

# Comparison of Target Identification by gCrisprTools and MAGeCK in CRISPR screens with RRA-alpha

Russell Bainer

10/27/2016

## Overview

This document clarifies the differences between *MAGECK*<sup>1</sup> and the *gCrisprTools* pipeline that may be relevant to target detection in CRISPR screens. Overall, *MAGECK* and *gCrisprTools* perform similarly when selecting top candidates and disqualifying poor ones, and top-ranked targets identified with one approach are top-ranked by the other method. The differences observed between the default *MAGECK* and *gCrisprTools* pipelines principally involve targets for which a single associated gRNA contributes a large outlier signal; these targets score somewhat better in *MAGECK*'s signal aggregation framework relative to *gCrisprTools*, but are nevertheless not usually among the top candidates identified by either algorithm. Conversely, *gCrisprTools* more strongly prioritizes targets where weak signals are shared across all associated gRNAs, possibly in the context of elevated variance. These differences in target-level performance are largely due to differences between the gRNA-level *P*-values estimated by *MAGECK*, which uses a standard Negative Binomial test, and those estimated in *gCrisprTools*, typically by *voom/limma*.

## Algorithmic Details

*MAGECK* and *gCrisprTools* identify targets whose abundance changes across experimental conditions in a similar manner. First, each algorithm normalizes the observed read counts of all gRNAs in the experiment (via median scaling by default, although other options may be available), and then estimates the significance of the observed differences in each gRNA's abundance to generate one-tailed *P*-values. Both methods then aggregate these gRNA signals to the target level using Robust Rank Alpha Aggregation ( $\alpha$ -RRA), a modification of the RRA algorithm.<sup>2</sup>

Differences in gRNA *P*-value estimation form the core difference between *MAGECK* and *gCrisprTools*. *gCrisprTools* estimates gRNA-level *P*-values from the t-statistics derived from the relevant coefficient estimate of linear model fit in the *voom/limma* framework<sup>3</sup>, whereas *MAGECK* assigns significance based on a gRNA-specific Negative Binomial distribution. The *P*-values assigned by each of these methods follow a roughly monotonic relationship but differ in terms of the exact *P*-value estimates and gRNA rankings.

In the RRA algorithm, the normalized gRNA signal ranks (i.e., gRNA rank/N, where N is the number of gRNAs in the screen, rank is from 1 to N, and the best scoring guide is assigned rank 1) associated with each target are assigned a statistic,  $\rho$ , based on their distribution on the unit interval. Specifically, each gRNA is assigned a score by comparing its normalized rank to a Beta distribution, appropriately parameterized for the gRNA's rank position within the set of all gRNAs associated with the target. The smallest observed gRNA  $\rho$  associated with each target is reported as the target-level  $\rho$  score, which is assigned a final significance level via permutation of the full set of gRNA labels relative to nominal membership in guide sets. Intuitively, the  $\rho$  score quantifies the extent to which the ranks of the gRNAs associated with a target are skewed toward the bottom of the overall distribution. It is possible for gRNA ranks to deviate from a uniform distribution in ways that are not of interest in the context of a screening experiment, however, as in the case where gRNA signals happen to cluster in the middle of the overall distribution. To deal with this possibility, the modified

<sup>1</sup>Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. *MAGECK* enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 2014;15(12):554. PMID:25476604

<sup>2</sup>Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28(4):573-80. PMID:22247279

<sup>3</sup>Law CW, Chen Y, Shi W, Smyth GK. *voom*: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014 Feb 3;15(2):R29. doi: 10.1186/gb-2014-15-2-r29.

$\alpha$ -RRA version introduces a significance cutoff parameter ( $\alpha$ ) to mitigate this concern. The  $\alpha$ -RRA algorithm proceeds identically to RRA, but only assigns  $\rho$  statistics to the gRNAs whose signals are “significant” by an external criterion (i.e., below  $\alpha$ ) for the purposes of target-level scoring.

*gCrisprTools* introduces an additional refinement upon *MAGECK*’s  $\alpha$ -RRA algorithm. *MAGECK* uses gRNA  $P$ -values for both ranking and disqualification of gRNAs during rho statistic calculation, but by default *gCrisprTools* uses a ‘Combined Scoring’ approach, where gRNAs are (i) ranked by their by fold change estimates, but (ii) the associated  $P$ -values are only used to disqualify gRNAs during the  $\rho$  statistic calculation. If desired, this functionality may be suppressed by the user to functionally replicate *MAGECK*’s aggregation approach.

