Comparing SCEPTRE and Schraivogel method on Schraivogel data (updated)

Gene and Tim

2023-04-08

I applied SCEPTRE (based on the new `sceptre` package) to the Schraivogel discovery and positive control pairs from chromosome 8, and collated the results with those of Schraivogel:

```
## # A tibble: 68,289 x 7
##    response_id grna_group  p_value_sceptre logFC_sceptre p_value_schraivogel
##    <chr>       <chr>               <dbl>         <dbl>              <dbl>
##  1 RIPK2       RIPK2-TSS         4.27e-73        -2.73           9.41e-145
##  2 FAM83A      FAM83A-TSS        8.43e-65        -3.08           1.93e-167
##  3 LRRCC1      LRRCC1-TSS        1   e-50        -3.94           1.12e- 69
##  4 CPQ         CPQ-TSS           1.96e-43        -2.78           3.54e- 86
##  5 DSCC1       DSCC1-TSS         2.76e-34        -1.62           2.62e- 53
##  6 STK3        STK3-TSS          6.43e-33        -1.64           4.45e- 39
##  7 UBR5        UBR5-TSS          6.75e-27        -1.65           7.74e- 29
##  8 OXR1        OXR1-TSS          2.05e-22        -0.721          9.22e- 28
##  9 PHF20L1     PHF20L1-TSS       4.14e-14        -3.02           1.76e- 16
## 10 E2F5        LRRCC1-TSS        3.74e-11        -0.537          1.39e- 12
## # i 68,279 more rows
## # i 2 more variables: logFC_schraivogel <dbl>, `Positive control` <lgl>
```

A few things immediately seem strange. There appear to be many pairs for which the SCEPTRE p-value is zero (in fact there are 0 such pairs). In the meantime, the Schraivogel p-values for those pairs are not even significant. These pairs are not positive controls. The signs and magnitudes of the log fold changes between the two methods do not line up very well either.

SCEPTRE no longer returns p-values equal to zero; the smallest p-values are on the order of 1e-75 and correspond to positive controls.

Let's first take a look at the p-value distributions for both methods:
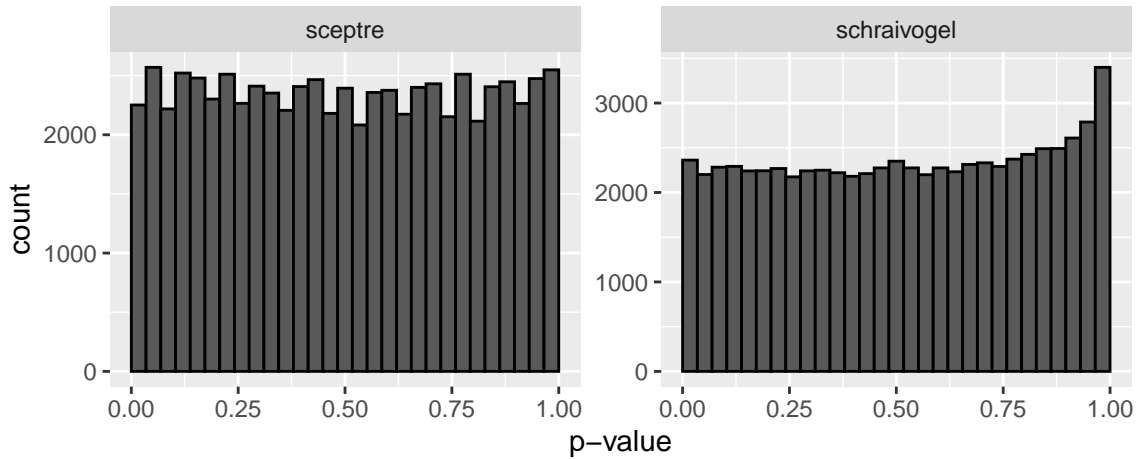
The SCEPTRE p-values have a massive spike near zero, unlike the Schraivogel p-values. Next, let's plot the two sets of p-values against one another:

There is no longer a massive spike near zero; the p-values are uniform, as would be expected in an enhancer-targeting screen.
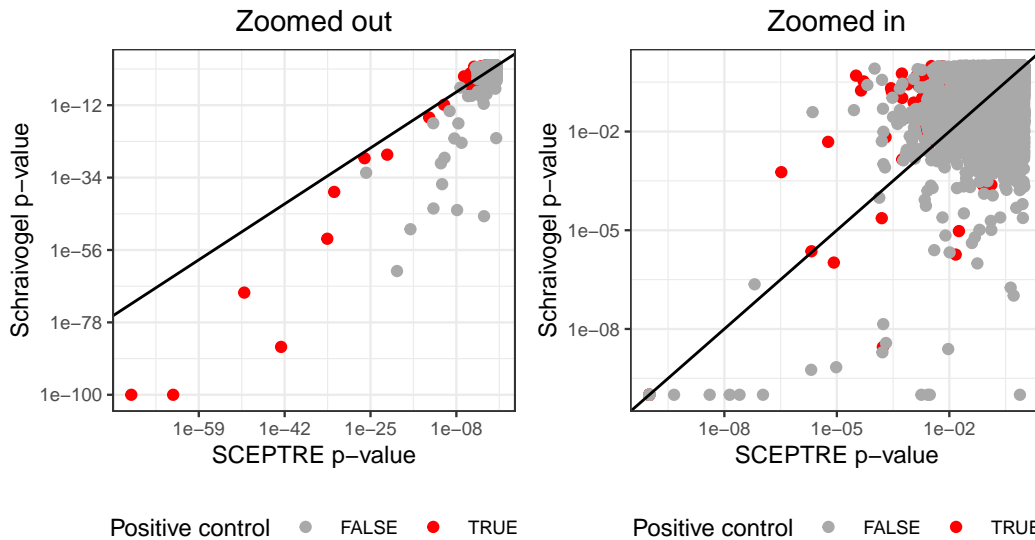
We see that the p-values from the two methods are generally concordant for the positive control pairs, but apparently not concordant at all for the other pairs. This reminds me of a plot Kaishu made for the Papalexi data, comparing SCEPTRE and Seurat p-values. There's a long tail of SCEPTRE p-values that are much smaller than their Schraivogel method counterparts.

No longer the case.

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

It appears the log fold-changes are on different scales. If 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 Papalexi data, comparing SCEPTRE and Seurat log fold-changes.

It is still not totally clear to me why there is a cross shape; I'd have to look into how Schraivogel Method computes log fold changes.