Visualisation of PTEN

Tiphaine Martin

April 19, 2018

```
library("coMET")

## Loading required package:  grid
## Loading required package:  biomaRt
## Loading required package:  Gviz
## Loading required package:  S4Vectors
## Loading required package:  stats4
## Loading required package:  BiocGenerics
## Loading required package:  parallel
##
## Attaching package:  'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##    clusterApply, clusterApplyLB, clusterCall, clusterEvalQ, clusterExport,
##    clusterMap, parApply, parCapply, parLapply, parLapplyLB, parRapply,
##    parSapply, parSapplyLB
## The following objects are masked from 'package:stats':
##
##    IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##    anyDuplicated, append, as.data.frame, cbind, colMeans, colnames, colSums,
##    do.call, duplicated, eval, evalq, Filter, Find, get, grep, grepl, intersect,
##    is.unsorted, lapply, lengths, Map, mapply, match, mget, order, paste, pmax,
##    pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rowMeans, rownames,
##    rowSums, sapply, setdiff, sort, table, tapply, union, unique, unsplit,
##    which, which.max, which.min
##
## Attaching package:  'S4Vectors'
## The following object is masked from 'package:base':
##
##    expand.grid
## Loading required package:  IRanges
## Loading required package:  GenomicRanges
## Loading required package:  GenomeInfoDb
## Loading required package:  psych
## Warning:  package 'psych' was built under R version 3.4.4
##
## Attaching package:  'psych'
## The following object is masked from 'package:IRanges':
##
##    reflect
```

```
## Loading required package:  ggbio
## Loading required package:  ggplot2
##
## Attaching package:  'ggplot2'
## The following objects are masked from 'package:psych':
##
##    %+%, alpha
## Need specific help about ggbio?  try mailing
##  the maintainer or visit http://tengfei.github.com/ggbio/
##
## Attaching package:  'ggbio'
## The following objects are masked from 'package:ggplot2':
##
##    geom_bar, geom_rect, geom_segment, ggsave, stat_bin, stat_identity, xlim
## The following object is masked from 'package:psych':
##
##    rescale
## Loading required package:  trackViewer
```

```
library("data.table")
```

```
##
## Attaching package:  'data.table'
## The following object is masked from 'package:GenomicRanges':
##
##    shift
## The following object is masked from 'package:IRanges':
##
##    shift
## The following objects are masked from 'package:S4Vectors':
##
##    first, second
```

```
library("rtracklayer")
library("gwascat")
```

```
## Loading required package:  Homo.sapiens
## Loading required package:  AnnotationDbi
## Loading required package:  Biobase
## Welcome to Bioconductor
##
##    Vignettes contain introductory material; view with 'browseVignettes()'.  To
##    cite Bioconductor, see 'citation("Biobase")', and for packages
##    'citation("pkgname")'.
## Loading required package:  OrganismDbi
## Loading required package:  GenomicFeatures
## Loading required package:  GO.db
##
## Loading required package:  org.Hs.eg.db
##
## Loading required package:  TxDb.Hsapiens.UCSC.hg19.knownGene
## gwascat loaded.  Use data(ebicat38) for hg38 coordinates;
##  data(ebicat37) for hg19 coordinates.
```

```
library("IRanges")
library("GenomicRanges")
```

PTEN

PTEN gene :

- hg19 chr10:89,623,195-89,728,532 <- visualisation 89423195-89828532 (-200,000/+100,000)
- hg38 chr10:87,863,113-87,971,930 <- visualisation 87663113-88071930 (-200,000/+100,000)

```
chrom <- "chr10"
chrom_ucsc <- "10"
#hg19
start_hg19 <- 89423195
end_hg19 <- 89828532
gen_hg19 <- "hg19"

#hg38
start_hg38 <- 87663113
end_hg38 <- 88071930
gen_hg38 <- "hg38"
strand <- "*"

BROWSER.SESSION="UCSC"
mySession <- browserSession(BROWSER.SESSION)
```

Peak P53 in MCF10A

data from paper Pappas et al, 2017, "p53 Maintains Baseline Expression of Multiple Tumor Suppressor Genes" http://mcr.aacrjournals.org/content/15/8/1051.long

```
cd /sc/prga/projects/parsonslab/P53_KyriePappas/Pappas_p53_manuscript/\
analysis_ChIPseq/macs_mcf10a_EV/
```

```
awk -F"\t" '$1 == "chr10" && $2 > 89423195 && $3 < 89828532 {print $0}' peaks.clean.bed > \
~/peaks.clean_MCF10A_EVp53_PTEN.bed
```

```
p53_MCF10A_path=paste0(dirfolder,"peaks.clean_MCF10A_EVp53_PTEN.bed")
p53_MCF10A_tab <- read.table(file=p53_MCF10A_path,header=FALSE,sep="\t")

df_p53_MCF10A_PTEN <-data.frame(chr="chr10",start=p53_MCF10A_tab$V2,
                      end=p53_MCF10A_tab$V3,
                      strand="*",score=0,
                      feature="hotspotP53",
                      id=1:nrow(p53_MCF10A_tab),
                      group=1:nrow(p53_MCF10A_tab))
genome(mySession) <- gen_hg19
ChIPseqP53HotSpotTrack <- AnnotationTrack(genome="hg19",range=df_p53_MCF10A_PTEN,
                                    chromosome=chrom_ucsc,
                                  name = "hotspot ChIP-seq P53 in MCF10A cell-line",
                                  stacking="dense",
                                  col.line = "black", col = NULL, collapse= FALSE)
displayPars(ChIPseqP53HotSpotTrack) <- list(rotation.title = 360, cex.title=0.5,
                                        stacking="dense", shape="box",
                                        stackHeight=0.8,
```

```
                                    background.title="transparent",
                                col.title="black", col=NULL,
                                hotspotP53 = "darkred")
```

Extract COSMIC data

Data were downloaded on the March,26th 2018 and are release v84 (13th February 2018)

1. Downlaod only mutations in Breast tissues in the section **COSMIC Mutation Data** in COSMIC database (`https://cancer.sanger.ac.uk/cosmic/download`)
2. Slip the value of the column **MUTATION_GENOME_POSITION** in 3 columns (chromosome, start and end)
3. Slip the value of the column **MUTATION_DESCRIPTION** in 2 columns (main and subtype)
4. extract only the mutation in the region of interest

Note as the most of data visualised around PTEN in the version hg19 and that COSMIC is in the version hg38, we need to liftover the position. More info at this `http://hgdownload-test.cse.ucsc.edu/goldenPath/hg38/liftOver/`

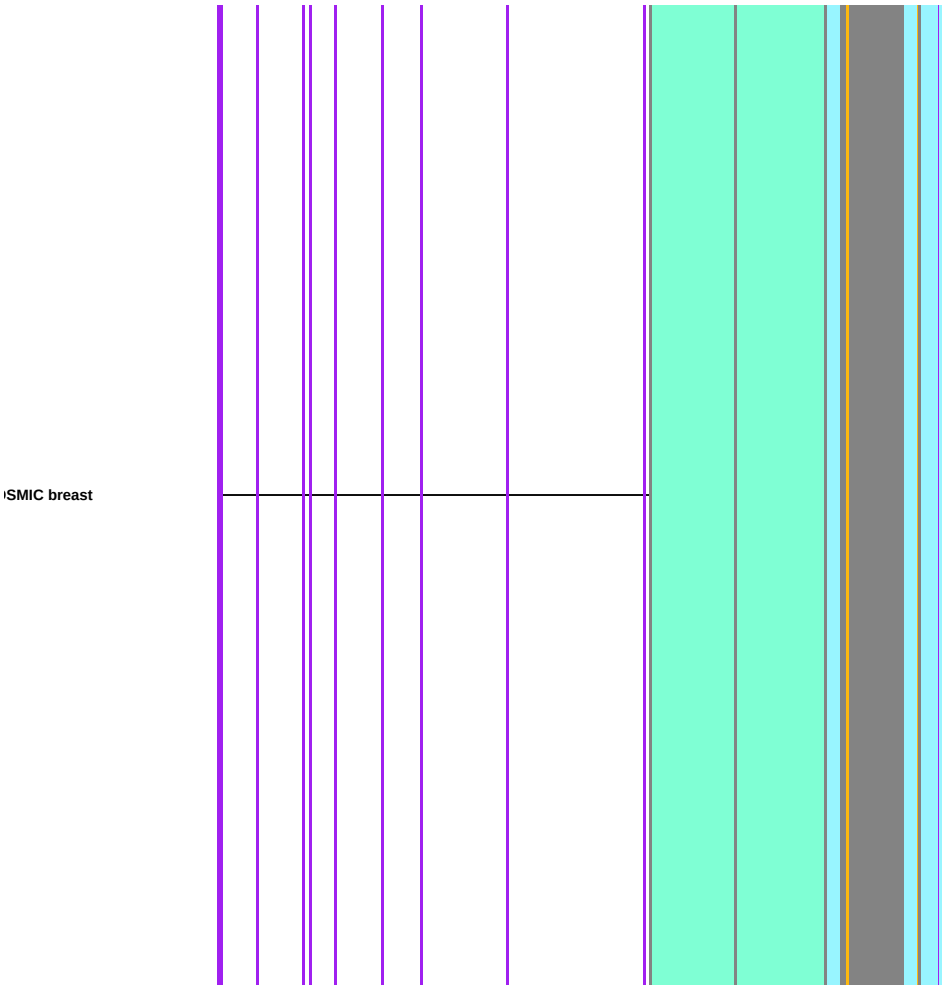
rsync -avzP rsync://hgdownload.cse.ucsc.edu/goldenPath/hg38/liftOver/hg38ToHg19.over.chain.gz . gunzip hg38ToHg19.over.chain.gz

color for different features in COSMIC:

Name	hex code	color
Complex	#556B2F	
Deletion	#98F5FF	
Frameshift	#FF7256	
Insertion	#FFB90F	
Nonstop extension	#ADFF2F	
Substitution	#A020F0	
Unknown	#838383	
Whole gene deletion	#7FFFD4	

```
genome(mySession) <- gen_hg38

cosmicfile_path<-paste0(dirfolder,"V84_38_MUTANT_BREAST_updated.csv")
cosmicMutation <- read.csv(file=cosmicfile_path,header=TRUE)

cosmicMutationPTEN <- cosmicMutation[which(cosmicMutation$MUTATION_GENOME_POSITION_CHROM %in% chrom_ucsc
                                           cosmicMutation$MUTATION_GENOME_POSITION_START >start_hg38 &
                                           cosmicMutation$MUTATION_GENOME_POSITION_STOP < end_hg38 ),]
dim(cosmicMutation)

## [1] 277553     40

dim(cosmicMutationPTEN)
```

```
## [1] 257  40

#head(cosmicMutationPTEN)

cosmicMutationPTEN_short <- unique(cosmicMutationPTEN[,which(colnames(cosmicMutationPTEN) %in%
                                                               c("MUTATION_GENOME_POSITION_CHROM",
                                                                 "MUTATION_GENOME_POSITION_START",
                                                                 "MUTATION_GENOME_POSITION_STOP",
                                                                 "MUTATION_STRAND",
                                                                 "MUTATION_DESCRIPTION_main"))])
dim(cosmicMutationPTEN_short)

## [1] 164   5

# list of different types to visualise
#remove space begin and end of string
cosmicMutationPTEN_short$MUTATION_DESCRIPTION_main <- trimws(cosmicMutationPTEN_short$MUTATION_DESCRIPTI
table(cosmicMutationPTEN_short$MUTATION_DESCRIPTION_main)

##
##             Complex            Deletion           Insertion        Substitution
##                   2                  37                  20                  84
##             Unknown Whole gene deletion
##                  20                   1

#strand=as.character(cosmicMutationPTEN_short$MUTATION_STRAND)
df_PTEN <-data.frame(chr="chr10",start=cosmicMutationPTEN_short$MUTATION_GENOME_POSITION_START,
                     end=cosmicMutationPTEN_short$MUTATION_GENOME_POSITION_STOP,
                     strand="*",score=0,
                     feature=cosmicMutationPTEN_short$MUTATION_DESCRIPTION_main,
                     id=1:nrow(cosmicMutationPTEN_short),group=1:nrow(cosmicMutationPTEN_short))

## liftover hg38 to hg19
chain <- import.chain(paste0(dirfolder,"hg38ToHg19.over.chain"))
gr <- makeGRangesFromDataFrame(df_PTEN, TRUE)
humcon <- liftOver(gr, chain)

genome(mySession) <- gen_hg19
cosmicTrack <- AnnotationTrack(genome="hg19",range=humcon,chromosome=chrom_ucsc,
                                    name = "COSMIC breast",
                                    stacking="full",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(cosmicTrack) <- list('Complex' = '#556B2F', 'Deletion' = '#98F5FF',
                                    'Frameshift' = '#FF7256', 'Insertion' = '#FFB90F',
                                    'Nonstop extension' = '#ADFF2F',
                                    'Substitution' = '#A020F0', 'Unknown' = '#838383',
                                    'Whole gene deletion' = '#7FFFD4')
displayPars(cosmicTrack) <- list(rotation.title = 360, cex.title=0.5,
                                   shape="box",stackHeight=0.8,
                                   background.title="transparent", col.title="black",
                                   col=NULL)

plotTracks(cosmicTrack,from=start_hg19,to=end_hg19)
```



Extact chromHMM around PTEN from 5 normal Breast cell-line

5 breast cell-lines published by Pellacani et al., 2017, "Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks" More info on data `https://martit26.u.hpc.mssm.edu/`

- CEMT0007 for MCF10A cell-line
- CEMT0035 for Basal human breast cell-line
- CEMT0036 for stromal human breast cell-line
- CEMT0037 for luminal mature human cell-line
- CEMT0038 for lumninal progenitor human cell-line

```
awk -F"\t" '$1 == "chr10" && $2 > 89423195 && $3 < 89828532 {print $0}' CEMT_7_18_segments.
~/CEMT_7_MCF10A_PTEN_18_segments.bed

gunzip CEMT_35_18_segments.bed.gz | awk -F"\t" '$1 == "chr10" && $2 > 89423195 && $3 < 8982
{print $0}' CEMT_7_18_segments.bed > ~/CEMT_35_Basal_PTEN_18_segments.bed

gunzip CEMT_36_18_segments.bed.gz | awk -F"\t" '$1 == "chr10" && $2 > 89423195 && $3 < 8982
{print $0}' CEMT_7_18_segments.bed > ~/CEMT_36_Stromal_PTEN_18_segments.bed

gunzip CEMT_37_18_segments.bed.gz | awk -F"\t" '$1 == "chr10" && $2 > 89423195 && $3 < 8982
{print $0}' CEMT_7_18_segments.bed > ~/CEMT_37_LuminalMature_PTEN_18_segments.bed

gunzip CEMT_38_18_segments.bed.gz | awk -F"\t" '$1 == "chr10" && $2 > 89523195 && $3 < 8982
{print $0}' CEMT_7_18_segments.bed > ~/CEMT_38_LuminalProgenitor_PTEN_18_segments.bed
```

list of different states:

```
Emission Name Short\_Name Order
E1 Enhancer Enh 7
E2 TSS\_Flank\_Down TSS\_Flnk\_D 4
E3 TSS\_Flank\_1 TSS\_Flnk\_1 2
E4 Active TSS TSS\_A 1
E5 TSS\_Flank\_2 TSS\_Flnk\_2 3
E6 TSS\_Bivalent TSS\_Biv 11
E7 Enhhancer\_Bivalent Enh\_Biv 12
E8 Polycomb\_Repressed Repr\_PC 13
E9 Quiescent Quies 17
E10 Repressed Repr 16
E11 Heterochromatin Het 15
E12 Znf\_Repeats Znf\_Rpts 14
E13 Transcribed\_3\_prime Tx\_3p 6
E14 Transcribed Tx 5
E15 Enhancer\_Genic Enh\_G 9
E16 Enhancer\_Genic\_Active Enh\_G\_A 10
E17 Enhancer\_Active Enh\_A 8
E18 Quiescent\_Genic Quies\_G 18
```

Color for different features:

Emission	Name	Short_Name	Order	hex code	color
E1	Enhancer	Enh	7	#FFFF00	
E2	TSS_Flank_Down	TSS_Flnk_D	4	#DA7B08	
E3	TSS_Flank_1	TSS_Flnk_1	2	#FF6E00	
E4	Active TSS	TSS_A	1	#FF0000	
E5	TSS_Flank_2	TSS_Flnk_2	3	#FF9300	
E6	TSS_Bivalent	TSS_Biv	11	#CD5C5C	
E7	Enhancer_Bivalent	Enh_Biv	12	#BDB76B	
E8	Polycomb_Repressed	Repr_PC	13	#AFAFAF	
E9	Quiescent	Quies	17	#DCDCDC	
E10	Repressed	Repr	16	#323232	
E11	Heterochromatin	Het	15	#8A91D0	
E12	Znf_Repeats	Znf_Rpts	14	#66CDAA	
E13	Transcribed_3_prime	Tx_3p	6	#006400	
E14	Transcribed	Tx	5	#008000	
E15	Enhancer_Genic	Enh_G	9	#C2FF05	
E16	Enhancer_Genic_Active	Enh_G_A	10	#FE00DB	
E17	Enhancer_Active	Enh_A	8	#FFA7D6	
E18	Quiescent_Genic	Quies_G	18	#F7F7F7	

```
# colorchromHMM <- list('E1' = '#FFFF00', 'E2' = '#FF4000','E3' = '#FF4000',
#                       'E4' = '#DF0101','E5' = '#FF4000','E6' = '#C43D3D',
#                       'E7' = '#688A08', 'E8' = '#585858','E9' = '#FFFFFF',
#                       'E10' = '#5F04B4','E11' = '#642EFE', 'E12' = '#01A9DB',
```

```
#                         'E13' = '#0E9000','E14' = '#13C000','E15' = '#A4F000',
#                         'E16' = '#F0B800', 'E17' = '#CC9C00', 'E18' = '#BDBDBD')
#AFAFAF
colorchromHMM <- list('E1' = '#FFFF00', 'E2' = '#DA7B08','E3' = '#FF6E00',
                      'E4' = '#FF0000','E5' = '#FF9300','E6' = '#CD5C5C',
                      'E7' = '#BDB76B', 'E8' = '#AFAFAF','E9' = '#DCDCDC',
                      'E10' = '#323232','E11' = '#8A91D0', 'E12' = '#66CDAA',
                      'E13' = '#006400','E14' = '#008000','E15' = '#C2FF05',
                      'E16' = '#FE00DB', 'E17' = '#FFA7D6', 'E18' = '#F7F7F7')

# MCF10A
chromHMMTrack_07_path<-paste0(dirfolder,"CEMT_7_18_MCF10A_PTEN_18_segments.bed")
chromHMMTrack_07_tab <- read.table(file=chromHMMTrack_07_path,header=FALSE,sep="\t")

chromHMMTrack_07_mat <- data.frame(chr="chr10",start=chromHMMTrack_07_tab[,2],
                                   end=chromHMMTrack_07_tab[,3],strand="*",
                                   score=0,feature=chromHMMTrack_07_tab[,4],
                                   id=1:nrow(chromHMMTrack_07_tab),
                                   group=1:nrow(chromHMMTrack_07_tab))
chromHMMTrack_07_gr <- makeGRangesFromDataFrame(chromHMMTrack_07_mat, TRUE)
chromHMMTrack_07 <- AnnotationTrack(genome="hg19",range=chromHMMTrack_07_gr,
                                    chromosome=chrom_ucsc,
                                    name = "chromHMM MCF10A cell-line",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(chromHMMTrack_07) <- colorchromHMM
displayPars(chromHMMTrack_07) <- list(rotation.title = 360, cex.title=0.5,
                                      shape="box",stackHeight=0.8,
                                      background.title="transparent", col.title="black",
                                      col=NULL)

# Basal
chromHMMTrack_35_path<-paste0(dirfolder,"CEMT_35_18_Basal_PTEN_18_segments.bed")
chromHMMTrack_35_tab <- read.table(file=chromHMMTrack_35_path,header=FALSE,sep="\t")

chromHMMTrack_35_mat <- data.frame(chr="chr10",start=chromHMMTrack_35_tab[,2],
                                   end=chromHMMTrack_35_tab[,3],strand="*",
                                   score=0,feature=chromHMMTrack_35_tab[,4],
                                   id=1:nrow(chromHMMTrack_35_tab),
                                   group=1:nrow(chromHMMTrack_35_tab))

chromHMMTrack_35_gr <- makeGRangesFromDataFrame(chromHMMTrack_35_mat, TRUE)
chromHMMTrack_35 <- AnnotationTrack(genome="hg19",range=chromHMMTrack_35_gr,
                                    chromosome=chrom_ucsc,
                                    name = "chromHMM basal cell fraction",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(chromHMMTrack_35) <- colorchromHMM
displayPars(chromHMMTrack_35) <- list(rotation.title = 360, cex.title=0.5,
                                      shape="box",stackHeight=0.8,
                                      background.title="transparent", col.title="black",
                                      col=NULL)

# stromal
```

```
chromHMMTrack_36_path<-paste0(dirfolder,"CEMT_36_18_Stromal_PTEN_18_segments.bed")
chromHMMTrack_36_tab <- read.table(file=chromHMMTrack_36_path,header=FALSE,sep="\t")

chromHMMTrack_36_mat <- data.frame(chr="chr10",start=chromHMMTrack_36_tab[,2],
                                   end=chromHMMTrack_36_tab[,3],strand="*",
                                   score=0,feature=chromHMMTrack_36_tab[,4],
                                   id=1:nrow(chromHMMTrack_36_tab),
                                   group=1:nrow(chromHMMTrack_36_tab))
chromHMMTrack_36_gr <- makeGRangesFromDataFrame(chromHMMTrack_36_mat, TRUE)
chromHMMTrack_36 <- AnnotationTrack(genome="hg19",range=chromHMMTrack_36_gr,
                                    chromosome=chrom_ucsc,
                                    name = "chromHMM stromal cell fraction",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(chromHMMTrack_36) <- colorchromHMM
displayPars(chromHMMTrack_36) <- list(rotation.title = 360, cex.title=0.5,
                                      shape="box",stackHeight=0.8,
                                      background.title="transparent", col.title="black",
                                      col=NULL)

# luminal mature
chromHMMTrack_37_path<-paste0(dirfolder,"CEMT_37_18_LuminalMature_PTEN_18_segments.bed")
chromHMMTrack_37_tab <- read.table(file=chromHMMTrack_37_path,header=FALSE,sep="\t")

chromHMMTrack_37_mat <- data.frame(chr="chr10",start=chromHMMTrack_37_tab[,2],
                                   end=chromHMMTrack_37_tab[,3],strand="*",
                                   score=0,feature=chromHMMTrack_37_tab[,4],
                                   id=1:nrow(chromHMMTrack_37_tab),
                                   group=1:nrow(chromHMMTrack_37_tab))
chromHMMTrack_37_gr <- makeGRangesFromDataFrame(chromHMMTrack_37_mat, TRUE)
chromHMMTrack_37 <- AnnotationTrack(genome="hg19",range=chromHMMTrack_37_gr,
                                    chromosome=chrom_ucsc,
                                    name = "chromHMM mature luminal cell fraction",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(chromHMMTrack_37) <- colorchromHMM
displayPars(chromHMMTrack_37) <- list(rotation.title = 360, cex.title=0.5,
                                      shape="box",stackHeight=0.8,
                                      background.title="transparent", col.title="black",
                                      col=NULL)

# luminal progenitor
chromHMMTrack_38_path<-paste0(dirfolder,"CEMT_38_18_LuminalProgenitor_PTEN_18_segments.bed")
chromHMMTrack_38_tab <- read.table(file=chromHMMTrack_38_path,header=FALSE,sep="\t")

chromHMMTrack_38_mat <- data.frame(chr="chr10",
                                   start=chromHMMTrack_38_tab[,2],
                                   end=chromHMMTrack_38_tab[,3],strand="*",
                                   score=0,feature=chromHMMTrack_38_tab[,4],
                                   id=1:nrow(chromHMMTrack_38_tab),
                                   group=1:nrow(chromHMMTrack_38_tab))
chromHMMTrack_38_gr <- makeGRangesFromDataFrame(chromHMMTrack_38_mat, TRUE)
chromHMMTrack_38 <- AnnotationTrack(genome="hg19",range=chromHMMTrack_38_gr,
```

```
                                    chromosome=chrom_ucsc,
                                    name = "chromHMM luminal progenitor cell fraction",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(chromHMMTrack_38) <- colorchromHMM
displayPars(chromHMMTrack_38) <- list(rotation.title = 360, cex.title=0.5,
                                      shape="box",stackHeight=0.8,
                                      background.title="transparent",
                                      col.title="black", col=NULL)
```