The differences between *gCrisprTools* and *MAGECK* can be summarized as follows:

1. Normalization may be handled differently by the two algorithms. By Default, *MAGECK* median-scales the libraries unless large numbers of gRNAs are not present in one or more of the samples and the median gRNA count is zero; in that case, it normalizes the libraries by equalizing the total read counts. *gCrisprTools* requires the user to explicitly normalize the data, and similarly recommends median scaling or total read count normalization if necessary.
2. *MAGECK* always considers all of the input gRNAs in its analysis but *gCrisprTools* removes gRNAs with low abundance in the control samples from consideration prior to analysis, as the abundances of these gRNAs are not expected to be estimated accurately. The user may suppress this behavior if they prefer, however.
3. gRNA  $P$ -values are calculated differently, as described above.
4. The two approaches differ in the manner in which individual gRNA signals are disqualified during signal aggregation, as described above. This is likely an unimportant difference because the inclusion or exclusion of marginal gRNA signals does not in practice have a strong effect on the  $\alpha$ -RRA results.
5. In the  $\alpha$ -RRA step, *MAGECK* uses the nominal gRNA  $P$ -values to rank the gRNAs and to check against  $\alpha$  to disqualify invariant gRNAs. By default, *gCrisprTools* ranks gRNAs on the basis of their fold change (unlike *MAGECK*), but disqualifies gRNAs on the basis of an  $\alpha$  cutoff on their one-sided  $P$  (similar to *MAGECK*).
6. When none of the signals associated with a target surpass the  $\alpha$  cutoff, *MAGECK* assigns the target a  $\rho$  score on the basis of the lowest normalized rank observed among the associated gRNAs (i.e., it allows one gRNA past the cutoff even though the  $P$ -value associated with that gRNA did not meet the  $\alpha$  criterion). *gCrisprTools* assigns such targets a  $P$ -value of 1.

## Methods

The following analyses were performed using a two-condition contrast comparing gRNA abundances estimated from a set of early-timepoint reference samples and a paired set of untreated late-timepoint expansion samples. This comparison was selected because it contains meaningful signals in the context of minimal selective pressure and library distortion, and consequently should be well suited for analysis by the *MAGECK* algorithm. In this experiment, triplicate Cas9-expressing cell cultures were infected with a lentiviral library containing 16902 distinct cassettes expressing gRNAs that target 2128 transcribed loci (‘targets’). After sequencing, reads were trimmed and aligned to the library annotation by exact sequence matching to produce a count matrix of gRNA cassette observations for each sample. Low abundance gRNAs were discarded from this matrix, and then sample medians were equalized and library sizes scaled so that all samples are prenormalized and contain equivalent numbers of gRNA counts. These count data were then processed by *gCrisprTools* in a manner mimicking *MAGECK*’s default behavior (i.e., using only gRNA-level  $P$ -values for ranking and  $\alpha$  disqualification), or by *MAGECK* 0.5.3 using the following command:

```
mageck test -k ControlSamp_Count_data_norm.txt -t 4,5,3 -c 1,2,0 -n normalized -norm-method none -variance-from-all-samples
```

