

# ChIPseeker\_analysis\_GSE108150

Nur-Taz Rahman

4/22/2020

## Install necessary packages/dependencies

```
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
```

```
BiocManager::install(version = "3.10")
```

```
BiocManager::install("TxDb.Hsapiens.UCSC.hg19.knownGene")
```

```
BiocManager::install("clusterProfiler")
```

```
BiocManager::install("org.Hs.eg.db")
```

```
BiocManager::install("ggupset")
```

```
BiocManager::install("ggimage")
```

```
BiocManager::install("ReactomePA")
```

## Load the installed packages

```
library(BiocManager)

library(ChIPseeker)

library(TxDb.Hsapiens.UCSC.hg19.knownGene)

library(clusterProfiler)

library(org.Hs.eg.db)

library(ggupset)

library(ggplot2)

library(ggimage)

library(ReactomePA)
```

## Set working directory

```
setwd ("/Users/nr267/Desktop/All Work/Classes/Spring 2020/ChIPseeker_2020/peak_files_be  
d/")
```

## Load and rename package for transcript-related features

```
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
```

## Loading the files as a list

```
all_files <- list(NY8_1 = "../ChIPseeker_2020/peak_files_bed/GSM2891281_NY8_HNF1A_ChIP_1_peaks.bed", NY8_2 = "../ChIPseeker_2020/peak_files_bed/GSM2891283_NY8_HNF1A_ChIP_2_peaks.bed", NY15_1 = "../ChIPseeker_2020/peak_files_bed/GSM2891285_NY15_HNF1A_ChIP_1_peaks.bed", NY15_2 = "../ChIPseeker_2020/peak_files_bed/GSM2891287_NY15_HNF1A_ChIP_2_peaks.bed")

all_files
```

```
## $NY8_1
## [1] "../ChIPseeker_2020/peak_files_bed/GSM2891281_NY8_HNF1A_ChIP_1_peaks.bed"
##
## $NY8_2
## [1] "../ChIPseeker_2020/peak_files_bed/GSM2891283_NY8_HNF1A_ChIP_2_peaks.bed"
##
## $NY15_1
## [1] "../ChIPseeker_2020/peak_files_bed/GSM2891285_NY15_HNF1A_ChIP_1_peaks.bed"
##
## $NY15_2
## [1] "../ChIPseeker_2020/peak_files_bed/GSM2891287_NY15_HNF1A_ChIP_2_peaks.bed"
```

## Loading one file at a time

```
NY8_2_single <- readPeakFile("../ChIPseeker_2020/peak_files_bed/GSM2891283_NY8_HNF1A_ChIP_2_peaks.bed")

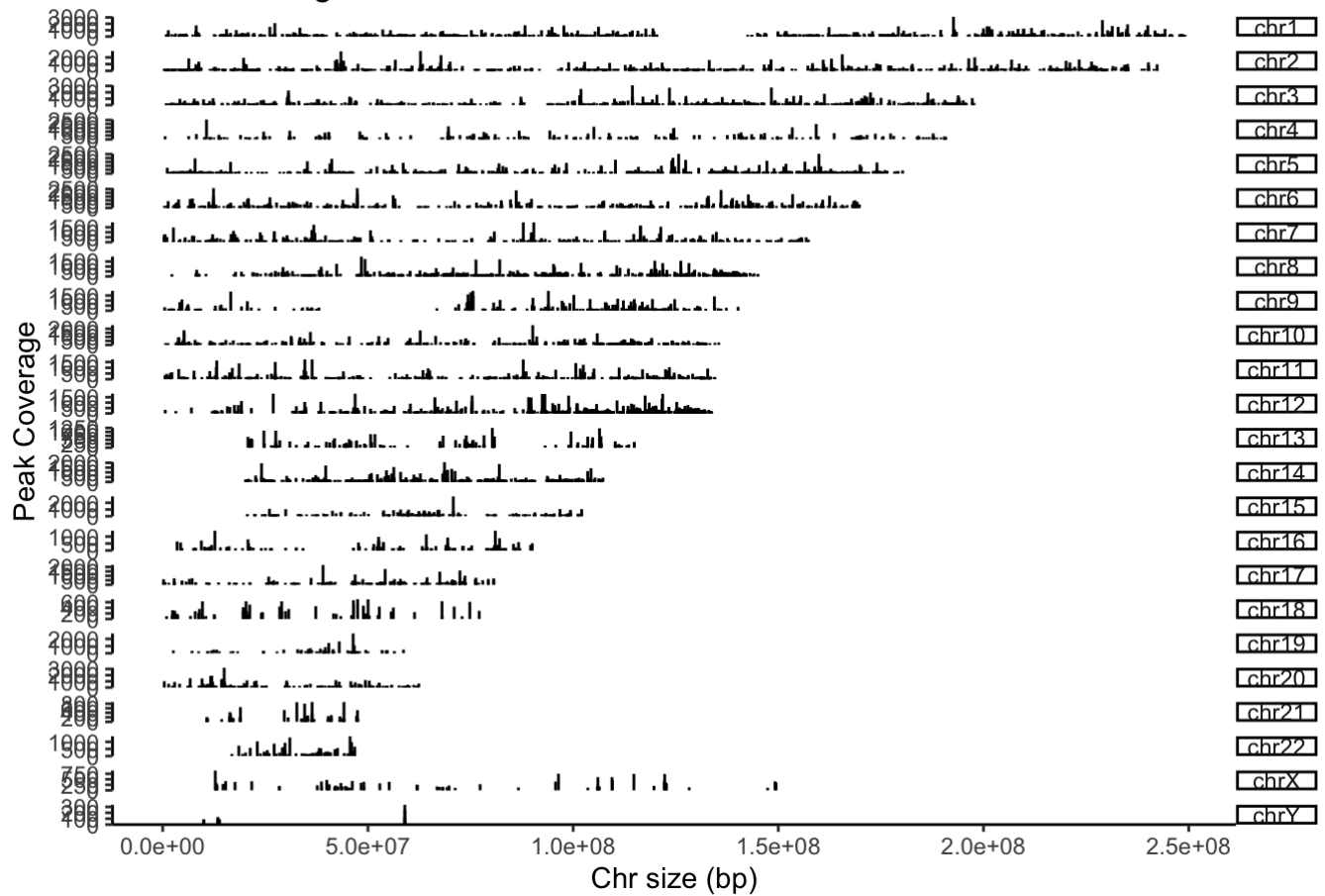
NY8_2_single
```

```
## GRanges object with 13311 ranges and 2 metadata columns:
##           seqnames           ranges strand |           V4
##           <Rle>           <IRanges> <Rle> |           <factor>
##      [1]      chr1      822578-823691      * |      NY8_HNF1A_ChIP_2_MACS_peak_1
##      [2]      chr1     1059764-1060972      * |      NY8_HNF1A_ChIP_2_MACS_peak_2
##      [3]      chr1     1099831-1101333      * |      NY8_HNF1A_ChIP_2_MACS_peak_3
##      [4]      chr1     1265654-1266811      * |      NY8_HNF1A_ChIP_2_MACS_peak_4
##      [5]      chr1     1282997-1283540      * |      NY8_HNF1A_ChIP_2_MACS_peak_5
##      ...      ...      ...      ...      ...
## [13307]      chrY 13838319-13838636      * | NY8_HNF1A_ChIP_2_MACS_peak_13307
## [13308]      chrY 58974432-58976586      * | NY8_HNF1A_ChIP_2_MACS_peak_13308
## [13309]      chrY 58976820-58981555      * | NY8_HNF1A_ChIP_2_MACS_peak_13309
## [13310]      chrY 58982295-58988963      * | NY8_HNF1A_ChIP_2_MACS_peak_13310
## [13311]      chrY 59001267-59004140      * | NY8_HNF1A_ChIP_2_MACS_peak_13311
##           V5
##           <numeric>
##      [1]      63.87
##      [2]      96.14
##      [3]     138.88
##      [4]     130.92
##      [5]      53.25
##      ...      ...
## [13307]      62.45
## [13308]     103.05
## [13309]     334.94
## [13310]     224.29
## [13311]      54.83
## -----
## seqinfo: 24 sequences from an unspecified genome; no seqlengths
```

