Comparing methods on Schraivogel discovery data
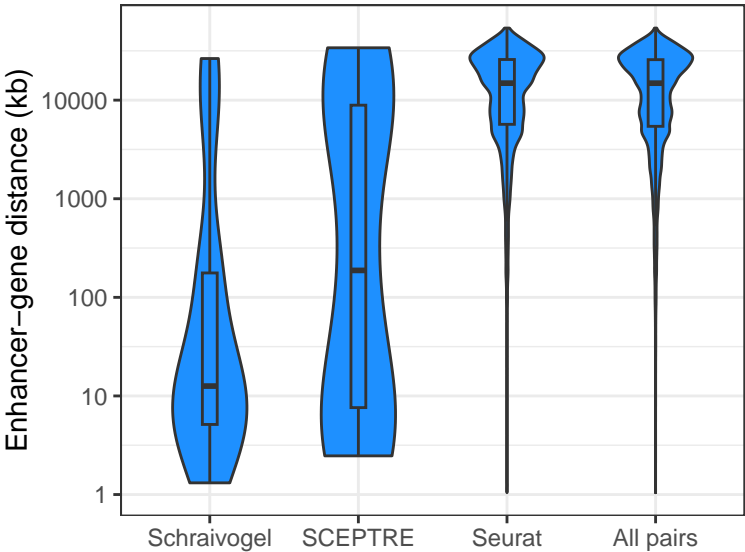
Gene

2023-04-18

In this writeup, I will compare the SCEPTRE (exact), Seurat DE, and the Schraivogel method (using 1000 randomly-selected cells from the complement set as the control group) on the Schraivogel discovery data from chromosome 8. I subject the p-values generated by each method to the BH correction at target level 0.1.

Numbers of discoveries

| Method	Number of discoveries
Schraivogel	31
SCEPTRE	21
Seurat	17005
All pairs	63864

SCEPTRE and Schraivogel make a few dozen discoveries each, whereas Seurat discovers about a quarter of all pairs tested.

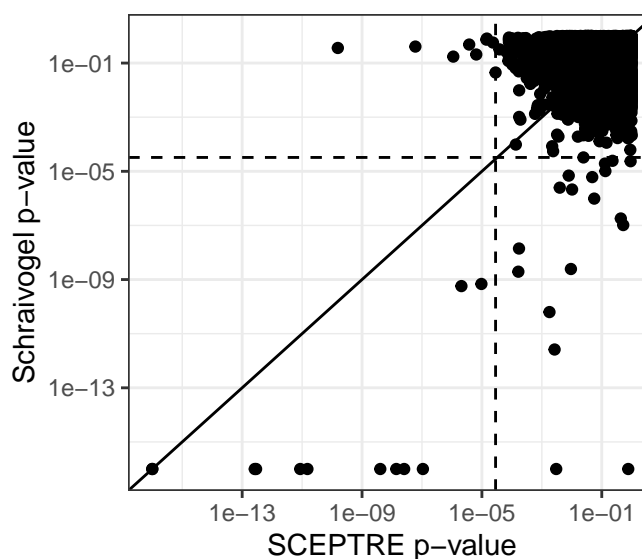## Enhancer-gene distance

Schraivogel has the smallest gene-enhancer distances, followed by SCEPTRE, followed distantly by Seurat.

HI-C interaction frequency
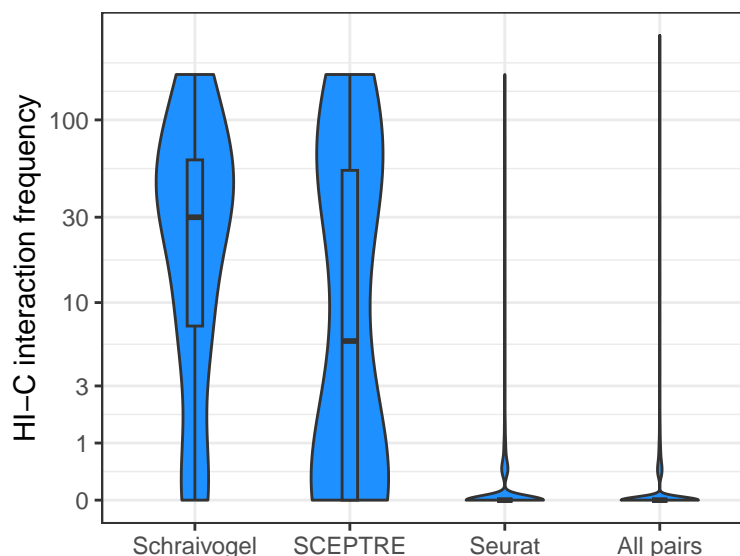
Schraivogel has the largest HI-C interaction frequencies, followed by SCEPTRE, followed distantly by Seurat.

Comparing SCEPTRE and Schraivogel p-values

The SCEPTRE and Schraivogel p-values are not particularly well aligned, although there is nontrivial overlap between the corresponding discovery sets.

Conclusion

The Schraivogel method is clearly the best. The SCEPTRE discoveries are clearly enriched for biological signal, but not as much so as Schraivogel's. The reason for Schraivogel's superior performance might be that it sidesteps the weird confounding issue by using the complement set rather than the NTC set as controls. Seurat is the worst, its discoveries not being enriched for biological signal. The reason for this is that it does not control for batch but uses the NTC set as controls, so it suffers from the extreme confounding issue.