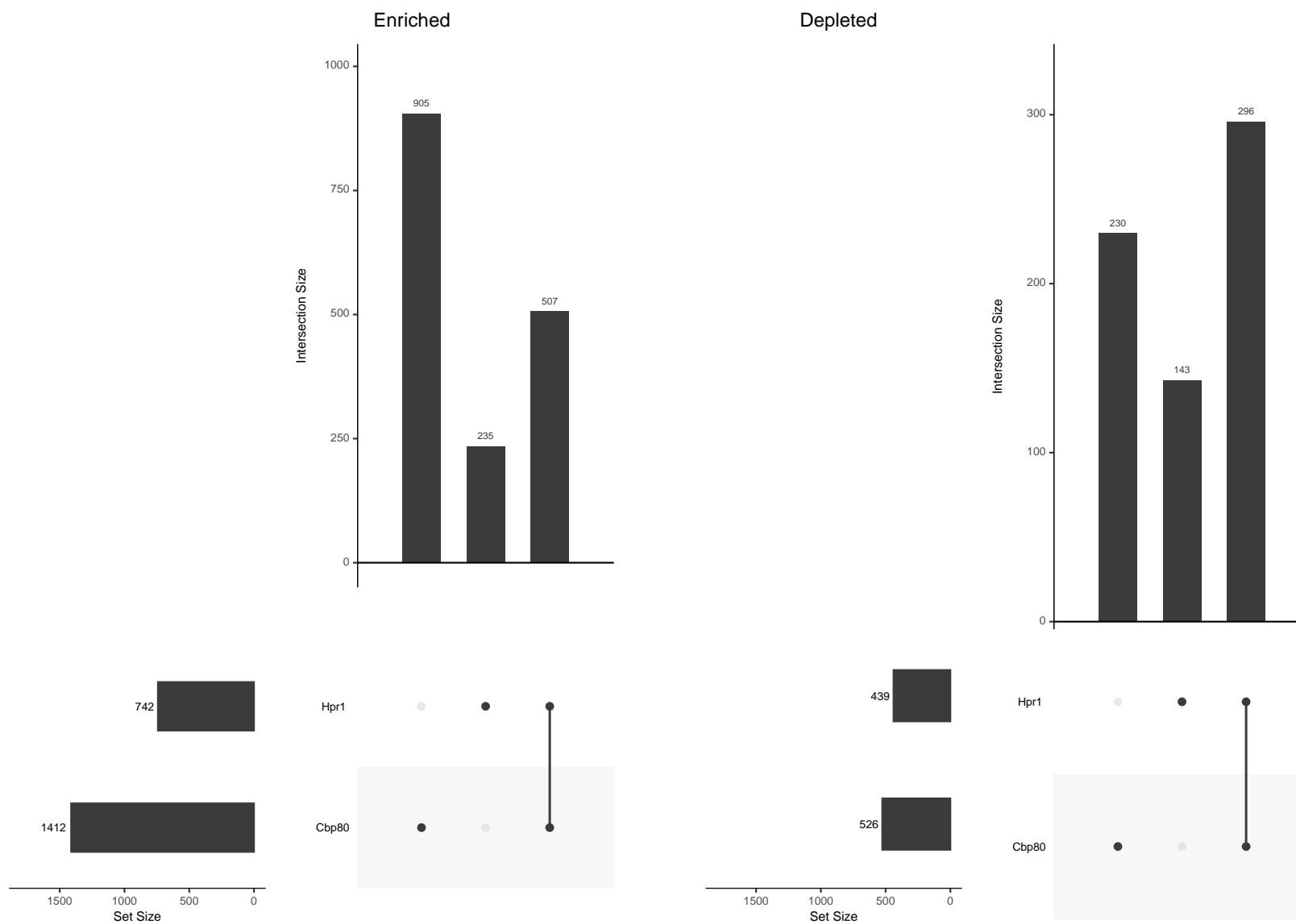


## Yra1 RNA-seq Transcript Characterization analysis

This report was run with a log2FC threshold of 1

## Genes enriched and depleted in Yra1 containing IPs



##FEATURES Here we are interested in exploring the features that correspond to specific IPs (introns, Poly(A) length, ... etc)

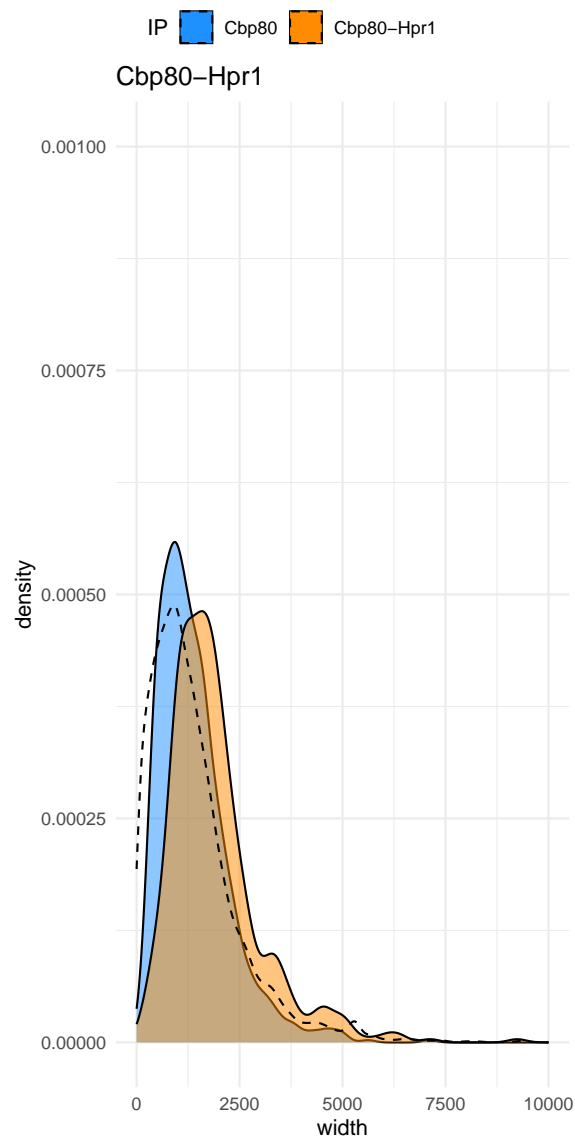
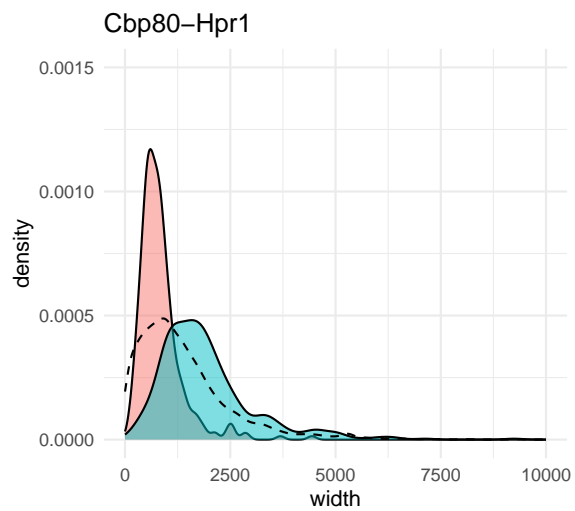
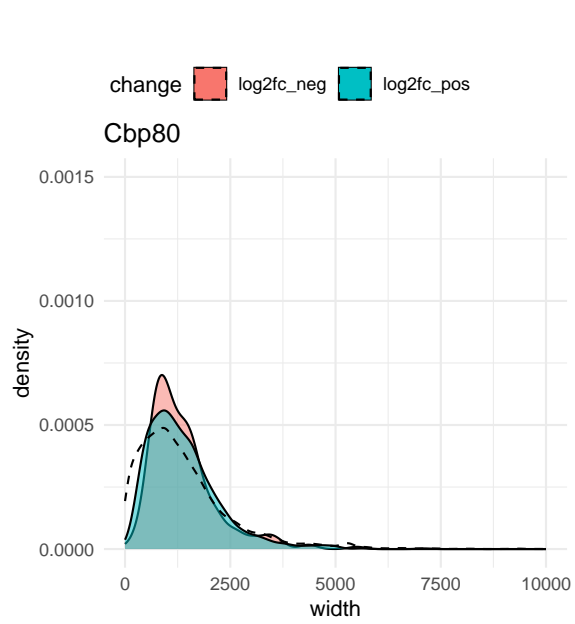
## Length of transcripts

