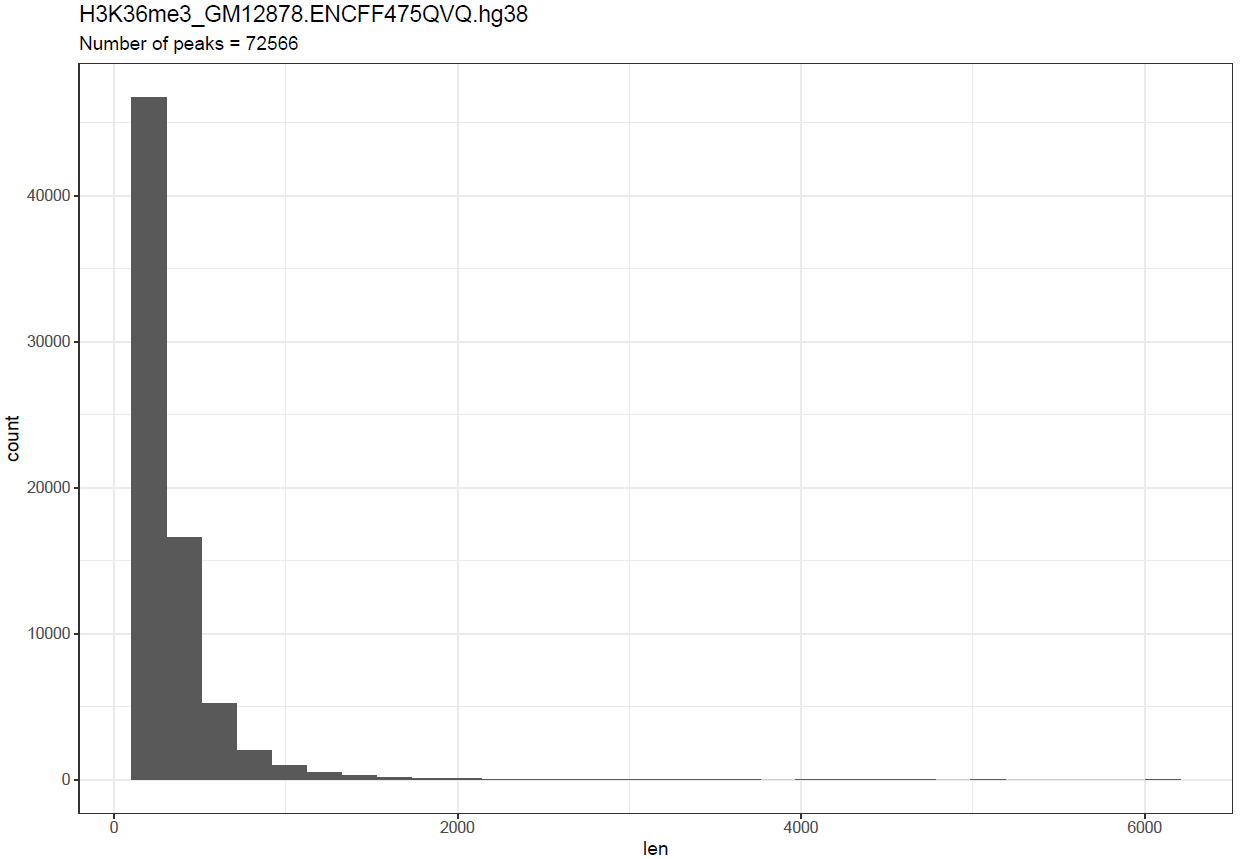
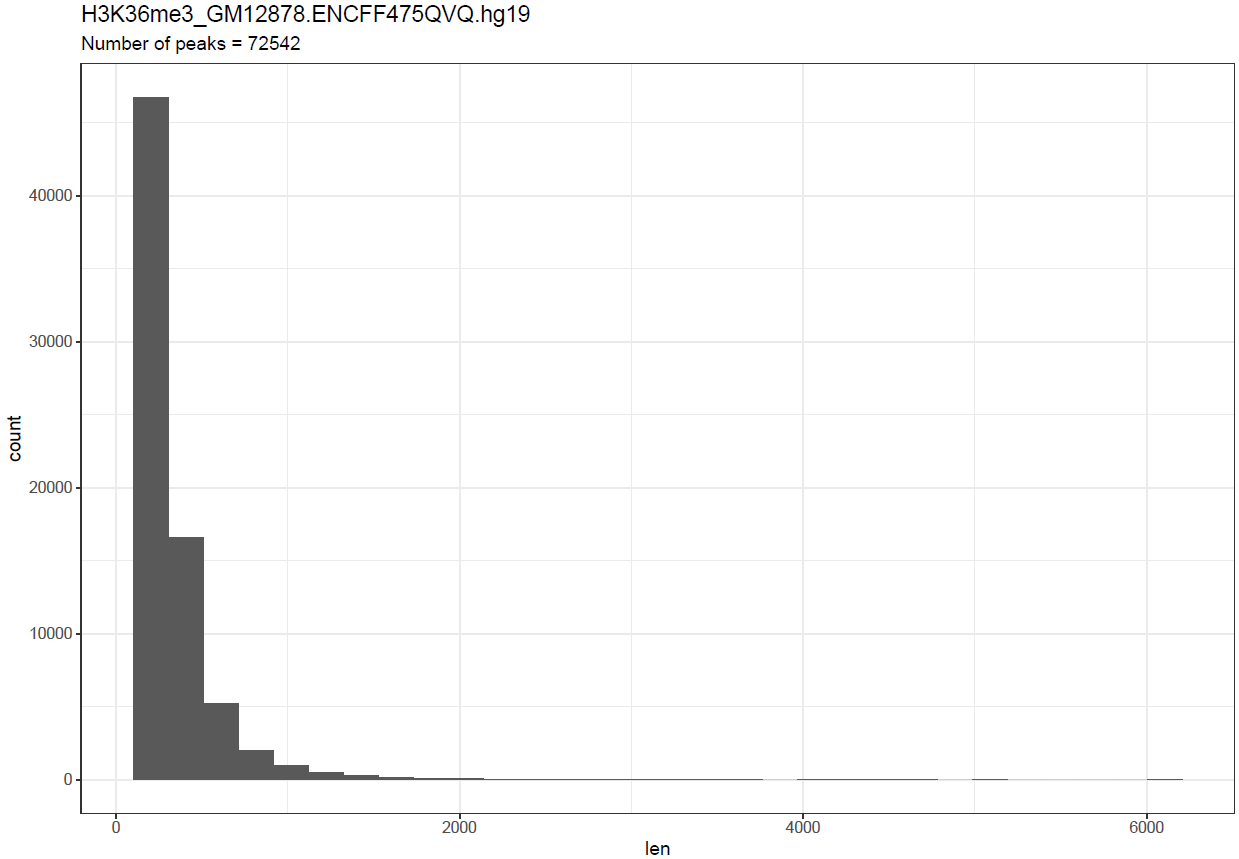
1. **Построение распределения длин пиков до и после конвертации в hg19, фильтрация длинных пиков, расположение пиков относительно аннотированных генов**