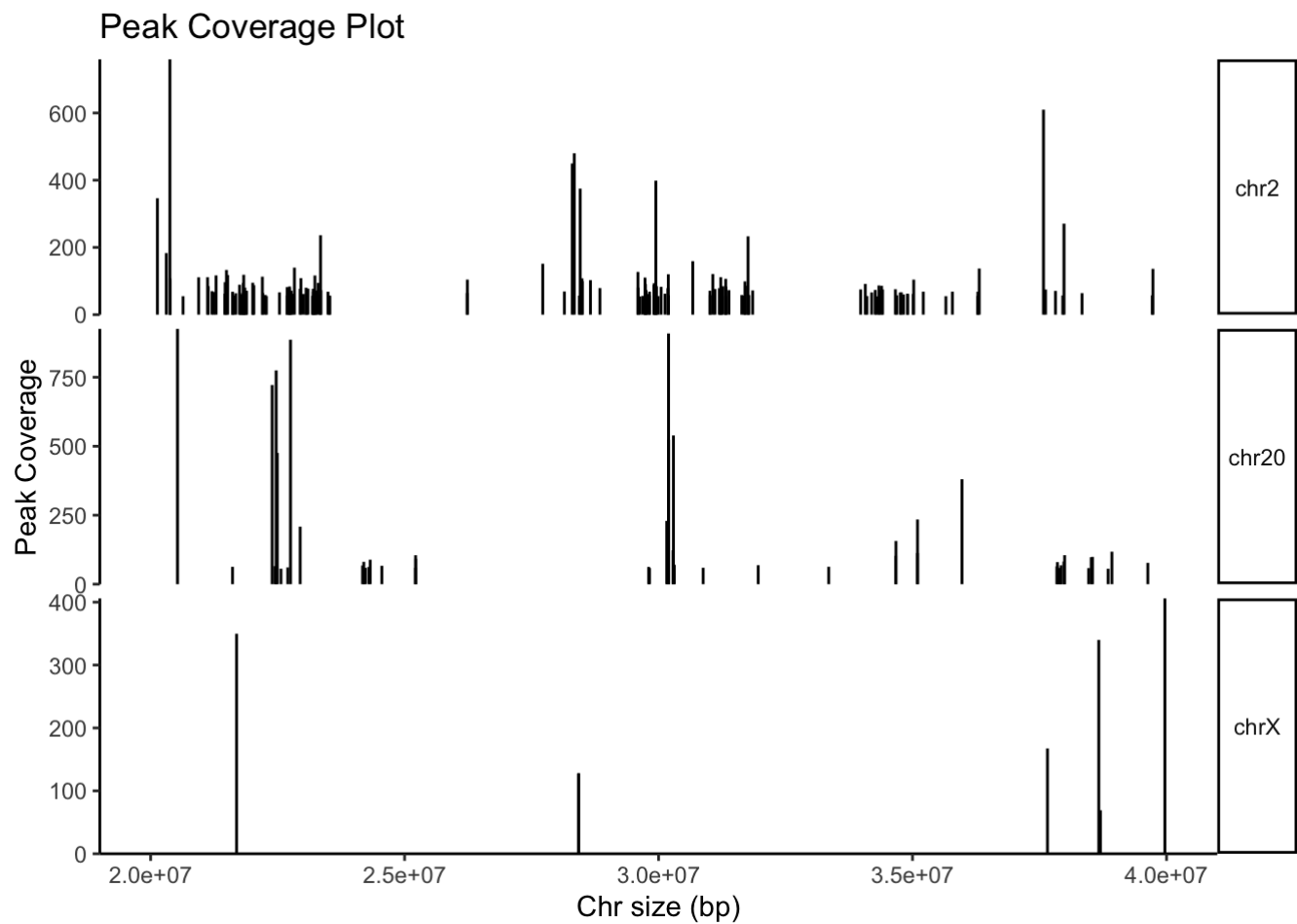
## Profiling the peaks from a single file

```
covplot (NY8_2_single, weightCol = "V5", title = "Peak Coverage Plot", xlab = "Chr size
(bp)", ylab = "Peak Coverage")
```

## Peak Coverage Plot

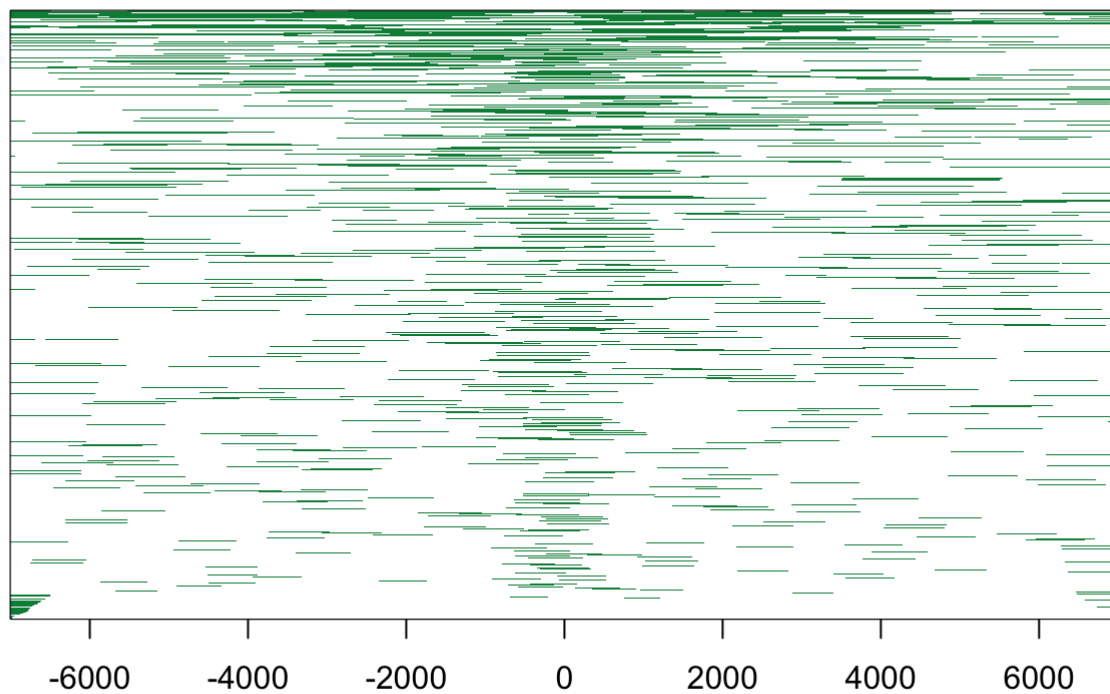


```
covplot (NY8_2_single, weightCol = "V5", title = "Peak Coverage Plot", xlab = "Chr size
(bp)", ylab = "Peak Coverage", chrs = c("chr2", "chr20", "chrX"), xlim = c(2e7, 4e7))
```



```
peakHeatmap(NY8_2_single, TxDb = txdb, upstream = 7000, downstream = 7000, color = "springgreen4")
```

