

position_within_UTR

Nikita Verheyden

2023-04-22

dir

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Figure2/Position_in_Gene")
```

packages

```
library(cliProfiler)
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
##      table, tapply, union, unique, unsplit, which.max, which.min
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      expand.grid, I, unname
```

```
library(rtracklayer)
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: IRanges
```

```
##
```

```
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':
```

```
##
```

```
## windows

## Loading required package: GenomeInfoDb
library(GenomicRanges)

data

#gff file
gff <- file.path( "D:/Krueger_Lab/Ribo_Profiling/run15112022M23/ref_genome", "gencode.vM23.annotation.gff3.gz" )
head(gff)

## [1] "D:/Krueger_Lab/Ribo_Profiling/run15112022M23/ref_genome/gencode.vM23.annotation.gff3.gz"

#binding site data
bsdat <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure1/mir181_binding_sites__venn_types/mir181_binding_sites__venn_types.rds")
rownames(bsdat) <- 1:length(bsdat$start)
head(bsdat)

##      seqnames      start      end width strand scoreSum scoreMean scoreMax
## 1 chr1 6245651 6245657      7      + 9.52553 4.762765 6.00678
## 2 chr1 6248341 6248347      7      + 92.68921 23.172303 48.76900
## 3 chr1 6248857 6248863      7      + 14.07133 7.035665 7.04425
## 4 chr1 6248918 6248924      7      + 38.91451 12.971503 20.65080
## 5 chr1 7170481 7170487      7      + 66.92218 13.384436 25.84490
## 6 chr1 9899605 9899611      7      + 25.15963 6.289907 8.61019
##      geneType geneName      geneID region BS_ID      mir_IP
## 1 protein_coding Rb1cc1 ENSMUSG00000025907 cds      5 mmu-miR-181a-5p
## 2 protein_coding Rb1cc1 ENSMUSG00000025907 cds      8 mmu-miR-181a-5p
## 3 protein_coding Rb1cc1 ENSMUSG00000025907 cds     10 mmu-miR-181a-5p
## 4 protein_coding Rb1cc1 ENSMUSG00000025907 cds     11 mmu-miR-181a-5p
## 5 protein_coding Pcmt1 ENSMUSG00000051285 utr3     19 mmu-miR-181a-5p
## 6 protein_coding Sgk3 ENSMUSG00000025915 utr3     23 mmu-miR-181a-5p
##      n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d      set WT KO
## 1      1      1      0      0      0 ago_bs_mir181_chi 1 1
## 2      5      5      0      0      0 ago_bs_mir181_chi 1 1
## 3      6      6      0      0      0 ago_bs_mir181_chi 1 0
## 4      6      6      0      0      0 ago_bs_mir181_chi 1 1
## 5      4      4      0      0      0 ago_bs_mir181_chi 1 1
## 6      1      1      0      0      0 ago_bs_mir181_chi NA NA
##      geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 1 ENSMUSG00000025907 Rb1cc1 cds      4      3
## 2 ENSMUSG00000025907 Rb1cc1 cds     28     32
## 3 ENSMUSG00000025907 Rb1cc1 cds     13     11
## 4 ENSMUSG00000025907 Rb1cc1 cds     15     15
## 5 ENSMUSG00000051285 Pcmt1 utr3     12     22
## 6      <NA>      <NA>      <NA>      NA      NA
##      counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 1      3      3      10      3
## 2     27     46     41     20
## 3      4     22     13     12
## 4     10     33     20     18
## 5     14     16     20      9
## 6     NA     NA     NA     NA
##      geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 1 ENSMUSG00000025907      1609      1973      1250
```