Оба эксперимента проводились со сборкой hg38, поэтому придется конвертировать оба файла

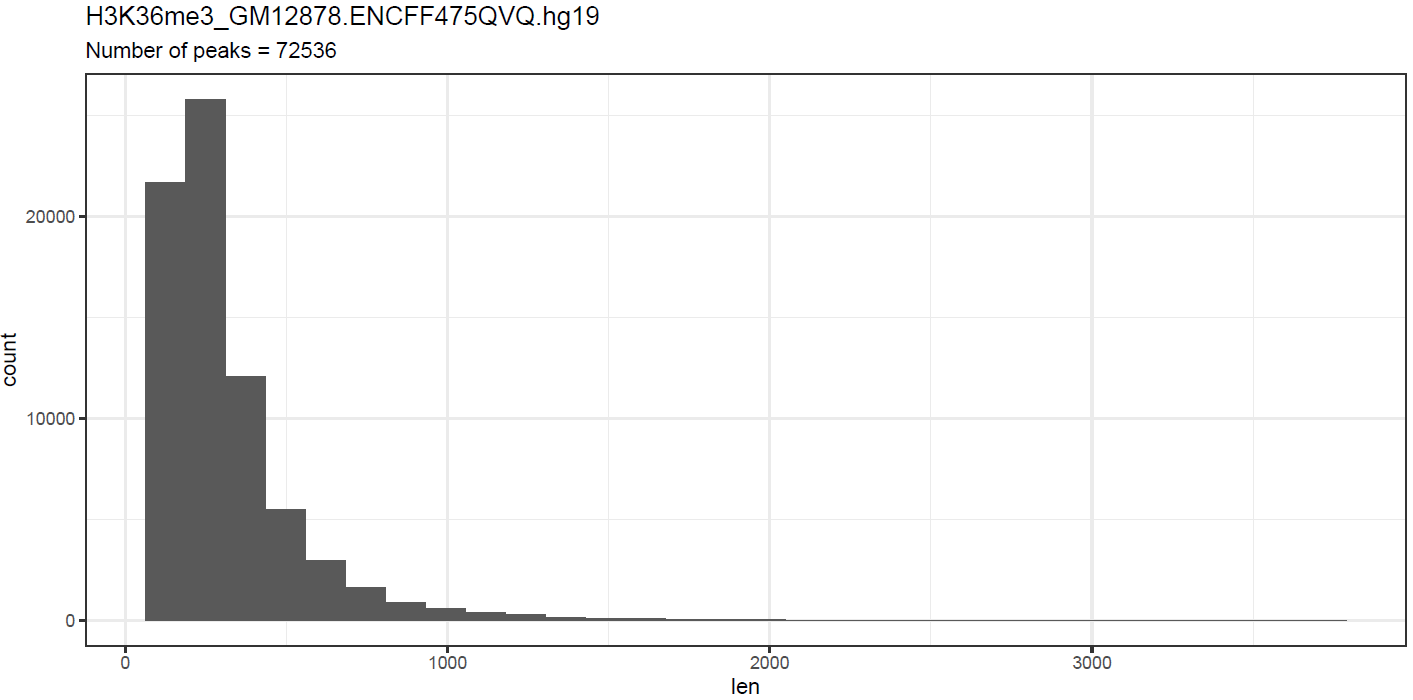
**Эксперимент ENCFF475QVQ**

Распределение до конвертации (число пиков равно 72566): 

