Figure S4

Zheng Zuo

04/16/2022

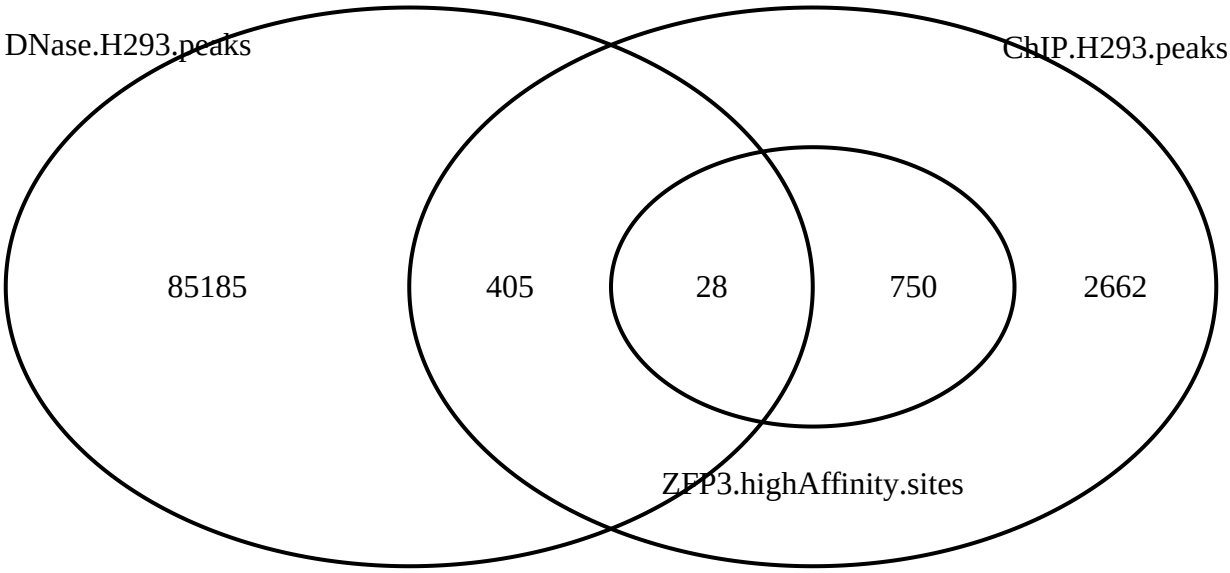
Contents


```
require(dplyr)
require(GenomicRanges)
require(TFCookbook)

load("ZFP3.motif.RData")
```

HEK293 cells data

```
ChIP.H293.peaks <- read.table("ENCFF107KSN.bed") %>%
                   GenomicRanges::makeGRangesFromDataFrame(seqnames.field = "V1",
                                                           start.field    = "V2",
                                                           end.field      = "V3") %>% unique()


ZFP3.sites <- TFCookbook::matchPEM(PEM = ZFP3.Core.PEM,
                                   subject = ChIP.H293.peaks,
                                   genome  = "hg38",
                                   out     = "positions",
                                   E.cutoff= -5)

ZFP3.sites$predicted.Core.Energy     <- ZFP3.sites$predicted.Energy
ZFP3.sites$predicted.Energy          <- NULL
ZFP3.sites$predicted.Upstream.Energy <- TFCookbook::predictEnergy(ZFP3.sites$Sequence, ZFP3.Upstream.PE
ZFP3.sites$predicted.Full.Energy     <- ZFP3.sites$predicted.Upstream.Energy + ZFP3.sites$predicted.Cor

ZFP3.highAffinity.sites <- subset(ZFP3.sites, predicted.Upstream.Energy < (-4))

DNase.H293.peaks <- read.table("DNase-seq signals/ENCFF285OXK.H293T.bed") %>%
                   GenomicRanges::makeGRangesFromDataFrame(seqnames.field = "V1",
                                                           start.field    = "V2",
                                                           end.field      = "V3") %>% unique()

ChIPpeakAnno::findOverlapsOfPeaks(ZFP3.highAffinity.sites,
```

```
                                ChIP.H293.peaks,
                                DNase.H293.peaks,
                                minoverlap = 0.85) %>%
ChIPpeakAnno::makeVennDiagram()
```

