

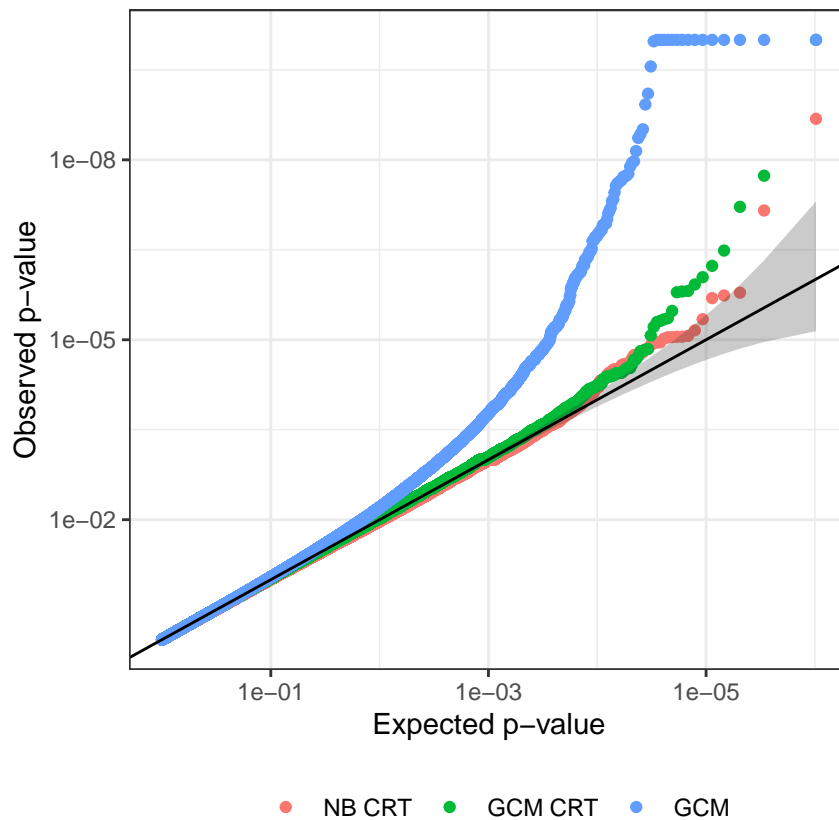
Benchmarking GCM-based variants of SCEPTRE

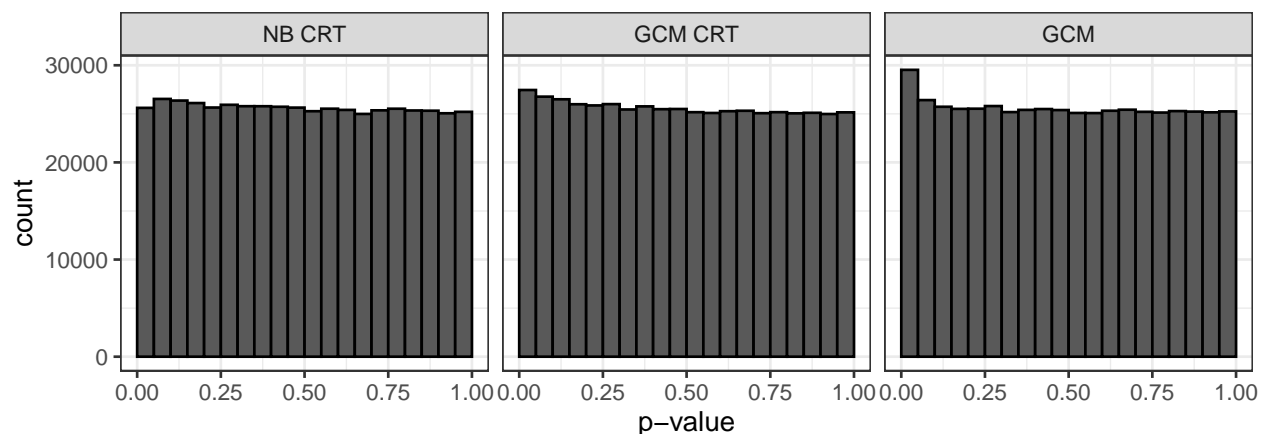
2022-08-02

In this writeup, we benchmark the performance of three SCEPTRE variants on the Gasperini negative and positive control data. The first is “NB CRT”, the original version. The second is “GCM CRT”, which is the CRT using the GCM test statistic based on Poisson regression of Y on Z. The third is “GCM”, which is the resampling-free GCM test.