ATAC-seq data in MCF10A

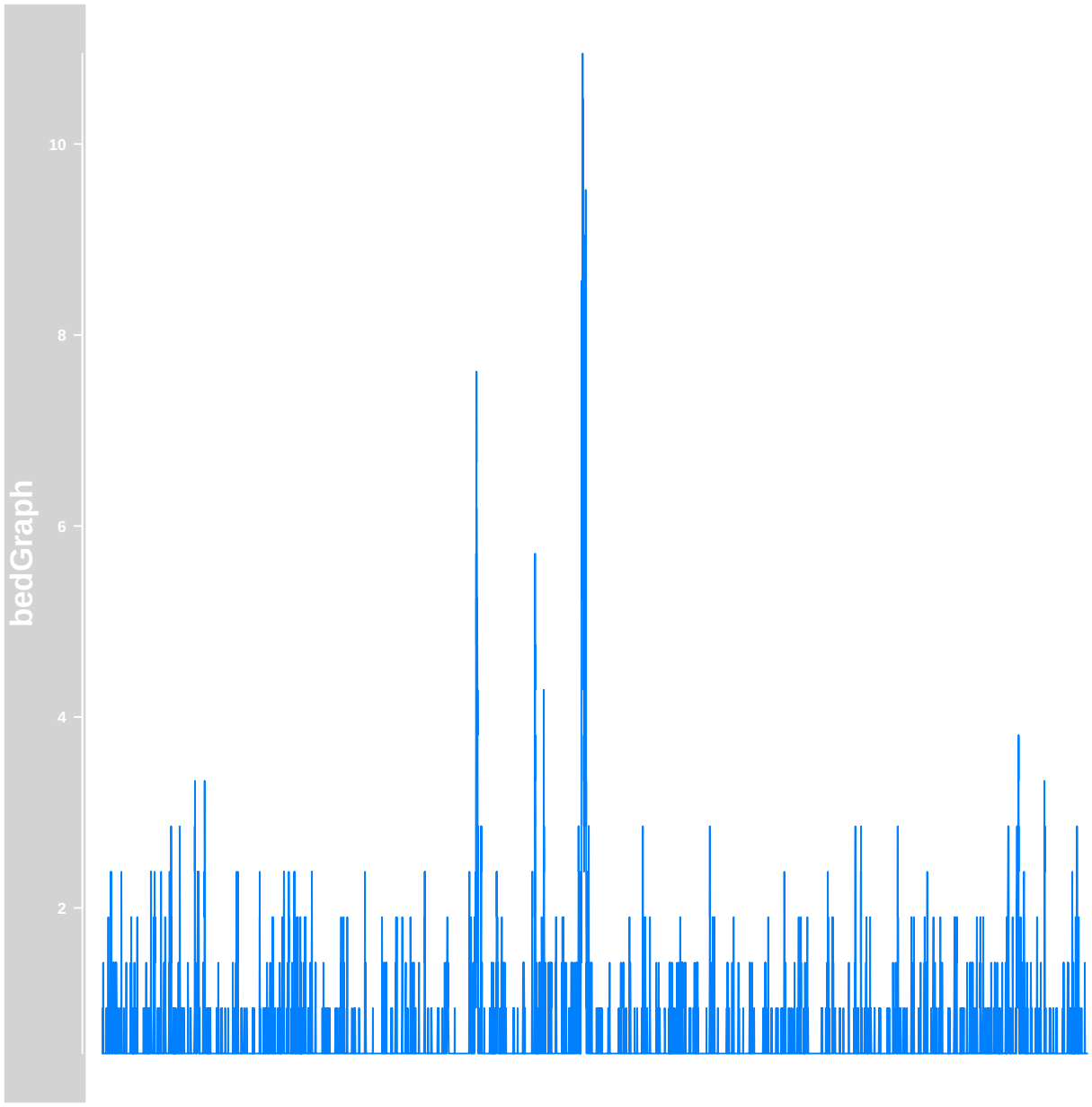
Data from paper Liu et al, 2017, "Identification of breast cancer associated variants that modulate transcription factor binding" http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006761