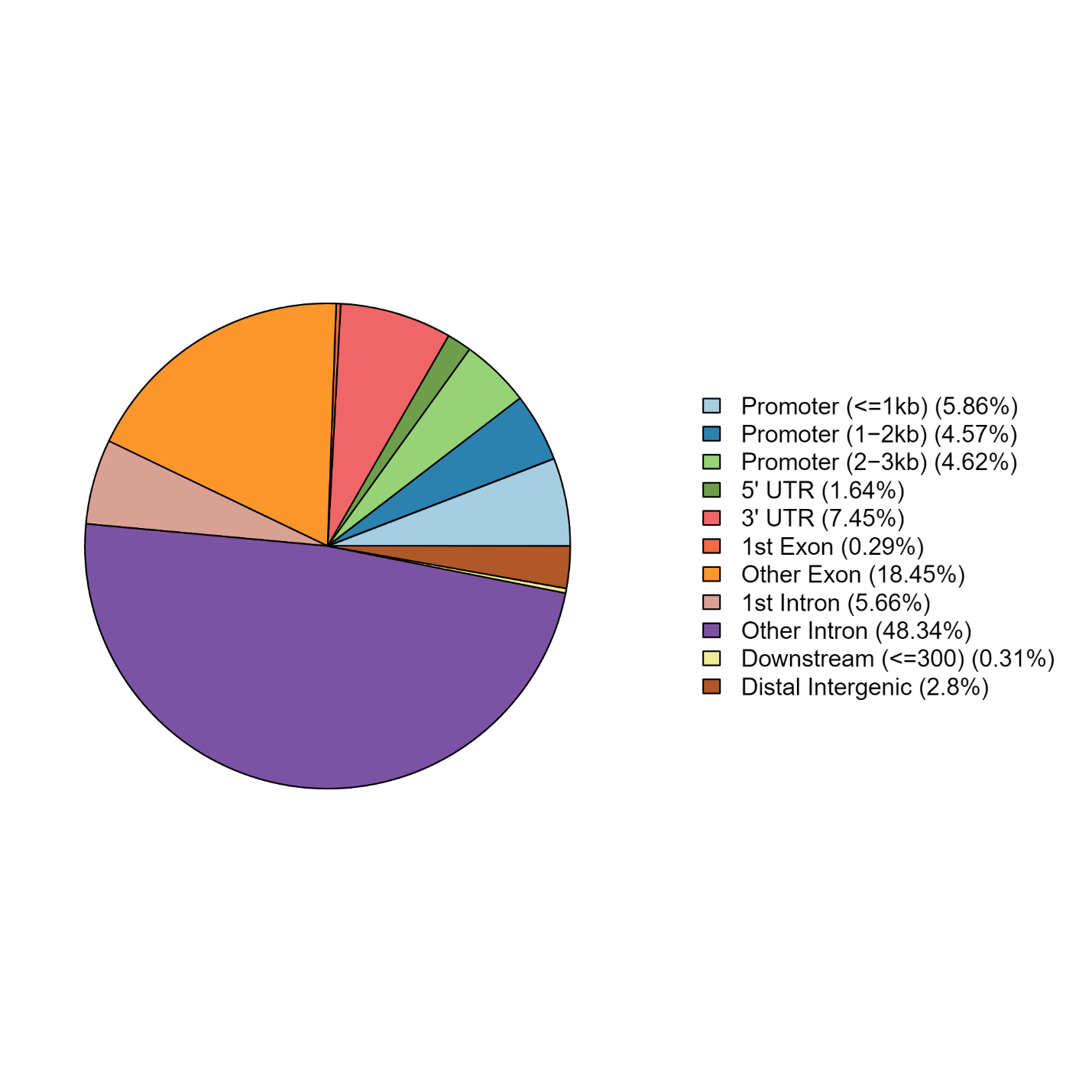
Распределение после конвертации из hg38 в hg19 (число пиков равно 72542):

Произошло незначительное сокращение числа пиков: не удалось перевести всего 14 из них

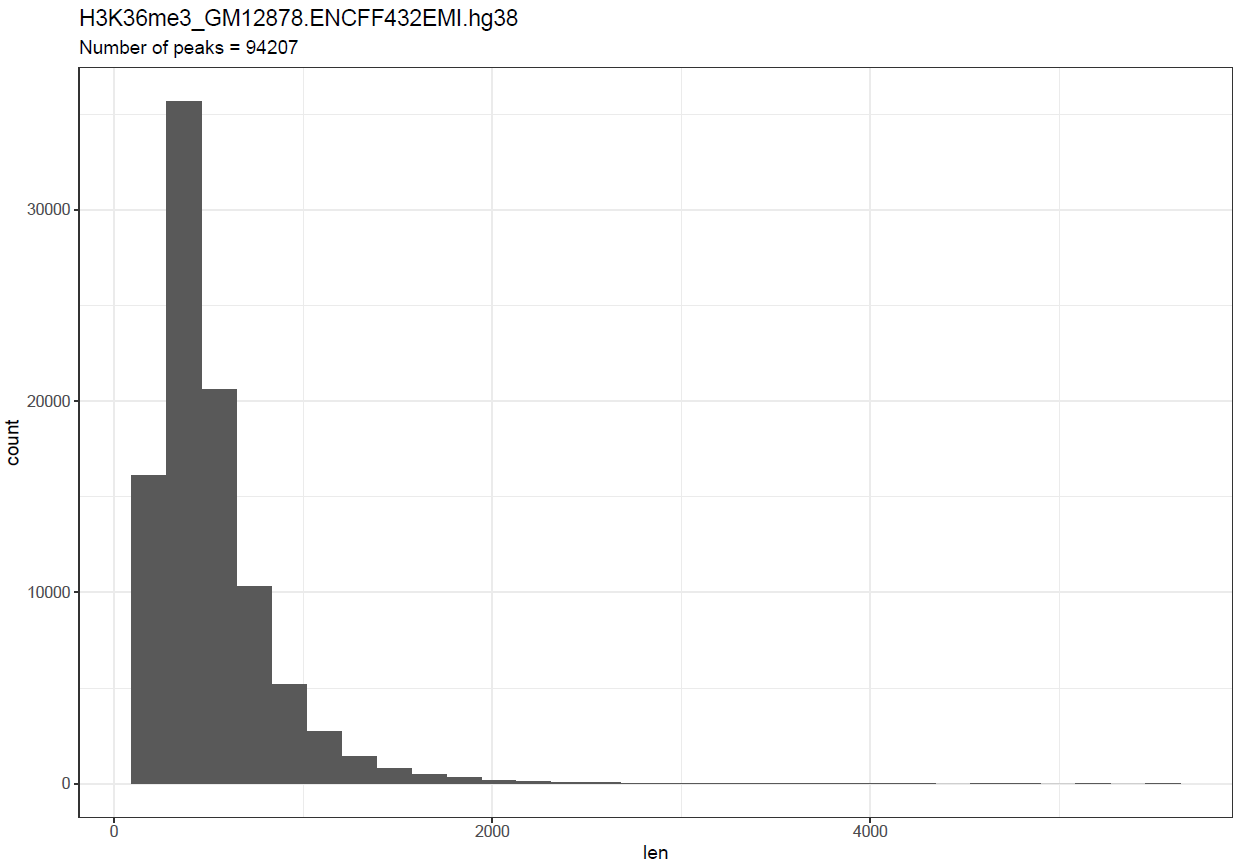
Пороговым значением для отсечения длинных пиков я выбрал 4000, т.к. после этого значения распределение перестаёт быть непрерывным. После фильтрации распределение длин пиков приобрело следующий вид, число пиков сократилось на 6:

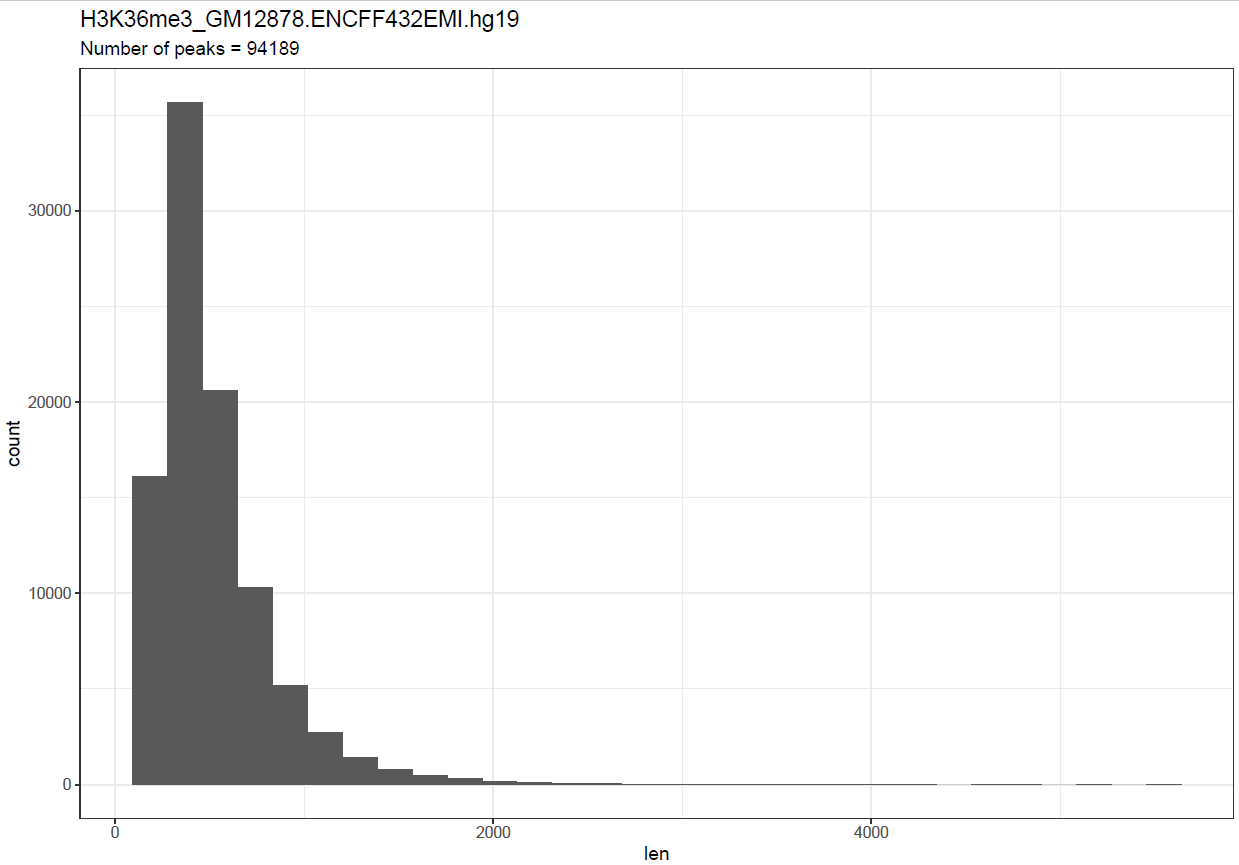


В результате работы ChIPseeker получилась такая круговая диаграмма:



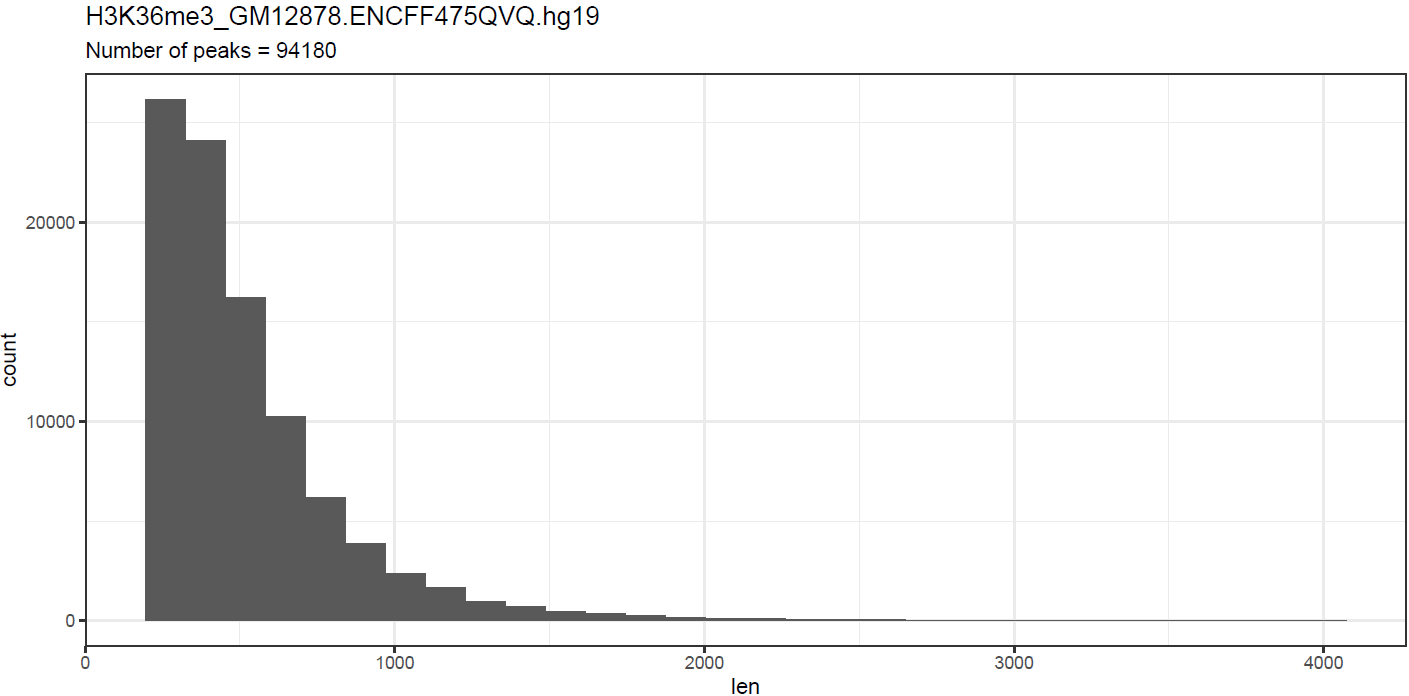
**Эксперимент ENCFF432EMI**

Распределение до конвертации (число пиков равно 94207): 

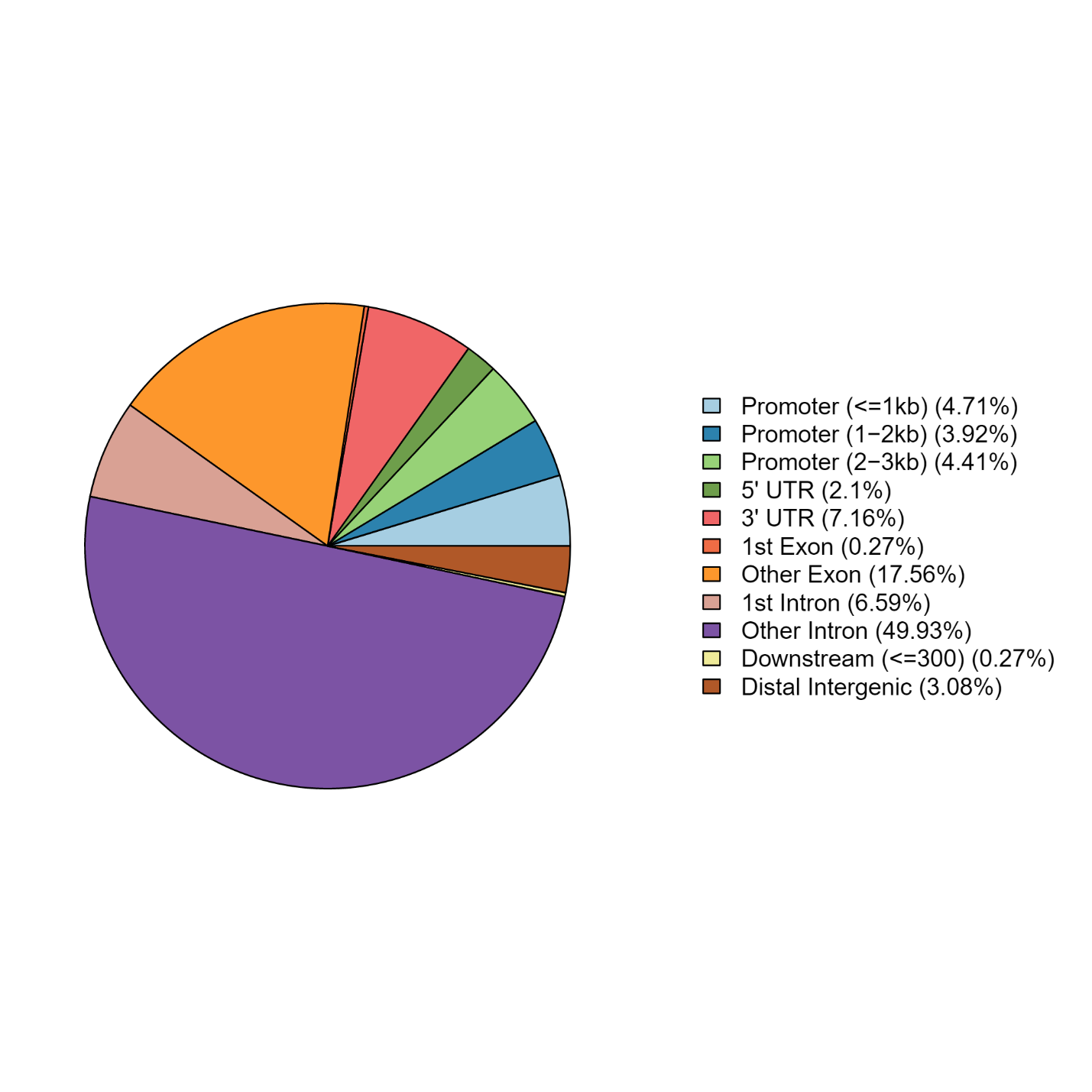
Распределение после конвертации (число пиков равно 94189): 