Negative control results

Let’s look at the distributions of the NTC p-values for the three methods.

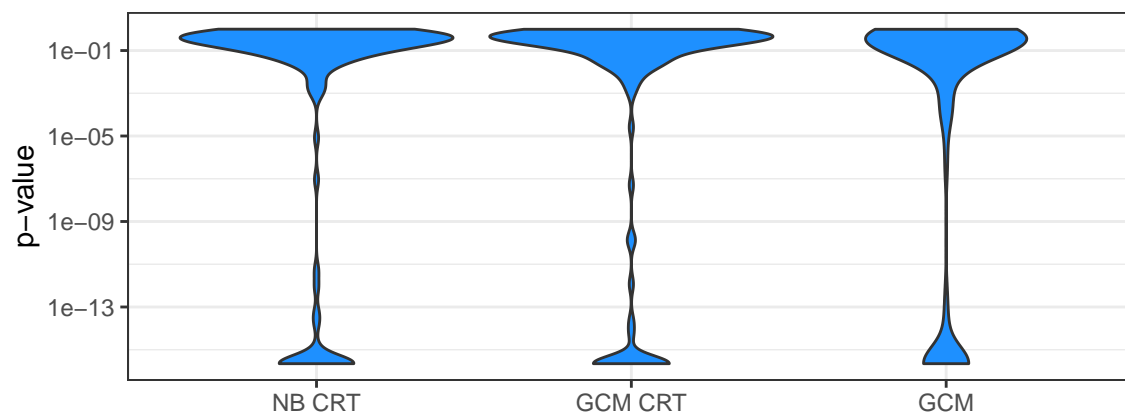




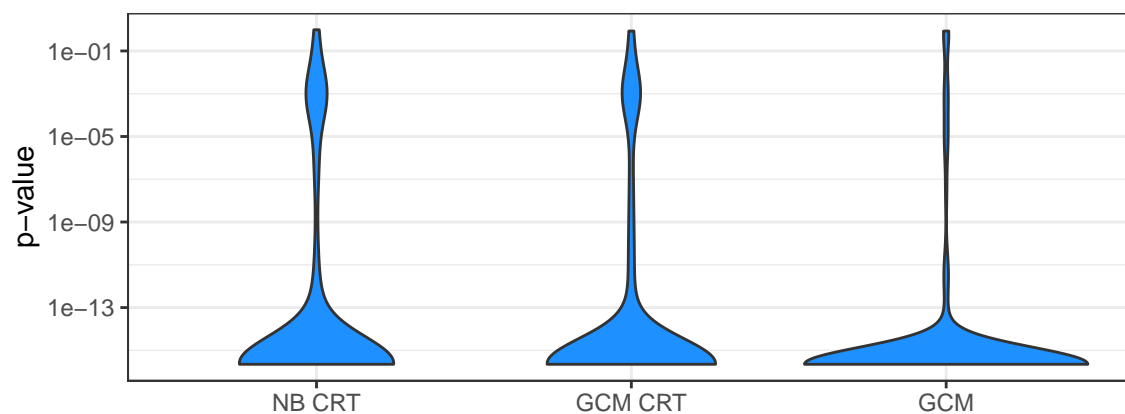
We see that the NB CRT has the best calibration, followed by GCM CRT, followed by GCM. The two CRT-based methods are markedly better than the asymptotic GCM method. It's likely that asymptopia has not kicked in yet. The slight miscalibration in the tails for the CRT-based methods is likely due to poor skew-t fit, induced by discreteness. This hypothesis remains to be verified.