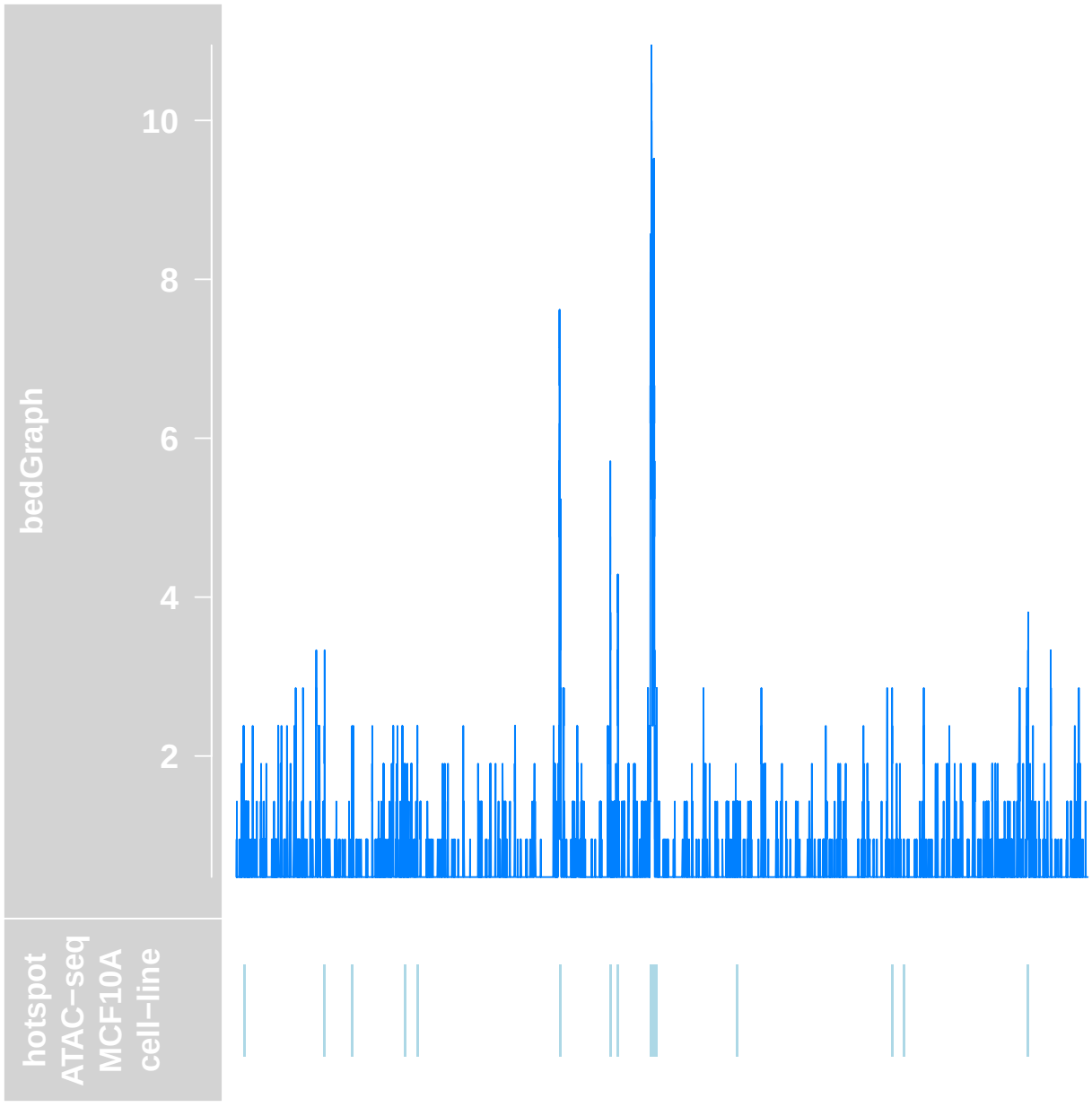
Download data

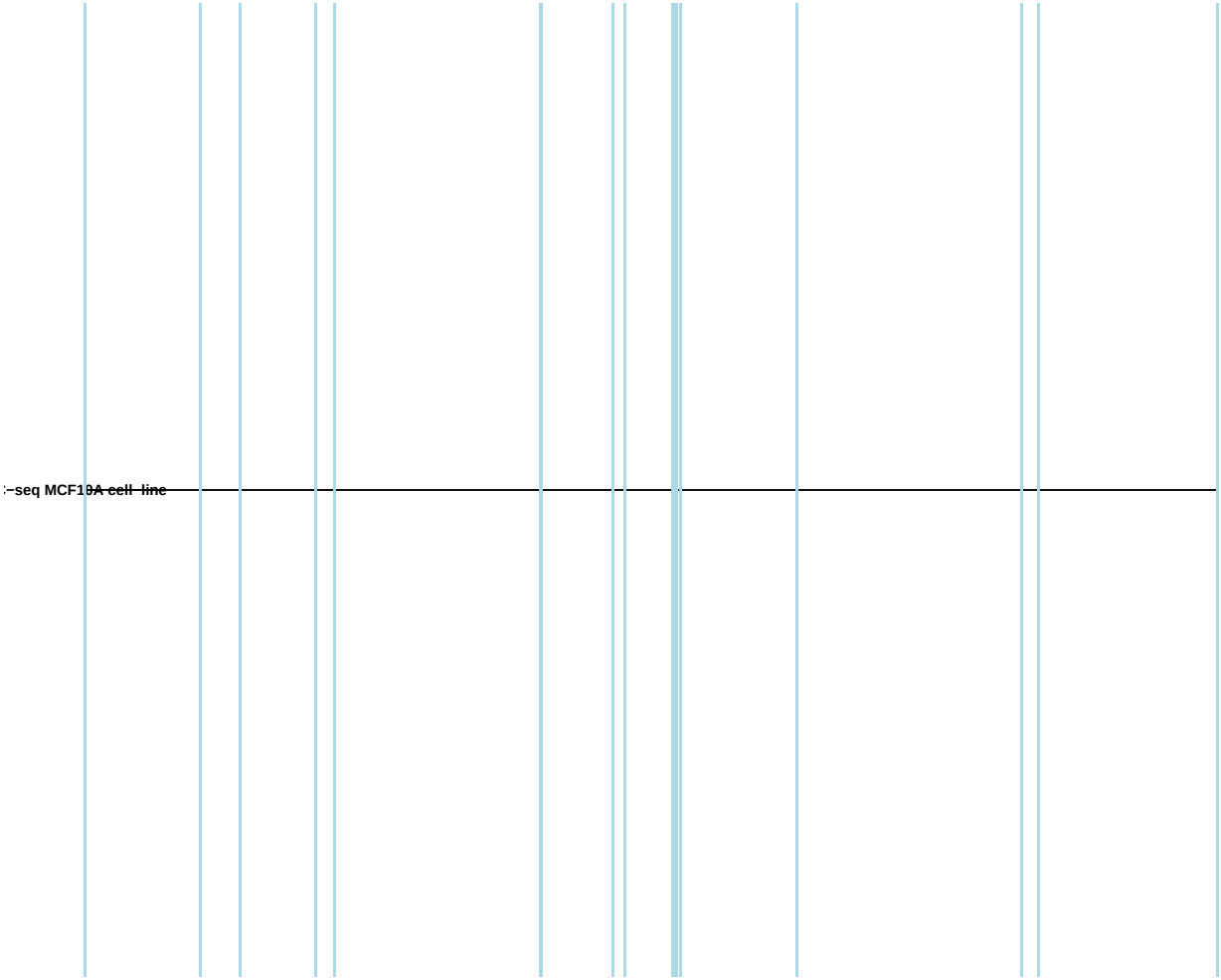
```
## Download data from GEO
wget ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE89nnn/GSE89013/suppl/GSE89013_MCF10a_merged.b

wget ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE89nnn/GSE89013/suppl/GSE89013_MCF10a_merged.b


wget ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE89nnn/GSE89013/suppl/GSE89013_MCF10a_merged.r

## extract only the regions around PTEN
gunzip GSE89013_MCF10a_merged.bedGraph | awk -F"\t" '$1 == "chr10" && $2 > 87763113 && $3 <
{print $0}' CEMT_7_18_segments.bed > ~/GSE89013_MCF10a_merged_PTEN_hg38.bedGraph

gunzip GSE89013_MCF10a_mergedrmdup.fdr0.01.hot.bed | awk -F"\t" '$1 == "chr10" && $2 > 8776
&& $3 < 88071930 {print $0}' CEMT_7_18_segments.bed > ~/GSE89013_MCF10a_mergedrmdup.fdr0.01
```

```
atacseq_MCF10A_path<-paste0(dirfolder,"GSE89013_MCF10a_merged_PTEN_hg38.bedGraph")
atacseq_MCF10A_tab <- read.table(file=atacseq_MCF10A_path,header=FALSE,sep="\t")

dTrack2 <- DataTrack(range = atacseq_MCF10A_path, genome = "hg38", type = "l",
                     chromosome = "chr10", name = "bedGraph")

plotTracks(dTrack2)


#hotspot
atacseq_MCF10A_hotstop_path<-paste0(dirfolder,"GSE89013_MCF10a_mergedrmdup.fdr0.01.hot_PTEN_hg38.bed")
atacseq_MCF10A_hotstop_tab <- read.table(file=atacseq_MCF10A_hotstop_path,header=FALSE,sep="\t")

df_atacseq_PTEN <-data.frame(chr="chr10",start=atacseq_MCF10A_hotstop_tab$V2,
                     end=atacseq_MCF10A_hotstop_tab$V3,
                     strand="*",score=0,
                     feature="hotspot",
                     id=1:nrow(atacseq_MCF10A_hotstop_tab),
                     group=1:nrow(atacseq_MCF10A_hotstop_tab))
```

```
## liftover hg38 to hg19
chain <- import.chain(paste0(dirfolder,"hg38ToHg19.over.chain"))
gr_hotspot <- makeGRangesFromDataFrame(df_atacseq_PTEN, TRUE)
humcon_hotspot <- liftOver(gr_hotspot, chain)

genome(mySession) <- gen_hg19
ATACseqHotSpotTrack <- AnnotationTrack(genome="hg19",range=humcon_hotspot,
                                       chromosome=chrom_ucsc,
                                    name = "hotspot ATAC-seq MCF10A cell-line",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(ATACseqHotSpotTrack) <- list(rotation.title = 360, cex.title=0.5,
                                         shape="box",stackHeight=0.8,
                                         background.title="transparent",
                                         col.title="black", col=NULL)


ATACseqHotSpotTrack_hg38 <- AnnotationTrack(genome="hg38",range=gr_hotspot,
                                            chromosome=chrom_ucsc,
                                    name = "hotspot ATAC-seq MCF10A cell-line",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)


atacseq_bigwig_File <- paste0(dirfolder,"GSE89013_MCF10a_merged.bigWig")
dTrack4 <- DataTrack(range = atacseq_bigwig_File, genome = "hg38",type = "l", name = "Coverage",
                  window = -1, chromosome = chrom_ucsc,importFunction = function(file) import(con = file

plotTracks(list(dTrack2,ATACseqHotSpotTrack_hg38),start=start_hg38,stop=end_hg38)
plotTracks(list(ATACseqHotSpotTrack))
```


