

Distance from Region to TAD

Spiro Stilianoudakis

August 14, 2018

Loading libraries

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.4
#library(MultiAssayExperiment)
library(GenomicRanges)

## 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
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:base':
##
##   expand.grid
## Loading required package: IRanges
## Loading required package: GenomeInfoDb
```

```

#library(IRanges)
library(caret)

## Warning: package 'caret' was built under R version 3.4.4
## Loading required package: lattice
library(data.table)

## Warning: package 'data.table' was built under R version 3.4.4
##
## 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(gbm)

## Loading required package: survival
##
## Attaching package: 'survival'
## The following object is masked from 'package:caret':
##
##     cluster
## Loading required package: splines
## Loaded gbm 2.1.3
library(pROC)

## Warning: package 'pROC' was built under R version 3.4.4
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
## The following objects are masked from 'package:IRanges':
##
##     cov, var
## The following objects are masked from 'package:S4Vectors':
##
##     cov, var
## The following object is masked from 'package:BiocGenerics':
##
##     var

```

```
## The following objects are masked from 'package:stats':
##
##      cov, smooth, var
library(plyr)

##
## Attaching package: 'plyr'
## The following object is masked from 'package:IRanges':
##
##      desc
## The following object is masked from 'package:S4Vectors':
##
##      rename
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.4
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:plyr':
##
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
##      summarize
## The following objects are masked from 'package:data.table':
##
##      between, first, last
## The following objects are masked from 'package:GenomicRanges':
##
##      intersect, setdiff, union
## The following object is masked from 'package:GenomeInfoDb':
##
##      intersect
## The following objects are masked from 'package:IRanges':
##
##      collapse, desc, intersect, setdiff, slice, union
## The following objects are masked from 'package:S4Vectors':
##
##      first, intersect, rename, setdiff, setequal, union
## The following objects are masked from 'package:BiocGenerics':
##
##      combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.4.4
```

Distance Summaries for each feature from region to TAD boundary

```
grlist <- readRDS("C:/Users/Spiro Stilianoudakis/Documents/TAD_data/RData/GM12878/grlist.rds")

meandist <- unlist(lapply(grlist, function(x){mean(mcols(x)$distance)}))
mediandist <- unlist(lapply(grlist, function(x){median(mcols(x)$distance)}))
rangedist <- unlist(lapply(grlist, function(x){max(mcols(x)$distance)})) - unlist(lapply(grlist, function(x){min(mcols(x)$distance)}))

meanlogdist <- unlist(lapply(grlist, function(x){mean(mcols(x)$logdistance)}))
medianlogdist <- unlist(lapply(grlist, function(x){median(mcols(x)$logdistance)}))
rangelogdist <- unlist(lapply(grlist, function(x){max(mcols(x)$logdistance, base = 2)})) - unlist(lapply(grlist, function(x){min(mcols(x)$logdistance, base = 2)}))

distancetab <- data.frame(Feature = names(unlist(lapply(grlist, function(x){mean(mcols(x)$distance)}))),
                          Mean = meandist,
                          Median = mediandist,
                          Range = rangedist,
                          MeanLog = meanlogdist,
                          MedianLog = medianlogdist,
                          RangeLog = rangelogdist)

rownames(distancetab) <- NULL

distancetab[,2:7] <- round(distancetab[,2:7], 1)