**Question:** Does transcript length impact transcript association with IP samples?

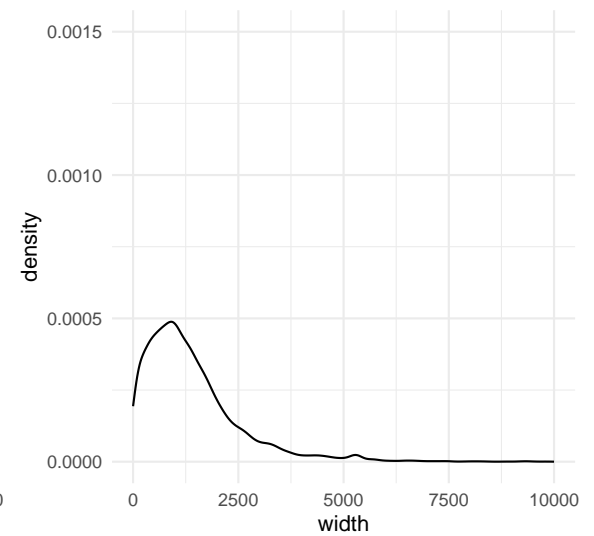
**Answer:** yes, genes that are enriched in Cbp80-Hpr1 pull downs are of longer length than those enriched in Cbp80-only pull downs

**Test used:** `wilcox.test()` – wilcoxon rank sum test. Is mean gene length different between up and down regulated genes? (data are not normally distributed)

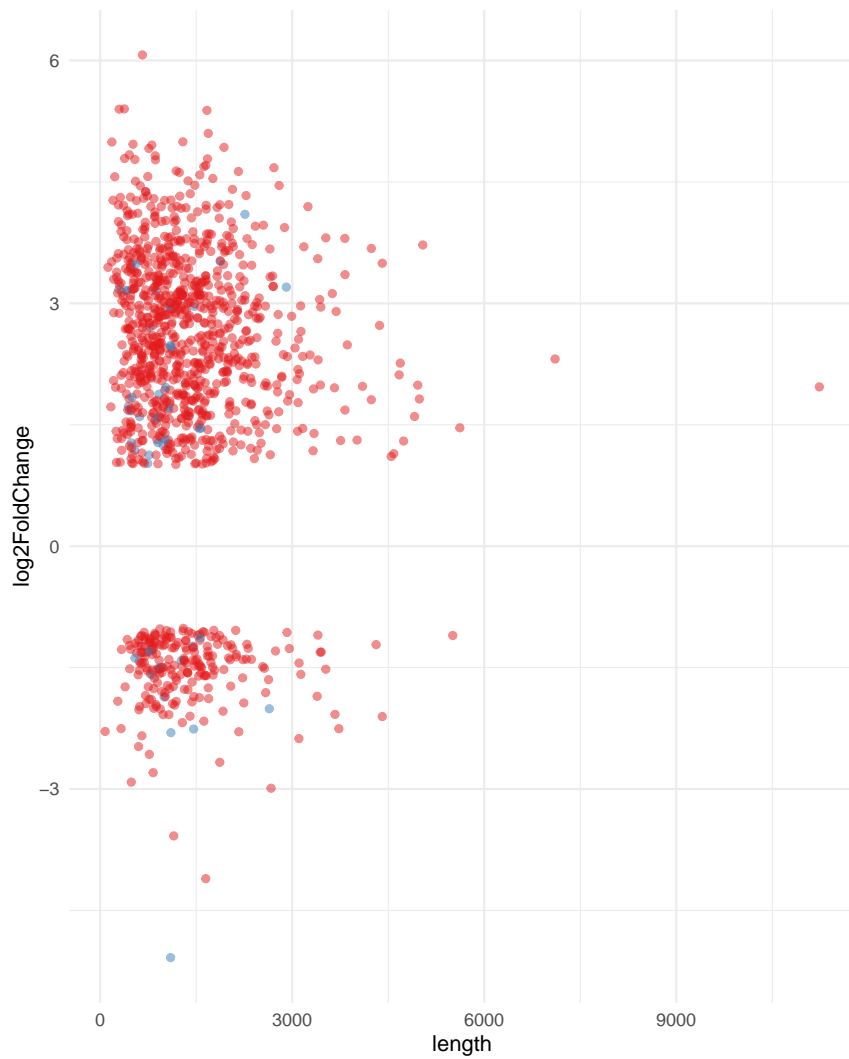
```
## $x
## [1] "transcript length"
##
## $title
## [1] "transcriptome distribution"
##
## attr("class")
## [1] "labels"
```



statistic	p.value	method	set
105598	7.713634e-01	wilcox.test()	Cbp80
20509	2.856618e-66	wilcox.test()	Hpr1
155691	6.705644e-24	wilcox.test()	Cbp80 v Cbp80-Hpr1



Cbp80 (n = 905 + 230)



Cbp80-Hpr1 (n = 507 + 296)



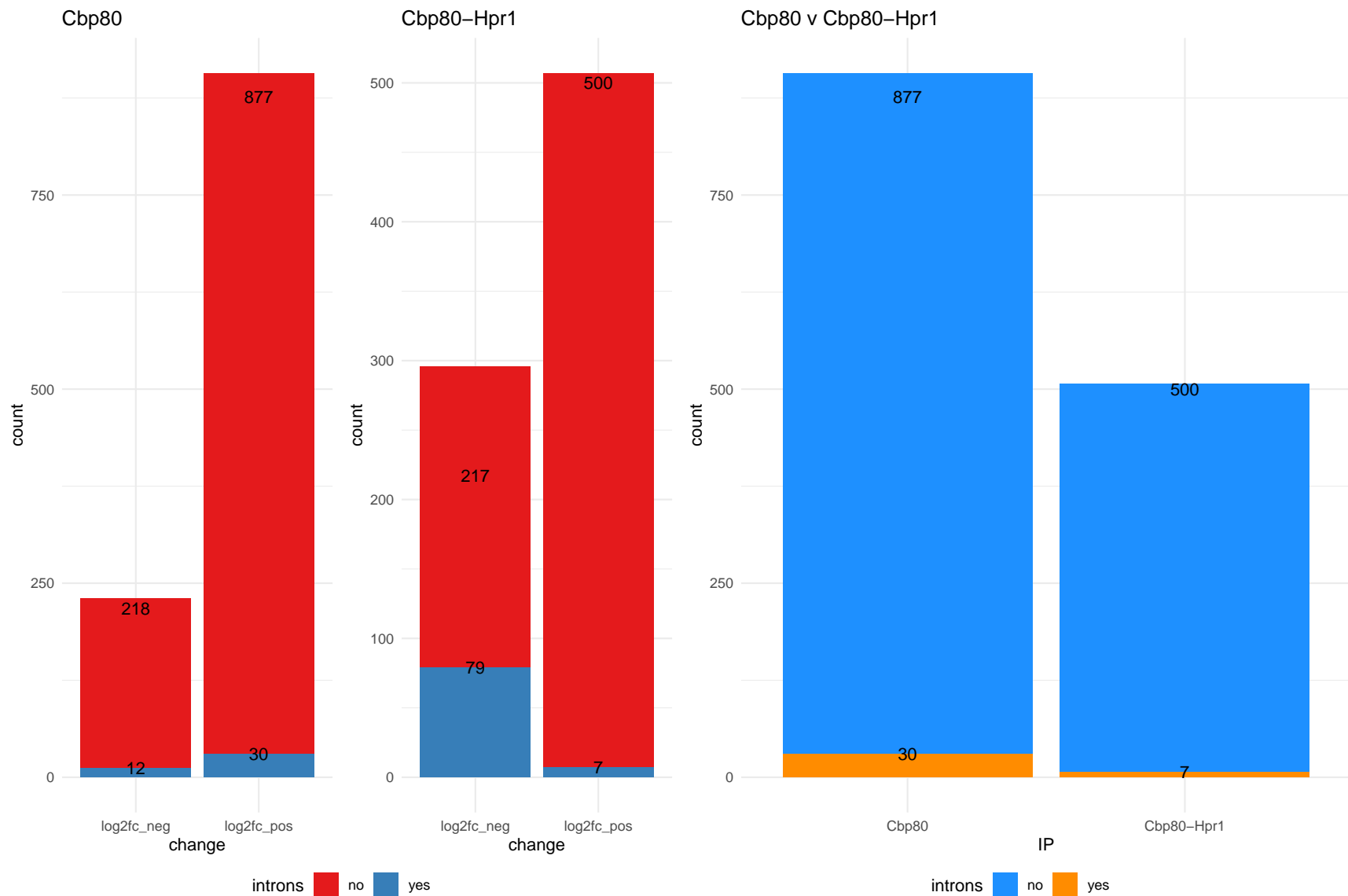
introns • no • yes

## **Introns**

**Question:** Are transcripts with introns associated with certain RBPs?

**Answer:** There is no significant difference in the intron containing genes that associate with Cbp80-only or Cbp80-Hpr1