bedGraph
10
8
6
4
2


bedGraph
10
8
6
4
2
hotspot
ATAC−seq
MCF10A
cell−line



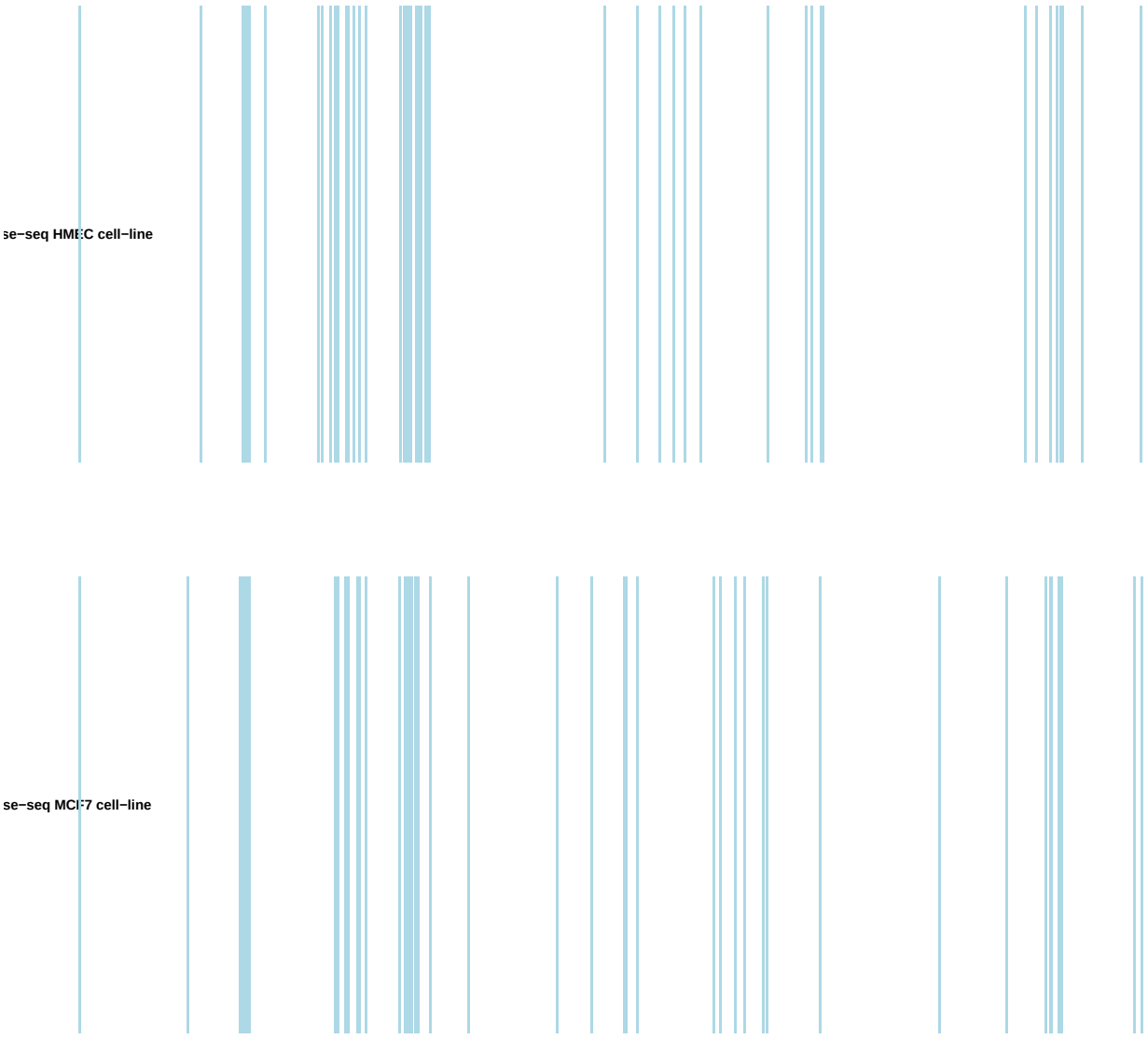
DNase-seq from ENCODE

DNase-seq identifies regulatory regions genome-wide based on their relative sensitivity to cleavage by the DNase I enzyme (Boyle et al., 2008). Each DNase-seq data used in this study was obtained from UCSC ENCODE consortium (`https://genome.ucsc.edu/ENCODE/`). DNase-seq data collected from breast cancer-relevant cells or tissue including MCF7 cells and cultured human mammary epithelial cells (HMEC). For MCF7 cells, there are data for two conditions. The treated MCF7 cells were incubated with 100 nM estradiol (in EtOH) for 1 hour; the control cells were incubated for 1 hour with EtOH. Two replicates were conducted for each condition. path: `http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/`

```
### MCF7
#boradPeak files
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncode
```

```
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncode
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncode
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncode

gunzip wgEncodeUwDnaseMcf7Est100nm1hHotspotsRep1.broadPeak.gz
gunzip wgEncodeUwDnaseMcf7Est100nm1hHotspotsRep2.broadPeak.gz
gunzip wgEncodeUwDnaseMcf7Estctrl0hHotspotsRep1.broadPeak.gz
gunzip wgEncodeUwDnaseMcf7Estctrl0hHotspotsRep2.broadPeak.gz


#bigWig files
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncode
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncodeU
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncodeU


###Preprcessing
##broadPeak
cat wgEncodeUwDnaseMcf7Est100nm1hHotspotsRep1.broadPeak wgEncodeUwDnaseMcf7Est100nm1hHotsp

#cat wgEncodeUwDnaseMcf7Estctrl0hHotspotsRep1.broadPeak
wgEncodeUwDnaseMcf7Estctrl0hHotspotsRep2.broadPeak > E2_0hr_Hotspots.broadPeak

module load bedtools/2.27.0
# sort and merge treatment group
sortBed -i E2_100nM_1hr_Hotspots.broadPeak > E2_100nM_1hr_Hotspots.sort.broadPeak
mergeBed -i E2_100nM_1hr_Hotspots.sort.broadPeak > E2_100nM_1hr_Hotspots.merge.broadPeak


#sort and merge control group
sortBed -i E2_0hr_Hotspots.broadPeak > E2_0hr_Hotspots.sort.broadPeak
mergeBed -i E2_0hr_Hotspots.sort.broadPeak > E2_0hr_Hotspots.merge.broadPeak

#combine different conditions
cat E2_0hr_Hotspots.merge.broadPeak E2_100nM_1hr_Hotspots.merge.broadPeak | sort -k1,1 -k2,

#extract only peak around PTEN

awk -F"t" '$1 == "chr10" && $2 > 89523195 && $3 < 89828532 {print $0}' E2_0hr_Hotspots.merg

awk -F"t" '$1 == "chr10" && $2 > 89523195 && $3 < 89828532 {print $0}' Hotspots_Either_E2_c

## HMEC
# download, unzip and sort peak
curl -OL http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/wgEncodeU

gunzip wgEncodeUwDnaseHmecHotspotsRep2.broadPeak.gz

sortBed -i wgEncodeUwDnaseHmecHotspotsRep2.broadPeak > wgEncodeUwDnaseHmecHotspotsRep2.sort


## Extract only arount PTEN
awk -F"t" '$1 == "chr10" && $2 > 89523195 && $3 < 89828532 {print $0}' wgEncodeUwDnaseHmec
```

```
#HMEC
DNAseq_HMEC_hotstop_path<-paste0(dirfolder,"wgEncodeUwDnaseHmecHotspotsRep2.sort_PTEN.broadPeak")
DNAseq_HMEC_hotstop_tab <- read.table(file=DNAseq_HMEC_hotstop_path,header=FALSE,sep="\t")

df_DNAseq_HMEC_PTEN <-data.frame(chr="chr10",start=DNAseq_HMEC_hotstop_tab$V2,
                      end=DNAseq_HMEC_hotstop_tab$V3,
                      strand="*",score=0,
                      feature="hotspot",
                      id=1:nrow(DNAseq_HMEC_hotstop_tab),group=1:nrow(DNAseq_HMEC_hotstop_tab))

gr_hotspot_HMEC <- makeGRangesFromDataFrame(df_DNAseq_HMEC_PTEN, TRUE)

genome(mySession) <- gen_hg19
DNaseqHotSpotTrack_HMEC <- AnnotationTrack(genome="hg19",range=gr_hotspot_HMEC,
                                           chromosome=chrom_ucsc,
                                    name = "hotspot DNase-seq HMEC cell-line",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(DNaseqHotSpotTrack_HMEC) <- list(rotation.title = 360, cex.title=0.5,
                                             shape="box",stackHeight=0.8,
                                             background.title="transparent",
                                             col.title="black", col=NULL)

#MCF7
DNAseq_MCF7_hotstop_path<-paste0(dirfolder,"E2_0hr_Hotspots.merge_PTEN.broadPeak")
DNAseq_MCF7_hotstop_tab <- read.table(file=DNAseq_MCF7_hotstop_path,header=FALSE,sep="\t")

df_DNAseq_MCF7_PTEN <-data.frame(chr="chr10",start=DNAseq_MCF7_hotstop_tab$V2,
                      end=DNAseq_MCF7_hotstop_tab$V3,
                      strand="*",score=0,
                      feature="hotspot",
                      id=1:nrow(DNAseq_MCF7_hotstop_tab),group=1:nrow(DNAseq_MCF7_hotstop_tab))

gr_hotspot_MCF7 <- makeGRangesFromDataFrame(df_DNAseq_MCF7_PTEN, TRUE)

genome(mySession) <- gen_hg19
DNaseqHotSpotTrack_MCF7 <- AnnotationTrack(genome="hg19",range=gr_hotspot_MCF7,
                                           chromosome=chrom_ucsc,
                                    name = "hotspot DNase-seq MCF7 cell-line",
                                    stacking="dense",
                                    col.line = "black", col = NULL, collapse= FALSE)
displayPars(DNaseqHotSpotTrack_MCF7) <- list(rotation.title = 360, cex.title=0.5,
                                             stacking="dense", shape="box",
                                             stackHeight=0.8, background.title="transparent",
                                             col.title="black", col=NULL)

plotTracks(list(DNaseqHotSpotTrack_HMEC,DNaseqHotSpotTrack_MCF7))
```



Creation annotation tracks around PTEN

PTEN gene :

- hg19 chr10:89,623,195-89,728,532
- hg38 chr10:87,863,113-87,971,930

Color for different features for chromHMM from ENCODE/BROAD

coMET Colour Scheme		
Omic Feature	**Colour**	**Hex Code**
1_Active_Promoter		#E31A1C
2_Weak_Promoter		#FB9A99
3_Poised_Promoter		#6A3D9A
4_Strong_Enhancer		#FF7F00
5_Strong_Enhancer		#CAB2D6
6_Weak_Enhancer		#FFFF99
7_Weak_Enhancer		#FDBF6F
8_Insulator		#1F78B4
9_Txn_Transition		#B2DF8A
10_Txn_Elongation		#33A02C
11_Weak_Txn		#00E1EF
12_Repressed		#FF00FF
13_Heterochrom/lo		#806000
14_Repetitive/CNV		#808080
15_Repetitive/CNV		#BFBFBF