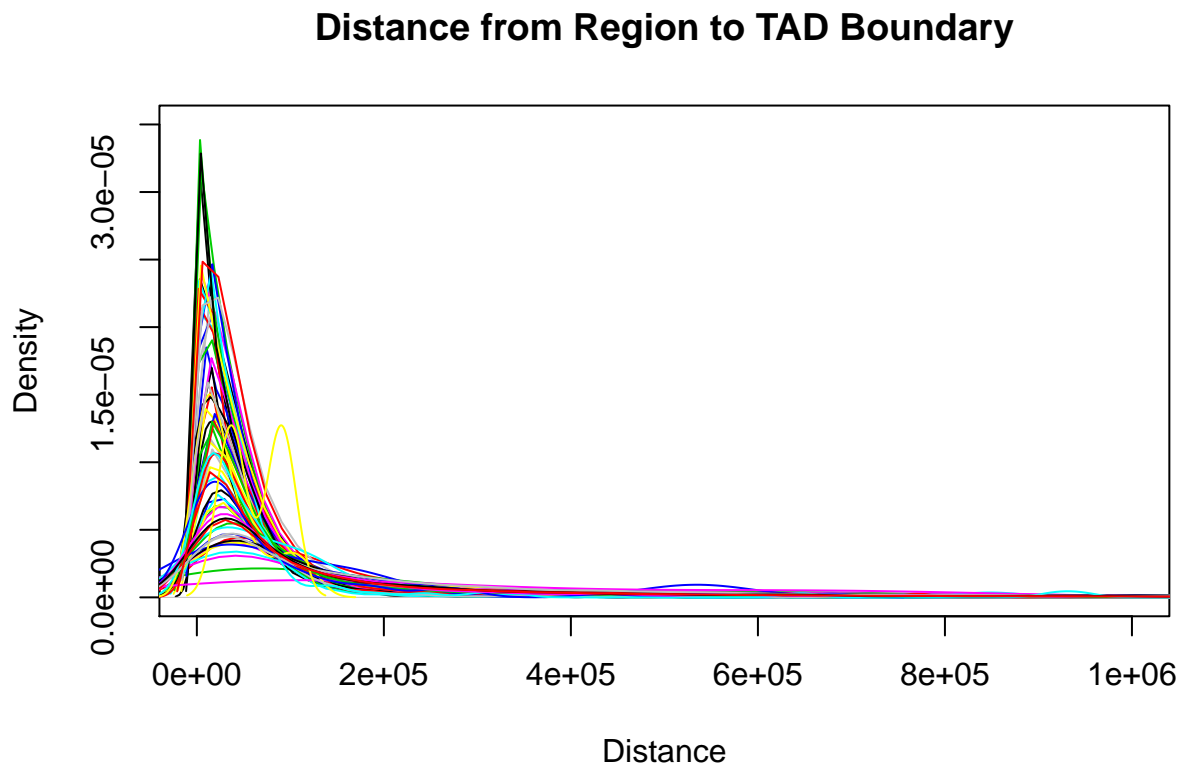
kable(distancetab)
```

Feature	Mean	Median	Range	MeanLog	MedianLog	RangeLog
A_gr_center	53463.6	39998.0	2009998	14.5	15.3	20.9
B_gr_center	243785.8	109998.0	6445000	16.5	16.7	22.6
complex_gr_center	435851.4	159183.5	6474051	17.0	17.3	20.6
deletion_gr_center	241916.2	71123.0	8656650	16.1	16.1	23.0
DNA_gr_center	174422.6	53897.0	8902653	15.7	15.7	21.1
duplication_gr_center	521044.4	111761.0	8909159	16.8	16.8	22.1
gain_loss_gr_center	384379.9	109600.0	8833893	16.7	16.7	21.1
gerp_gr_center	186412.5	58439.0	8909204	15.8	15.8	23.1
Gm12878_ActivePromoter_gr_center	56963.4	19403.0	8439204	13.8	14.2	23.0
Gm12878_CTCF_gr_center	118693.5	32466.0	8799400	14.9	15.0	23.1
Gm12878_DNaseI_gr_center	58602.8	27638.0	7037275	14.3	14.8	22.7
Gm12878_E_gr_center	51768.8	32089.0	5436164	14.6	15.0	22.4
Gm12878_H2az_gr_center	70235.3	32766.0	7324423	14.7	15.0	22.8
Gm12878_H3k27ac_gr_center	67868.4	32233.0	8519585	14.6	15.0	23.0
Gm12878_H3k27me3_gr_center	125774.1	50820.0	6588236	15.5	15.6	20.7
Gm12878_H3k36me3_gr_center	78210.5	30182.0	6472936	14.6	14.9	22.6
Gm12878_H3k4me1_gr_center	62327.9	32536.0	7035557	14.6	15.0	22.7
Gm12878_H3k4me2_gr_center	65637.0	32478.0	7324454	14.6	15.0	22.8
Gm12878_H3k4me3_gr_center	83111.4	32462.0	7324321	14.6	15.0	22.8
Gm12878_H3k79me2_gr_center	61017.5	29198.5	7324298	14.4	14.8	20.8

Feature	Mean	Median	Range	MeanLog	MedianLog	RangeLog
Gm12878_H3k9ac_gr_center	67714.1	28350.0	7324327	14.3	14.8	21.8
Gm12878_H3k9me3_gr_center	133194.8	47104.0	8781218	15.4	15.5	22.1
Gm12878_H4k20me1_gr_center	97795.0	34583.0	6588238	14.9	15.1	22.7
Gm12878_Heterochromlo_gr_center	135751.9	46296.0	8847780	15.4	15.5	23.1
Gm12878_Insulator_gr_center	113386.3	27196.0	8764980	14.7	14.7	23.1
Gm12878_PF_gr_center	53031.3	24034.0	7324843	14.2	14.6	22.8
Gm12878_PoisedPromoter_gr_center	104476.5	34922.0	6995304	14.8	15.1	22.7
Gm12878_R_gr_center	173436.4	53152.0	8862600	15.7	15.7	23.1
Gm12878_RepetitiveCNV14_gr_center	470858.4	112295.0	8782380	16.8	16.8	23.1
Gm12878_RepetitiveCNV15_gr_center	476516.3	123135.0	8630678	16.9	16.9	21.5
Gm12878_Repressed_gr_center	99046.2	43298.0	8503704	15.2	15.4	23.0
Gm12878_StrongEnhancer4_gr_center	47868.5	29828.5	6950404	14.4	14.9	22.7
Gm12878_StrongEnhancer5_gr_center	53246.1	33699.0	4494118	14.7	15.0	22.1
Gm12878_T_gr_center	177812.9	53361.0	8812302	15.7	15.7	23.1
Gm12878_TSS_gr_center	43053.5	22098.0	7324461	13.9	14.4	22.8