```

## 2 ENSMUSG00000025907      1609      1973      1250
## 3 ENSMUSG00000025907      1609      1973      1250
## 4 ENSMUSG00000025907      1609      1973      1250
## 5 ENSMUSG00000051285      1355      1706      1064
## 6      <NA>      NA      NA      NA
## counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 1      2638      2231      1352      92.10645
## 2      2638      2231      1352      281.53271
## 3      2638      2231      1352      145.51107
## 4      2638      2231      1352      186.74162
## 5      1654      1348      755      151.36245
## 6      NA      NA      NA      NA
## resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue resBs.padj
## 1      -0.1093039      0.5923673      0.03419066      0.8533018      0.9652601
## 2      0.2749428      0.2351157      1.35874137      0.2437557      0.6729889
## 3      -0.1805519      0.3623758      0.25017050      0.6169550      0.8961239
## 4      -0.2606282      0.3062717      0.73169661      0.3923338      0.7868678
## 5      0.1466485      0.3122905      0.22052922      0.6386370      0.9013566
## 6      NA      NA      NA      NA      NA
## resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 1      NA      NA      NA      NA      NA
## 2      NA      NA      NA      NA      NA
## 3      NA      NA      NA      NA      NA
## 4      NA      NA      NA      NA      NA
## 5      NA      NA      NA      NA      NA
## 6      NA      NA      NA      NA      NA
## resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 1      NA      133.7259      117.9980      129.8669
## 2      NA      133.7259      117.9980      129.8669
## 3      NA      133.7259      117.9980      129.8669
## 4      NA      133.7259      117.9980      129.8669
## 5      NA      248.6210      225.2505      244.0445
## 6      NA      NA      NA      NA
## tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1      139.8635      146.2855      163.5360
## 2      139.8635      146.2855      163.5360
## 3      139.8635      146.2855      163.5360
## 4      139.8635      146.2855      163.5360
## 5      193.5994      195.1330      201.6149
## 6      NA      NA      NA
## BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported down
## 1 ENSMUSG00000025907.bs5      3      3      TRUE FALSE
## 2 ENSMUSG00000025907.bs8      3      3      TRUE FALSE
## 3 ENSMUSG00000025907.bs10      3      3      TRUE FALSE
## 4 ENSMUSG00000025907.bs11      3      3      TRUE FALSE
## 5 ENSMUSG00000051285.bs4      3      3      TRUE FALSE
## 6      <NA>      NA      NA      NA      NA

```

```
bsrange <- makeGRangesFromDataFrame(bsdats, keep.extra.columns = T)
```

metaGeneProfile

```
meta <- metaGeneProfile(object = bsrange, annotation = gff)
```

```
meta
```

```
## $Peaks
```

```
## GRanges object with 13846 ranges and 61 metadata columns:
```

```
##           seqnames           ranges strand |   scoreSum scoreMean   scoreMax
##           <Rle>             <IRanges> <Rle> | <numeric> <numeric> <numeric>
##      [1]    chr1 10027158-10027164     - |    14.9100    7.45499    8.38098
##      [2]    chr1 10144758-10144764     - |    24.6005   12.30026   19.23430
##      [3]    chr1 13142515-13142521     - |    29.5267   14.76335   15.00270
##      [4]    chr1 13142708-13142714     - |    16.3183    5.43943    6.05726
##      [5]    chr1 13165078-13165084     - |    12.5979    6.29894    8.00921
##      ...      ...                ...   ... .      ...      ...      ...
## [13842]   chrM      2747-2753         + |   164.5827   54.8609   106.8490
## [13843]   chrM      3908-3914         + |   161.0789   80.5395   116.2640
## [13844]   chrM      5011-5017         + |    53.0906   26.5453   30.1305
## [13845]   chrM      9455-9461         + |   590.5809  196.8603  314.0320
## [13846]   chrM     15351-15357        + |    83.2476   41.6238   52.4650
##           geneType   geneName       geneID       region       BS_ID
##           <character> <character>   <character> <character> <character>
##      [1] protein_coding    Cops5 ENSMUSG000000025917      utr3         920
##      [2] protein_coding    Arfgef1 ENSMUSG000000067851       cds         924
##      [3] protein_coding      Ncoa2 ENSMUSG000000005886      utr3         929
##      [4] protein_coding      Ncoa2 ENSMUSG000000005886      utr3         930
##      [5] protein_coding      Ncoa2 ENSMUSG000000005886       cds         932
##      ...      ...                ...   ...      ...      ...
## [13842]      Mt_tRNA      mt-Tl1 ENSMUSG000000064340      <NA>        27491
## [13843]      Mt_tRNA      mt-Tm ENSMUSG000000064344      <NA>        27492
## [13844]      Mt_tRNA      mt-Tw ENSMUSG000000064346      <NA>        27493
## [13845]      Mt_tRNA      mt-Tg ENSMUSG000000064359      <NA>        27494
## [13846]      Mt_tRNA      mt-Tt ENSMUSG000000064371      <NA>        27496
##           mir_IP  n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
##           <character> <integer> <integer> <integer> <integer> <integer>
##      [1] mmu-miR-181a-5p      1         1         0         0         0
##      [2] mmu-miR-181a-5p      6         3         3         0         0
##      [3] mmu-miR-181c-5p      2         1         0         1         0
##      [4] mmu-miR-181a-5p      7         6         1         0         0
##      [5] mmu-miR-181a-5p      6         5         1         0         0
##      ...      ...                ...   ...      ...      ...
## [13842] mmu-miR-181a-5p      3         2         1         0         0
## [13843] mmu-miR-181b-5p     11         8         3         0         0
## [13844] mmu-miR-181a-5p      4         4         0         0         0
## [13845] mmu-miR-181b-5p     13        11         1         1         0
## [13846] mmu-miR-181b-5p      9         0         9         0         0
##           set      WT      KO      geneID.2
##           <character> <integer> <integer>   <character>
##      [1] ago_bs_mir181_chi <NA>      <NA>      <NA>
##      [2] ago_bs_mir181_chi      1         1 ENSMUSG000000067851
##      [3] ago_bs_mir181_chi      1         1 ENSMUSG000000005886
##      [4] ago_bs_mir181_chi      1         1 ENSMUSG000000005886
##      [5] ago_bs_mir181_chi      1         1 ENSMUSG000000005886
##      ...      ...                ...   ...      ...
## [13842] ago_bs_mir181_chi&mi..      1         1 ENSMUSG000000064340
```