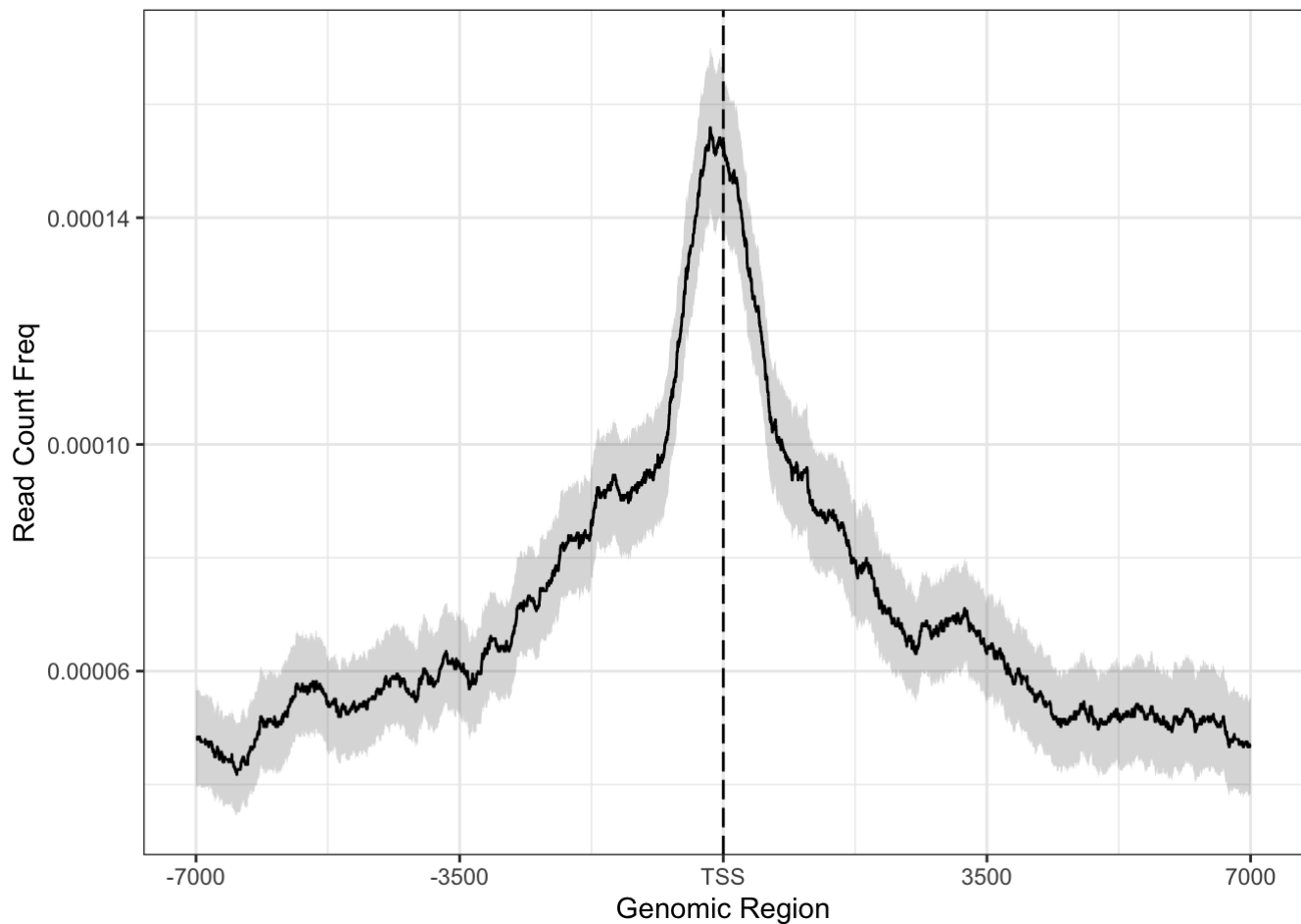
```
## >> preparing promoter regions... 2020-04-22 01:43:27
## >> preparing tag matrix...      2020-04-22 01:43:28
## >> generating figure...         2020-04-22 01:43:47
```



```
## >> done...                2020-04-22 01:43:50
```

```
plotAvgProf2(NY8_2_single, TxDb = txdb, upstream = 7000, downstream = 7000, xlab = "Genomic Region", ylab = "Read Count Freq", conf = 0.95, resample = 1000)
```

```
## >> preparing promoter regions... 2020-04-22 01:43:50
## >> preparing tag matrix...       2020-04-22 01:43:50
## >> plotting figure...            2020-04-22 01:44:03
## >> Running bootstrapping for tag matrix... 2020-04-22 01:44:26
```



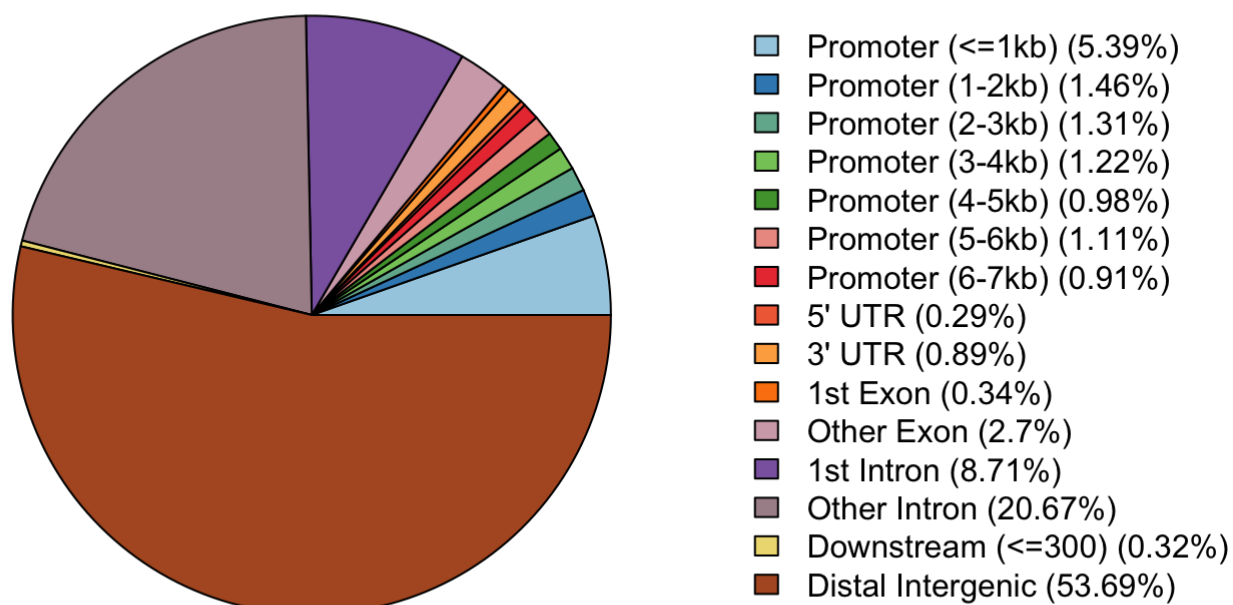
# BREAK

## Annotation & visualization of peaks

```
Anno_NY8_2 <- annotatePeak(NY8_2_single, tssRegion = c (-7000, 7000), TxDb = txdb, annoDb = "org.Hs.eg.db")
```

```
## >> preparing features information...      2020-04-22 01:44:33
## >> identifying nearest features...      2020-04-22 01:44:34
## >> calculating distance from peak to TSS... 2020-04-22 01:44:34
## >> assigning genomic annotation...      2020-04-22 01:44:34
## >> adding gene annotation...           2020-04-22 01:44:47
## >> assigning chromosome lengths       2020-04-22 01:44:47
## >> done...                           2020-04-22 01:44:47
```

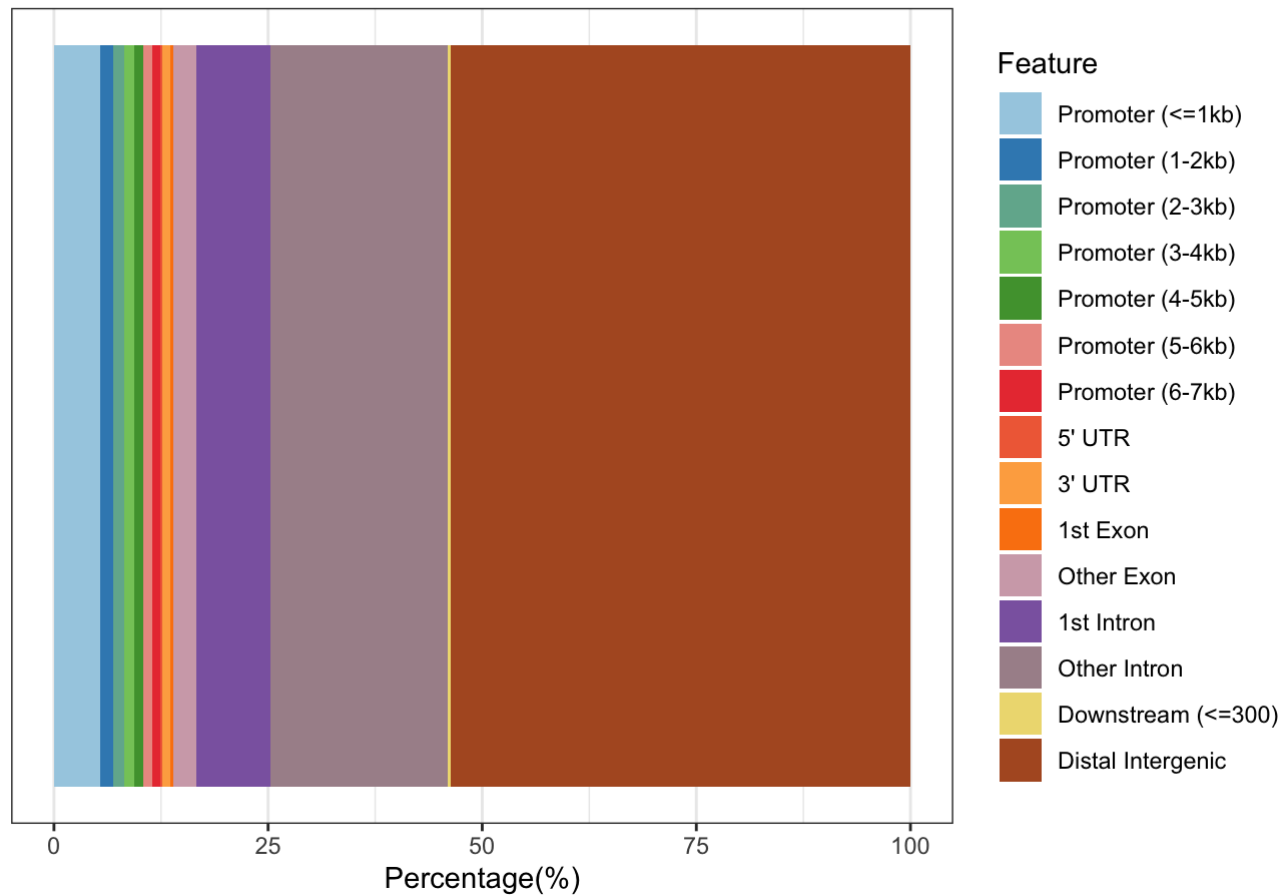
```
plotAnnoPie(Anno_NY8_2, title = "Distribution of ChIP peaks")
```



```
plotAnnoBar(Anno_NY8_2, title = "Distribution of ChIP peaks")
```



