

Федеральное государственное автономное образовательное учреждение  
высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

ИТОГОВЫЙ ПРОЕКТ ПО МАЙНОРУ «БИОИНФОРМАТИКА»

Выполнила студентка группы №2:

Ямпольская София Андреевна, ББИ184

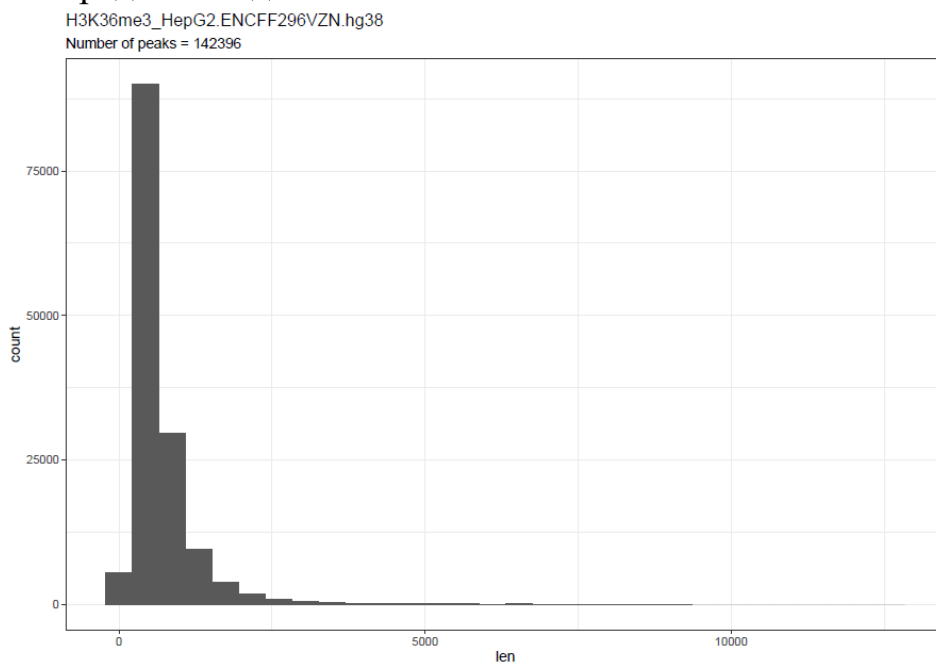
Москва 2021

1. Построение распределения длин пиков до и после конвертации в hg19, фильтрация длинных пиков, расположение пиков относительно аннотированных генов.

Оба файла были в сборке hg38, поэтому их надо преобразовать в hg19 с помощью liftOver.

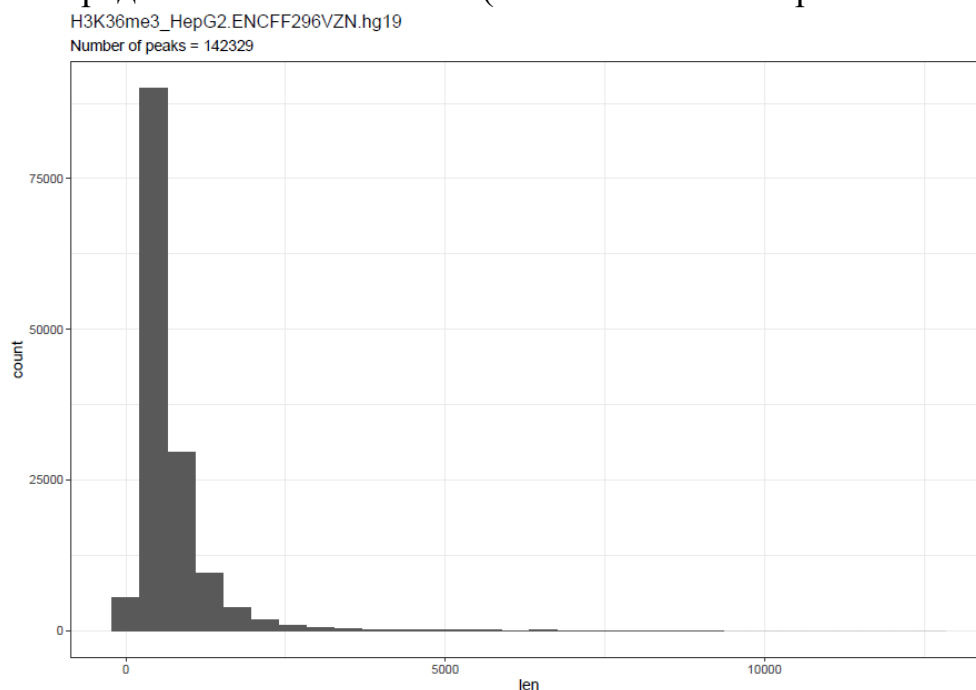
## Эксперимент ENCFF296VZN

Распределение до liftOver:



Число пиков равно 142396.