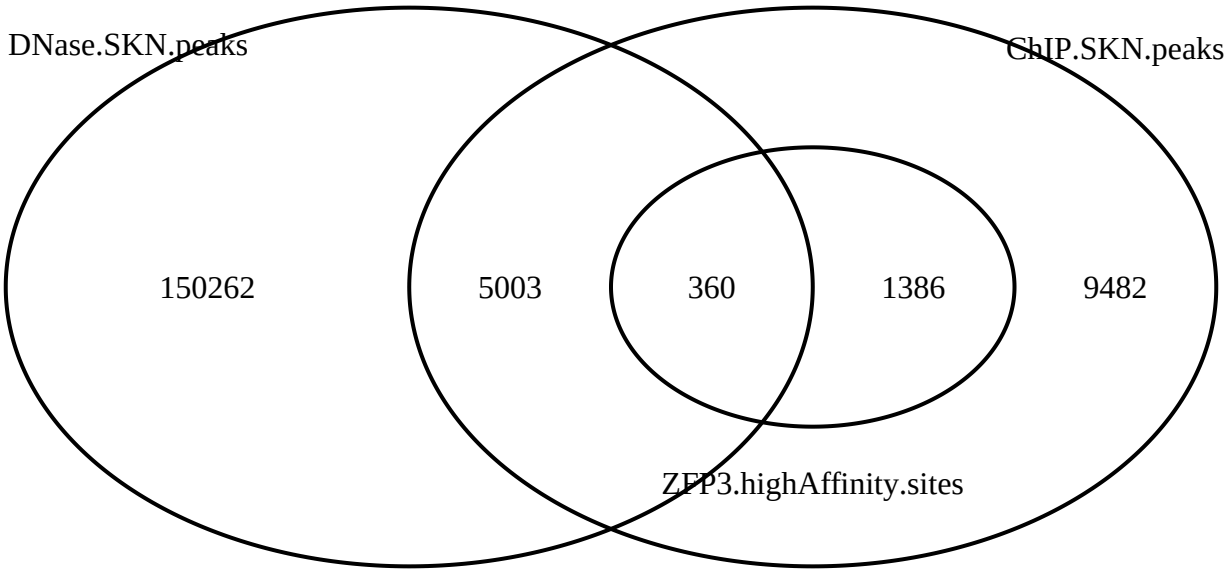



```
## $p.value
##      ZFP3.highAffinity.sites ChIP.H293.peaks DNase.H293.peaks pval
## [1,]                       0               1                1    1
## [2,]                       1               0                1    1
## [3,]                       1               1                0    0
##
## $vennCounts
##      ZFP3.highAffinity.sites ChIP.H293.peaks DNase.H293.peaks Counts
## [1,]                       0               0                0      0
## [2,]                       0               0                1  85185
## [3,]                       0               1                0   2662
## [4,]                       0               1                1    405
## [5,]                       1               0                0      0
## [6,]                       1               0                1      0
## [7,]                       1               1                0    750
## [8,]                       1               1                1     28
##      count.ZFP3.highAffinity.sites count.ChIP.H293.peaks count.DNase.H293.peaks
## [1,]                             0                    0                     0
## [2,]                             0                    0                 85185
## [3,]                             0                 2662                     0
```

```
## [4,]                             0                   406                   409
## [5,]                             0                     0                     0
## [6,]                             0                     0                     0
## [7,]                           814                   757                     0
## [8,]                            28                    29                    30
## attr(,"class")
## [1] "VennCounts"
```

SK-N-SH cells data

```
ChIP.SKN.peaks <- read.table("ENCFF049VST.bed") %>%
                  GenomicRanges::makeGRangesFromDataFrame(seqnames.field = "V1",
                                                          start.field    = "V2",
                                                          end.field      = "V3") %>% unique()


ZFP3.sites <- TFCookbook::matchPEM(PEM = ZFP3.Core.PEM,
                                   subject = ChIP.SKN.peaks,
                                   genome  = "hg38",
                                   out     = "positions",
                                   E.cutoff= -5)

ZFP3.sites$predicted.Core.Energy     <- ZFP3.sites$predicted.Energy
ZFP3.sites$predicted.Energy          <- NULL
ZFP3.sites$predicted.Upstream.Energy <- TFCookbook::predictEnergy(ZFP3.sites$Sequence, ZFP3.Upstream.PE
ZFP3.sites$predicted.Full.Energy     <- ZFP3.sites$predicted.Upstream.Energy + ZFP3.sites$predicted.Cor

ZFP3.highAffinity.sites <- subset(ZFP3.sites, predicted.Upstream.Energy < (-4))

DNase.SKN.peaks <- read.table("DNase-seq signals/ENCFF752OZB.SKN.bed") %>%
                   GenomicRanges::makeGRangesFromDataFrame(seqnames.field = "V1",
                                                           start.field    = "V2",
                                                           end.field      = "V3") %>% unique()

ChIPpeakAnno::findOverlapsOfPeaks(ZFP3.highAffinity.sites,
                                  ChIP.SKN.peaks,
                                  DNase.SKN.peaks,
                                  minoverlap = 0.85) %>%
ChIPpeakAnno::makeVennDiagram()
```




```
## $p.value
##      ZFP3.highAffinity.sites ChIP.SKN.peaks DNase.SKN.peaks pval
## [1,]                       0              1               1    1
## [2,]                       1              0               1    1
## [3,]                       1              1               0    0
##
## $vennCounts
##      ZFP3.highAffinity.sites ChIP.SKN.peaks DNase.SKN.peaks Counts
## [1,]                       0              0               0      0
## [2,]                       0              0               1 150262
## [3,]                       0              1               0   9482
## [4,]                       0              1               1   5003
## [5,]                       1              0               0      0
## [6,]                       1              0               1      0
## [7,]                       1              1               0   1386
## [8,]                       1              1               1    360
##      count.ZFP3.highAffinity.sites count.ChIP.SKN.peaks count.DNase.SKN.peaks
## [1,]                             0                   0                     0
## [2,]                             0                   0                150262
## [3,]                             0                9482                     0
## [4,]                             0                5119                  5088
## [5,]                             0                   0                     0
## [6,]                             0                   0                     0
## [7,]                          1479                1430                     0
## [8,]                           389                 408                   372
## attr(,"class")
```

```
## [1] "VennCounts"
```