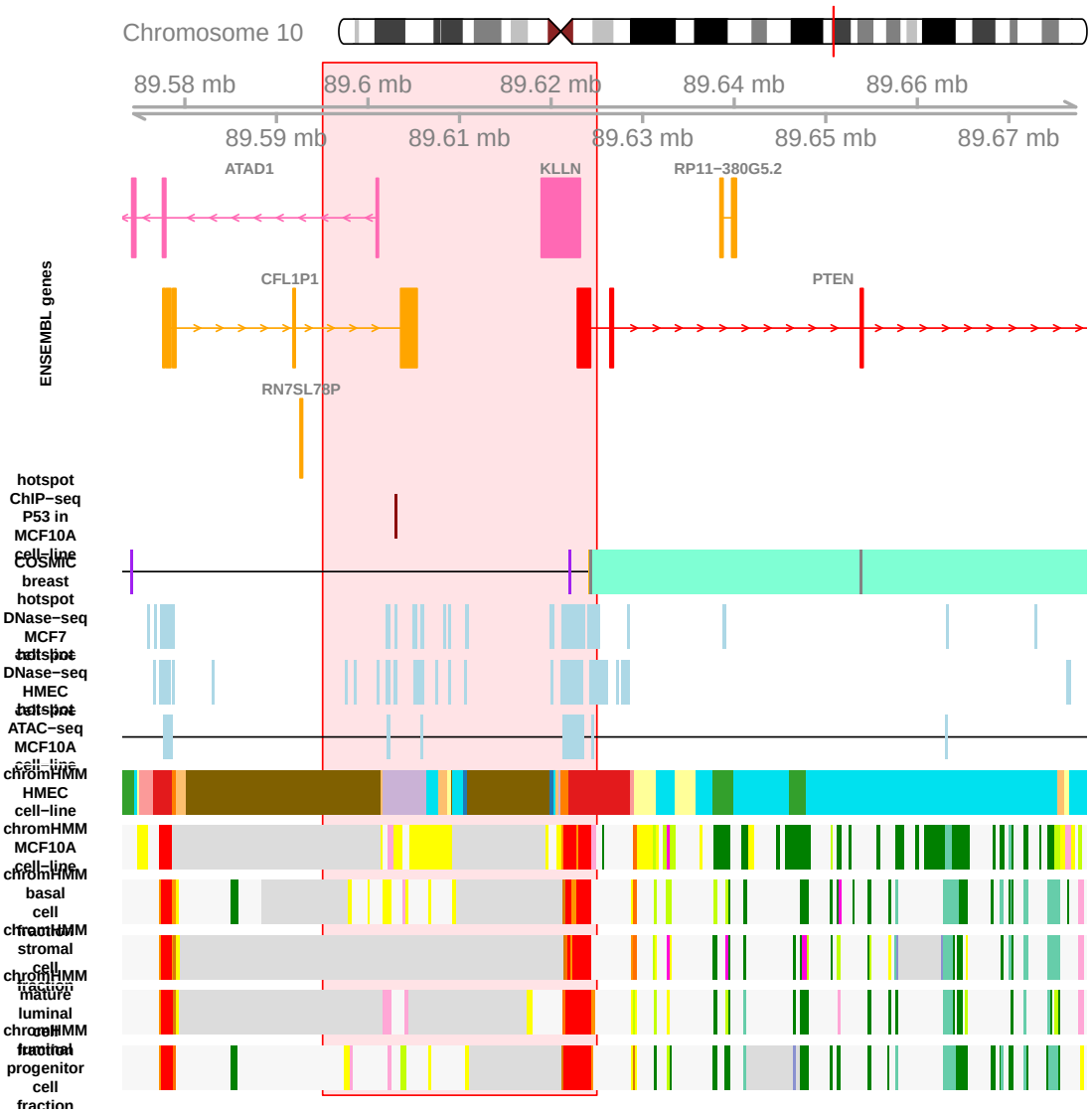
```
track.name="Broad ChromHMM"
tablestrack<-tableNames(ucscTableQuery(mySession, track=track.name))
table.name<-tablestrack[1]
chromhmmtrackone<-chromatinHMMOne_UCSC(gen_hg19,chrom,start_hg19,
                                       end_hg19,mySession,color="coMET",table.name)

dnasetrack<-DNAse_UCSC(gen_hg19,chrom,start_hg19,end_hg19,mySession)
#regulationENSEMBLtrack<-regulationBiomart_ENSEMBL(gen_hg19,chrom,start_hg19,end_hg19)
```

```
genome(mySession) <- gen_hg19

ENSMBLtrack<-genes_ENSEMBL(gen_hg19,chrom,start_hg19,end_hg19,showId=TRUE)

#PTEN=ENSG00000171862
geneTrackShow_PTEN <- ENSMBLtrack[gene(ENSMBLtrack) %in% c("ENSG00000171862")]
start_exonPTEN <- start(geneTrackShow_PTEN@range@ranges)
end_exonPTEN <- end(geneTrackShow_PTEN@range@ranges)

# ATAD1=ENSG00000138138
geneTrackShow_ATAD1 <- ENSMBLtrack[gene(ENSMBLtrack) %in% c("ENSG00000138138")]
start_exonATAD1 <- start(geneTrackShow_ATAD1@range@ranges)
end_exonATAD1 <- end(geneTrackShow_ATAD1@range@ranges)

# KLLN=ENSG00000227268
geneTrackShow_KLLN <- ENSMBLtrack[gene(ENSMBLtrack) %in% c("ENSG00000227268")]
start_exonKLLN <- start(geneTrackShow_KLLN@range@ranges)
end_exonKLLN <- end(geneTrackShow_KLLN@range@ranges)
```

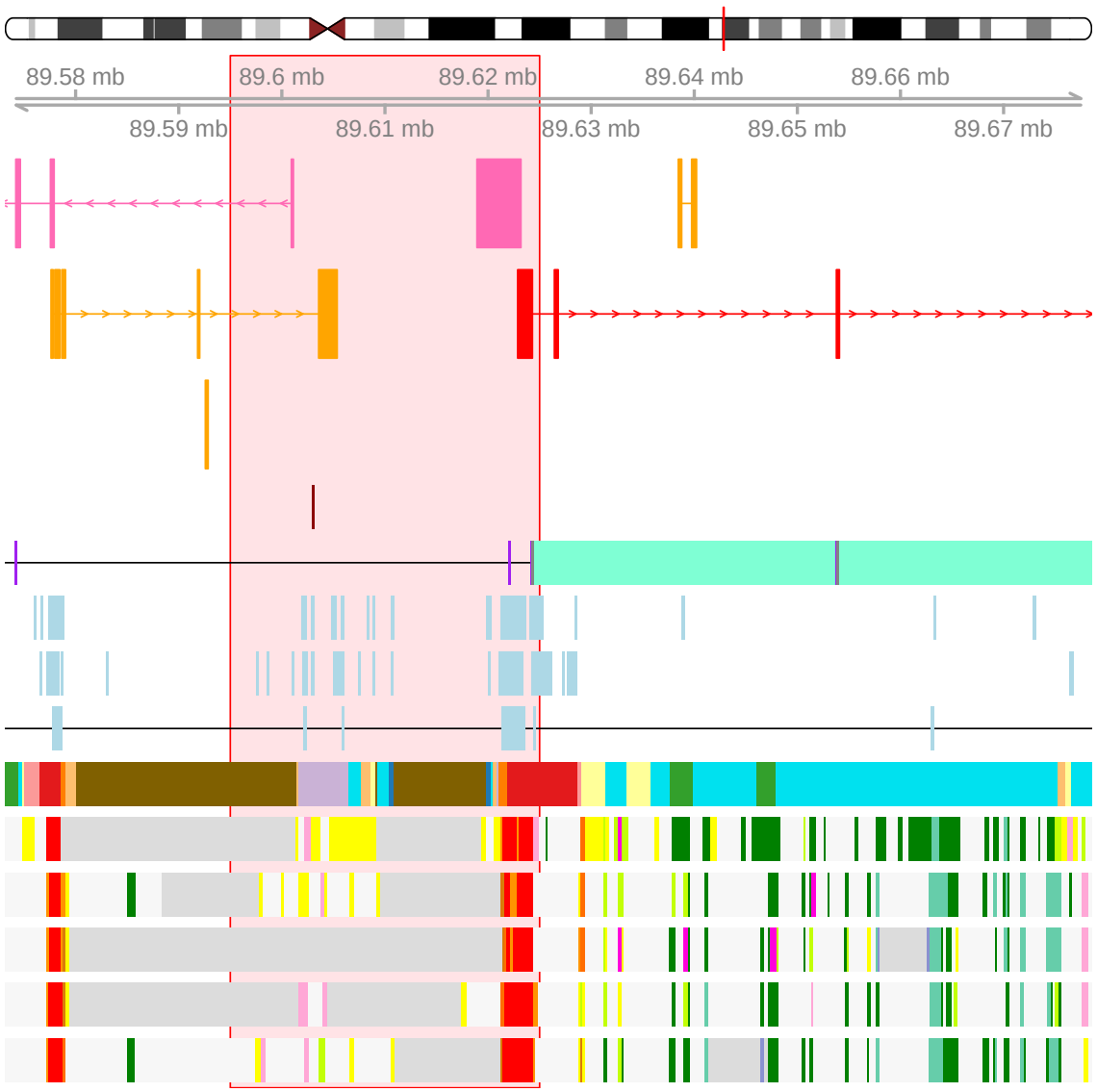
```
interestfeatures <- rbind(cbind(start_exonPTEN,end_exonPTEN,"PTEN"),
                          cbind(start_exonATAD1,end_exonATAD1,"ATAD1"),
                          cbind(start_exonKLLN,end_exonKLLN,"KLLN"))

interestcolor <- list("PTEN"="red", "ATAD1"="hotpink", "KLLN"="hotpink")

PTENhighlighted_track <- interestGenes_ENSEMBL(gen_hg19,chrom,
                                               start_hg19,end_hg19,
                                               interestfeatures,interestcolor,
                                               showId=TRUE)
#rotation.title = 360,
displayPars(PTENhighlighted_track) <- list(cex.title=0.5,
                                           stackHeight=0.8, background.title="transparent",
                                           col.title="black", col=NULL, name = "ENSEMBL genes")

cpgIstrack<-cpgIslands_UCSC(gen_hg19, chrom, start_hg19, end_hg19)

displayPars(cpgIstrack) <- list(rotation.title = 360, cex.title=0.5, stacking="dense",
                                shape="box",stackHeight=0.8, background.title="transparent",
                                col.title="black", col=NULL)



# chromHMM in HMEC "wgEncodeBroadHmmHmecHMM"
track.name="Broad ChromHMM"
tablestrack<-tableNames(ucscTableQuery(mySession, track=track.name))
table.name<-"wgEncodeBroadHmmHmecHMM" # one from tablestrack
chromhmmtrack_EncodeBroadHmmHmecHMM<-chromatinHMMOne_UCSC(gen_hg19,chrom,start_hg19,end_hg19,mySession,c
displayPars(chromhmmtrack_EncodeBroadHmmHmecHMM) <- list(rotation.title = 360,
                                                         cex.title=0.5,
                                           stacking="dense", shape="box",
                                           stackHeight=0.8,
                                           background.title="transparent",
                                col.title="black", col=NULL)


IdType <- "name"
#genesUcsctrack<-refGenes_UCSC(gen_hg19,chrom,start_hg19,end_hg19,IdType)

##general position
itrack <- IdeogramTrack(genome = gen_hg19, chromosome = chrom)
gtrack <- GenomeAxisTrack()

#highlight the region of interest
listgviz_to_highlight <- list(gtrack, PTENhighlighted_track,
                              ChIPseqP53HotSpotTrack,
                              cosmicTrack,
                  DNaseqHotSpotTrack_MCF7,
                  DNaseqHotSpotTrack_HMEC,ATACseqHotSpotTrack,
                  chromhmmtrack_EncodeBroadHmmHmecHMM,
                  chromHMMTrack_07,chromHMMTrack_35,chromHMMTrack_36,
                  chromHMMTrack_37,chromHMMTrack_38)
```

```
## region to highlight 89423195
ht <- HighlightTrack(trackList = listgviz_to_highlight, start = 89595000, width = 30000,
                     chromosome = chrom_ucsc)

#combine different annotation tracks
listgviz <- list(itrack,ht)
```

```
#150000 or 80000
zoom=150000
start_hg19_vis <- start_hg19 + zoom
end_hg19_vis <- end_hg19 - zoom

plotTracks(listgviz,from=start_hg19_vis, to=end_hg19_vis,fontsize=14)
```




```
plotTracks(listgviz,from=start_hg19_vis, to=end_hg19_vis,fontsize=14,showId = FALSE,panel.only=TRUE)
```



