

# Comparing SCEPTRE and Schraivogel method on Schraivogel data

Gene; 2023-04-08

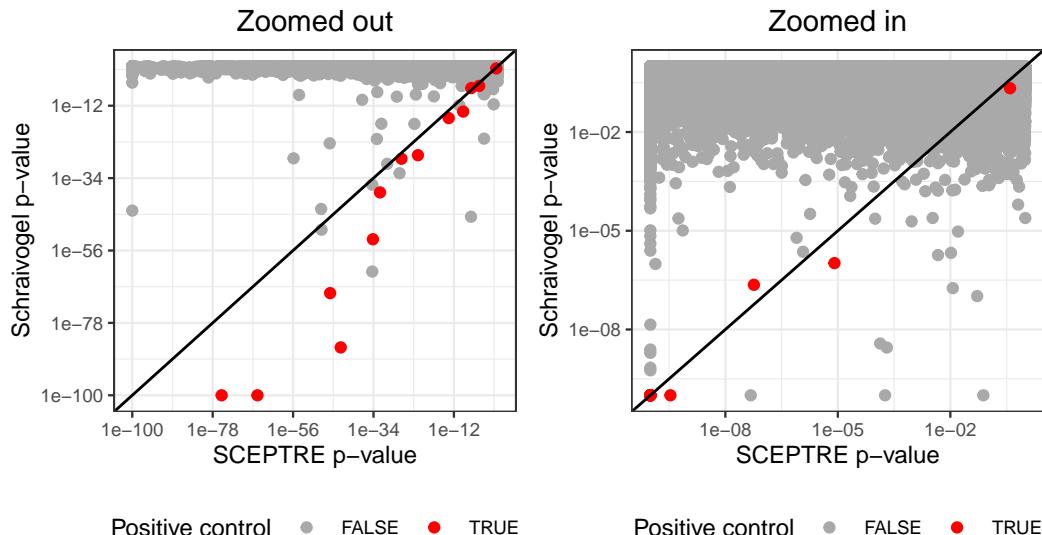
I applied SCEPTRE (based on the new `sceptre` package) to the Schraivogel discovery pairs from chromosome 8, and collated the results with those of Schraivogel. The result is the following data frame:

`results_merged`

```
## # A tibble: 68,289 x 7
##   response_id grna_group      p_val~1 logFC~2 p_val~3 logFC_~4 Posit~5
##   <chr>        <chr>          <dbl>     <dbl>     <dbl>     <dbl>    <lgl>
## 1 AZIN1        chr8:129076474-129076835 0 -0.628  0.189  1.97e-4 FALSE
## 2 TATDN1       chr8:101450301-101450732 0  2.76   0.417  2.56e-5 FALSE
## 3 TATDN1       chr8:101773933-101774280 0  3.13   0.936  2.21e-6 FALSE
## 4 TATDN1       chr8:101920208-101920392 0  2.34   0.675  -4.62e-5 FALSE
## 5 TATDN1       chr8:101997946-101999135 0  2.11   0.603  -6.27e-6 FALSE
## 6 TATDN1       chr8:102121715-102122062 0  1.99   0.0877 4.06e-4 FALSE
## 7 TATDN1       chr8:102137163-102137307 0  2.37   0.711  2.37e-5 FALSE
## 8 TATDN1       chr8:102345144-102345514 0  3.38   0.545  9.16e-5 FALSE
## 9 TATDN1       chr8:103404832-103405546 0  2.02   0.525  2.91e-4 FALSE
## 10 TATDN1      chr8:103409162-103409492 0  3.16   0.915  1.89e-4 FALSE
## # ... with 68,279 more rows, and abbreviated variable names 1: p_value_sceptre,
## #   2: logFC_sceptre, 3: p_value_schraivogel, 4: logFC_schraivogel,
## #   5: `Positive control`
```

A few things immediately seem strange. There appear to be many pairs for which the SCEPTRE p-value is zero (in fact there are 643 such pairs). In the meantime, the Schraivogel p-values for those pairs are not even significant. The signs and magnitudes of the log fold changes between the two methods do not line up very well either.

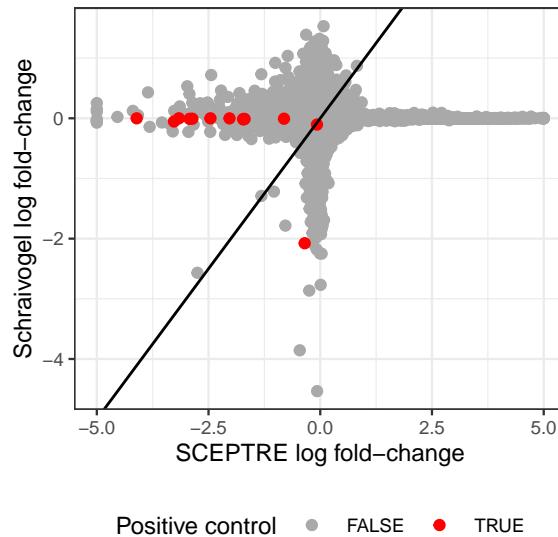
Let's plot the two sets of p-values against one another:



Note that the set of pairs considered includes not just the discovery pairs but a small number of positive control pairs, highlighted in red. We see that the p-values from the two methods are generally concordant for the positive control pairs, but apparently not concordant at all for the other pairs. This reminds me of a plot Kaishu made for the Papalex data, comparing SCEPTRE and Seurat p-values. There's a long tail of SCEPTRE p-values that are much smaller than their Schraivogel method counterparts.

Next, let's take a look at the effect sizes:

```
results_merged |>
  mutate(across(starts_with("logFC"), ~ pmin(pmax(., -5), 5))) |>
  ggplot(aes(x = logFC_sceptre, y = logFC_schraivogel, color = `Positive control`)) +
  geom_point() +
  geom_abline() +
  scale_colour_manual(values = c("darkgray", "red")) +
  labs(x = "SCEPTRE log fold-change",
       y = "Schraivogel log fold-change") +
  theme_bw(base_size = 9) +
  theme(legend.position = "bottom")
```



It appears the log fold-changes are on different scales. It would be good to figure out the source of this discrepancy. Aside from the scaling issue, there seems not to be a good agreement between the estimated log fold-changes of the two methods. The cross-shaped pattern in this plot also reminds me of a plot Kaishu made for the Papalex data, comparing SCEPTRE and Seurat log fold-changes.