Gm12878_TxnElongation_gr_center	51506.3	28929.0	7054504	14.6	14.8	22.8
Gm12878_TxnTransition_gr_center	44890.8	25868.0	6947004	14.3	14.7	22.7
Gm12878_WE_gr_center	71366.6	36470.5	7329362	14.9	15.2	22.8
Gm12878_WeakEnhancer6_gr_center	78580.2	32144.0	8644480	14.6	15.0	23.0
Gm12878_WeakEnhancer7_gr_center	71222.9	36747.0	8643980	14.9	15.2	23.0
Gm12878_WeakPromoter_gr_center	66892.8	22442.0	8456704	14.1	14.5	23.0
Gm12878_WeakTxn_gr_center	72938.0	32950.0	8640680	14.7	15.0	23.0
insertion_gr_center	224811.1	71612.0	7015057	16.1	16.1	20.7
inversion_gr_center	300800.3	93708.5	6472381	16.4	16.5	20.6
line_gr_center	231407.6	71210.5	8911165	16.1	16.1	23.1
low_complexity_gr_center	268021.2	60581.0	5931558	16.0	15.9	20.5
LTR_gr_center	264982.8	76211.0	8866156	16.3	16.2	22.1
mobile_element_insertion_gr_center	268654.4	93892.0	3483807	16.5	16.5	19.7
novel_sequence_insertion_gr_center	185711.6	60457.0	4709202	15.9	15.9	22.2
other_gr_center	98121.3	42182.0	1698284	15.2	15.4	18.7
RC_gr_center	204934.2	70209.5	5381317	16.1	16.1	20.4
RNA_gr_center	137251.7	41000.0	517384	16.0	15.3	17.0
rRNA_gr_center	197697.8	70102.5	2583941	15.8	16.1	19.3
satellite_gr_center	841805.4	308190.0	8358672	18.1	18.2	21.0
scRNA_gr_center	36227.7	22157.5	126075	14.1	14.4	14.9
se_GM12878_gr_center	57051.1	31517.0	6950727	14.5	14.9	20.7
sequence_alteration_gr_center	281898.3	91742.0	3740977	16.5	16.5	19.8
simple_repeat_gr_center	278579.3	77809.5	8783303	16.2	16.2	21.1
SINE_gr_center	133906.0	44229.0	8878105	15.4	15.4	23.1
snRNA_gr_center	93478.4	36665.0	678470	15.1	15.2	17.4
sprRNA_gr_center	185361.1	34018.0	2698941	15.2	15.1	19.4
tandem_duplication_gr_center	190883.8	59338.0	3465544	15.8	15.9	19.7
tRNA_gr_center	63495.0	63495.0	53730	15.8	15.8	14.5
UCNE_gr_center	216210.4	106782.0	2253608	16.4	16.7	19.1
unknown_gr_center	196398.8	66212.0	2965941	16.0	16.0	19.5
VMR_gr_center	209957.4	60414.0	8909829	15.8	15.9	23.1

Plots

```
d <- density(mcols(grlist[[1]])$distance)
plot(d, xlab="Distance",
     main="Distance from Region to TAD Boundary",
     xlim=c(0,1000000),
     ylim=c(0,3.5e-05))
for(i in 1:(length(grlist)-1)){
  lines(density(mcols(grlist[[i+1]])$distance), col=i+1)
}
```



```
d <- density(mcols(grlist[[1]])$logdistance)
plot(d, xlab="Distance",
     main="Log Distance from Region to TAD Boundary")
for(i in 1:(length(grlist)-1)){
  lines(density(mcols(grlist[[i+1]])$logdistance), col=i+1)
}
```

Log Distance from Region to TAD Boundary