## gRNA P-values (One Sided)

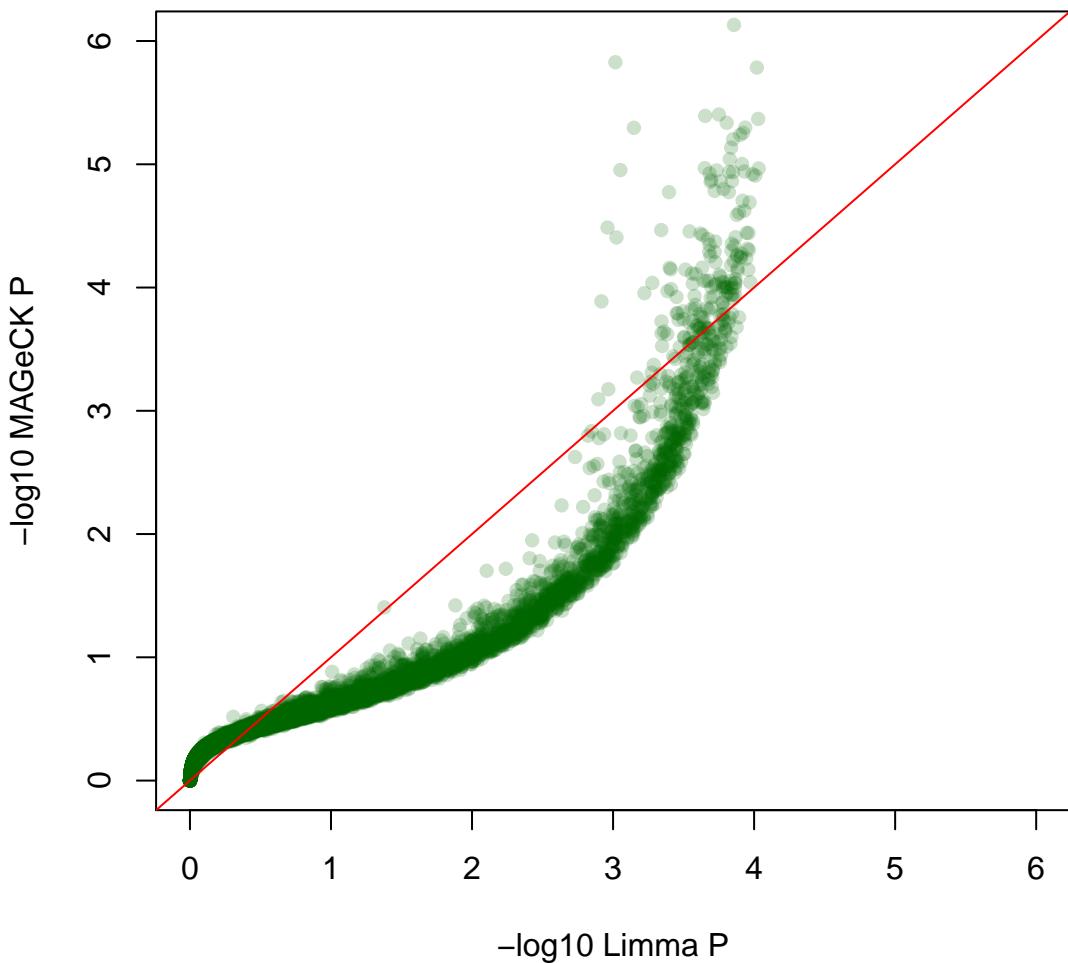


Figure 1: Comparison of gRNA P-values estimated by MAGeCK and voom/limma.

### MAGeCK Estimates gRNA P-values Differently

The differences in target prioritization between *MAGECK* and *gCrisprTools* are primarily driven by differences in the gRNA *P*-value estimation step. Generally speaking, the standard Negative Binomial model employed by *MAGECK* produces relatively smaller *P*-values for gRNAs with very large or very small effects, while *voom/limma* assigns more significance to gRNA signals in the intermediate range (Figure 1).

This is a consequence of the manner in which each algorithm calculates one-sided statistics, as described above. This difference has two consequences that directly impact signal aggregation by  $\alpha$ -RRA. The first is that some strong gRNA signals are assigned more significance by *MAGECK*; these gRNAs typically have unusually low abundance and are given small precision weights by *voom/limma* (e.g., have means that are not expected to be accurately estimated; Figure 2, red dots).

Intuitively, *voom/limma* downweights gRNAs that deplete to abundances at which the mean is not likely to be accurately estimated (Figure 2, left panel), whereas *MAGECK* assigns a *P*-value strictly on the basis of the mean gRNA counts in the test condition relative to the gRNA's null distribution estimate. In practice this means that the estimated effect size plays a larger role in *MAGECK*'s *P*-value estimates relative to those of *voom/limma* such that gRNAs with large effect estimates more consistently have small *P*-values in

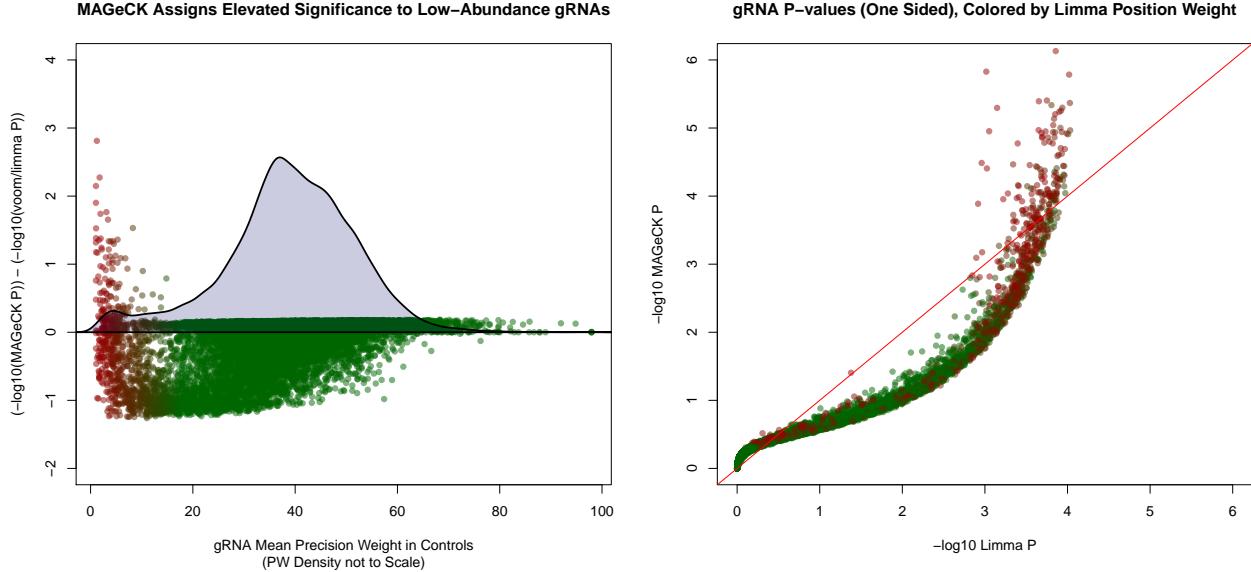


Figure 2: Abundance changes in gRNAs with low counts are assigned higher significance by MAGECK relative to voom/limma. gRNAs assigned low precision weights by voom, indicating low numbers of counts, are drawn in red.