**Test Used:** `chisq.test()` – chi squared test. Are genes with introns more or less likely to be associated with a different RBPs?

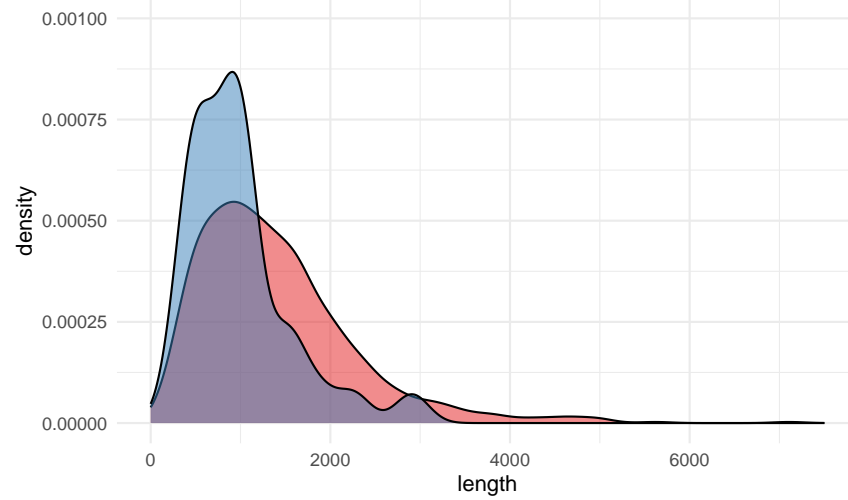


**Question:** What is the length of the intron-containing genes for genes enriched in the IP samples? are

**Answer:** Of the genes that contain introns, there is no significant difference in length between those that associate with Cbp80-only and those that

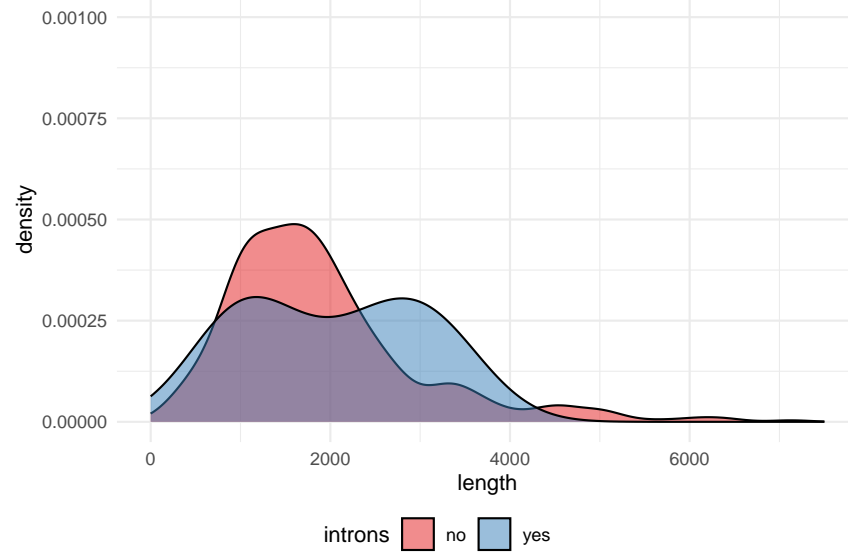
associate with Cbp80-Hpr1

Cbp80 Only, log2FC>1

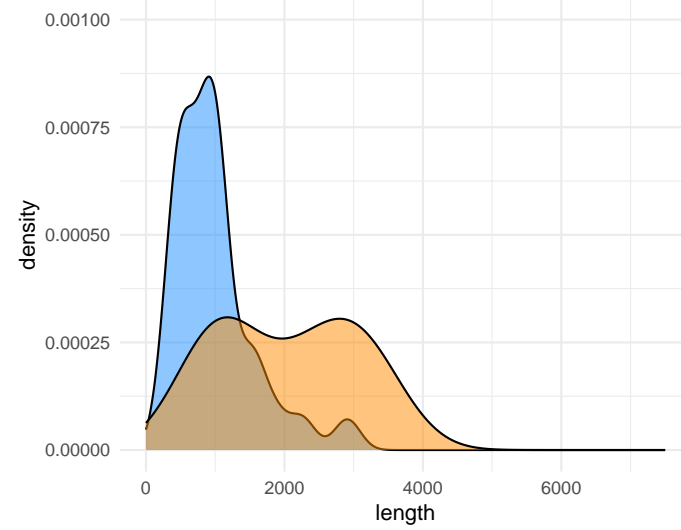


statistic	p.value	method	set
37	0.3767391	chisq.test()	Cbp80 v Cbp80-Hpr1

Cbp80-Hpr1, log2FC>1



Cbp80 v Cbp80-Hpr1, +introns, log2FC>1





```
## # A tibble: 1 x 12
##   r.squared adj.r.squa~1 sigma stati~2 p.value    df logLik   AIC   BIC devia~3
##   <dbl>      <dbl> <dbl>   <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1    0.288      0.286 1.60    104. 4.71e-74     4 -1936. 3884. 3914. 2610.
## # ... with 2 more variables: df.residual <int>, nobs <int>, and abbreviated
## #   variable names 1: adj.r.squared, 2: statistic, 3: deviance
```

```
## # A tibble: 5 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    1.09    0.419      2.61 9.15e- 3
## 2 width          0.000144 0.0000567    2.55 1.11e- 2
## 3 introns2       -0.393    0.281     -1.40 1.62e- 1
## 4 polya_length   0.0592    0.00896     6.61 6.27e-11
## 5 mean_polya_vst -0.506    0.0340    -14.9 1.46e-45
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squa~1 sigma stati~2 p.value    df logLik   AIC   BIC devia~3
##   <dbl>      <dbl> <dbl>   <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1    0.519      0.516 1.49    200. 1.69e-116     4 -1359. 2730. 2757. 1657.
## # ... with 2 more variables: df.residual <int>, nobs <int>, and abbreviated
## #   variable names 1: adj.r.squared, 2: statistic, 3: deviance
```

```
## # A tibble: 5 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)   -2.72    0.446     -6.11 1.58e- 9
## 2 width          0.000731 0.0000531    13.8 1.56e-38
## 3 introns2       -1.00    0.191     -5.26 1.87e- 7
## 4 polya_length   0.103    0.00928    11.1 1.26e-26
## 5 mean_polya_vst -0.234    0.0357     -6.56 1.03e-10
```