Positive control results

Enhancer-targeting perturbations



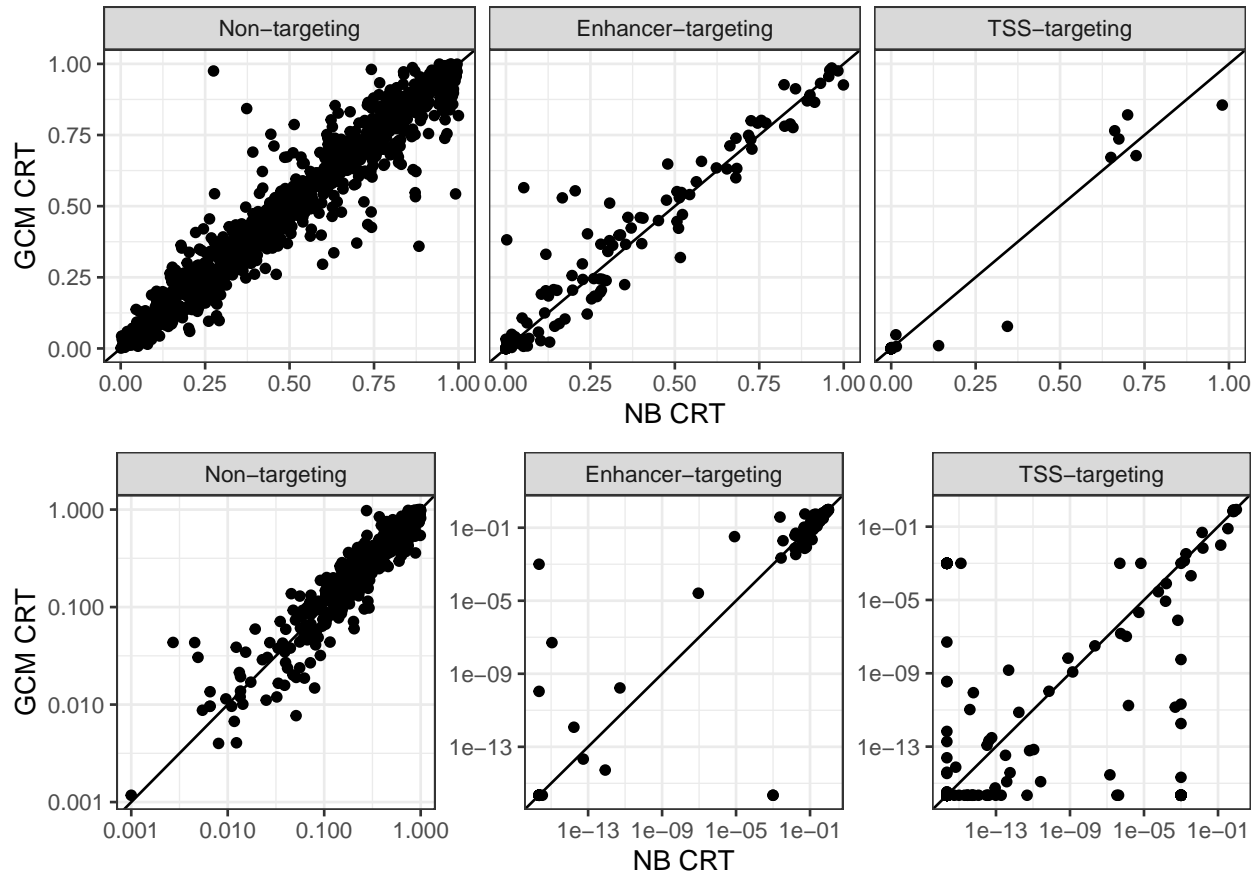
TSS-targeting perturbations



GCM has smaller p-values for positive control perturbations, while the two CRT-based methods perform similarly.

Comparing the two CRT-based methods

Let's see how the p-values from the two CRT-based methods compare with each other.



It looks like there's decent good agreement in the bulk of the distribution, but less good agreement as you get farther into the tail.