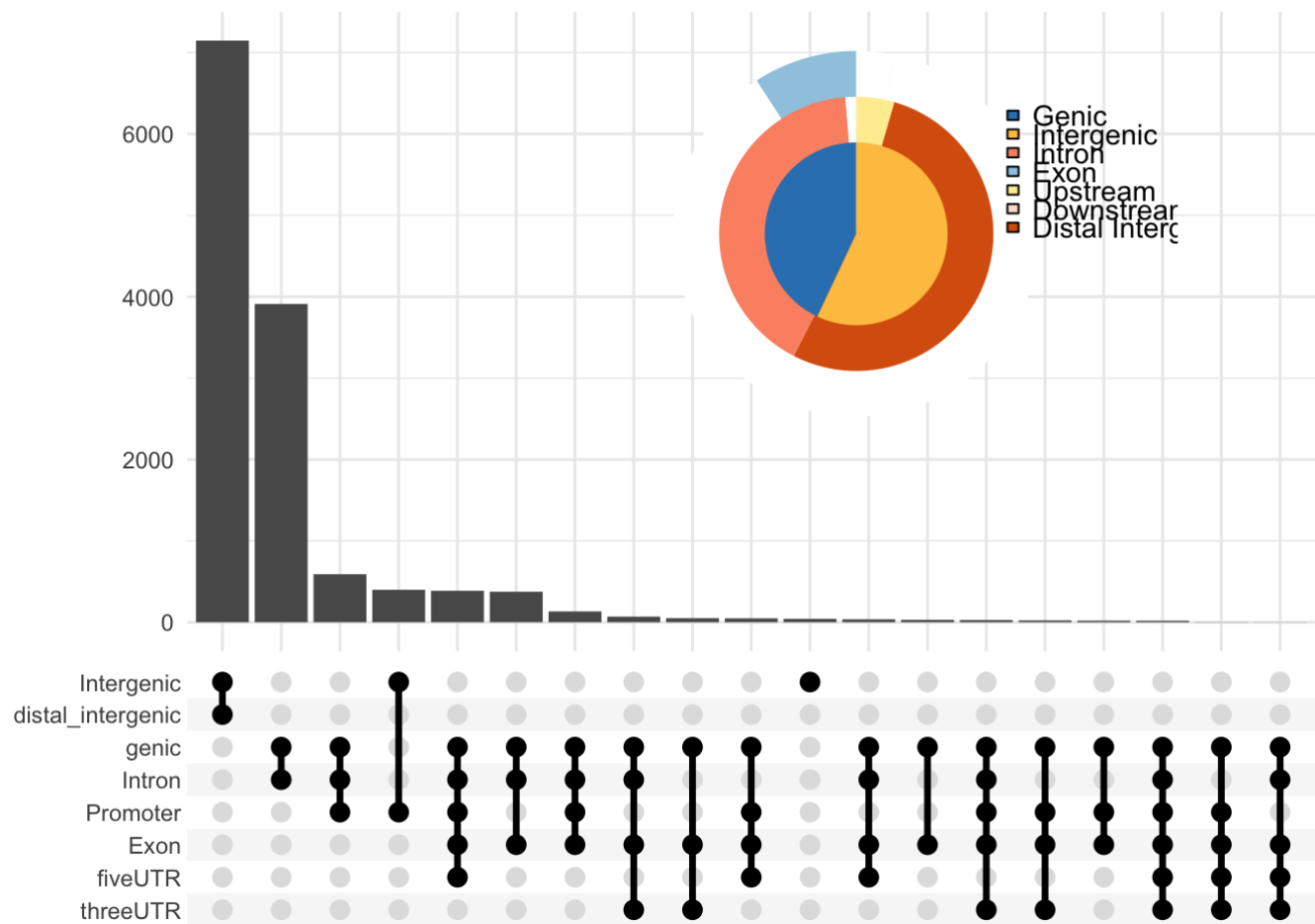
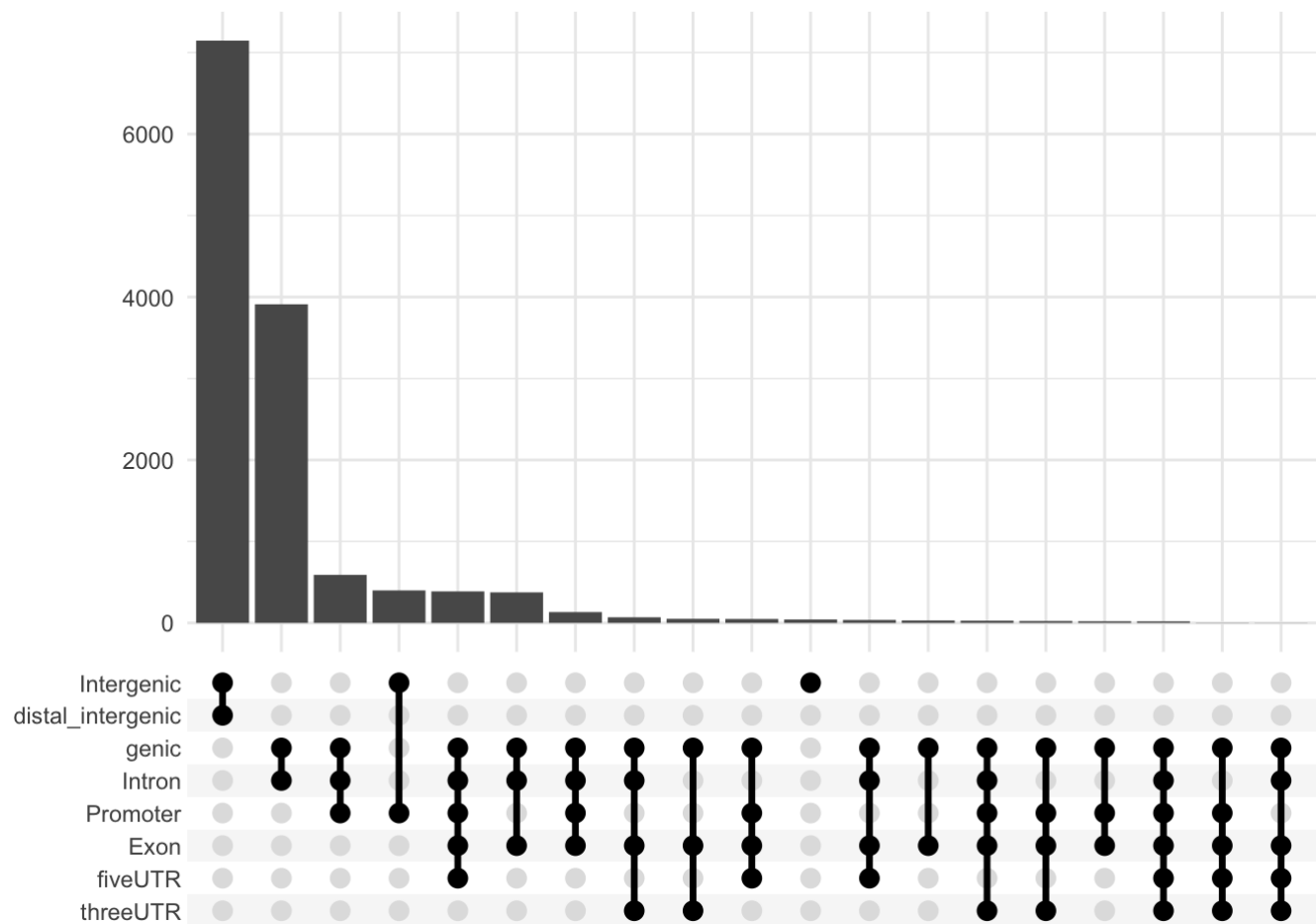
## Distribution of ChIP peaks



```
vennpie(Anno_NY8_2)
```

