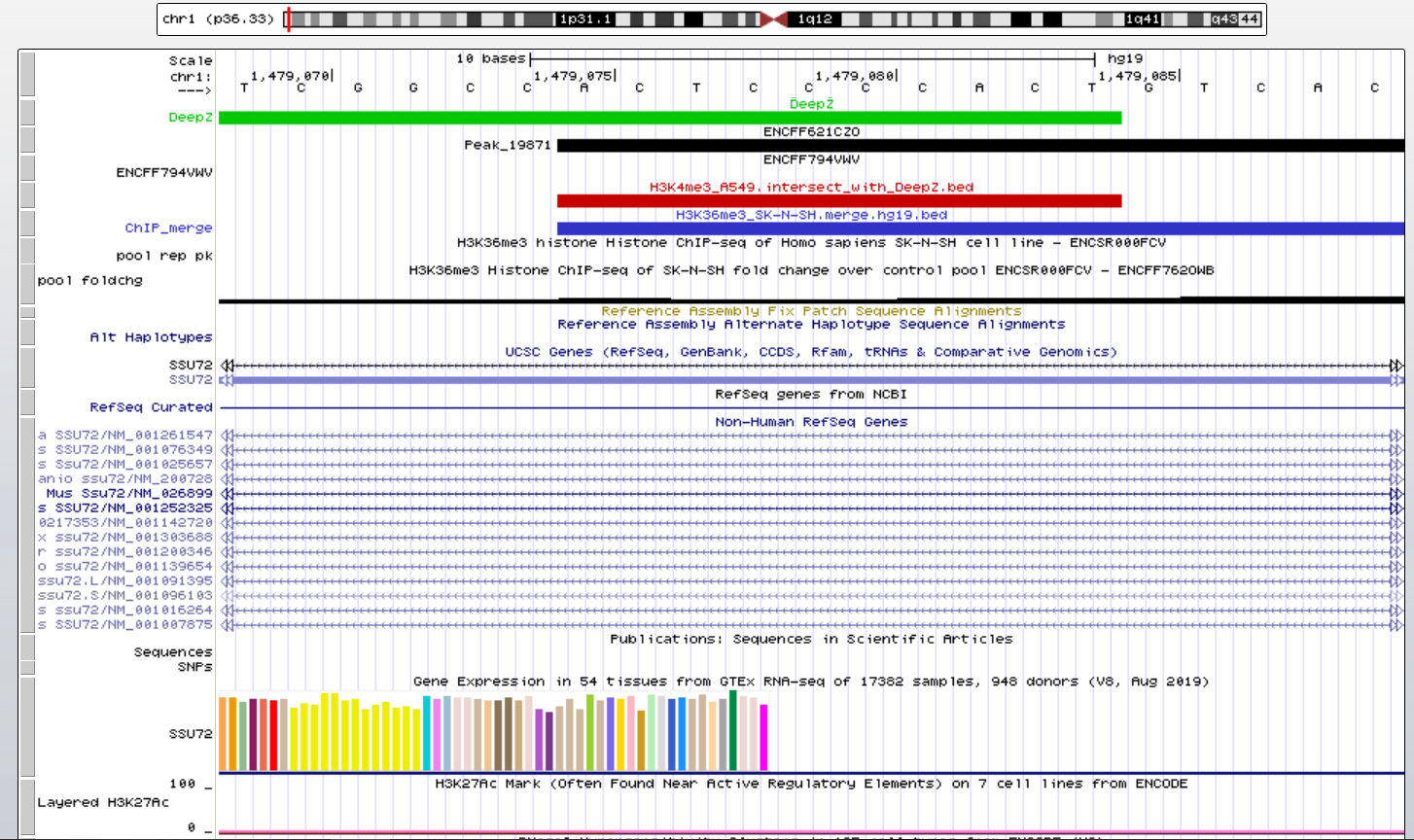
Использовался то же код, что и в предыдущих заданиях, только переменной NAME присваивалось другое значение (название merge-файла).

На пересечении с ZDNA\_DeepZ обнаруживаем 277 пиков.

Возвращаемся в genome browser:







**Position:** [chr1:1479075-1479084](http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=1124321089_AaQNSECDPrClbZtpecd8UVcMEmie&db=hg19&position=chr1%3A1479075-1479084)  
**Band:** 1p36.33  
**Genomic Size:** 10

Аннотированный ген: [SSU72](https://www.genecards.org/cgi-bin/carddisp.pl?gene=SSU72)

Сессия в браузере сохранена в папке src на github.

Проассоциировав полученные пересечения с ближайшими генами, обнаружили 39 пиков, ассоциированных с 32 уникальными генами.

GO-анализ не обнаружил статистически значимых результатов:

