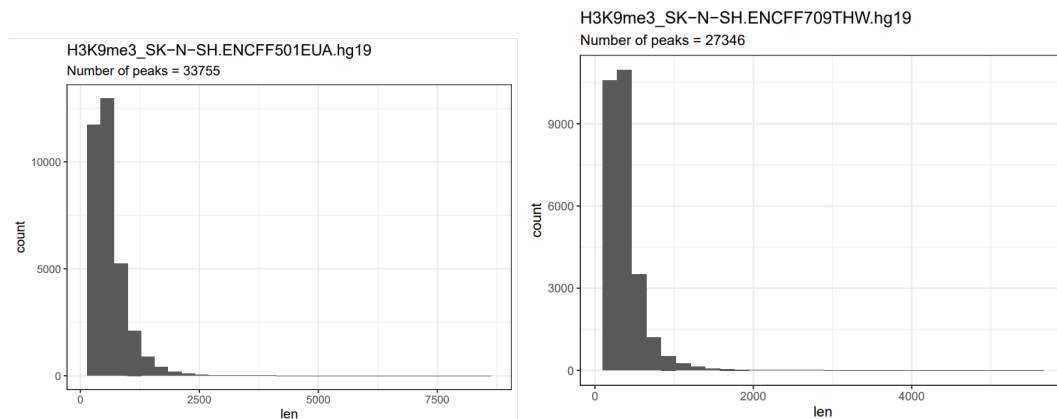
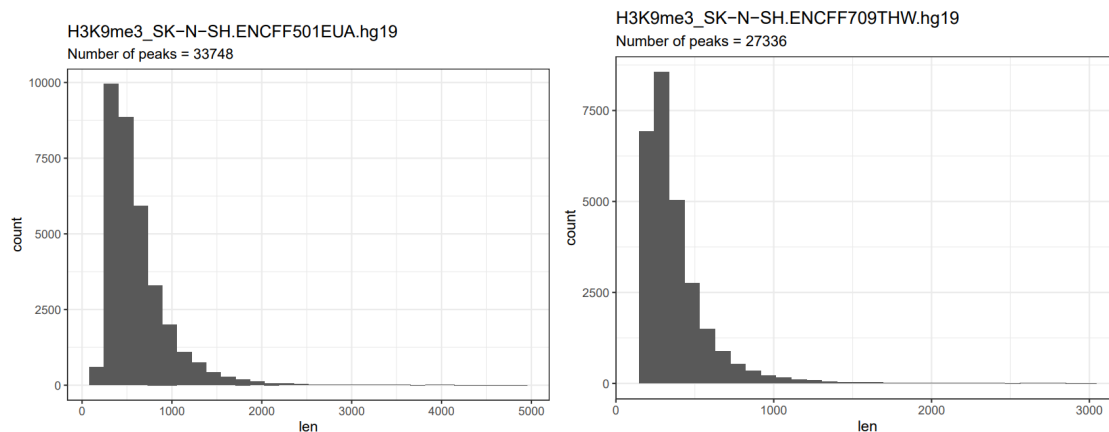


Я сразу скачала файлы версии hg19, поэтому часть с изменением версии и сравнением гистограмм пропустим.

Дальше построила гистограммы распределения пиков:

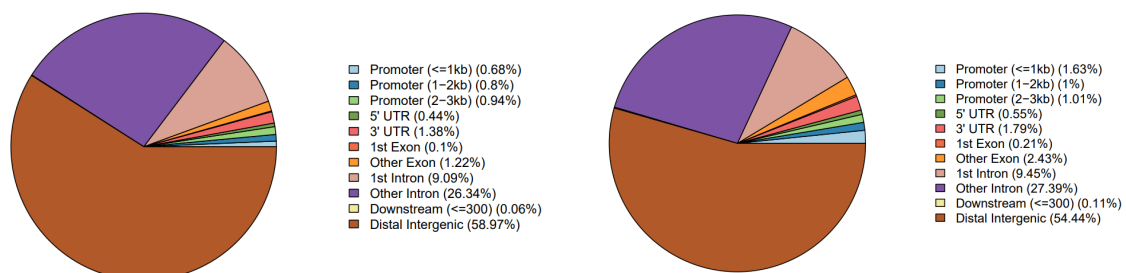


Получаем, что пиков в файлах 33755 и 27346, соответственно. Чтобы убрать слишком длинные, я установила пороги в 5000кб и 3000кб. После фильтрации получаем гистограммы:



Осталось 33748 и 27336 пиков, то есть 7 и 10 пиков отфильтровались.

Дальше строим график того, где располагаются пики:



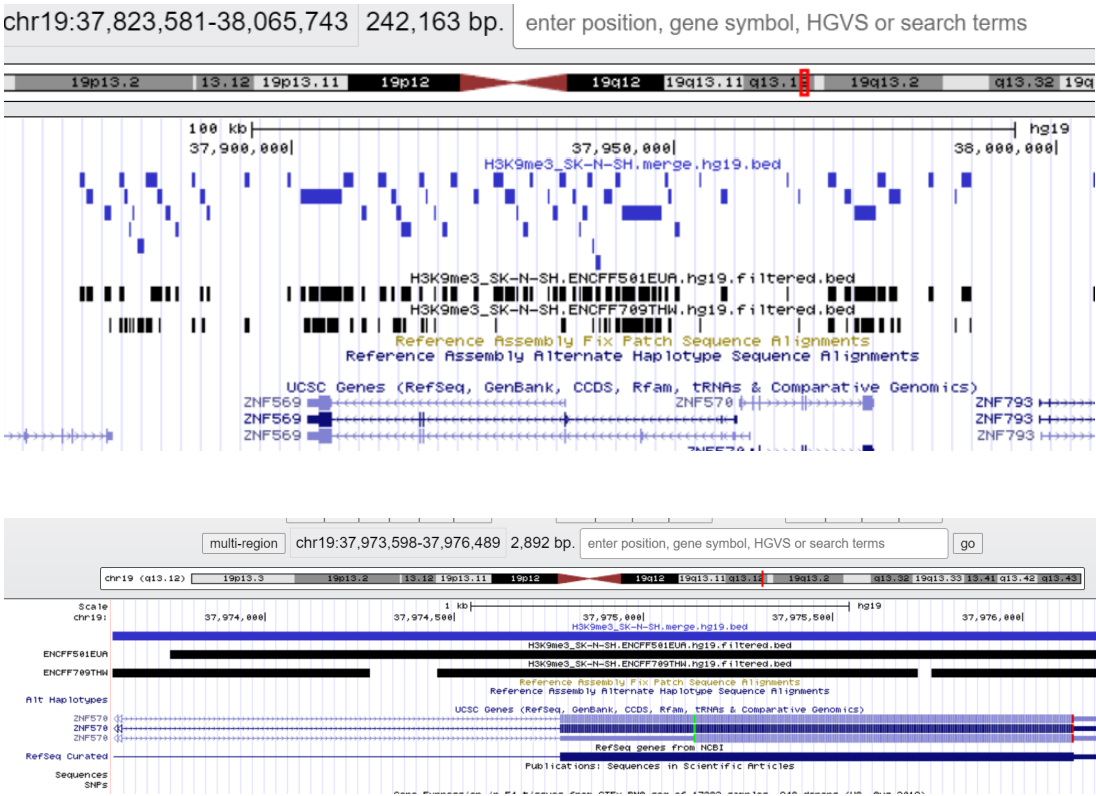
Теперь объединяем два отфильтрованных файла командой:

```
cat *.filtered.bed | sort -k1,1 -k2,2n | bedtools merge > H3K9me3_SK-N-SH.merge.hg19.bed
```

Загружаем все на гитхаб и оттуда в геномный браузер:

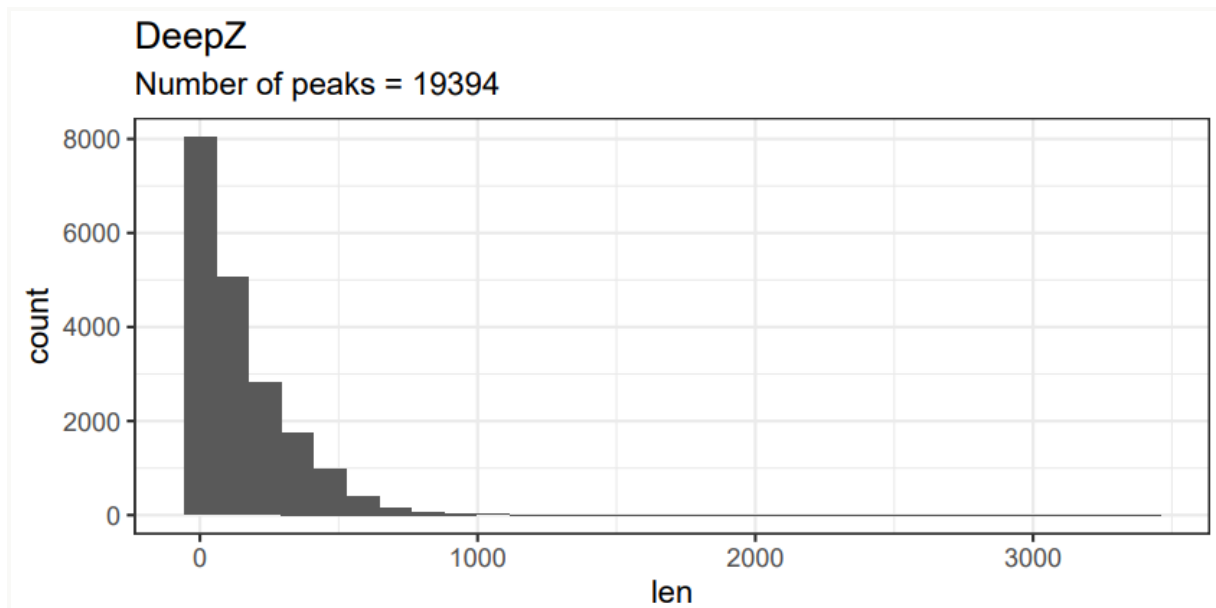
Manage Custom Tracks						
genome: Human assembly: Feb. 2009 (GRCh37/hg19) [hg19]						
Name	Description	Type	Doc	Items	Pos	delete
ENCFF709THW	H3K9me3_SK-N-SH.ENCFF709THW.hg19.filtered.bed	bed		27336	chr19:	<input type="checkbox"/>
ENCFF501EUA	H3K9me3_SK-N-SH.ENCFF501EUA.hg19.filtered.bed	bed		33748	chr7:	<input type="checkbox"/>
ChIP_merge	H3K9me3_SK-N-SH.merge.hg19.bed	bed		53766	chr1:	<input type="checkbox"/>

Видим, что есть участки, где данные пересекаются, например, на 19 хромосоме:

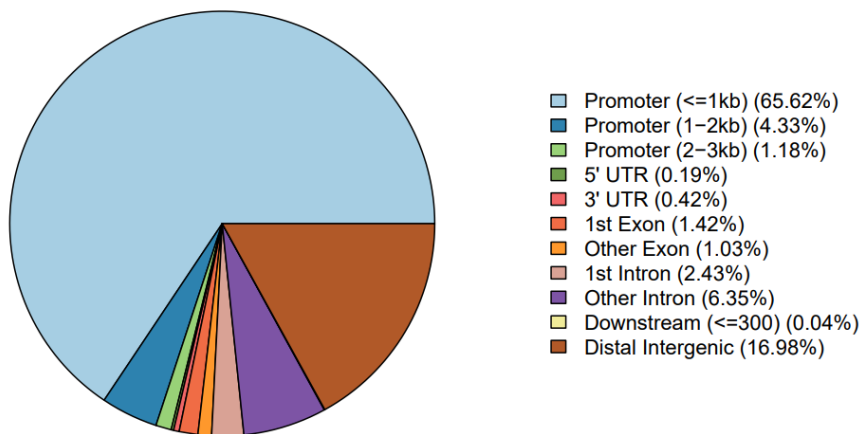


Запишем место: chr19:37,973,598-37,976,489

Далее строим гистограмму длин для вторичной структуры ДНК. Получилось 19394 пика.



Строим график того, куда попадает вторичная структура ДНК:



(Очень странно, что БОльшая часть попала в промотеры, а на предыдущих графиках было <=2%...)

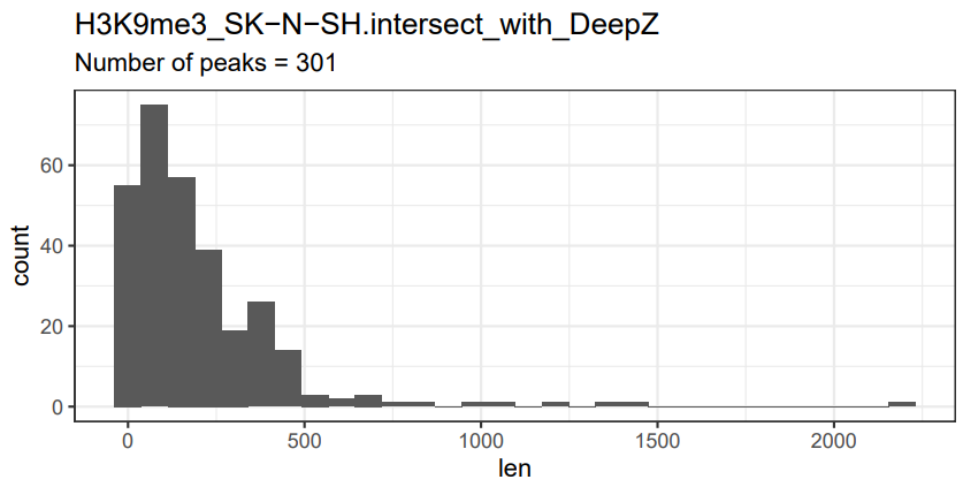
Пересекаем merge с DeepZ с помощью команды:

```
bedtools intersect -a DeepZ.bed -b H3K9me3_SK-N-SH.merge.hg19.bed >
H3K9me3_SK-N-SH.intersect_with_DeepZ.bed
```

У меня как-то особенно неудачно получилось....

```
asbogdanova_2@laboratory01:~/github/hse21_H3K9me3_G4_human/data$ wc -l *
19394 DeepZ.bed
33755 H3K9me3_SK-N-SH.ENCFF501EUA.hg19.bed
33748 H3K9me3_SK-N-SH.ENCFF501EUA.hg19.filtered.bed
27346 H3K9me3_SK-N-SH.ENCFF709THW.hg19.bed
27336 H3K9me3_SK-N-SH.ENCFF709THW.hg19.filtered.bed
  301 H3K9me3_SK-N-SH.intersect_with_DeepZ.bed
53766 H3K9me3_SK-N-SH.merge.hg19.bed
195646 total
```

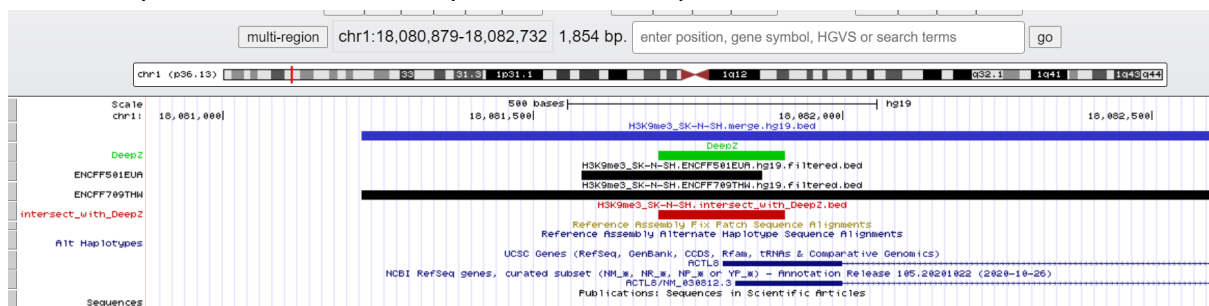
Ну, ладно, всего 301 пик.



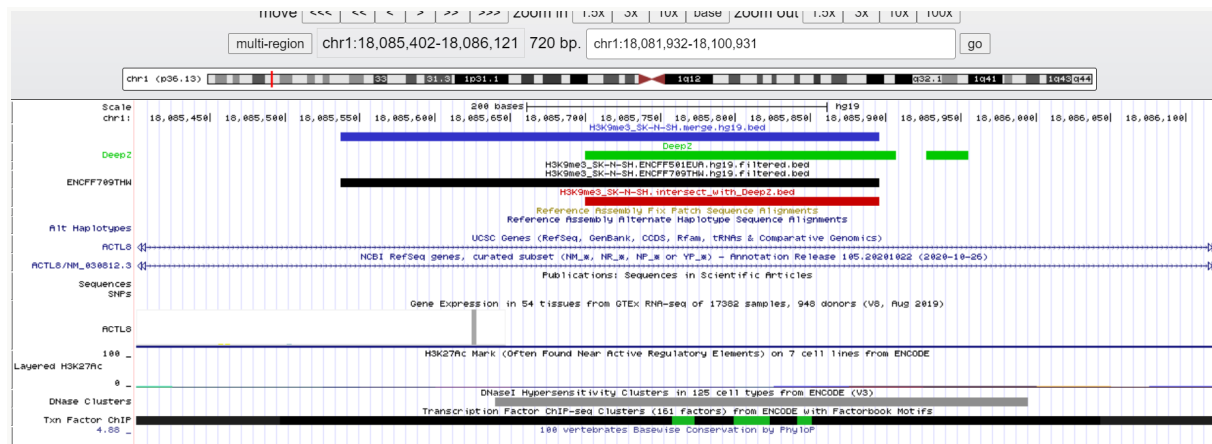
Добавляем данные в геномный браузер:

Manage Custom Tracks								
genome: Human		assembly: Feb. 2009 (GRCh37/hg19) [hg19]						
Name	Description	Type	Doc	Items	Pos	delete	view in	
intersect_with_DeepZ	H3K9me3_SK-N-SH.intersect_with_DeepZ.bed	bed		301	chr1:	<input type="checkbox"/>	Genome Browser	
DeepZ	DeepZ	bed		19394	chr1:	<input type="checkbox"/>	add c	
ENCFF709THW	H3K9me3_SK-N-SH.ENCFF709THW.hg19.filtered.bed	bed		27336	chr19:	<input type="checkbox"/>		
ENCFF501EUA	H3K9me3_SK-N-SH.ENCFF501EUA.hg19.filtered.bed	bed		33748	chr7:	<input type="checkbox"/>		
ChIP_merge	H3K9me3_SK-N-SH.merge.hg19.bed	bed		53766	chr1:	<input type="checkbox"/>		
check all / clear all						<input type="checkbox"/>		

Вот на 1 хромосоме осталось пересечение с DeepZ



chr1:18,081,703-18,081,908



chr1:18,085,402-18,086,121

Ссылка на сессию в геномном браузере:

http://genome.ucsc.edu/s/bbogdanovaalina/hse21_H3K9me3_G4_hum

Проассоциируем пики с генами, получилось всего 24 уникальных гена и 36 пиков, которые с ними ассоциируются.