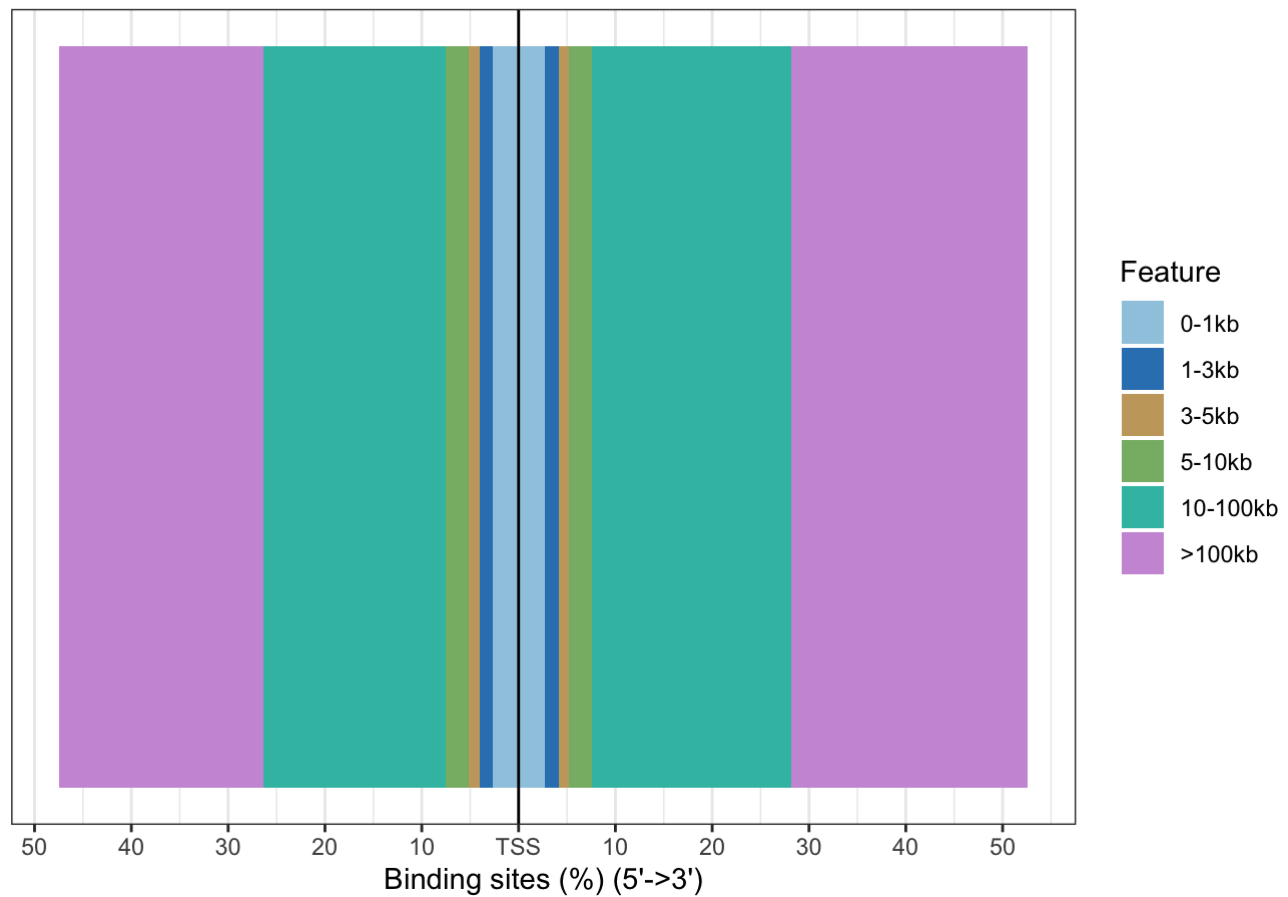
```
upsetplot(Anno_NY8_2)
```

```
upsetplot(Anno_NY8_2, vennpie = TRUE)
```



```
plotDistToTSS(Anno_NY8_2, title = "Distribution of TF-binding peaks relative to TSS")
```

### Distribution of TF-binding peaks relative to TSS



# BREAK

## Functional Enrichment Analysis

```
pathway_NY8_2 <- enrichPathway(as.data.frame(Anno_NY8_2)$geneId)
head(pathway_NY8_2, 5)
```

##	ID	Description	GeneRatio
## R-HSA-1500931	R-HSA-1500931	Cell-Cell communication	49/1896
## R-HSA-446728	R-HSA-446728	Cell junction organization	36/1896
## R-HSA-112316	R-HSA-112316	Neuronal System	110/1896
## R-HSA-9006934	R-HSA-9006934	Signaling by Receptor Tyrosine Kinases	122/1896
## R-HSA-383280	R-HSA-383280	Nuclear Receptor transcription pathway	21/1896

##	BgRatio	pvalue	p.adjust	qvalue
## R-HSA-1500931	129/10654	3.995355e-08	0.0000546165	5.240223e-05
## R-HSA-446728	91/10654	7.908045e-07	0.0005405149	5.186013e-04
## R-HSA-112316	413/10654	3.420530e-06	0.0015586215	1.495432e-03
## R-HSA-9006934	473/10654	5.666873e-06	0.0019366537	1.858138e-03
## R-HSA-383280	52/10654	1.085634e-04	0.0296812280	2.847789e-02

##  
geneID

## R-HSA-1500931  
55243/3918/3914/3688/56288/1308/1500/23705/5818/84623/5829/10979/87/1006/1009/1012/1366/3691/5296/81/7408/6711/3655/6714/84612/1002/1499/961/25945/51208/5010/9076/1016/1008/1004/5295/23607/667/2534/4301/221935/8935/9069/8976/137075/1015/5339/8573/9075

## R-HSA-446728  
3918/3914/3688/56288/1308/1500/23705/5818/5829/10979/87/1006/1009/1012/1366/3691/7408/3655/84612/1002/1499/25945/51208/5010/9076/1016/1008/1004/667/4301/221935/9069/137075/1015/5339/9075

## R-HSA-112316  
2035/2899/6512/57554/3749/3752/777/3756/3775/1103/347731/161/8495/3739/6506/1740/2915/2900/4900/341359/81539/54407/112/23426/3747/6857/8499/217/5564/8874/54331/27133/9369/56659/2562/2567/57369/5923/4916/2903/5579/3768/81033/5578/3773/3759/2788/3790/805/9378/80059/347730/57669/90134/2744/1385/23236/2036/22839/6714/6538/6529/131096/55799/22865/590/22871/10242/55584/10611/817/108/5562/348980/5924/3781/2741/2890/2561/2559/2566/347733/89822/57497/4646/2898/2037/6582/5879/2770/5577/3751/3084/1142/9312/83988/27012/114/3786/51305/169522/5789/10497/116443/2742/11141/8573/4128/1741/2892

## R-HSA-9006934  
10076/8503/8569/255631/1301/10451/10768/4803/117145/3918/4688/127124/3914/5867/3688/8829/10718/2263/161/5795/1500/10413/80310/5800/3709/3071/5908/10818/4254/1848/3479/10019/5829/10810/10253/8660/8874/382/23768/3320/2562/2567/7057/4734/102/4916/5046/5579/9564/2886/84951/535/5578/116/9306/5296/558/5581/805/200734/5899/2591/8828/1385/2335/3667/23767/6714/7074/1291/135/23543/1499/256076/131873/30849/152831/374/685/2247/10252/56034/2255/55914/5295/4208/2246/26999/2561/2566/8817/2264/10048/1432/7422/1297/2534/3910/5796/23118/7058/5154/5879/9844/64759/2887/1956/3912/4233/1846/3084/2260/51606/55824/11059/6461/5125/5774/79109/2889/5900/1536

## R-HSA-383280  
8431/6258/2494/2104/7181/9971/9612/2103/6095/4929/3172/5468/5915/4306/2908/5467/3174/8013/2649/6256/367

##	Count
## R-HSA-1500931	49
## R-HSA-446728	36
## R-HSA-112316	110
## R-HSA-9006934	122
## R-HSA-383280	21

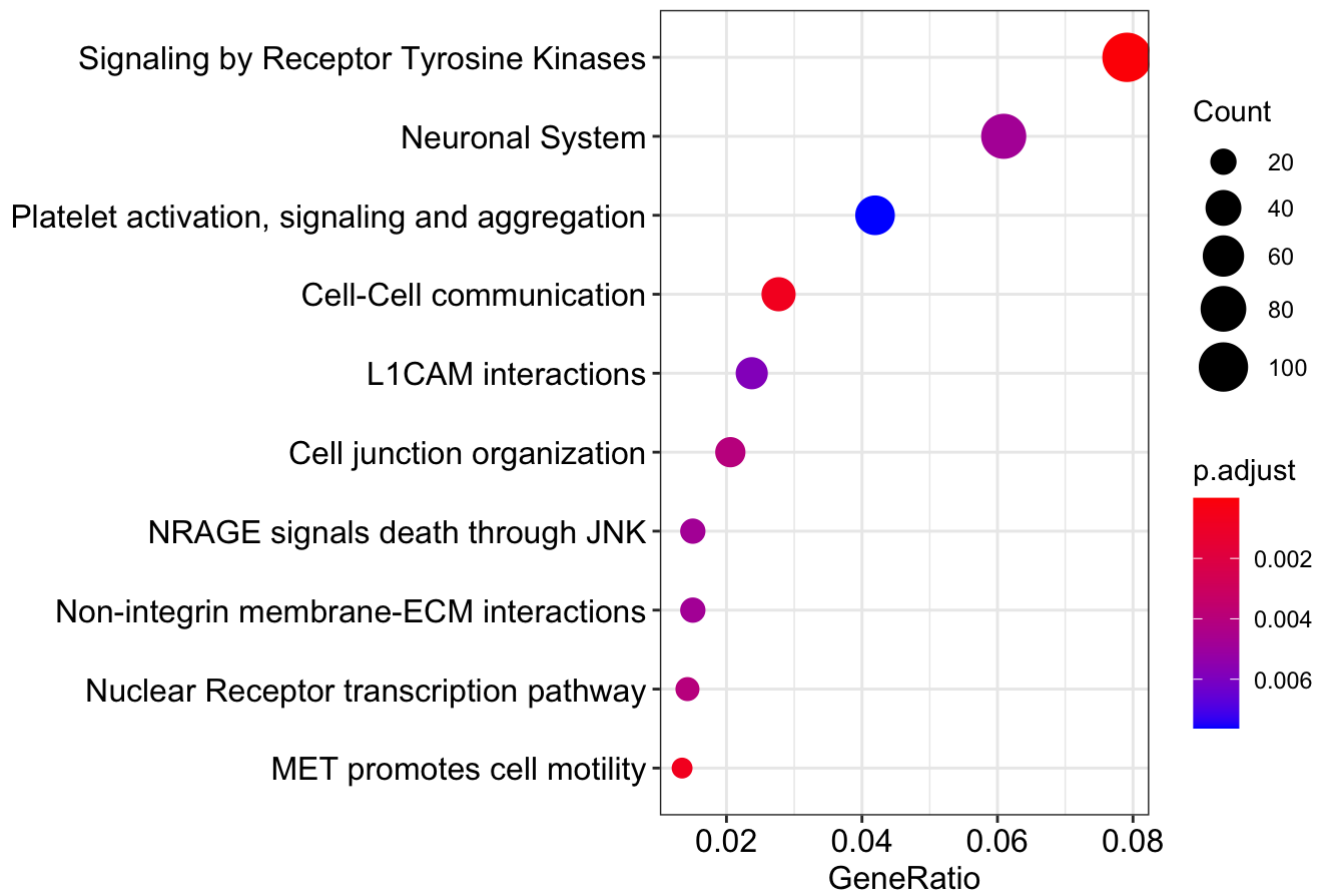
```

gene_NY8_2 <- seq2gene(NY8_2_single, tssRegion = c(-2000, 2000), flankDistance = 5000, T
xDb = txdb)
pathway_NY8_2_gene <- enrichPathway(gene_NY8_2)
head(pathway_NY8_2_gene, 5)

```

##	ID	Description	GeneRatio		
##	R-HSA-9006934	R-HSA-9006934 Signaling by Receptor Tyrosine Kinases	100/1264		
##	R-HSA-8875878	R-HSA-8875878 MET promotes cell motility	17/1264		
##	R-HSA-1500931	R-HSA-1500931 Cell-Cell communication	35/1264		
##	R-HSA-446728	R-HSA-446728 Cell junction organization	26/1264		
##	R-HSA-383280	R-HSA-383280 Nuclear Receptor transcription pathway	18/1264		
##	BgRatio	pvalue	p.adjust	qvalue	
##	R-HSA-9006934	473/10654	2.992588e-09	3.884379e-06	3.606856e-06
##	R-HSA-8875878	41/10654	1.524661e-06	6.811564e-04	6.324906e-04
##	R-HSA-1500931	129/10654	1.574321e-06	6.811564e-04	6.324906e-04
##	R-HSA-446728	91/10654	1.245539e-05	4.039737e-03	3.751114e-03
##	R-HSA-383280	52/10654	1.556139e-05	4.039737e-03	3.751114e-03
##	geneID				
##	R-HSA-9006934	255631/1301/10451/3918/127124/5867/8829/10413/3071/1848/3479/2886/1213/200734/1290/6714/1399/55615/256076/131873/152831/2264/64759/4233/1536/8503/85440/4803/26052/10000/3688/10718/161/5795/5241/80310/3709/5829/1284/8874/3320/2562/2567/4734/4916/3480/5046/5579/9564/51741/3728/535/5578/3643/1785/558/5581/5899/2591/8828/3667/7074/23543/3708/5156/374/2247/56034/2255/5295/4208/26999/2561/2566/10048/1432/2534/3910/5796/3908/23118/5154/5879/9844/1956/3084/51606/55824/11059/6461/5125/5774/79109/1759/2889/5900/3914/30849/3912/2335			
##	R-HSA-8875878	255631/1301/3918/1290/6714/1399/64759/4233/85440/3688/3910/3908/5879/2889/3914/3912/2335			
##	R-HSA-1500931	3918/1308/6714/23607/667/9069/1015/5339/3688/56288/23705/5818/84623/5829/10979/87/1012/3728/81/7408/6711/3655/84612/1002/253559/5010/1016/5295/2534/4301/221935/8935/3914/8976/1366			
##	R-HSA-446728	3918/1308/667/9069/1015/5339/3688/56288/23705/5818/5829/10979/87/1012/3728/7408/3655/84612/1002/253559/5010/1016/4301/221935/3914/1366			
##	R-HSA-383280	8431/2494/9971/9612/3172/5468/3174/2649/6258/2104/5241/2103/6095/5915/2908/5467/6256/367			
##	Count				
##	R-HSA-9006934	100			
##	R-HSA-8875878	17			
##	R-HSA-1500931	35			
##	R-HSA-446728	26			
##	R-HSA-383280	18			

```
dotplot (pathway_NY8_2_gene)
```

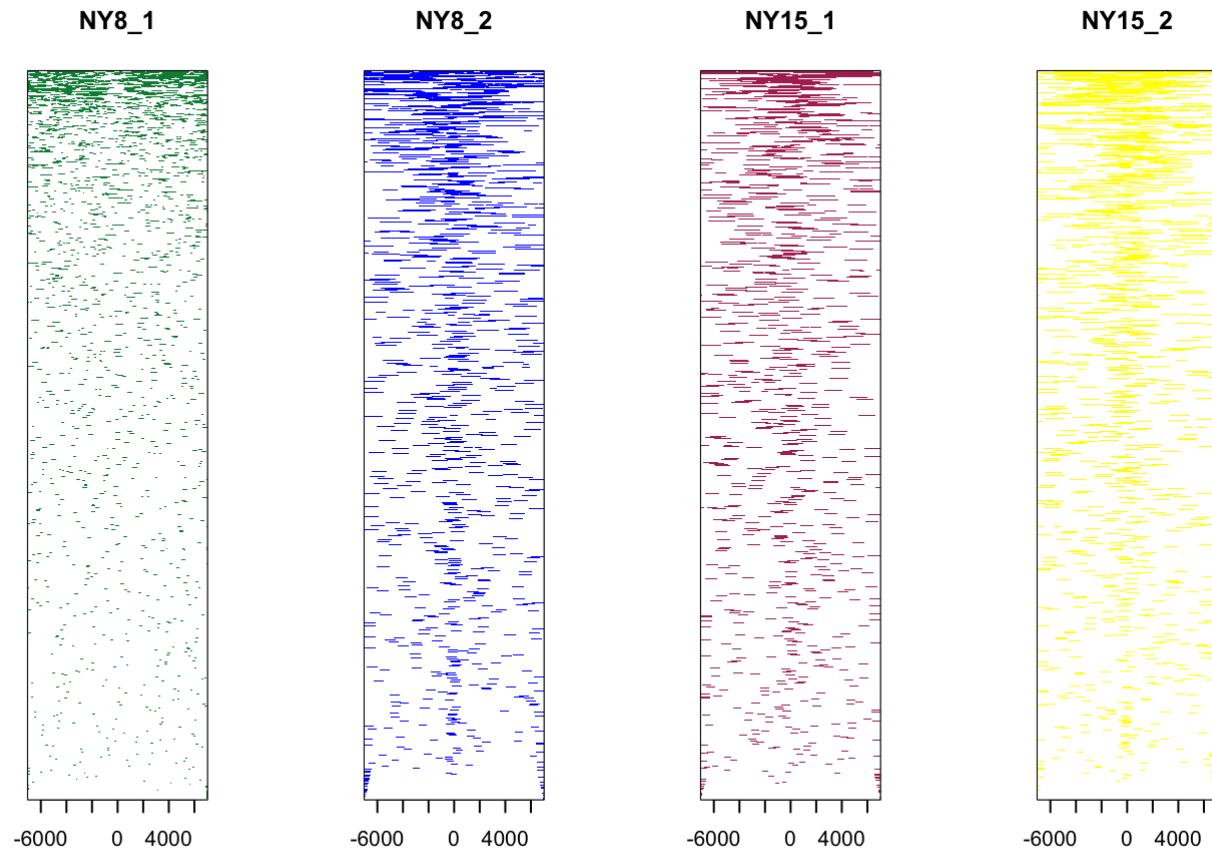


# BREAK

## Analyzing multiple files

```
peakHeatmap(all_files, TxDb = txdb, upstream = 7000, downstream = 7000, color = c("springgreen4", "blue", "maroon", "yellow"))
```

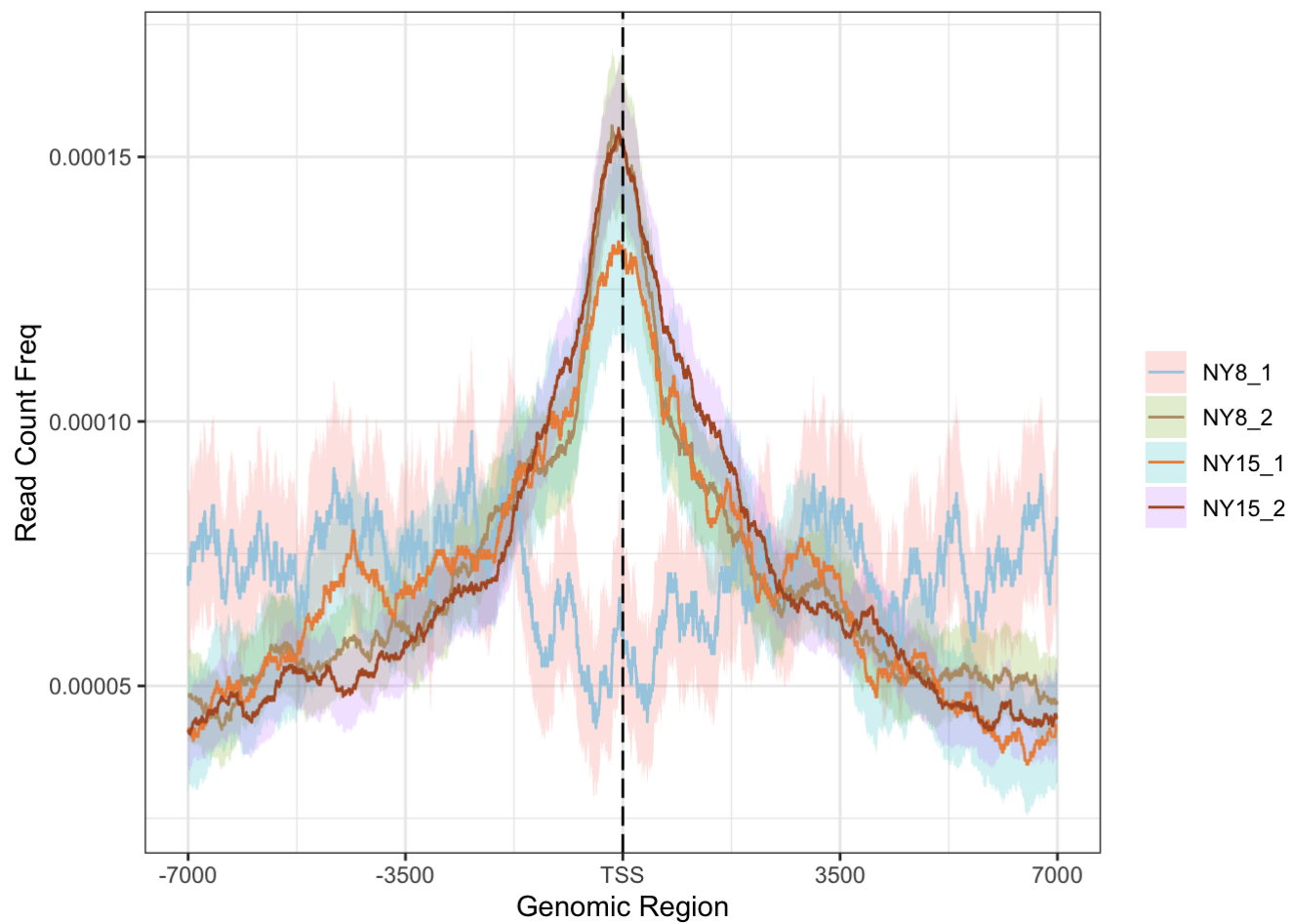
```
## >> preparing promoter regions... 2020-04-22 01:45:04
## >> preparing tag matrix...      2020-04-22 01:45:04
## >> generating figure...         2020-04-22 01:45:57
```



```
## >> done...                2020-04-22 01:46:05
```

```
plotAvgProf2(all_files, TxDb = txdb, upstream = 7000, downstream = 7000, xlab = "Genomic
Region", ylab = "Read Count Freq", conf = 0.95, resample = 1000)
```

```
## >> preparing promoter regions... 2020-04-22 01:46:06
## >> preparing tag matrix...       2020-04-22 01:46:06
## >> plotting figure...            2020-04-22 01:46:57
## >> Running bootstrapping for tag matrix... 2020-04-22 01:47:13
## >> Running bootstrapping for tag matrix... 2020-04-22 01:47:41
## >> Running bootstrapping for tag matrix... 2020-04-22 01:48:00
## >> Running bootstrapping for tag matrix... 2020-04-22 01:48:30
```

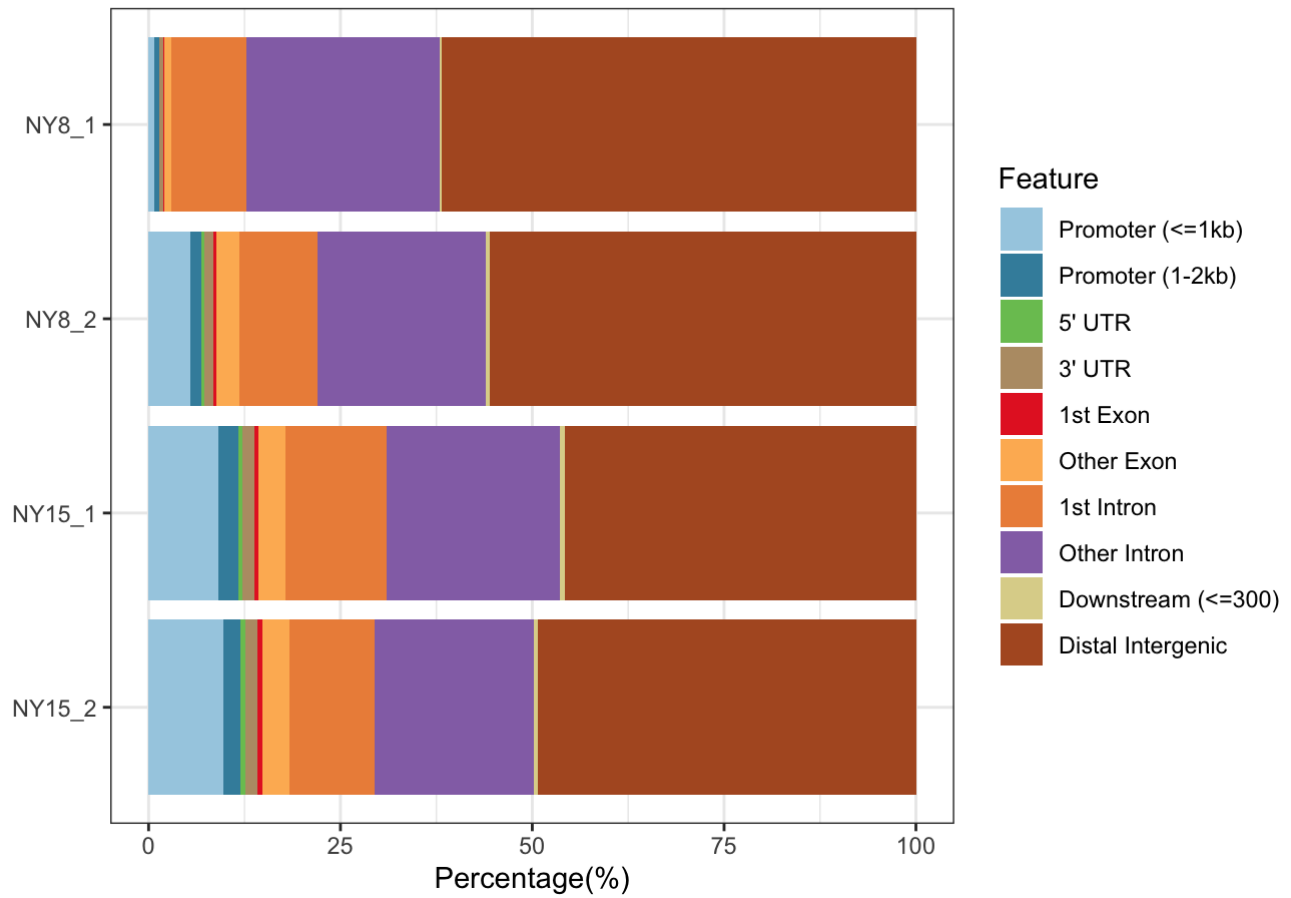


```
Annotation_all <- lapply (all_files, annotatePeak, TxDb = txdb, tssRegion = c(-2000, 2000), verbose = FALSE)

plotAnnoBar (Annotation_all, title = "Distribution of ChIP peaks")
```