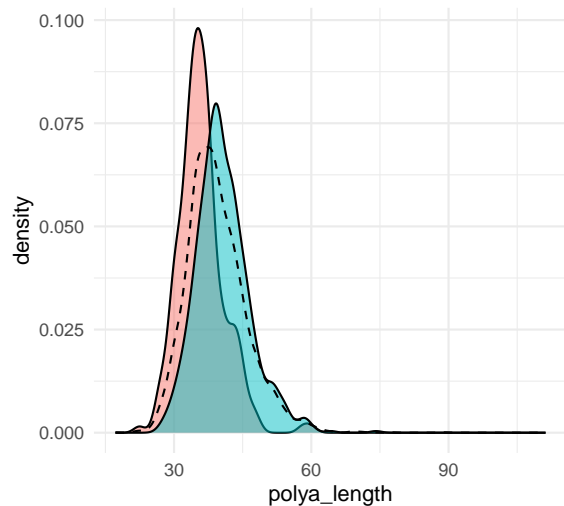
## **Poly(A) tail length**

**Question:** Do the transcripts that associate with a particular IP have longer/shorter tail lengths

**Answer:** Poly(A) tails are not significantly longer/shorter for transcripts in either IP

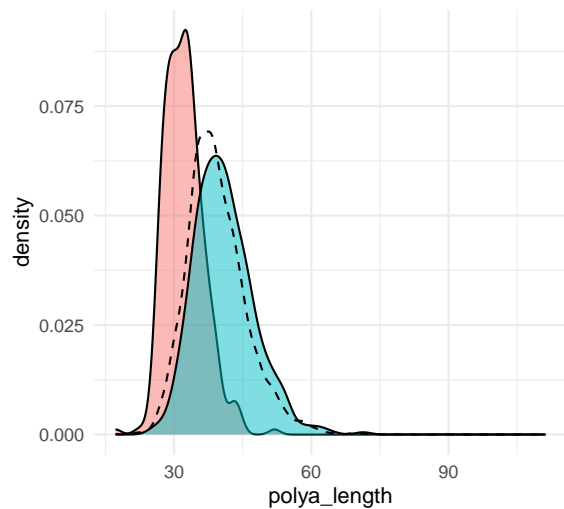
**Test used:** `wilcox.test()` – wilcoxon rank sum test. Is mean poly(A) tail length different between two groups? (data are not normally distributed)

Cbp80 (n = 905 + 230)



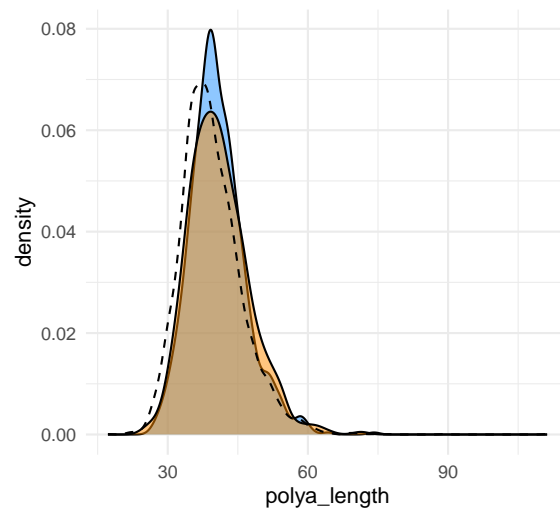
statistic	p.value	method	set
44036.0	1.312951e-29	wilcox.test()	Cbp80
15428.0	5.427098e-69	wilcox.test()	Cbp80Hpr1
187184.5	8.074055e-01	wilcox.test()	Cbp80 v Cbp80Hpr1

Cbp80Hpr1 (n = 507 + 296)



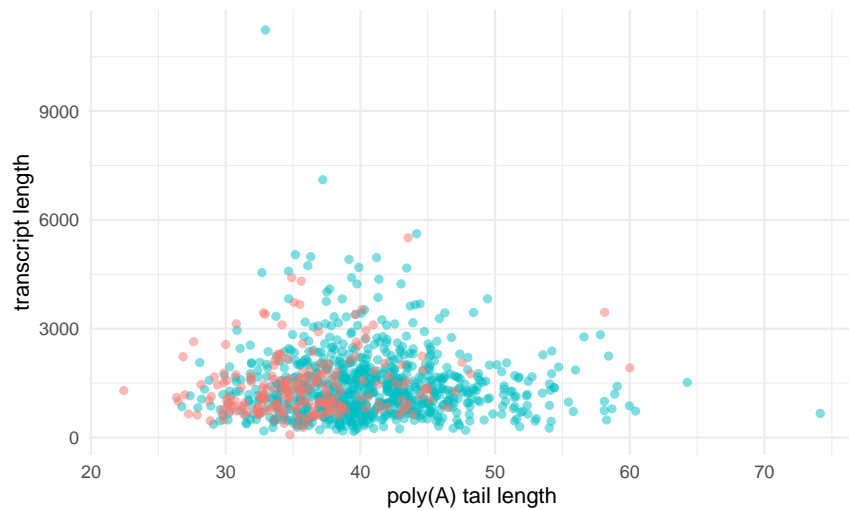
change ■ log2fc\_neg ■ log2fc\_pos

Cbp80 v Cbp80Hpr1 (n = 905 + 507)

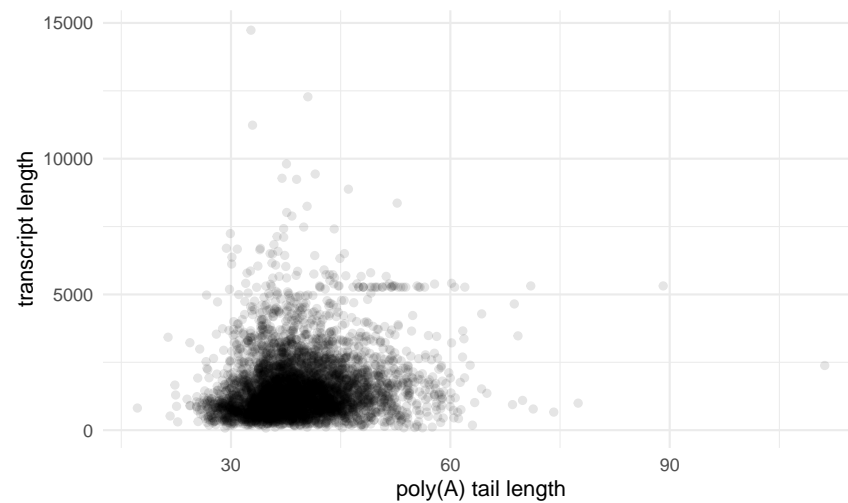


IP ■ Cbp80 ■ Cbp80-Hpr1

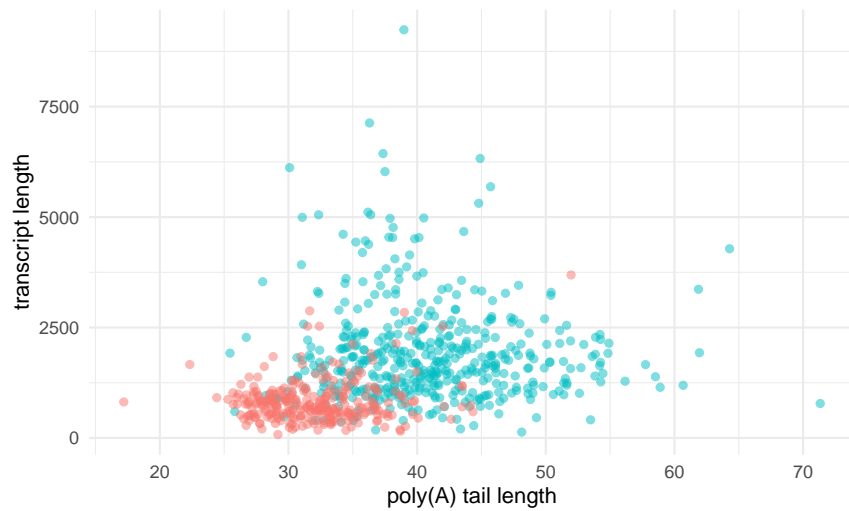
Cbp80 (n = 905 + 230)