Произошло незначительное сокращение числа пиков: не удалось перевести всего 18 из них

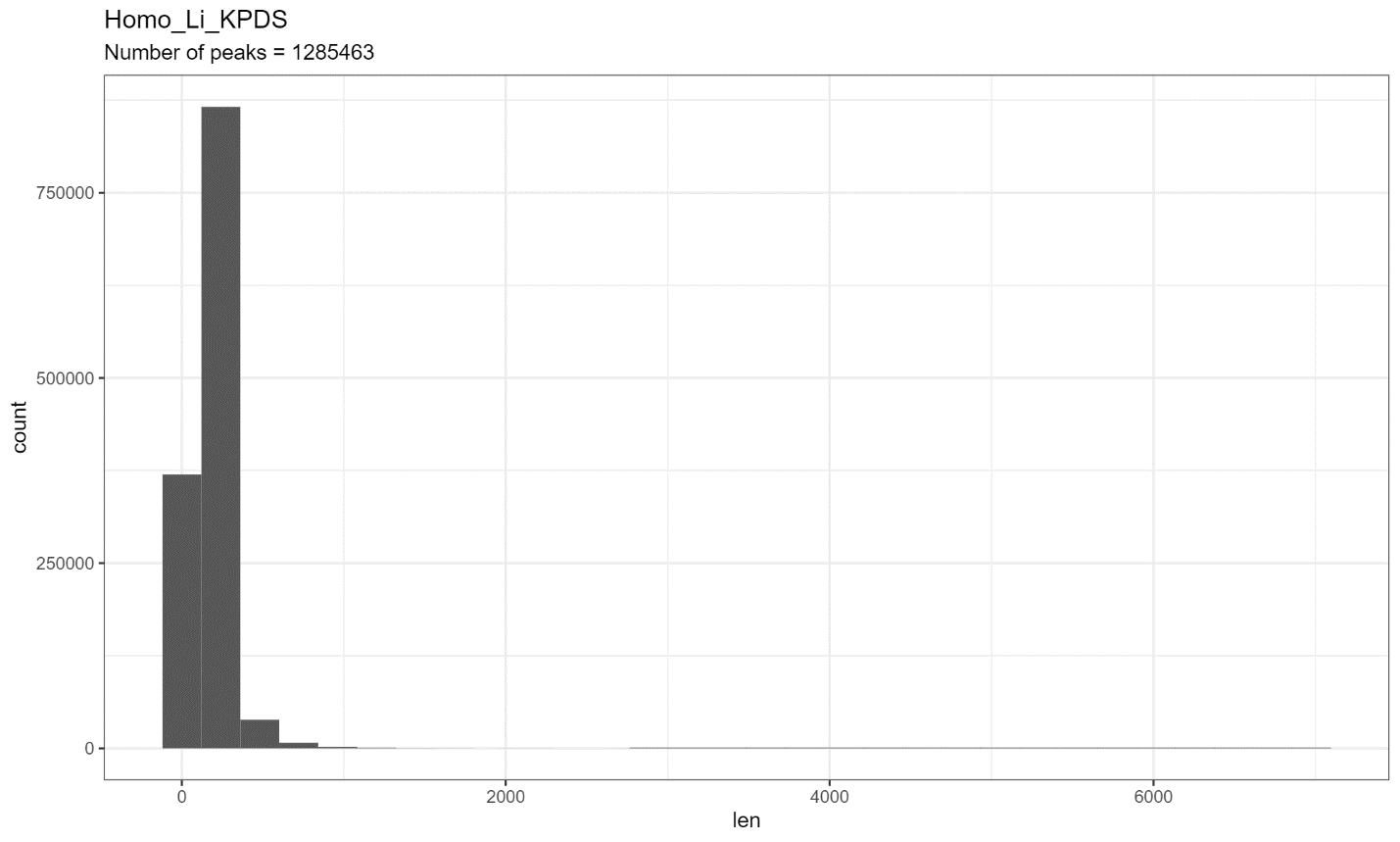
Пороговым значением для отсечения длинных пиков я выбрал 4000, т.к. после этого значения распределение перестаёт быть непрерывным. После фильтрации распределение длин пиков приобрело следующий вид, число пиков сократилось на 9:



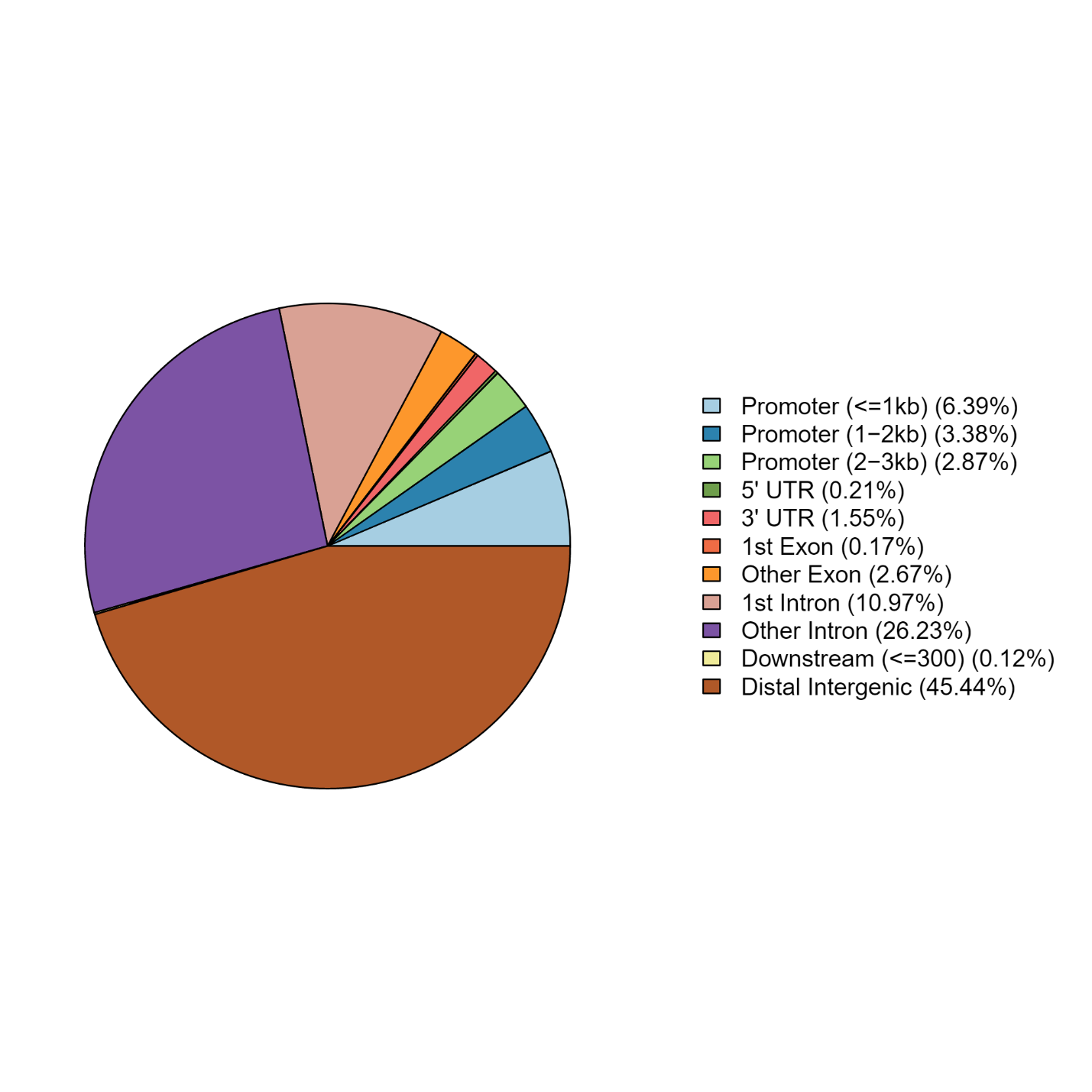
В результате работы ChIPseeker получилась такая круговая диаграмма:



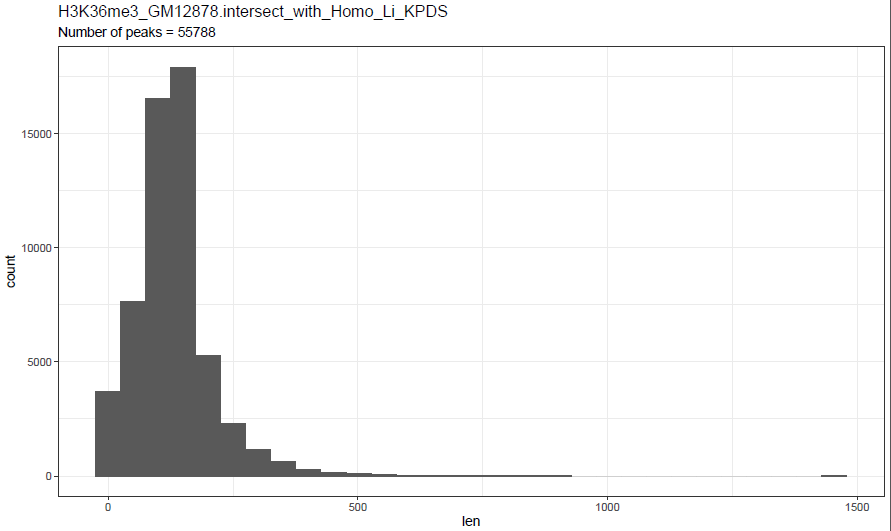
1. **Анализ участков вторичной структуры ДНК**