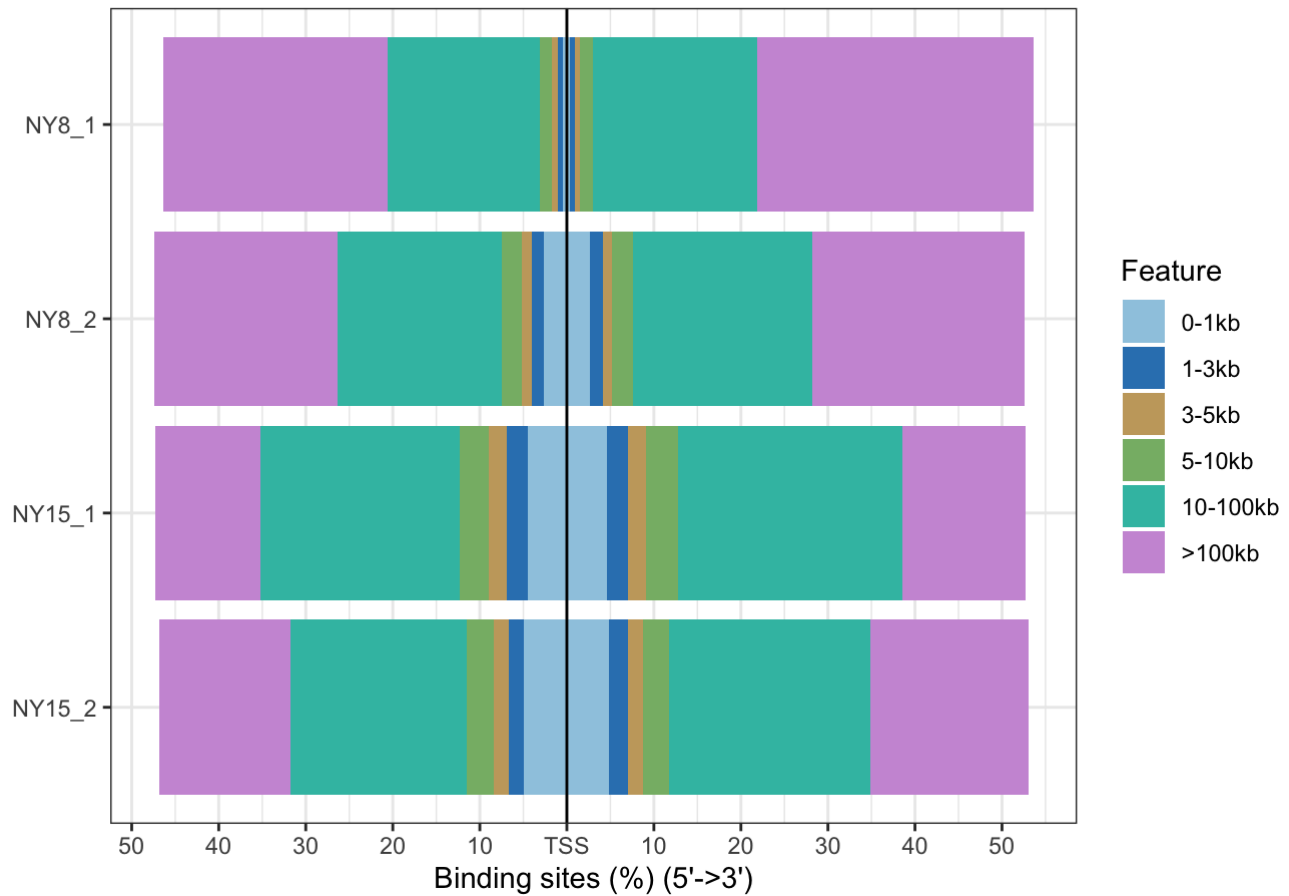


Distribution of ChIP peaks



```
plotDistToTSS(Annotation_all, title = "Distribution of TF-binding peaks relative to TSS"
)
```

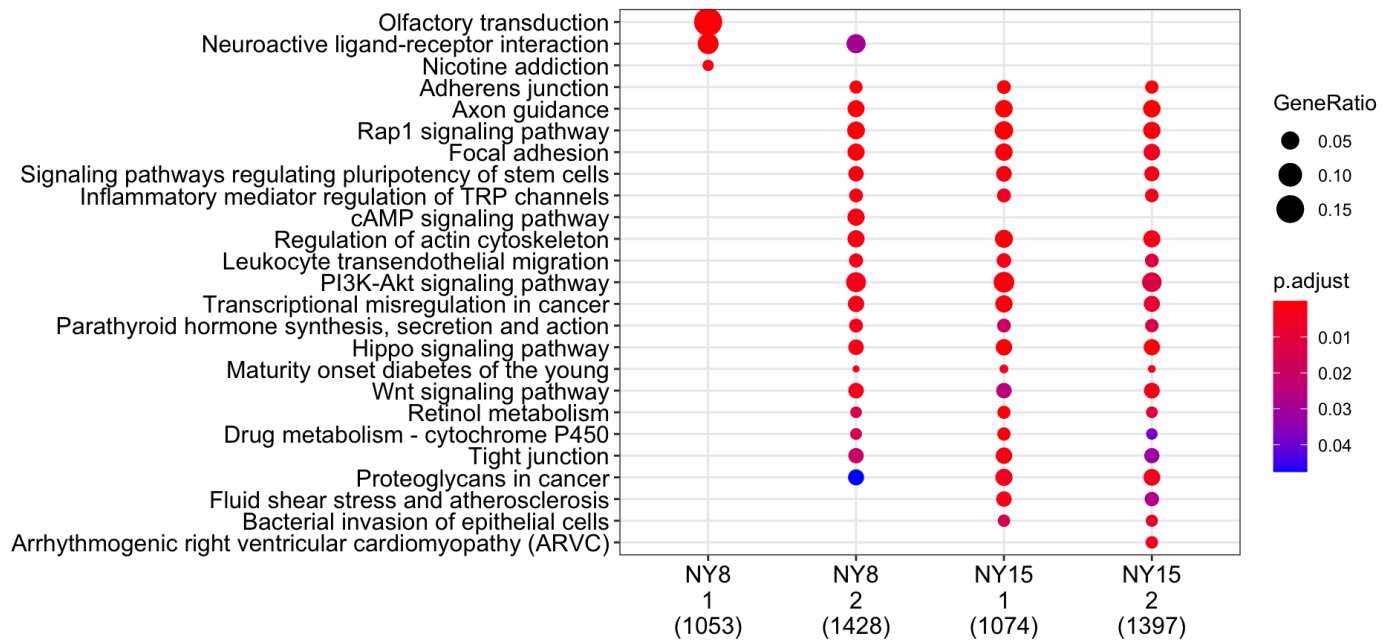
## Distribution of TF-binding peaks relative to TSS



```
all_genes <- lapply (Annotation_all, function(i) as.data.frame(i)$geneId)
names(all_genes) = sub("_", "\n", names(all_genes))
compareKEGG <- compareCluster(geneClusters = all_genes, fun = "enrichKEGG", pvalueCutoff
= 0.05, pAdjustMethod = "BH")
```

```
dotplot(compareKEGG, showCategory = 15, title = "Comparison of KEGG Pathway Enrichment A
nalysis")
```

Comparison of KEGG Pathway Enrichment Analysis



THE END