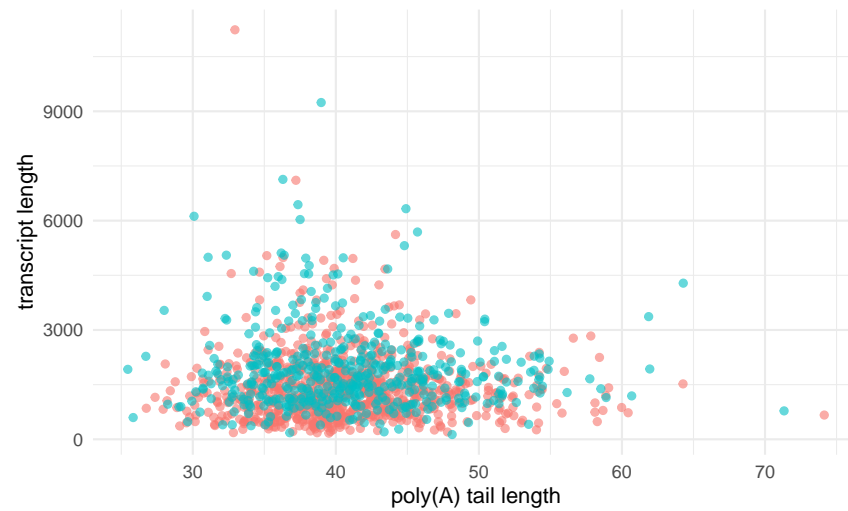
whole transcriptome



Cbp80Hpr1 (n = 507 + 296)



Cbp80 v Cbp80Hpr1 (n = 905 + 507)



change ● log2fc\_neg ● log2fc\_pos

IP ● Cbp80 ● Cbp80-Hpr1

## Synthesis rate

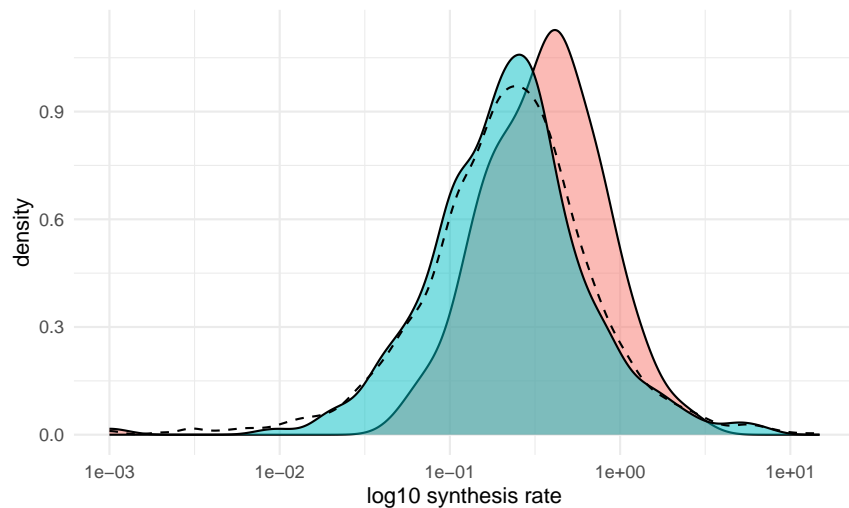
“Note because the x axis is log10 transformed, synthesis time = 0 are NOT on the density plots. I performed the statistics on non-log10 transformed synthesis rates.” - Taylor R

**Question:** Are the synthesis rates unique to different RBPs?

**Answer:** transcripts associated with Cbp80-Hpr1 have a significantly higher synthesis rate

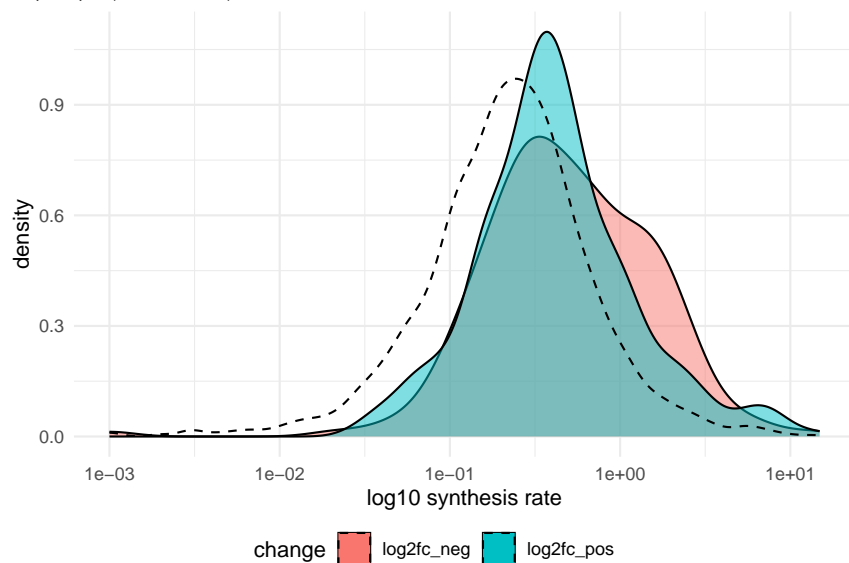
**Tested Used:** `wilcox.test()` – wilcoxon rank sum test. Is the mean synthesis rate different between two groups? (data are not normally distributed)

Cbp80 (n = 905 + 230)

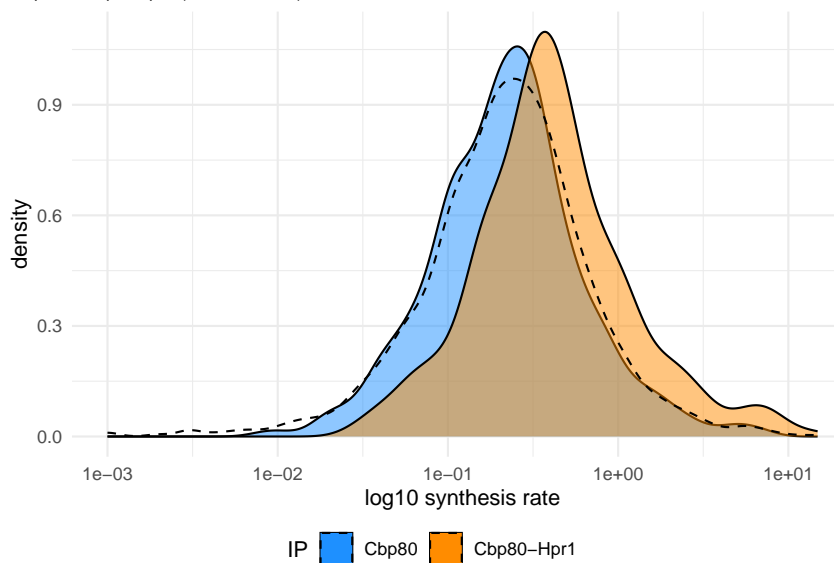


statistic	p.value	method	set
122626.0	1.422265e-12	wilcox.test()	Cbp80
50897.5	1.842068e-02	wilcox.test()	Cbp80Hpr1
125222.0	1.867869e-23	wilcox.test()	Cbp80 v Cbp80Hpr1

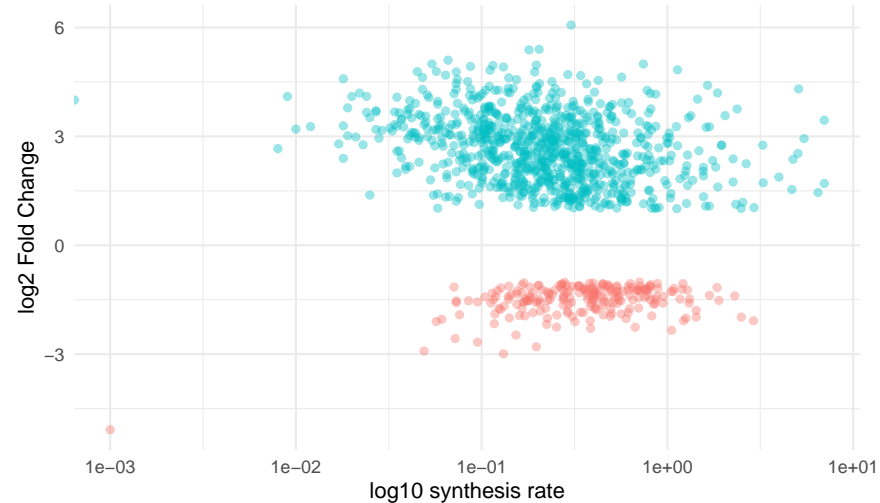
Cbp80Hpr1 (n = 507 + 296)



