

Comparing effective sample sizes across negative control and discovery pairs

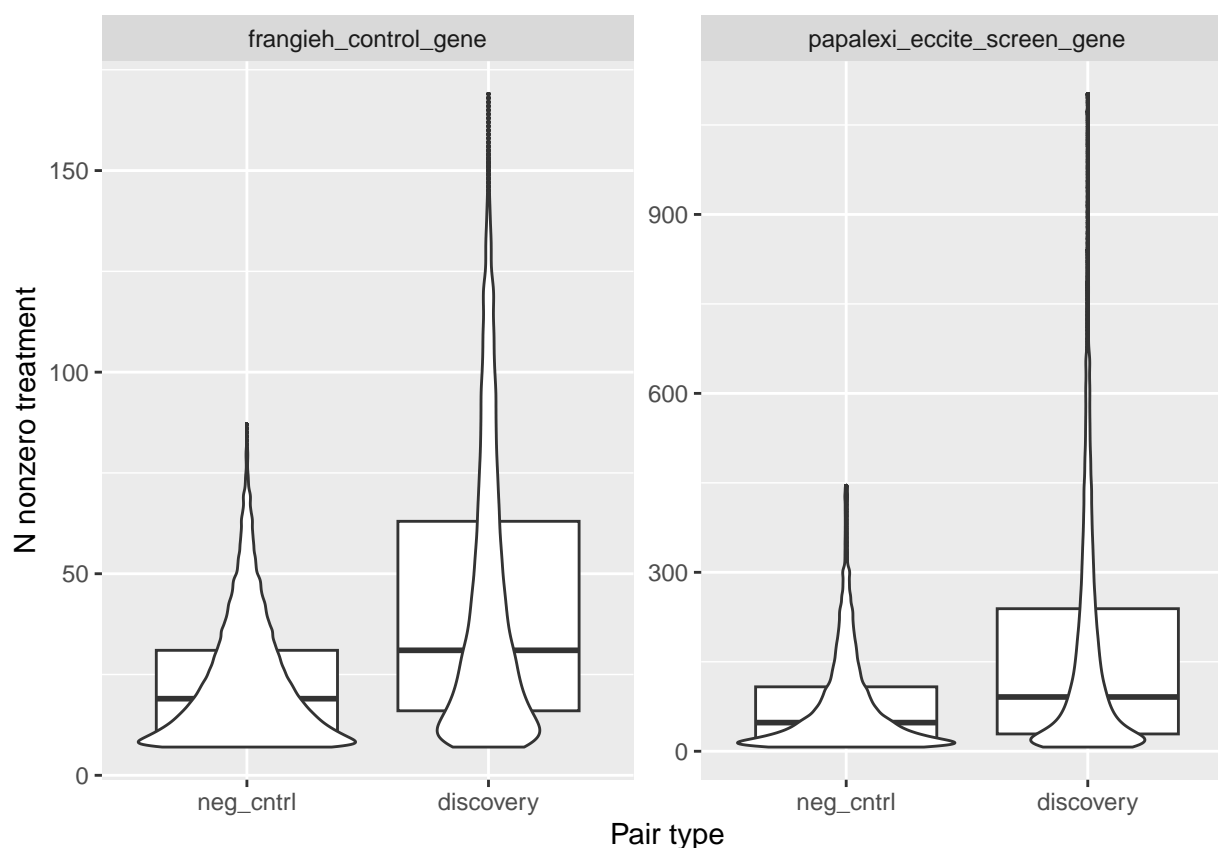
Tim

2023-04-24

R Markdown

In this short writeup I compare the number of nonzero treatment cells (per pair) across negative control and discovery sets.

Below are boxplots/violin plots of the number of nonzero treatment cells per pair, stratified by dataset (i.e., Frangieh control, Papalexi gene) and pair type (i.e., negative control, discovery). Visually, we see that the discovery pairs have a greater number of nonzero treatment cells.



I compute the mean and median number of nonzero treatment cells in each of these four categories.

```
## # A tibble: 4 x 4
## # Groups:   dataset [2]
##   dataset pair_type mean_n_nonzero_trt median_n_nonzero_trt
##   <fct>    <fct>          <dbl>              <dbl>
```

## 1	frangieh_control_gene	neg_cntrl	23.2	19
## 2	frangieh_control_gene	discovery	43.3	31
## 3	papalexi_eccite_screen_gene	neg_cntrl	78.9	48
## 4	papalexi_eccite_screen_gene	discovery	167.	91

We see that the discovery pairs have roughly twice as many nonzero treatment cells as the negative control pairs.