```

## [13843] ago_bs_mir181_chi&mi..      1      1 ENSMUSG00000064344
## [13844] ago_bs_mir181_chi&mi..      <NA>    <NA>    <NA>
## [13845] ago_bs_mir181_chi&mi..      1      0 ENSMUSG00000064359
## [13846] ago_bs_mir181_chi&mi..      1      1 ENSMUSG00000064371
##      geneName.1      region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO
##      <character> <character>      <numeric>      <numeric>      <numeric>
##      [1]      <NA>      <NA>      NA      NA      NA
##      [2]      Arfgef1      cds      3      3      2
##      [3]      Ncoa2      utr3      3      2      4
##      [4]      Ncoa2      utr3      2      5      2
##      [5]      Ncoa2      cds      11     9      4
##      ...      ...      ...      ...      ...
## [13842]      mt-Tl1      <NA>      70     80     67
## [13843]      mt-Tm      <NA>      63     59     58
## [13844]      <NA>      <NA>      NA     NA     NA
## [13845]      mt-Tg      <NA>      11     12     14
## [13846]      mt-Tt      <NA>      9      18     8
##      counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT      geneID.1
##      <numeric>      <numeric>      <numeric>      <character>
##      [1]      NA      NA      NA      <NA>
##      [2]      10     13      5 ENSMUSG00000067851
##      [3]      14     11      3 ENSMUSG00000005886
##      [4]      9      6      4 ENSMUSG00000005886
##      [5]      9      3      2 ENSMUSG00000005886
##      ...      ...      ...      ...
## [13842]      81     69     33 ENSMUSG00000064340
## [13843]      63     64     48 ENSMUSG00000064344
## [13844]      NA     NA     NA      <NA>
## [13845]     211    170    126 ENSMUSG00000064359
## [13846]      23     16     11 ENSMUSG00000064371
##      counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO counts.bg.4_WT
##      <numeric>      <numeric>      <numeric>      <numeric>
##      [1]      NA      NA      NA      NA
##      [2]     1126    1392     857    1596
##      [3]     4296    5887    3299    5947
##      [4]     4296    5887    3299    5947
##      [5]     4296    5887    3299    5947
##      ...      ...      ...      ...
## [13842]      37     53     47     70
## [13843]      66    114     45    100
## [13844]      NA     NA     NA     NA
## [13845]      11      9     11     30
## [13846]      10     15      8     20
##      counts.bg.5_WT counts.bg.6_WT resBs.baseMean resBs.log2FoldChange
##      <numeric>      <numeric>      <numeric>      <numeric>
##      [1]      NA      NA      NA      NA
##      [2]     1350     819     78.2351    -1.442887
##      [3]     4876    2740    189.9235    -1.424019
##      [4]     4876    2740    181.8876    -0.866460
##      [5]     4876    2740    195.2222     0.990124
##      ...      ...      ...      ...
## [13842]      45      23    465.1285     0.444960
## [13843]      69      45    428.5487     0.172623
## [13844]      NA      NA      NA      NA

```