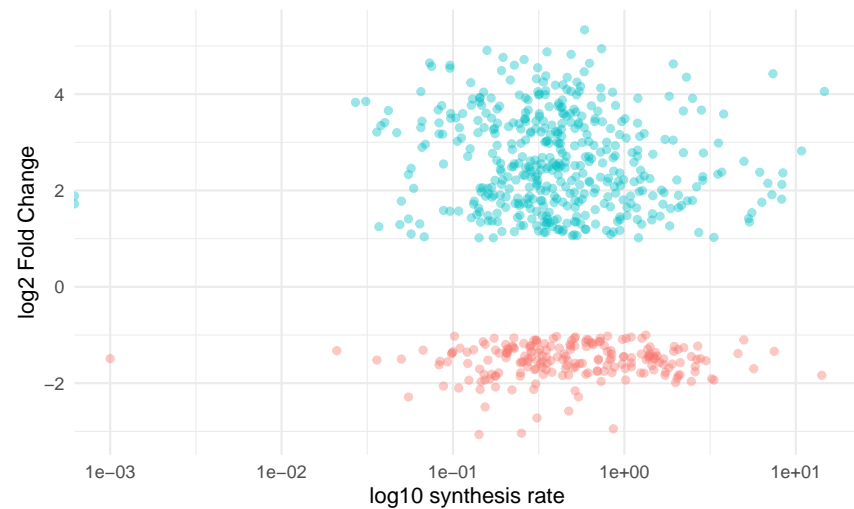
Cbp80 v Cbp80Hpr1 (n = 905 + 507)



Cbp80 (n = 905 + 230)

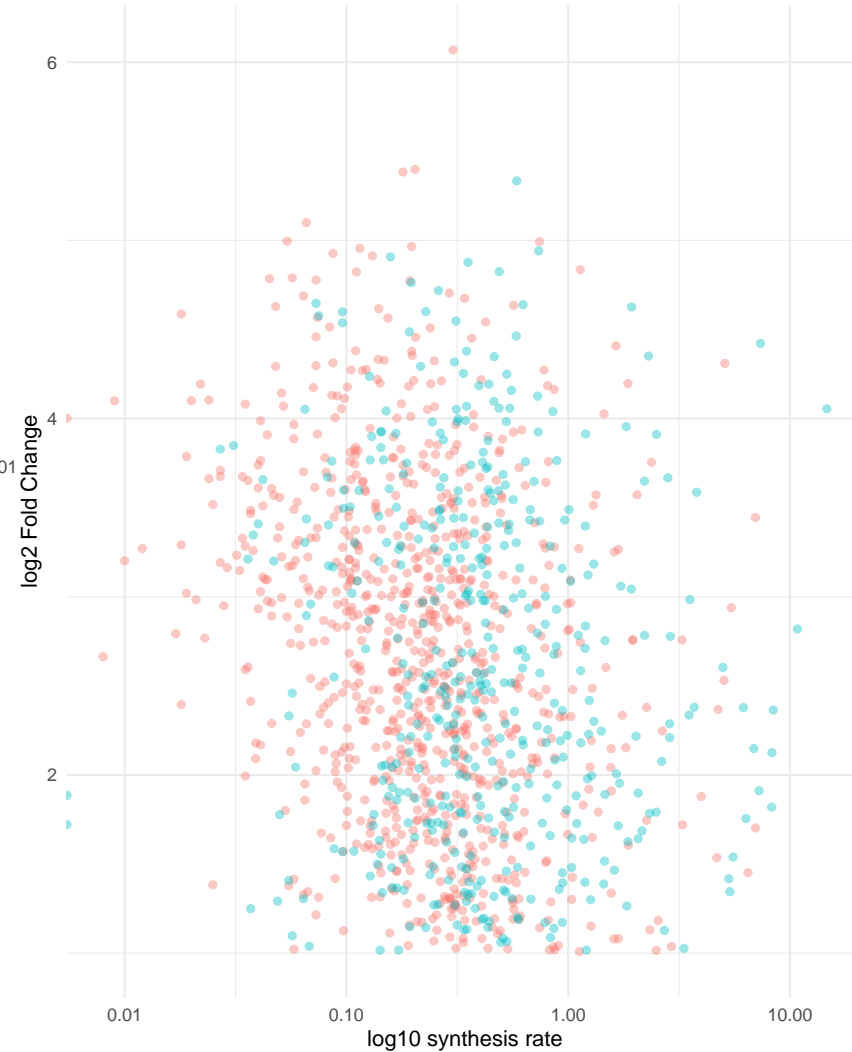


Cbp80Hpr1 (n = 507 + 296)



change ● log2fc\_neg ● log2fc\_pos

Cbp80 v Cbp80Hpr1 (n = 905 + 507)



IP ● Cbp80 ● Cbp80-Hpr1

## Half life

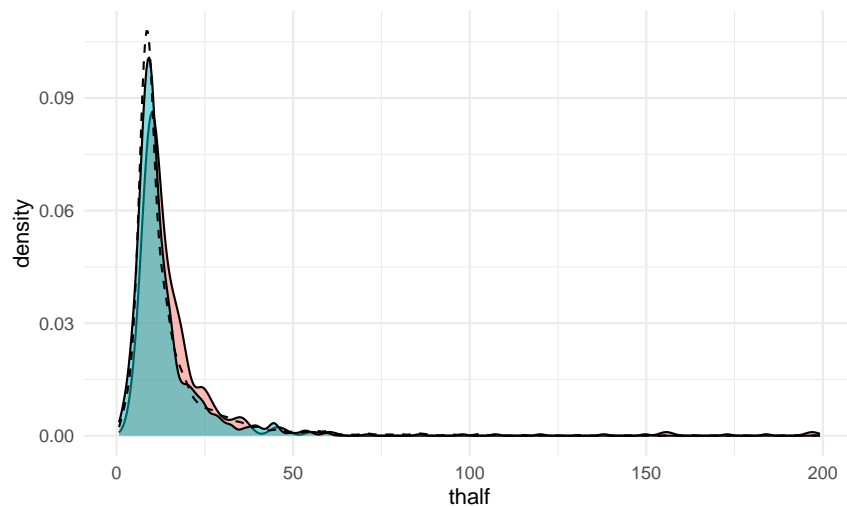
**Question:** Are transcripts with longer/shorter half-life more likely to be associated with specific RBP subsets.

**Answer:** Cbp80-only associate transcripts have a significantly longer half-life

**Test Used:** `wilcox.test()` – wilcoxon rank sum test. Is the mean half-life different between two groups? (data are not normally distributed)

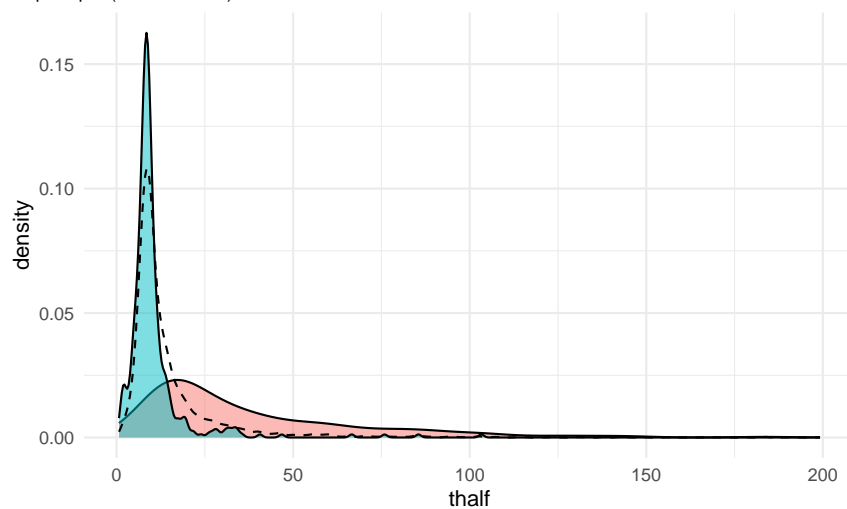


Cbp80 (n = 905 + 230)