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

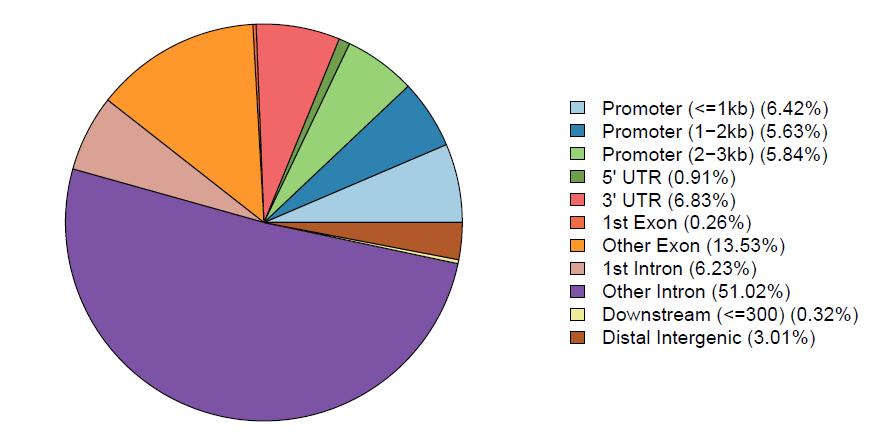
Расположение пиков относительно аннотированных генов:



1. **Анализ пересечений результатов эксперимента**

Распределение длин пиков пересечения вторичной структуры ДНК и результатов эксперимента (всего пиков - 55788):

Расположение результатов пересечения относительно аннотированных генов:



**Работа с геномным браузером** (использовавшиеся параметры кастомных треков):

track visibility=dense name="ENCFF432EMI" description="H3K36me3\_GM12878.ENCFF432EMI.hg19.filtered.bed"

https://raw.githubusercontent.com/amamedov/hse21\_H3K36me3\_G4\_human/main/data/H3K36me3\_GM12878.ENCFF432EMI.hg19.filtered.bed

track visibility=dense name="ENCFF475QVQ" description="H3K36me3\_GM12878.ENCFF475QVQ.hg19.filtered.bed"

https://raw.githubusercontent.com/amamedov/hse21\_H3K36me3\_G4\_human/main/data/H3K36me3\_GM12878.ENCFF475QVQ.hg19.filtered.bed

track visibility=dense name="ChIP\_merge" description="H3K36me3\_GM12878.merge.hg19.bed"

<https://raw.githubusercontent.com/amamedov/hse21_H3K36me3_G4_human/main/data/H3K36me3_GM12878.merge.hg19.bed>

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

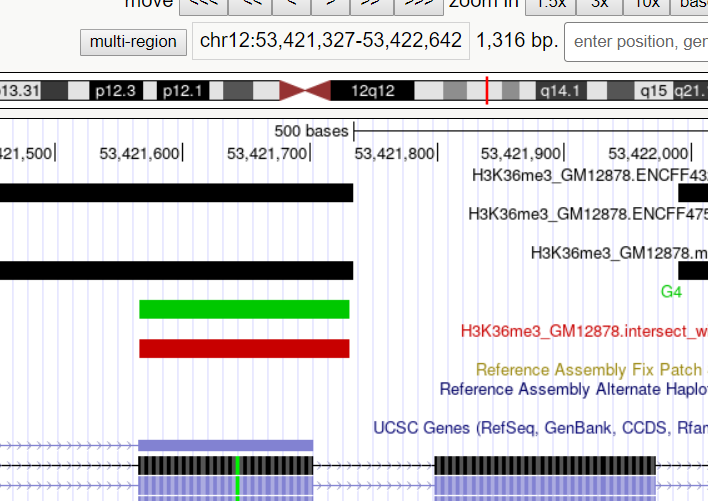
https://raw.githubusercontent.com/amamedov/hse21\_H3K36me3\_G4\_human/main/data/Homo\_Li\_KPDS.bed

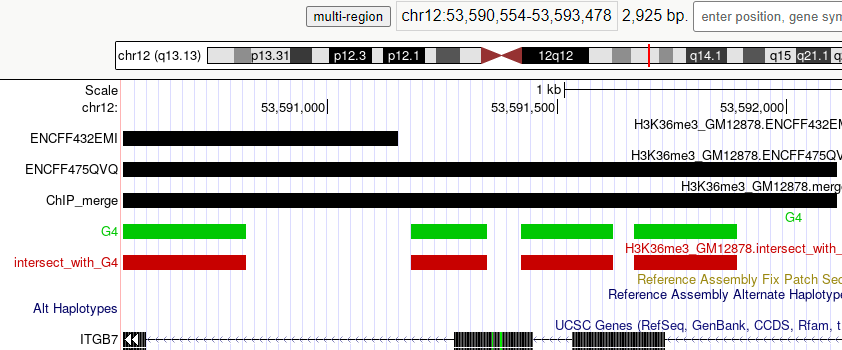
track visibility=dense name="intersect\_with\_G4" color=200,0,0 description="H3K36me3\_GM12878.intersect\_with\_Homo\_Li\_KPDS.bed"

<https://raw.githubusercontent.com/amamedov/hse21_H3K36me3_G4_human/main/data/H3K36me3_GM12878.intersect_with_Homo_Li_KPDS.bed>

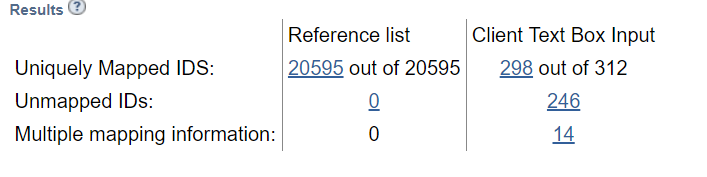
Ссылка на сессию в геномном браузере: <http://genome.ucsc.edu/s/mamedov%2Daa/hg19_H3K36me3_GM12878>

Примеры пересечения структуры ДНК с гистоновой меткой:

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

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

**GO-анализ уникальных генов:**



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

**Результаты ассоциации пиков с генами**

С генами удалось ассоциировать 1044 пика. Всего получилось 544 уникальных ассоциированных гена.