```

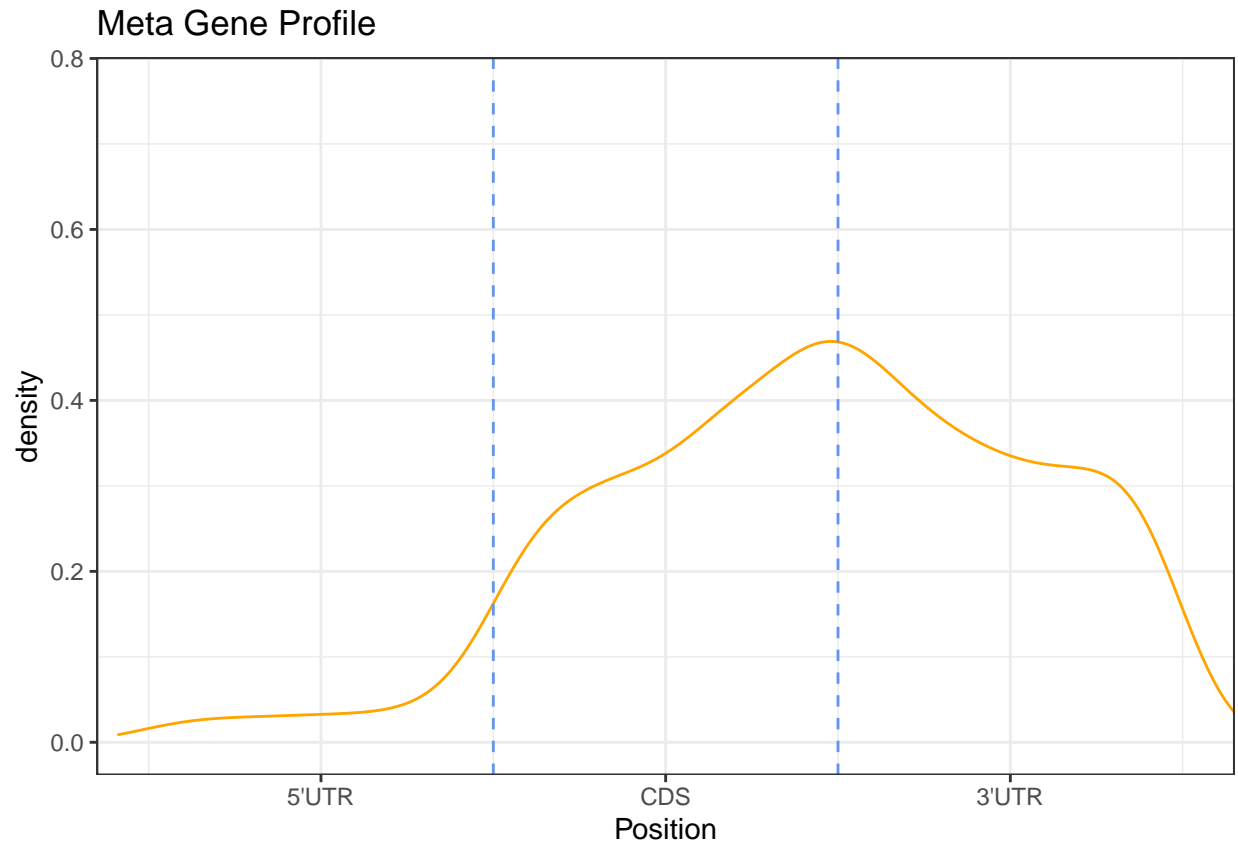
## [13845]          18          15      569.2355      -2.601586
## [13846]          14          14      93.7243      0.226665
##      resBs.lfcSE resBs.stat resBs.pvalue resBs.padj resBg.baseMean
##      <numeric> <numeric> <numeric> <numeric> <numeric>
##      [1]      NA      NA      NA      NA      NA
##      [2] 0.588541 6.96015 0.00833450 0.128045      NA
##      [3] 0.556464 7.50217 0.00616248 0.106099      NA
##      [4] 0.587308 2.30499 0.12895943 0.524353      NA
##      [5] 0.489466 4.26246 0.03896356 0.303616      NA
##      ...      ...      ...      ...      ...
## [13842] 0.273340 2.647084 1.03740e-01 4.77771e-01 44.1533
## [13843] 0.308778 0.310463 5.77396e-01 8.82032e-01 69.3062
## [13844]      NA      NA      NA      NA      NA
## [13845] 0.445210 32.529359 1.17405e-08 3.04740e-06 15.3460
## [13846] 0.473372 0.229200 6.32118e-01 8.99037e-01 13.3516
##      resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue resBg.padj
##      <numeric> <numeric> <numeric> <numeric> <numeric>
##      [1]      NA      NA      NA      NA      NA
##      [2]      NA      NA      NA      NA      NA
##      [3]      NA      NA      NA      NA      NA
##      [4]      NA      NA      NA      NA      NA
##      [5]      NA      NA      NA      NA      NA
##      ...      ...      ...      ...      ...
## [13842] 0.229298 0.281511 0.814527 0.4153434 0.645856
## [13843] 0.217153 0.226625 0.958205 0.3379592 0.570024
## [13844]      NA      NA      NA      NA      NA
## [13845] -0.816044 0.470339 -1.735013 0.0827384 0.228944
## [13846] -0.425541 0.489552 -0.869245 0.3847129 0.617400
##      tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
##      <numeric> <numeric> <numeric>
##      [1]      NA      NA      NA
##      [2] 121.465 108.054 115.564
##      [3] 335.315 330.652 321.885
##      [4] 335.315 330.652 321.885
##      [5] 335.315 330.652 321.885
##      ...      ...      ...
## [13842] 602.312 620.846 956.416
## [13843] 1167.821 1451.527 995.345
## [13844]      NA      NA      NA
## [13845] 197.499 116.279 246.885
## [13846] 182.224 196.692 182.232
##      tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
##      <numeric> <numeric> <numeric>
##      [1]      NA      NA      NA
##      [2] 109.828 114.891 128.580
##      [3] 296.113 300.258 311.255
##      [4] 296.113 300.258 311.255
##      [5] 296.113 300.258 311.255
##      ...      ...      ...
## [13842] 726.922 577.929 544.910
## [13843] 1128.762 963.216 1158.836
## [13844]      NA      NA      NA
## [13845] 343.608 254.969 391.959
## [13846] 232.491 201.269 371.289

```