Session Info

```
sessionInfo()

## R version 3.4.3 (2017-11-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.3
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
```

```
##  [1] parallel  stats4    grid      stats     graphics  grDevices utils     datasets
##  [9] methods   base
##
## other attached packages:
##  [1] gwascat_2.10.0                          Homo.sapiens_1.3.1
##  [3] TxDb.Hsapiens.UCSC.hg19.knownGene_3.2.2 org.Hs.eg.db_3.5.0
##  [5] GO.db_3.5.0                             OrganismDbi_1.20.0
##  [7] GenomicFeatures_1.30.3                  AnnotationDbi_1.40.0
##  [9] Biobase_2.38.0                          rtracklayer_1.38.3
## [11] data.table_1.10.4-3                     coMET_1.11.3
## [13] trackViewer_1.14.1                      ggbio_1.26.1
## [15] ggplot2_2.2.1                           psych_1.8.3.3
## [17] Gviz_1.22.3                             GenomicRanges_1.30.3
## [19] GenomeInfoDb_1.14.0                     IRanges_2.12.0
## [21] S4Vectors_0.16.0                        BiocGenerics_0.24.0
## [23] biomaRt_2.34.2                          knitr_1.20
##
## loaded via a namespace (and not attached):
##   [1] backports_1.1.2               fastmatch_1.1-0
##   [3] Hmisc_4.1-1                   AnnotationHub_2.10.1
##   [5] gQTLstats_1.10.1              corrplot_0.84
##   [7] plyr_1.8.4                    lazyeval_0.2.1
##   [9] splines_3.4.3                 BatchJobs_1.7
##  [11] BiocParallel_1.12.0           digest_0.6.15
##  [13] foreach_1.4.4                 BiocInstaller_1.28.0
##  [15] ensembldb_2.2.2               htmltools_0.3.6
##  [17] magrittr_1.5                  checkmate_1.8.5
##  [19] memoise_1.1.0                 BBmisc_1.11
##  [21] BSgenome_1.46.0               cluster_2.0.7-1
##  [23] doParallel_1.0.11             limma_3.34.9
##  [25] Biostrings_2.46.0             matrixStats_0.53.1
##  [27] prettyunits_1.0.2             colorspace_1.3-2
##  [29] blob_1.1.1                    dplyr_0.7.4
##  [31] RCurl_1.95-4.10               jsonlite_1.5
##  [33] graph_1.56.0                  ffbase_0.12.3
##  [35] bindr_0.1.1                   sendmailR_1.2-1
##  [37] brew_1.0-6                    survival_2.41-3
##  [39] VariantAnnotation_1.24.5      iterators_1.0.9
##  [41] glue_1.2.0                    hash_2.2.6
##  [43] gtable_0.2.0                  zlibbioc_1.24.0
##  [45] XVector_0.18.0                DelayedArray_0.4.1
##  [47] scales_0.5.0                  DBI_0.8
##  [49] GGally_1.3.2                  Rcpp_0.12.16
##  [51] viridisLite_0.3.0             xtable_1.8-2
##  [53] progress_1.1.2                htmlTable_1.11.2
##  [55] foreign_0.8-69                bit_1.1-12
##  [57] Formula_1.2-2                 erma_0.10.1
##  [59] htmlwidgets_1.0               httr_1.3.1
##  [61] RColorBrewer_1.1-2            acepack_1.4.1
##  [63] ff_2.2-13                     pkgconfig_2.0.1
##  [65] reshape_0.8.7                 XML_3.98-1.10
##  [67] nnet_7.3-12                   rlang_0.2.0
##  [69] reshape2_1.4.3                munsell_0.4.3
```

```
##  [71] tools_3.4.3                    RSQLite_2.1.0
##  [73] evaluate_0.10.1                stringr_1.3.0
##  [75] yaml_2.1.18                    bit64_0.9-7
##  [77] purrr_0.2.4                    AnnotationFilter_1.2.0
##  [79] bindrcpp_0.2.2                 pbapply_1.3-4
##  [81] RBGL_1.54.0                    nlme_3.1-137
##  [83] mime_0.5                       compiler_3.4.3
##  [85] rstudioapi_0.7                 beeswarm_0.2.3
##  [87] plotly_4.7.1                   curl_3.2
##  [89] interactiveDisplayBase_1.16.0 tibble_1.4.2
##  [91] stringi_1.1.7                  highr_0.6
##  [93] GenomicFiles_1.14.0            lattice_0.20-35
##  [95] ProtGenerics_1.10.0            Matrix_1.2-14
##  [97] pillar_1.2.1                   snpStats_1.28.0
##  [99] bitops_1.0-6                   grImport_0.9-0
## [101] httpuv_1.3.6.2                 R6_2.2.2
## [103] latticeExtra_0.6-28            RMySQL_0.10.14
## [105] gridExtra_2.3                  vipor_0.4.5
## [107] codetools_0.2-15               dichromat_2.0-0
## [109] assertthat_0.2.0               SummarizedExperiment_1.8.1
## [111] GenomicAlignments_1.14.2       Rsamtools_1.30.0
## [113] mnormt_1.5-5                   GenomeInfoDbData_1.0.0
## [115] mgcv_1.8-23                    gQTLBase_1.10.0
## [117] colortools_0.1.5               rpart_4.1-13
## [119] tidyr_0.8.0                    biovizBase_1.26.0
## [121] shiny_1.0.5                    base64enc_0.1-3
## [123] ggbeeswarm_0.6.0
```