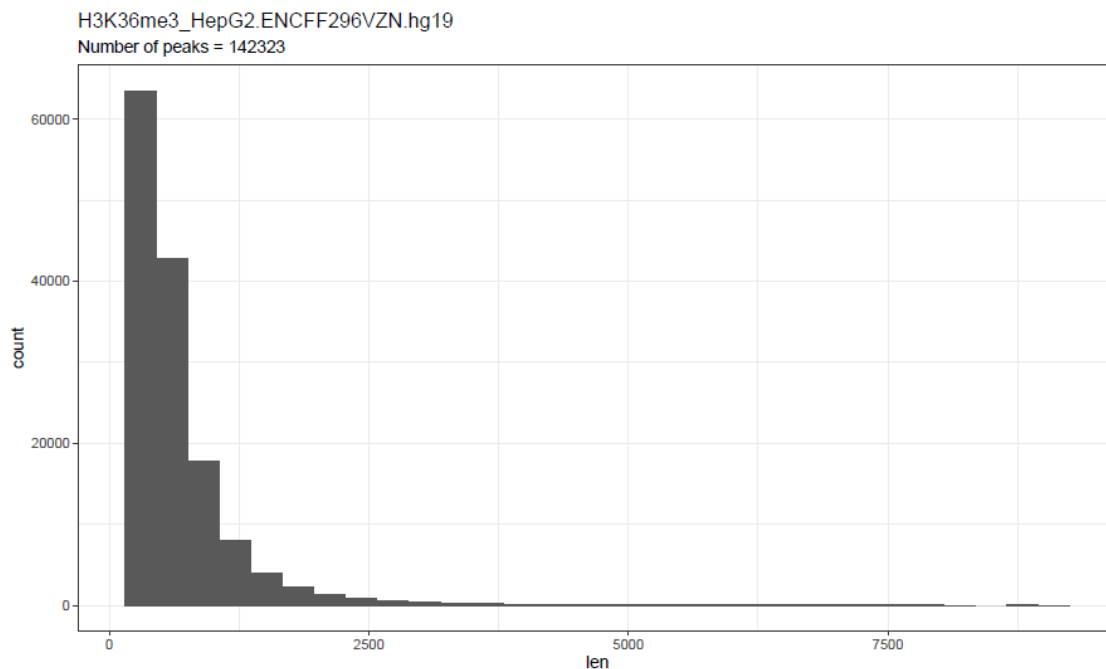
Распределение после liftOver (число пиков стало равно 142329):



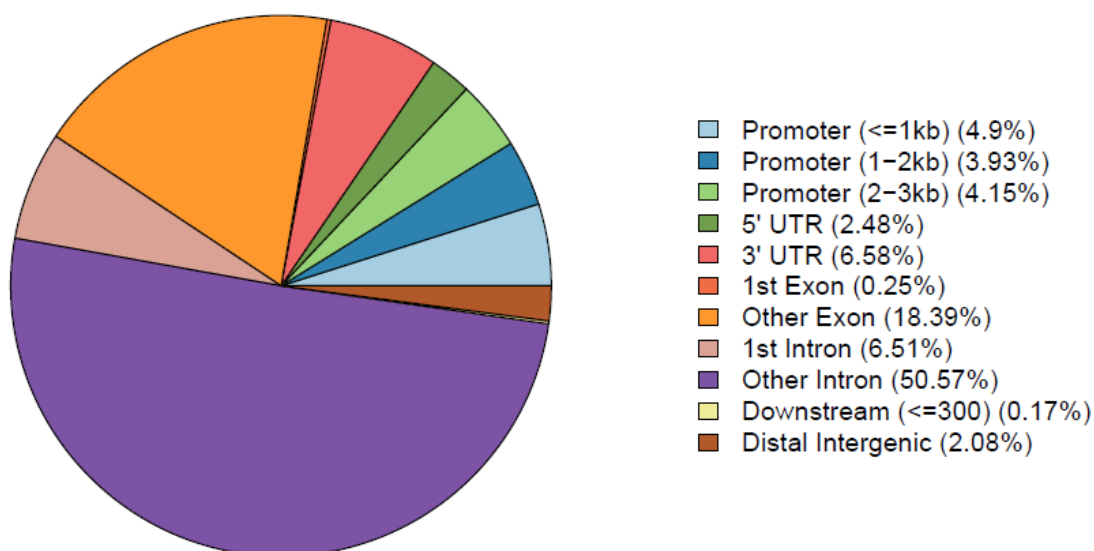
Число пиков уменьшилось на 67.

Пороговым значением для отсеечения длинных пиков было выбрано равным 9000.