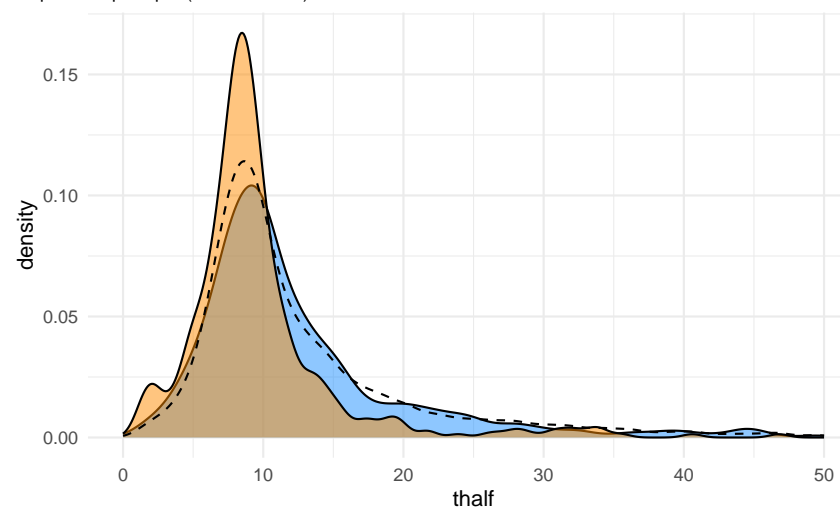
statistic	p.value	method	set
110656.5	3.061938e-05	wilcox.test()	Cbp80
83626.5	1.326980e-65	wilcox.test()	Cbp80-Hpr1
238766.5	1.107246e-14	wilcox.test()	Cbp80 v Cbp80-Hpr1

Cbp80Hpr1 (n = 507 + 296)



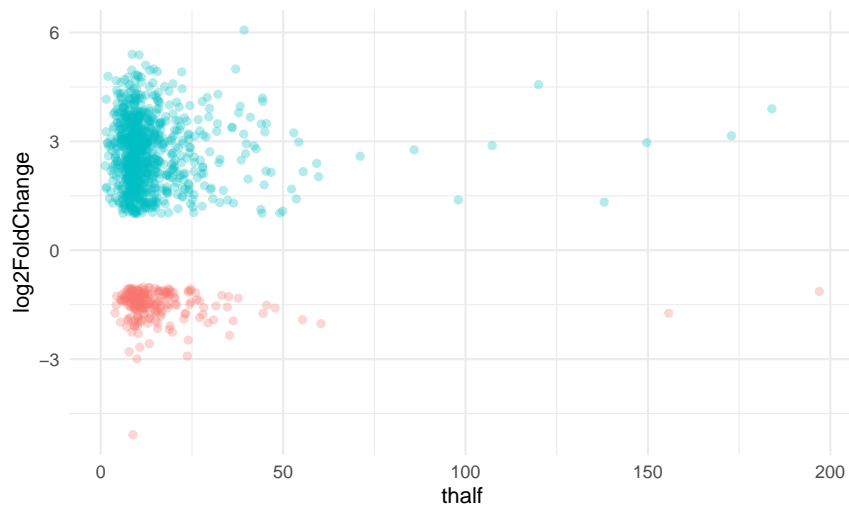
change ■ log2fc\_neg ■ log2fc\_pos

Cbp80 v Cbp80Hpr1 (n = 905 + 507)

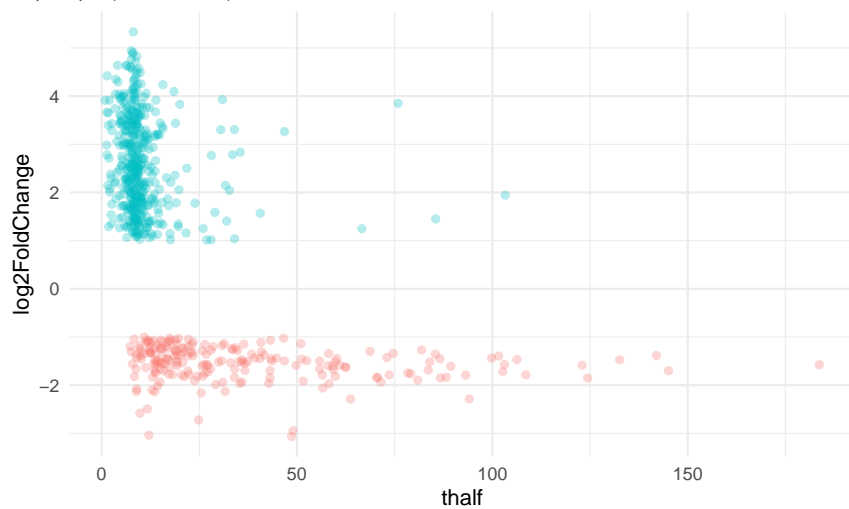


IP ■ Cbp80 ■ Cbp80-Hpr1

Cbp80 (n = 905 + 230)

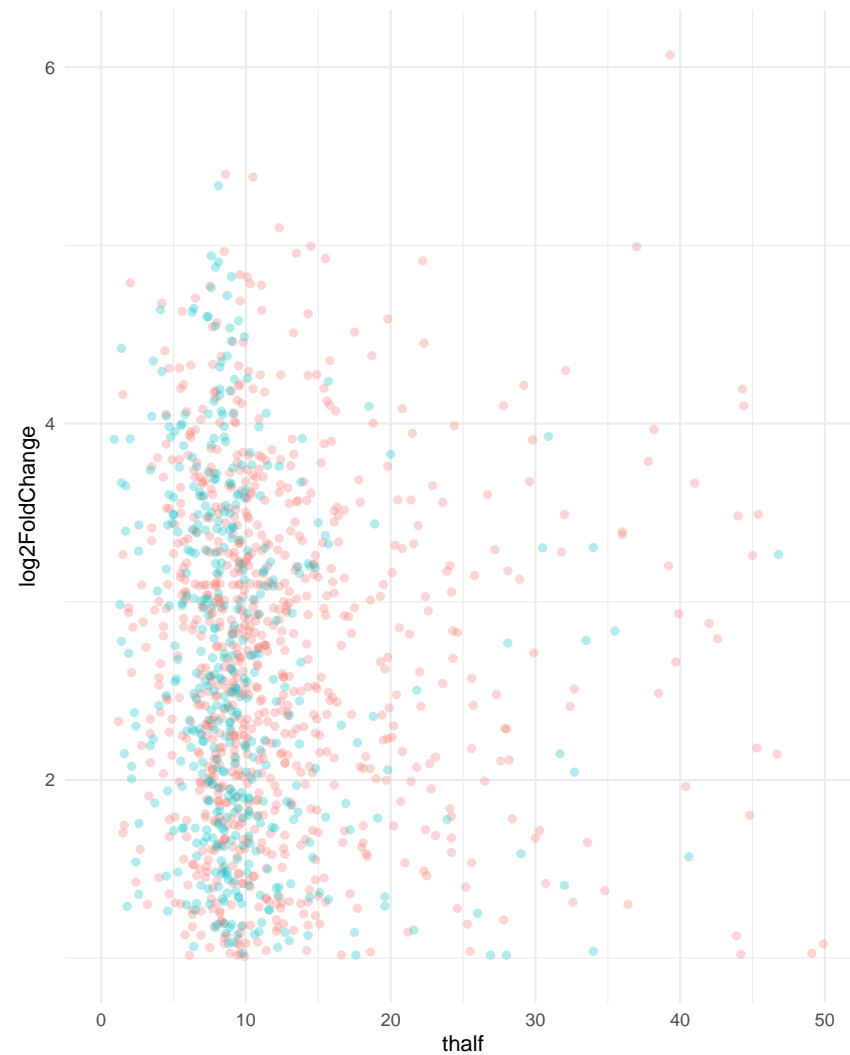


Cbp80Hpr1 (n = 507 + 296)



change • log2fc\_neg • log2fc\_pos

Cbp80 v Cbp80Hpr1 (n = 905 + 507)