```

##          BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported
##          <character>      <integer>      <integer>      <logical>
##      [1]          <NA>          <NA>          <NA>          <NA>
##      [2] ENSMUSG000000067851.bs4           3           3          TRUE
##      [3] ENSMUSG000000005886.bs4           3           3          TRUE
##      [4] ENSMUSG000000005886.bs5           3           3          TRUE
##      [5] ENSMUSG000000005886.bs7           3           3          TRUE
##      ...          ...          ...          ...          ...
## [13842] ENSMUSG000000064340.bs1           3           3          TRUE
## [13843] ENSMUSG000000064344.bs1           3           3          TRUE
## [13844]          <NA>          <NA>          <NA>          <NA>
## [13845] ENSMUSG000000064359.bs1           3           3          TRUE
## [13846] ENSMUSG000000064371.bs1           3           3          TRUE
##          down      center      location      Gene_ID
##          <logical> <integer> <character>      <character>
##      [1]          <NA>    10027162          CDS ENSMUSG000000025917.9
##      [2]         FALSE    10144762          CDS ENSMUSG000000067851.11
##      [3]         FALSE    13142519          UTR3 ENSMUSG000000005886.14
##      [4]         FALSE    13142712          UTR3 ENSMUSG000000005886.14
##      [5]         FALSE    13165082          CDS ENSMUSG000000005886.14
##      ...          ...          ...          ...          ...
## [13842]         FALSE         2751          CDS ENSMUSG000000064341.1
## [13843]         FALSE         3912           NO          Nan
## [13844]          <NA>         5015           NO          Nan
## [13845]          TRUE         9459          CDS ENSMUSG000000064360.1
## [13846]         FALSE        15355           NO          Nan
##          Transcript_ID      Position
##          <character>      <numeric>
##      [1] ENSMUST000000027050.9      0.8597015
##      [2] ENSMUST000000088615.10      0.8859412
##      [3] ENSMUST000000081713.10      0.0670131
##      [4] ENSMUST000000081713.10      0.0142232
##      [5] ENSMUST00000006037.12      0.5989975
##      ...          ...          ...
## [13842] ENSMUST000000082392.1 -0.00417973
## [13843]          <NA>      5.00000000
## [13844]          <NA>      5.00000000
## [13845] ENSMUST000000082411.1 -0.01149425
## [13846]          <NA>      5.00000000
## -----
## seqinfo: 22 sequences from an unspecified genome; no seqlengths
##
## $Plot

```



```
testpath <- system.file("extdata", package = "cliProfiler") ## loading the test GRanges object test <-
readRDS(file.path(testpath, "test.rds")) ## Show an example of GRanges object test
```

the path for the test gff3 file

```
test_gff3 <- file.path(testpath, "annotation_test.gff3") ## the gff3 file can be loaded by import.gff3 function
in rtracklayer package shown_gff3 <- rtracklayer::import.gff3(test_gff3) ## show the test gff3 file shown_gff3
meta <- metaGeneProfile(object = test, annotation = test_gff3) meta[[1]]
```