После фильтрации распределение длин пиков приобрело следующий вид, число пиков сократилось на 6:



В результате работы ChIPseeker получилась такая круговая диаграмма:

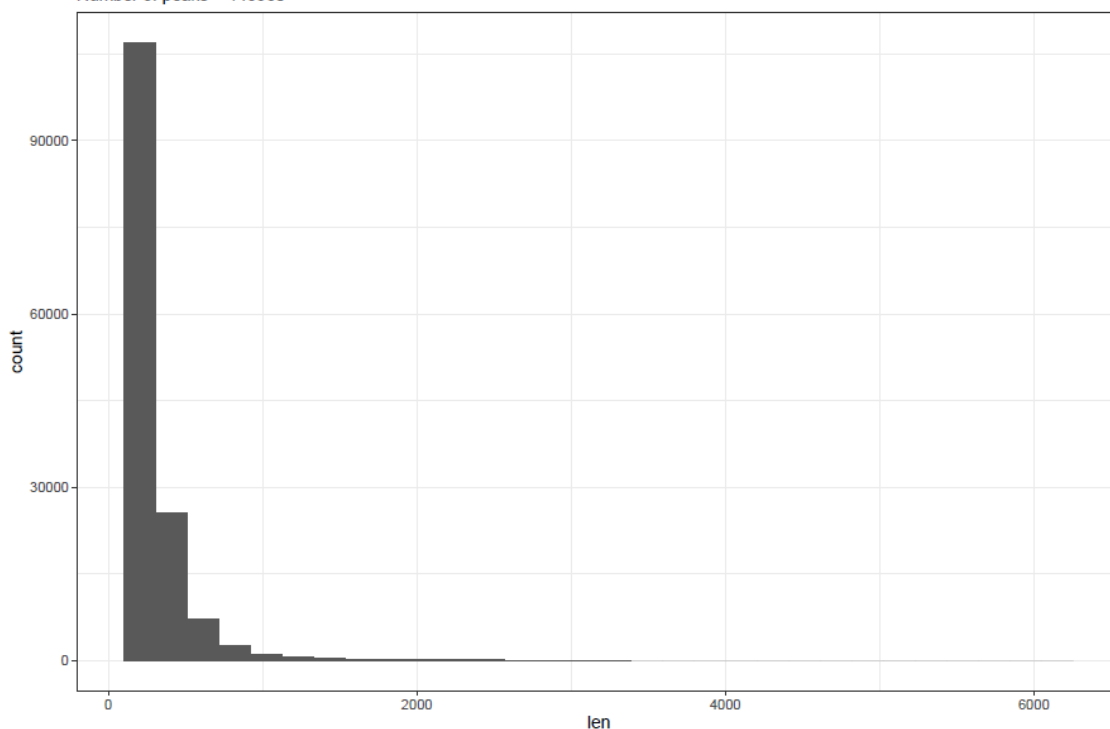


## Эксперимент *ENCFF924TKB*

Распределение до liftOver:

H3K36me3\_HepG2.ENCFF924TKB.hg38

Number of peaks = 143980

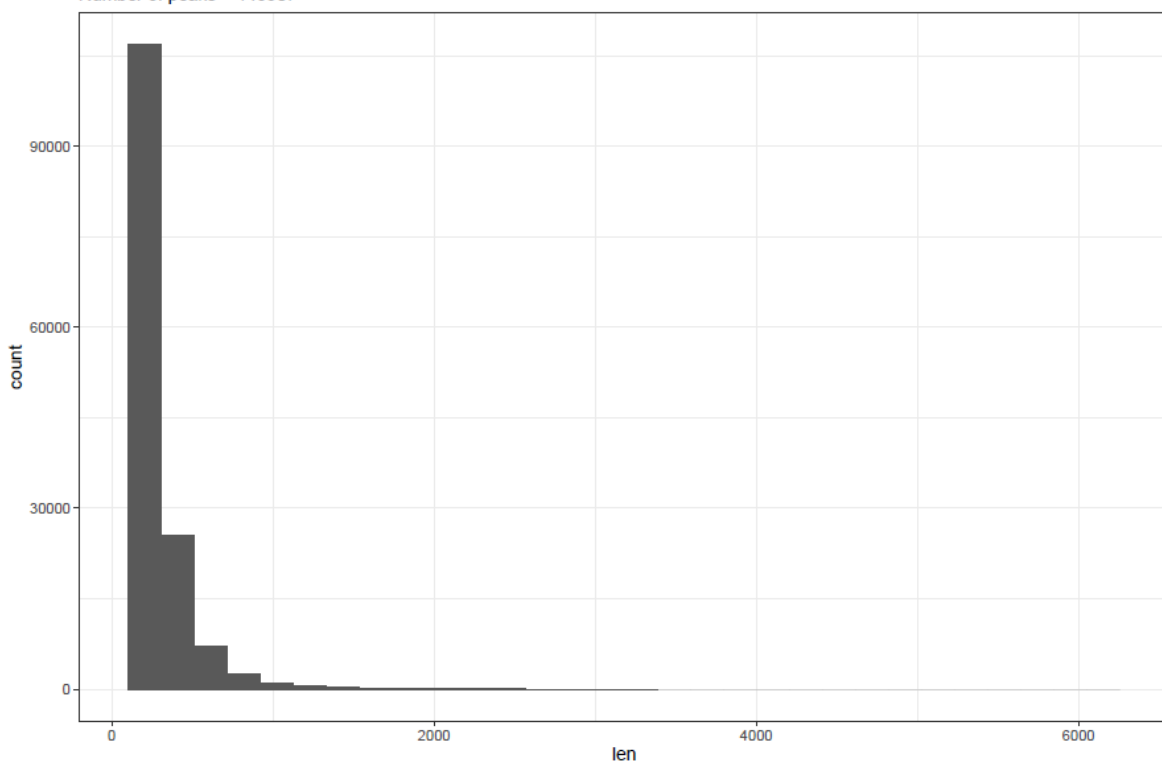


Число пиков равно 143980.

Распределение после liftOver (число пиков стало равно 143907):

H3K36me3\_HepG2.ENCFF924TKB.hg19

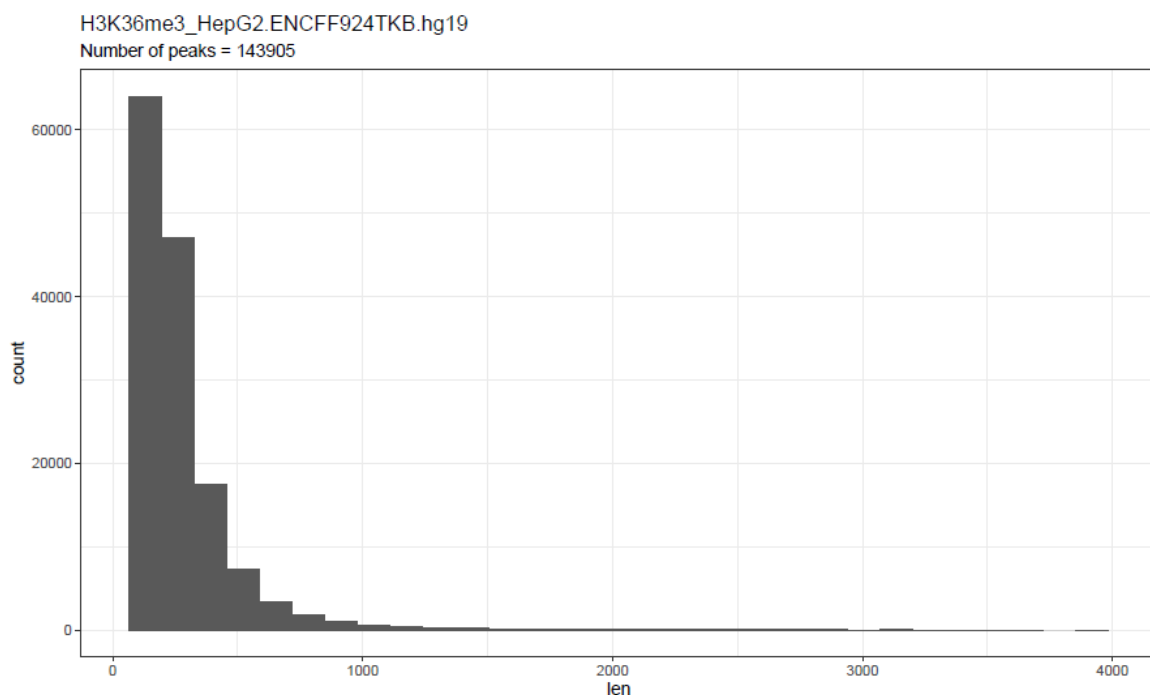
Number of peaks = 143907



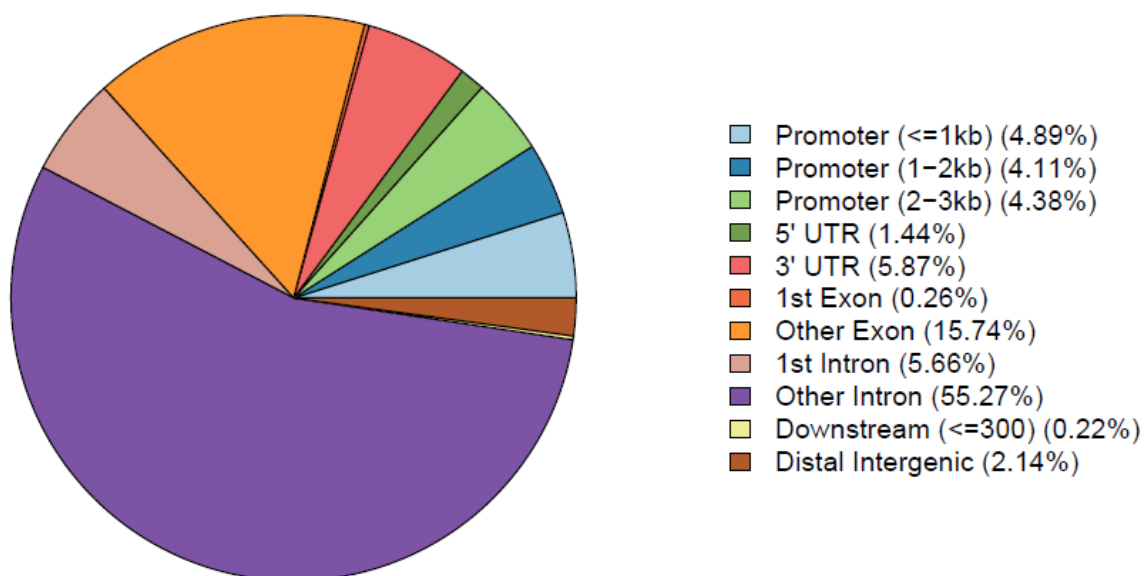
Число пиков уменьшилось на 73.