H3K4me3_A549.intersect_with_DeepZ.genes_uniq.txt – Блокнот

Файл Правка Формат Вид Справка

ACTL8
FAAH
KDM3A
MSL3P1
LRRN1
FEM1C
RPS14
RNF130
ELOVL2
LOC650226
RNF32
EBLN3P
C10orf88
DHX32
FANK1
YPEL3
LINC00273
GPC5C
ZSCAN1
ZNF544
MIR663A
MIR663AHG
NA
MIR3648-1

Data		
anno	Formal class	GRanges
..@ seqnames	: Formal class 'Rle' [pa	
.. .. @ values	: Factor w/ 23 lev	
.. .. @ lengths	: int [1:13] 2 3	
.. .. @ elementMetadata	: NULL	
.. .. @ metadata	: list()	
..@ ranges	: Formal class 'IRanges'	
.. .. @ start	: int [1:36] 1808	
.. .. @ width	: int [1:36] 205	
.. .. @ NAMES	: chr [1:36] "X00	
.. .. @ elementType	: chr "ANY"	
.. .. @ elementMetadata	: NULL	
.. .. @ metadata	: list()	
..@ strand	: Formal class 'Rle' [pa	
.. .. @ values	: Factor w/ 3 lev	
.. .. @ lengths	: int 36	

В целом, по такому маленькому кол-ву генов вряд ли можно получить статистически значимый результат...

	Homo sapiens (REF)	Client Text Box Input (Hierarchy) NEW!					
	#	#	expected	Fold Enrichment	+/-	raw P value	Δ FDR
3Q biological process complete	7	2	.01	> 100	+	2.89E-05	1.14E-01
atty acid elongation, monounsaturated fatty acid	7	2	.01	> 100	+	2.89E-05	1.52E-01
atty acid elongation, polyunsaturated fatty acid	13	2	.01	> 100	+	8.39E-05	1.89E-01
atty acid elongation	99	3	.09	32.85	+	1.08E-04	2.13E-01
ery long-chain fatty acid biosynthetic process	13	2	.01	> 100	+	8.39E-05	2.21E-01
atty acid metabolic process	330	4	.30	13.14	+	2.16E-04	2.28E-01
atty acid elongation, saturated fatty acid	7	2	.01	> 100	+	2.89E-05	2.28E-01
ong-chain fatty acid metabolic process	117	3	.11	27.79	+	1.74E-04	2.29E-01
atty acid biosynthetic process	124	3	.11	26.22	+	2.06E-04	2.32E-01
nsaturated fatty acid metabolic process	107	3	.10	30.39	+	1.35E-04	2.36E-01
ong-chain fatty-acyl-CoA biosynthetic process	19	2	.02	> 100	+	1.67E-04	2.40E-01
noleic acid metabolic process	21	2	.02	> 100	+	2.01E-04	2.45E-01
lefinic compound metabolic process	113	3	.10	28.78	+	1.58E-04	2.49E-01
membrane lipid biosynthetic process	138	3	.13	23.56	+	2.80E-04	2.60E-01
lpha-linolenic acid metabolic process	13	2	.01	> 100	+	8.39E-05	2.65E-01
ong-chain fatty-acyl-CoA metabolic process	25	2	.02	86.72	+	2.79E-04	2.75E-01
atty-acyl-CoA biosynthetic process	31	2	.03	69.93	+	4.18E-04	3.67E-01
phingolipid metabolic process	164	3	.15	19.83	+	4.60E-04	3.83E-01

Судя по количеству генов, самая значимая категория - процесс расщепления и синтеза жирных кислот.