MAGECK's framework (Figure 3).

### MAGECK and *gCrisprTools* Prioritize the Same Targets

*gCrisprTools* allows the user to specify an *MArrayLM* object containing comprehensive gRNA model estimates, and so to compare MAGECK's target ranking to that of *gCrisprTools*, we ran both algorithms with the gRNA *P*-value estimates and  $\alpha$  settings defined by the MAGECK algorithm. We note that MAGECK truncates gRNA *P*-values from their full precision before they are output to the user, and so in this scenario both algorithms disqualify the same gRNA but the gRNA *P*-value ranking available to *gCrisprTools* is presumably not strictly identical to the one used internally by MAGECK. The algorithms also still differ in their handling of cases where all gRNAs associated with a gene are disqualified (see above).

At the target level, *gCrisprTools* largely returns similar *P*-values in this context to those estimated by MAGECK (Figure 4). The target-level *P*-values that MAGECK assigns are systematically smaller than those returned by *gCrisprTools* when MAGECK-style gRNA estimates are supplied, but the high-priority target rankings are similar.

As shown below, there is a set of targets that both methods identify as equivalently optimal hits (Figure 5, left panel, dot in the lower left corner corresponding to lowest-ranked targets), and high-priority targets identified in one method are always among the best candidates of the other. Notably, MAGECK identifies a number of targets with weak to moderate signals that *gCrisprTools* eliminates from consideration completely (Figure 5, vertical stripe in right panel). This only applies to targets that are not prioritized by either algorithm, however, and is due to differences in the treatment of targets with no associated gRNAs passing the  $\alpha$  threshold (see above).

As mentioned above, *gCrisprTools* uses combined target scoring by default. When applied to the gRNA data generated by MAGECK, the consistency of the methods improves and the systematic bias in target *P*-value point estimates disappears (Figure 6). This is likely because fold change is a major contributor to MAGECK's gRNA *P*-value estimates, and including this information somewhat improves *gCrisprTools*' inference of MAGECK's internal ranking of gRNA significance estimates.

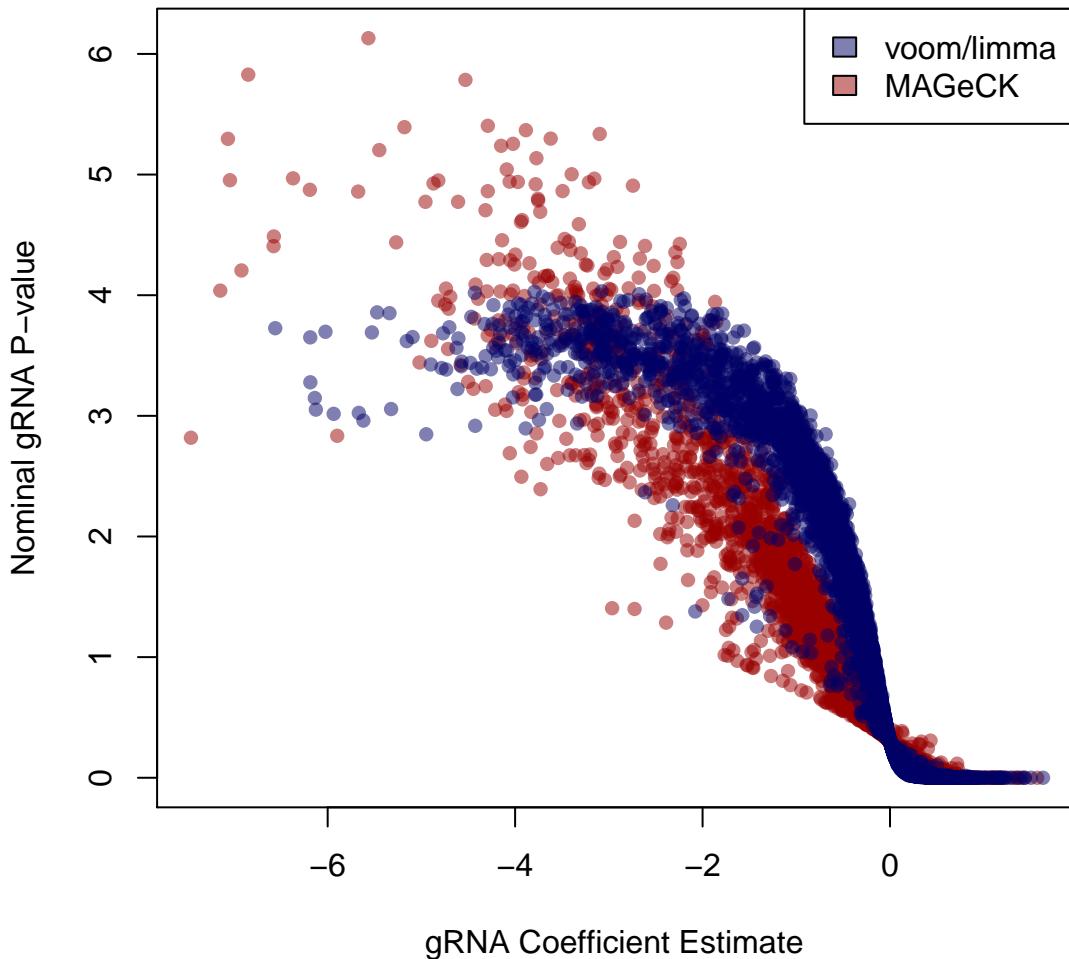


Figure 3: Fold changes in abundance (x-axis) are more closely associated with significance by MAGeCK relative to voom/limma.

## Target Depletion P–values with MAGeCK and gCrisprTools

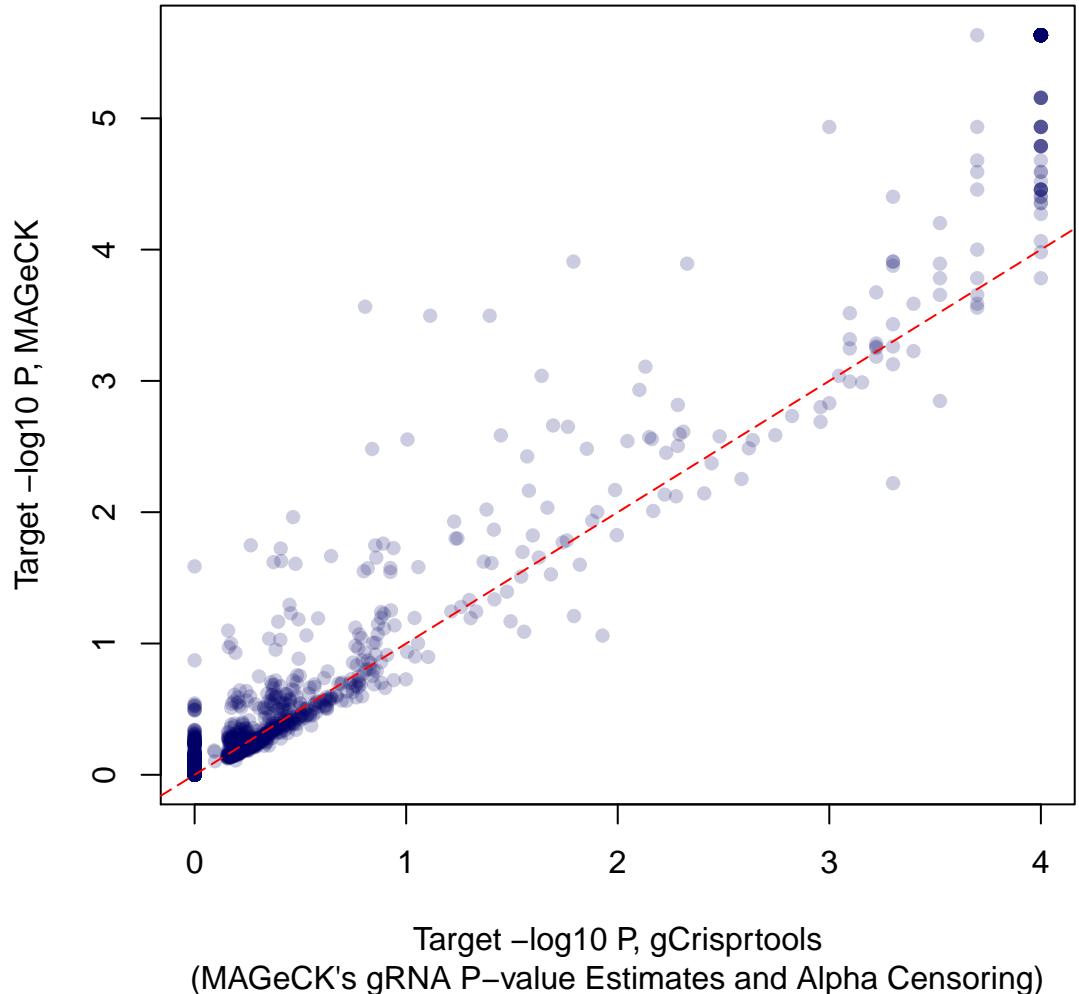


Figure 4: MAGeCK and gCrisprTools similarly estimate the significance of target-level depletion.

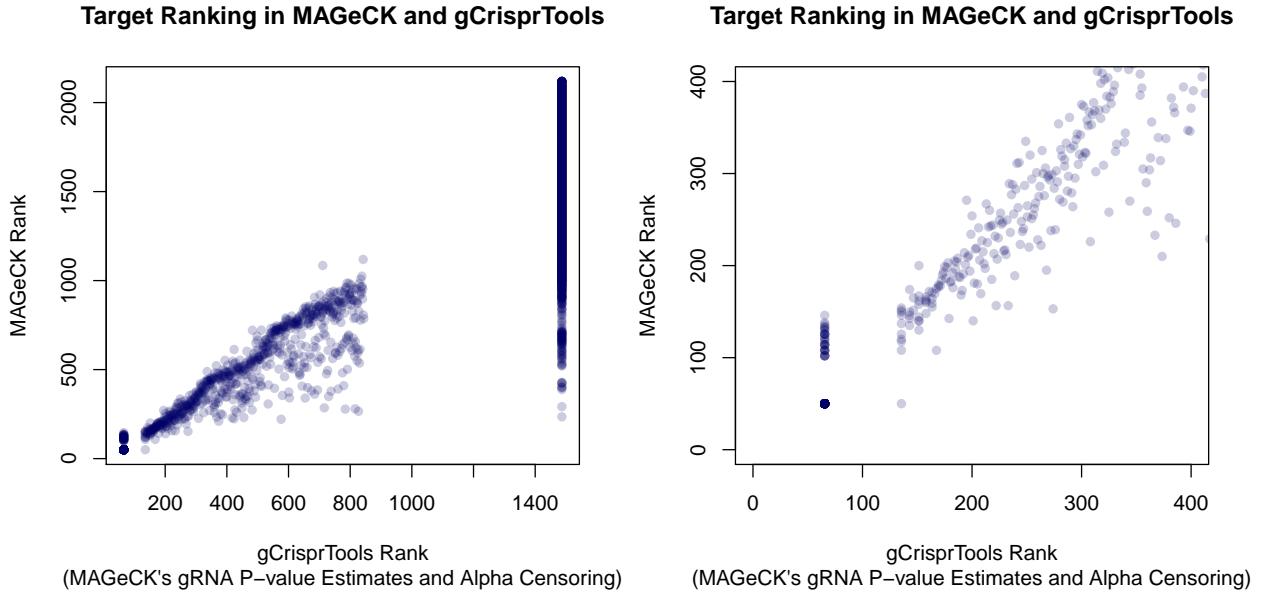


Figure 5: Similar ranking of targets by MAGeCK and gCrisprTools. The panels differ in the scale of the axes.

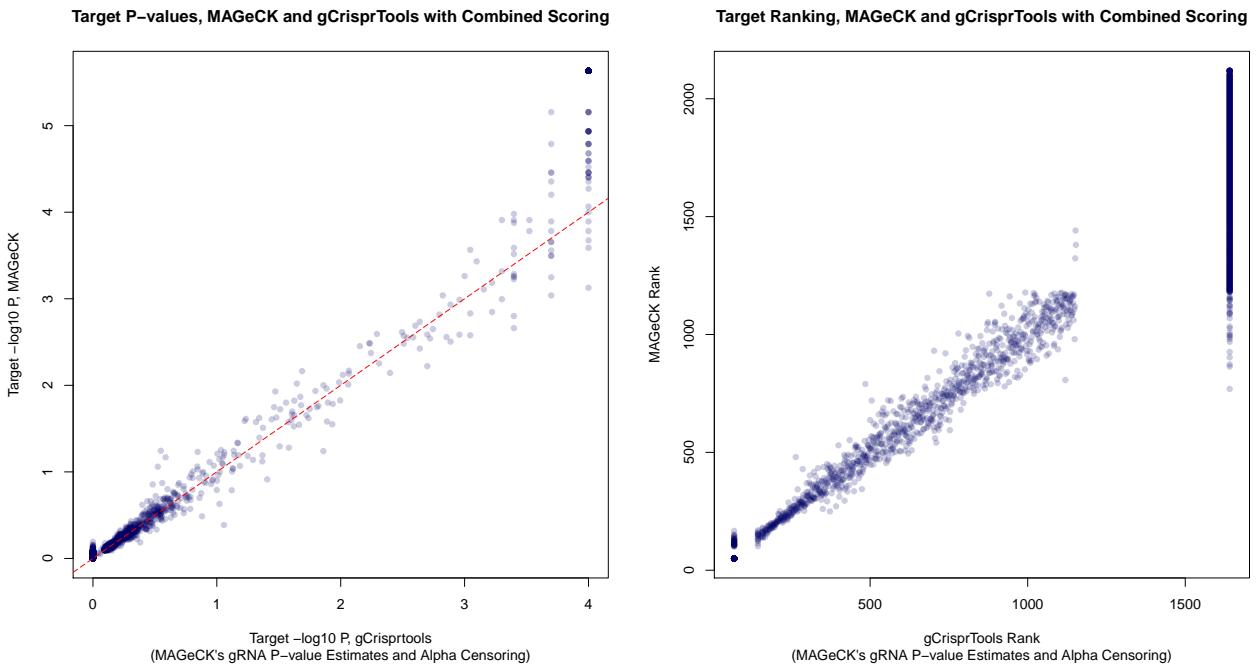


Figure 6: Combined scoring improves the consistency of gCrisprTools and MAGeCK target-level significance estimates.

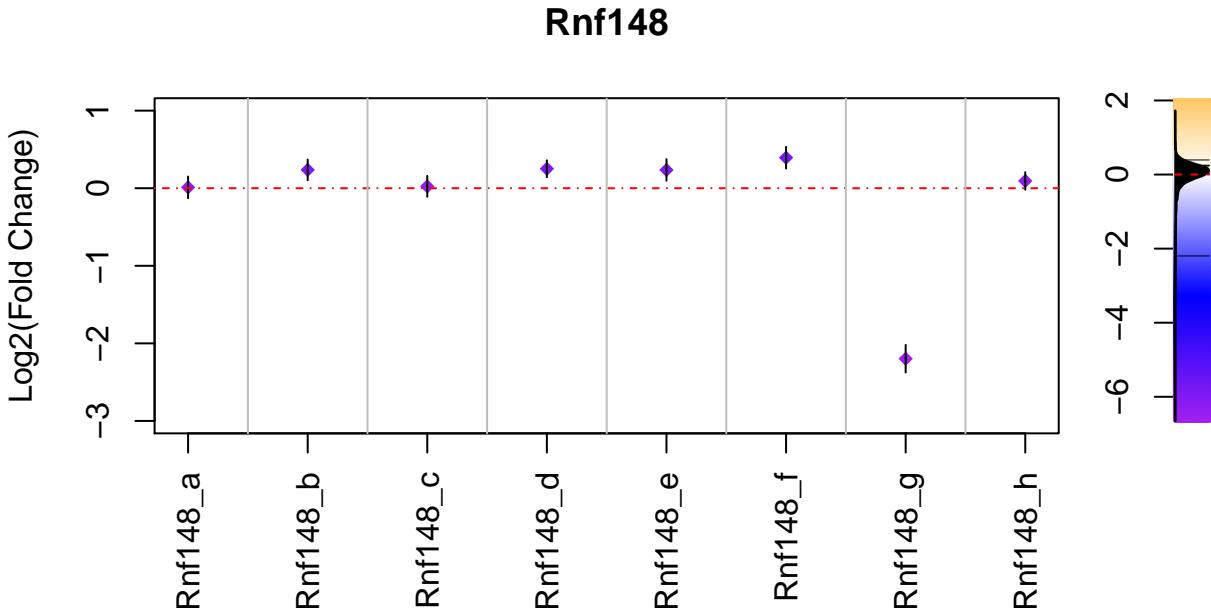


Figure 7: Example of a target whose depletion is assigned greater significance with MAGECK’s P-value scoring.

### Practical Consequences

In practice, *MAGECK* and *gCrisprTools* largely identify similar targets, but the subtle differences described in this document imply that certain patterns of gRNA behavior are favored by *MAGECK* more than *gCrisprTools* and vice versa. It is worth noting that these differences largely manifest within the set of targets that are not highly prioritized by either scoring method, and the following examples represent weak effects that are not likely to influence the prioritization of the top candidates.

When run using identical gRNA-level statistics (that is, the  $P$ -values,  $\alpha$  cutoffs, and fold changes estimated by *MAGECK*), *MAGECK*’s target-level  $P$ -value scoring tends to prefer targets for which one of the gRNA signals is particularly strong (Figure 7).

Conversely, the combined scoring approach somewhat upweights targets associated with weaker but more consistent gRNA signals that might have elevated variances (Figure 8).

This effect can be visualized by performing a one-sided Grubbs’ test on the set of coefficient estimates associated with each target. In this context, Grubbs’ test calculates a  $P$ -value quantifying the evidence that the most extreme coefficient estimate among the gRNAs associated with each target is an outlier relative to the others on the basis of the mean and variance of the remaining gRNAs’ observed effect sizes. Focusing on the targets whose significance estimates ( $P$ -values) change by more than 0.2 on the log<sub>10</sub> scale between the two scoring methods, it is clear that outliers are disproportionately present among the targets prioritized by *MAGECK*’s  $P$ -value scoring (Figure 9).

### Conclusion

In the direct comparison provided above, *MAGECK* and *gCrisprTools* largely identify the same high-priority targets. The major differences between the algorithms are driven by their respective methods for calculating gRNA-level  $P$ -values. *gCrisprTools* provides the *MAGECK*  $P$ -value scoring method for generating target-level statistics, as well as a combined scoring approach that may have modest advantages over *MAGECK*’s default scoring methods by limiting the influence of outliers on target prioritization.

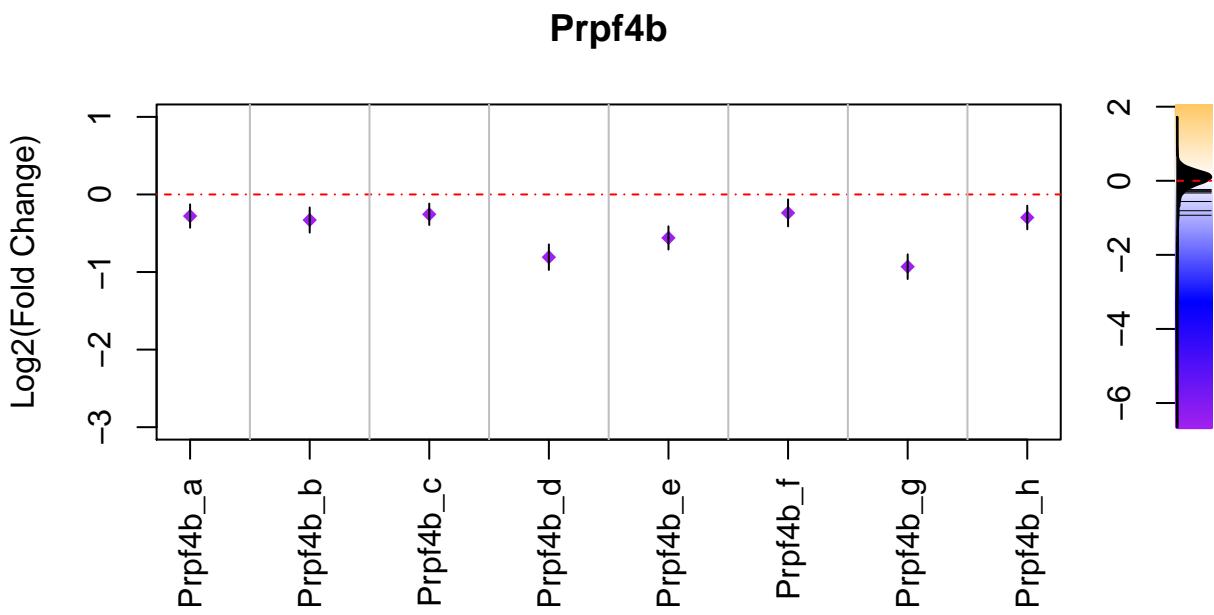


Figure 8: Example of a target whose depletion is assigned greater significance with gCrisprTools' Combined scoring.

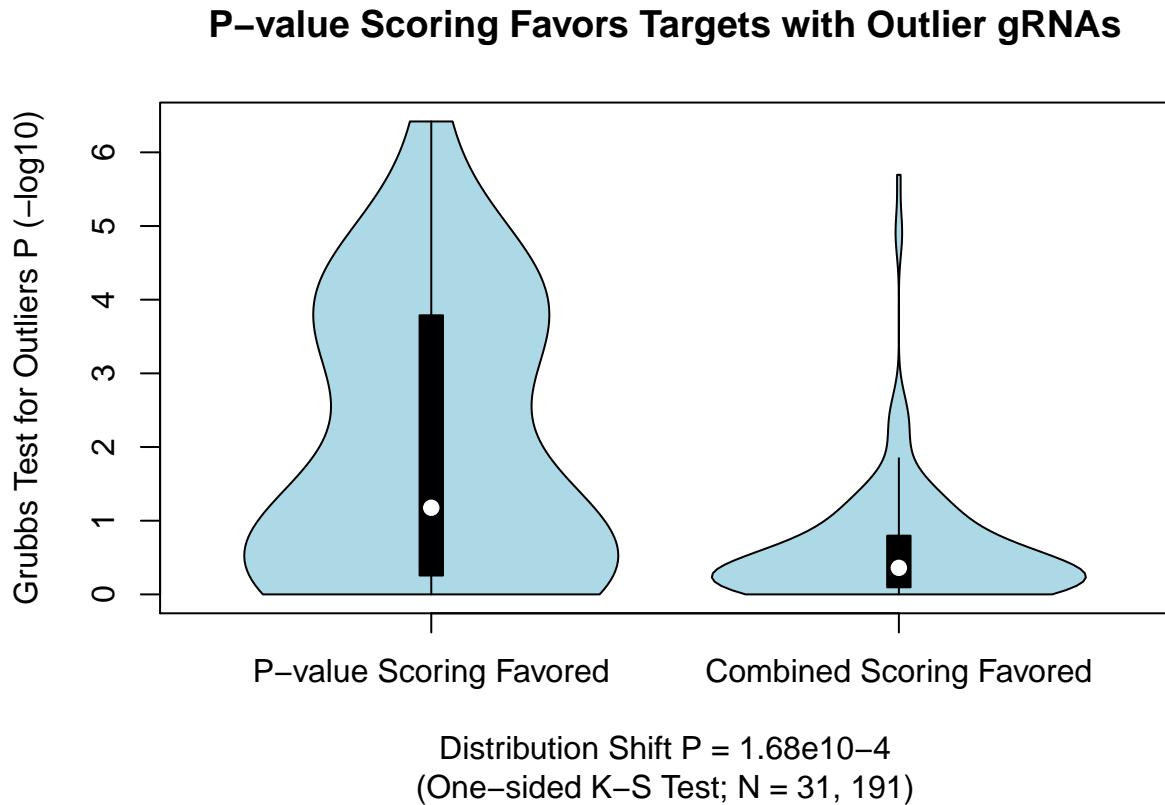


Figure 9: P-value scoring increases detection of targets with outlier signals.