Пороговым значением для отсека длинных пиков было выбрано равным 4000.

После фильтрации распределение длин пиков приобрело следующий вид, число пиков сократилось на 2:

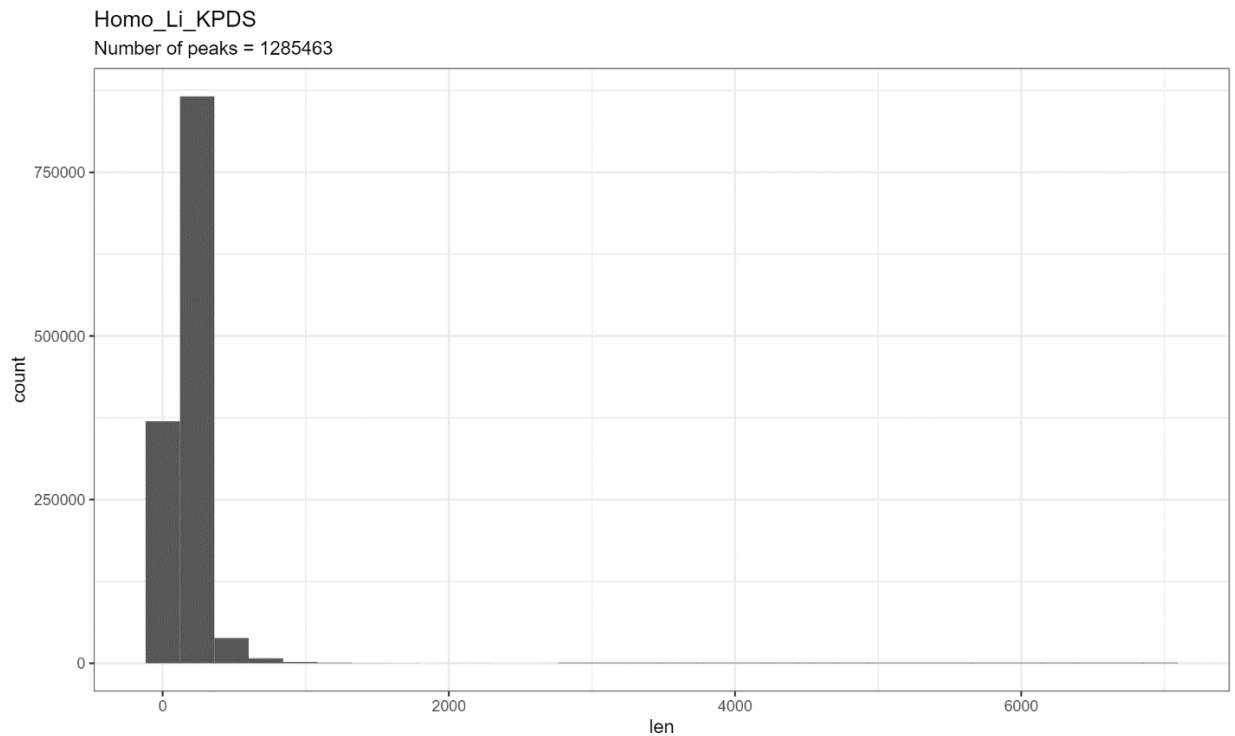


В результате работы ChIPseeker получилась такая круговая диаграмма:

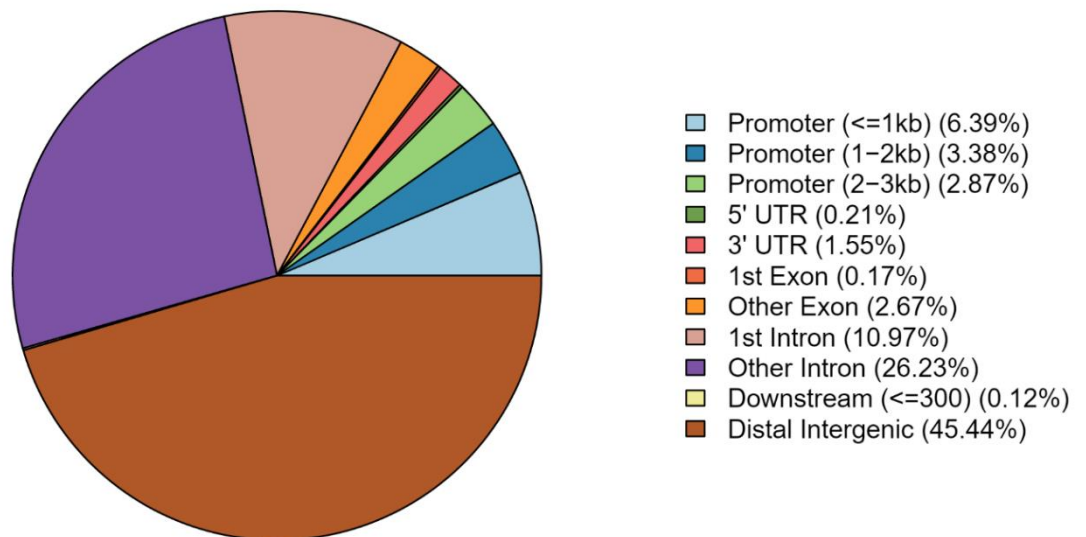


## 2. Анализ участков вторичной структуры ДНК.

Распределение длин пиков вторичной структуры (всего пиков - 1285463):

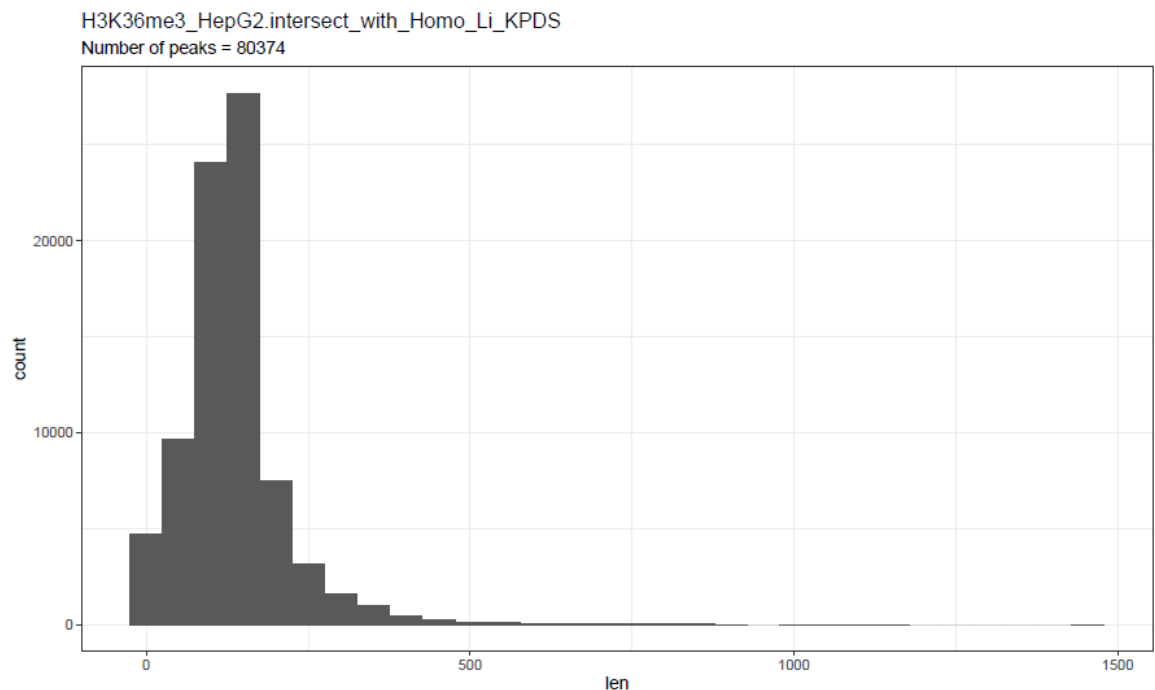


Расположение пиков относительно аннотированных генов:

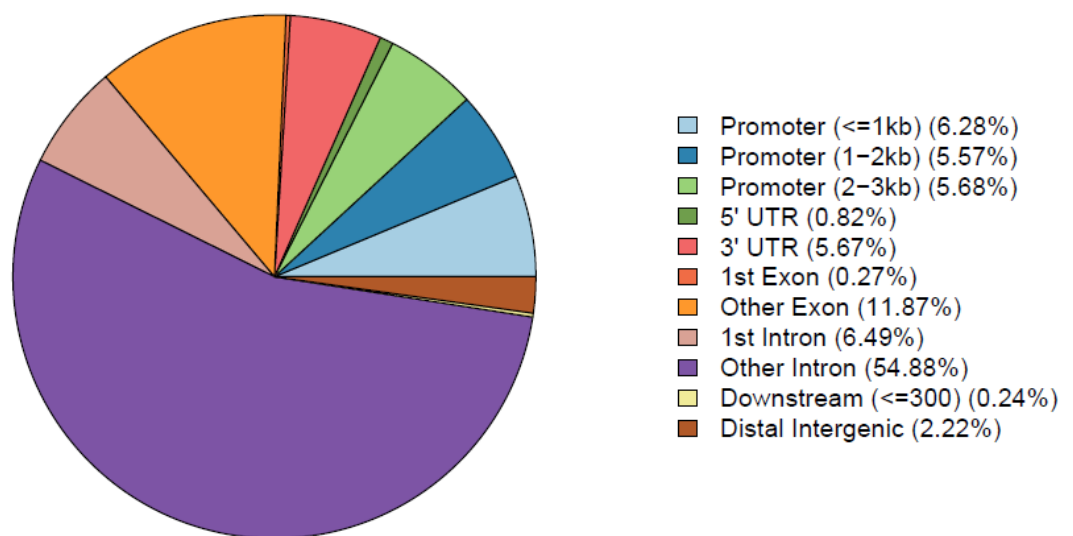


### 3. Анализ пересечений результатов эксперимента.

Распределение длин пиков пересечения вторичной структуры ДНК и результатов эксперимента (всего пиков - 80374):



Расположение результатов пересечения относительно аннотированных генов:



### *Работа с геномным браузером*

track visibility=dense name="ENCFF296VZN"

description="H3K36me3\_HepG2.ENCFF296VZN.hg19.filtered.bed"

[https://raw.githubusercontent.com/Sophie178/hse21\\_H3K36me3\\_G4\\_human/main/data/H3K36me3\\_HepG2.ENCFF296VZN.hg19.filtered.bed](https://raw.githubusercontent.com/Sophie178/hse21_H3K36me3_G4_human/main/data/H3K36me3_HepG2.ENCFF296VZN.hg19.filtered.bed)

track visibility=dense name="ENCFF924TKB"  
description="H3K36me3\_HepG2.ENCFF924TKB.hg19.filtered.bed"

[https://raw.githubusercontent.com/Sophie178/hse21\\_H3K36me3\\_G4\\_human/main/data/H3K36me3\\_HepG2.ENCFF924TKB.hg19.filtered.bed](https://raw.githubusercontent.com/Sophie178/hse21_H3K36me3_G4_human/main/data/H3K36me3_HepG2.ENCFF924TKB.hg19.filtered.bed)

track visibility=dense name="ChIP\_merge" description="H3K36me3\_HepG2.merge.hg19.bed"

[https://raw.githubusercontent.com/Sophie178/hse21\\_H3K36me3\\_G4\\_human/main/data/H3K36me3\\_HepG2.merge.hg19.bed](https://raw.githubusercontent.com/Sophie178/hse21_H3K36me3_G4_human/main/data/H3K36me3_HepG2.merge.hg19.bed)

track visibility=dense name="G4" color=0,200,0 description="G4"

[https://raw.githubusercontent.com/Sophie178/hse21\\_H3K36me3\\_G4\\_human/main/data/Homo\\_Li\\_KPDS.bed](https://raw.githubusercontent.com/Sophie178/hse21_H3K36me3_G4_human/main/data/Homo_Li_KPDS.bed)

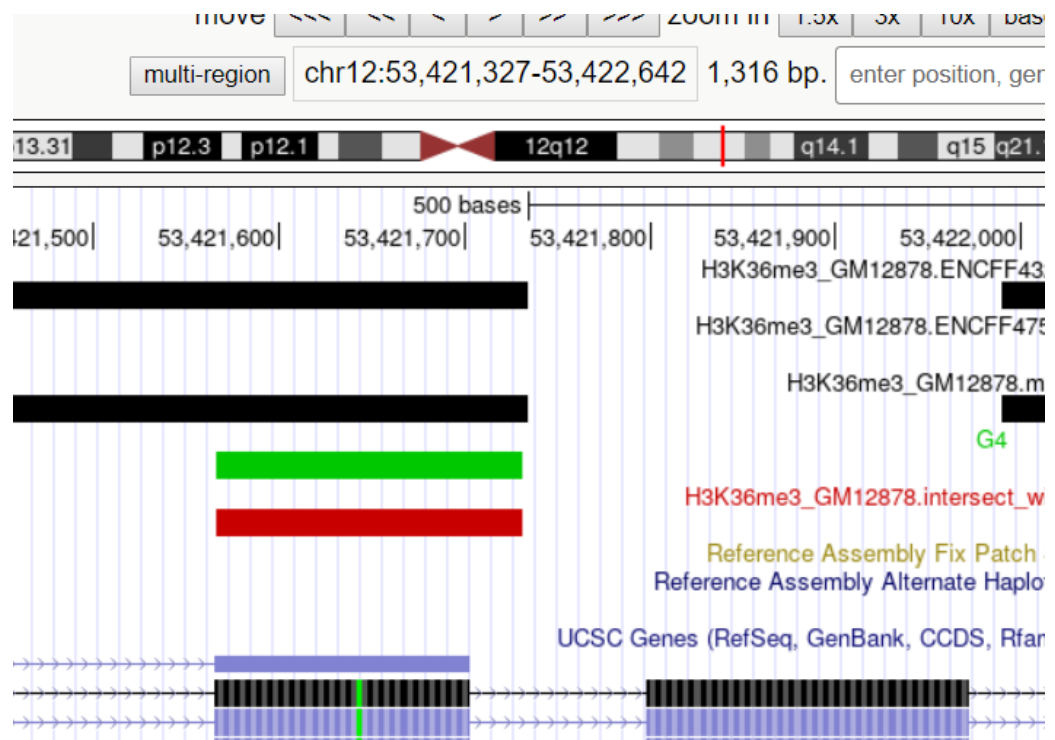
track visibility=dense name="intersect\_with\_G4" color=200,0,0  
description="H3K36me3\_HepG2.intersect\_with\_Homo\_Li\_KPDS.bed"