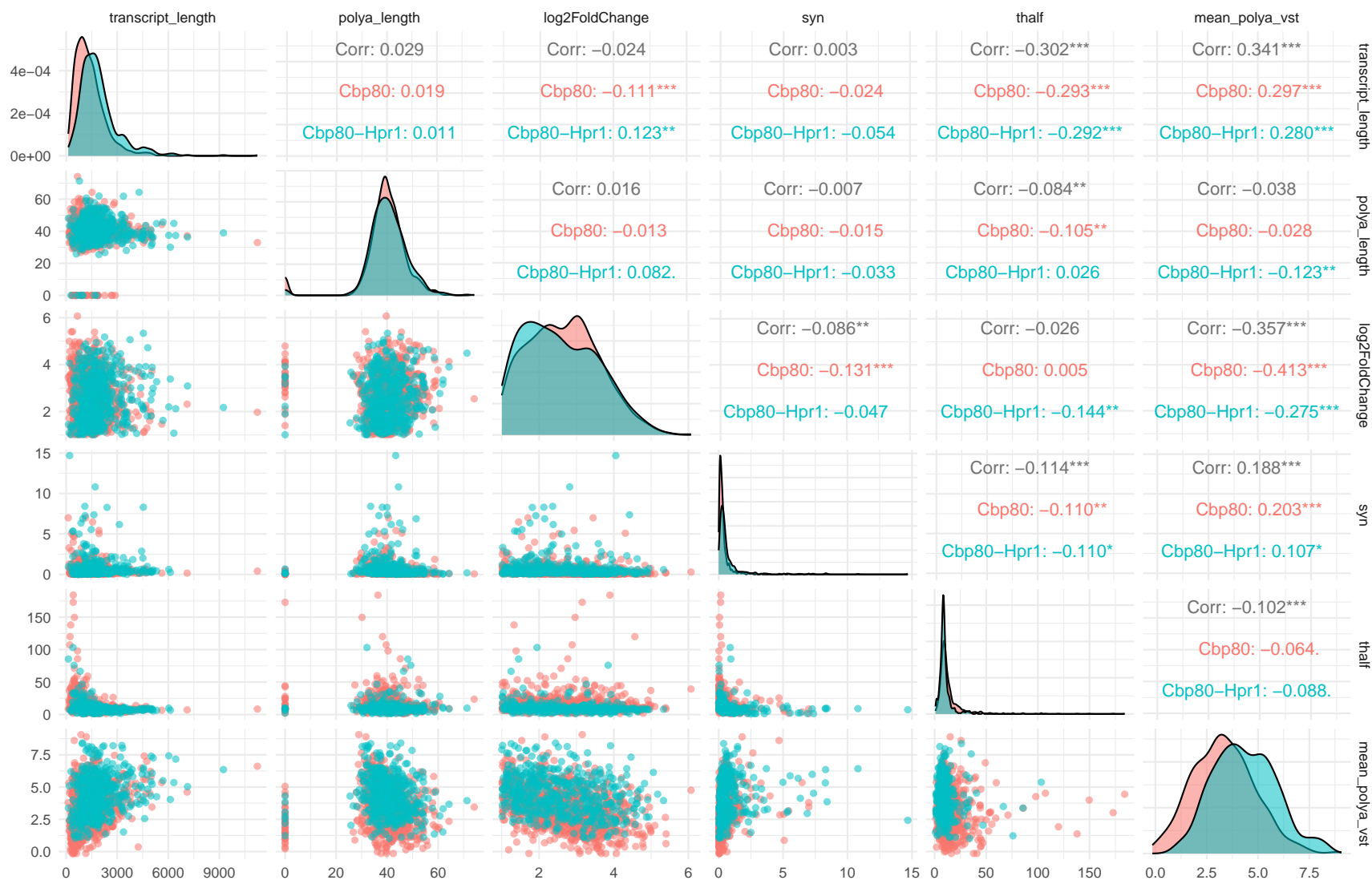
IP • Cbp80 • Cbp80-Hpr1

## Do all of these features correlate?

(E.g. are we pulling the same pattern over and over?)

Looks like most features overlap between Cbp80 and Cbp80-Hpr1

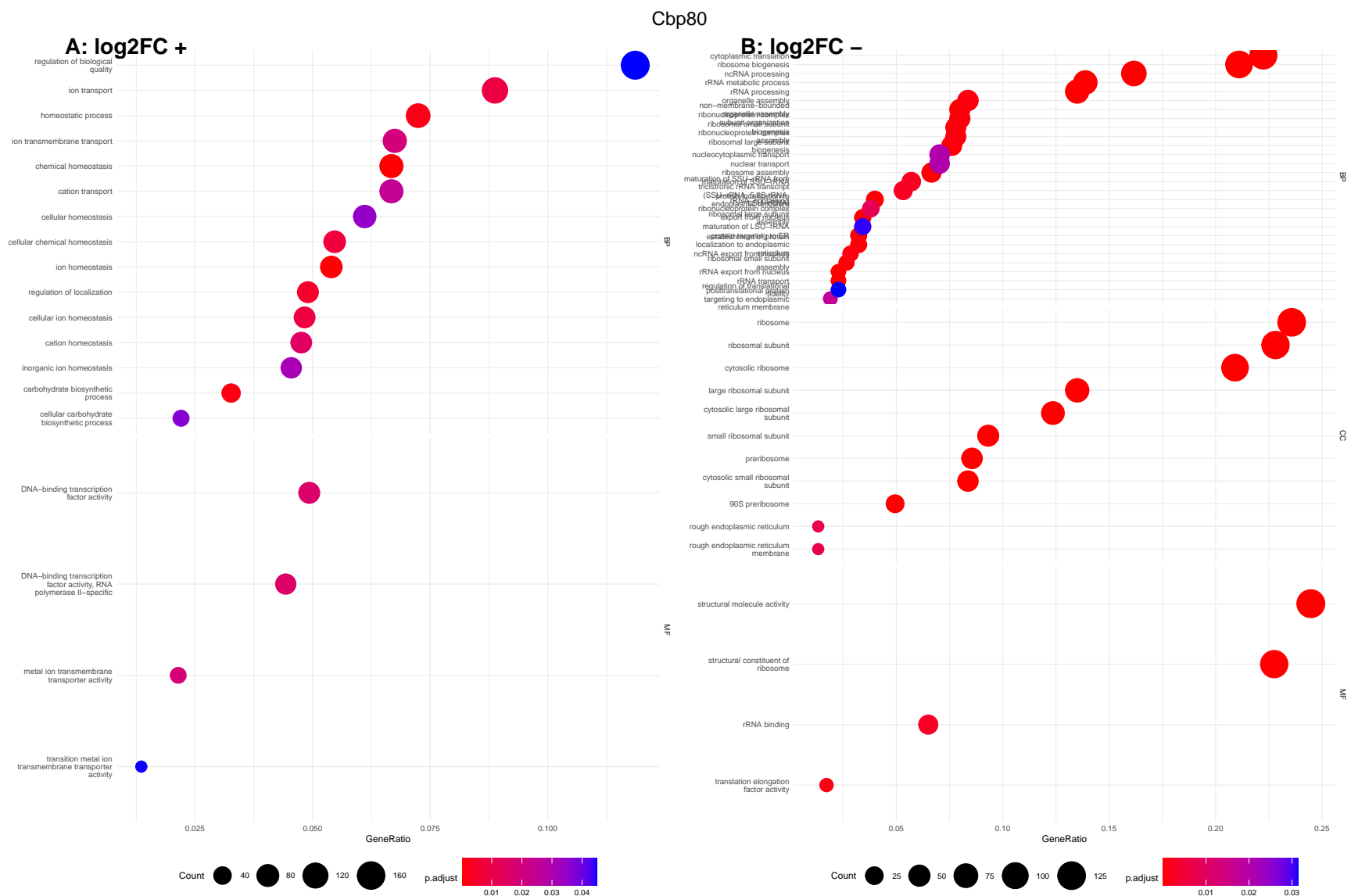
# Cbp80 v Cbp80Hpr1 (n = 905 + 507)



## Enrichment

**Question:** What are the GO terms enriched with each RBPs subset?

full sets



Cbp80-Hpr1