[https://raw.githubusercontent.com/Sophie178/hse21\\_H3K36me3\\_G4\\_human/main/data/H3K36me3\\_HepG2.intersect\\_with\\_Homo\\_Li\\_KPDS.bed](https://raw.githubusercontent.com/Sophie178/hse21_H3K36me3_G4_human/main/data/H3K36me3_HepG2.intersect_with_Homo_Li_KPDS.bed)

Ссылка на сессию в геномном браузере:

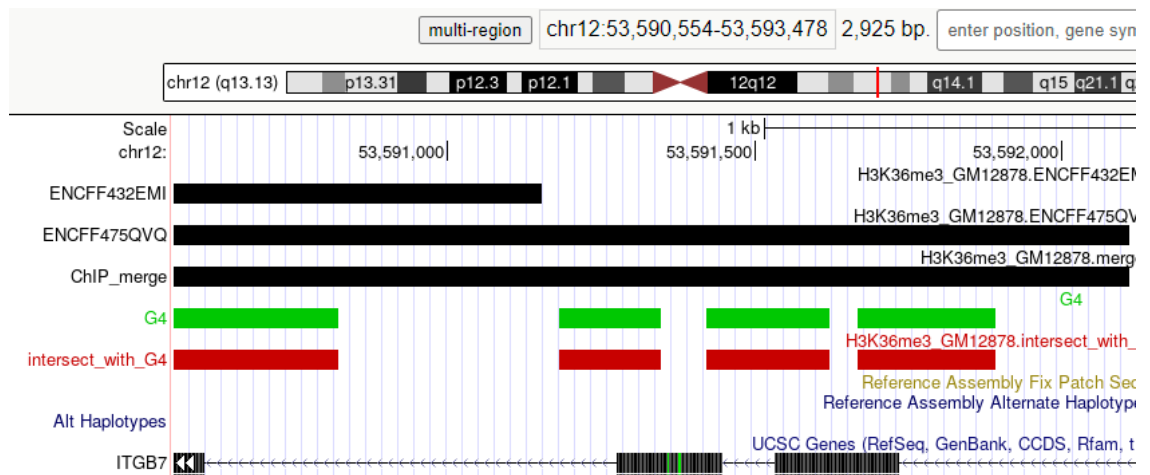
[http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A53412273%2D53451752&hgside=1126589641\\_9ahtwaq4LrY1yXanCoCwDmbQ5J5z](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A53412273%2D53451752&hgside=1126589641_9ahtwaq4LrY1yXanCoCwDmbQ5J5z)

Примеры пересечения структуры ДНК с гистоновой меткой:



Координаты - chr12:53,421,327-53,422,642





Координаты - chr12:53,590,554-53,593,478

## GO-анализ уникальных генов:

Results ?

	Reference list	Client Text Box Input
Uniquely Mapped IDs:	<a href="#">20595</a> out of 20595	<a href="#">369</a> out of 382
Unmapped IDs:	<a href="#">0</a>	<a href="#">325</a>
Multiple mapping information:	<a href="#">0</a>	<a href="#">13</a>

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)

No statistically significant results. [Click to see all results.](#)

Значения FDR одинаковые, равные единице, поэтому нельзя однозначно сказать, какие категории наиболее значимы. Пример некоторых встретившихся категорий:

[ion transmembrane transport](#)

[respiratory burst involved in defense response](#)

[response to lead ion](#)

[positive regulation of chronic inflammatory response](#)

[negative regulation of chronic inflammatory response](#)

[heat acclimation](#)

## Результаты ассоциации пиков с генами

С генами удалось ассоциировать 1274 пика. Всего получилось 694 уникальных ассоциированных гена.