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

Russell Bainer

10/27/2016

Overview

This document clarifies the differences between *MAGeCK*[1] 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 differences between the gRNA-level *P*-values estimated by *MAGeCK*, which uses a standard Negative Binomial test, and those used by *gCrisprTools*, which are typically estimated by *voom/limma*.

When the algorithms are run using identical gRNA statistics the concordance between the algorithms improves significantly. In that scenario, *gCrisprTools*' signal aggregation framework more strongly prioritizes targets where somewhat weaker signals are shared across all associated gRNAs, possibly in the context of elevated variance. In contrast, *MAGeCK* assigns somewhat better scores to a small number of targets for which a single associated gRNA contributes a large outlier. These differences are subtle, however, and typically manifest among targets that are not prioritized by either algorithm.

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 Alpha Robust Rank Aggregation (α-RRA), a modification of the RRA algorithm.[2]

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 object, typically generated in the *voom/limma* framework[3], 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, ρ, 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 ρ associated with each target is reported as the target-level ρ 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

[1] 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

[2] 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

[3] 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.

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

gCrisprTools introduces an additional refinement upon *MAGeCK*'s α-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 ρ 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 α-RRA results.
5. In the α-RRA step, *MAGeCK* uses the nominal gRNA *P*-values to rank the gRNAs and to check against α 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 of an α cutoff on their one-sided P (similar to *MAGeCK*).
6. When none of the signals associated with a target surpass the α cutoff, *MAGeCK* assigns the target a ρ 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 α 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

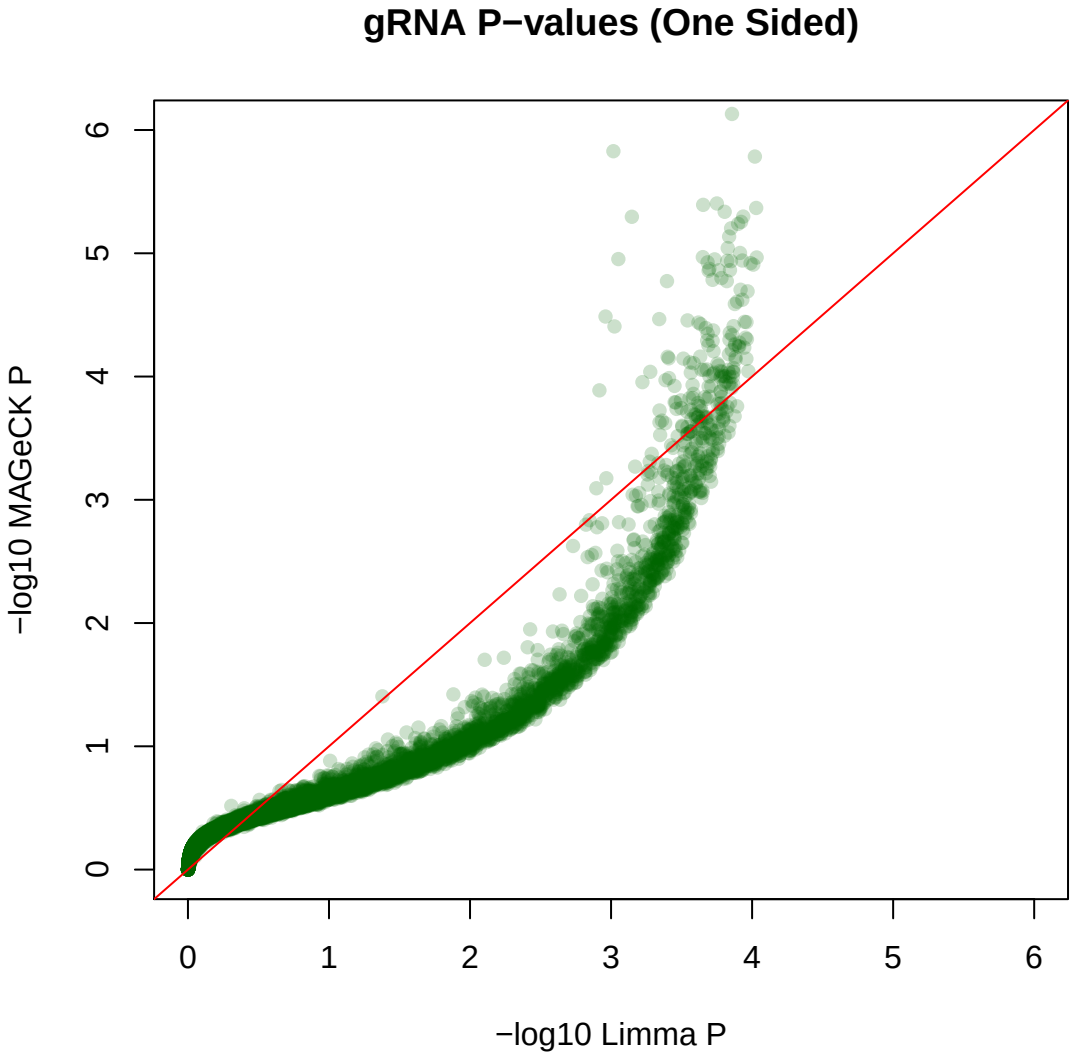

Figure 1: Comparison of gRNA P-values estimated by MAGeCK and voom/limma.

a manner mimicking *MAGeCK*'s default behavior (i.e., using only gRNA-level P-values for ranking and α 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

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 α-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).

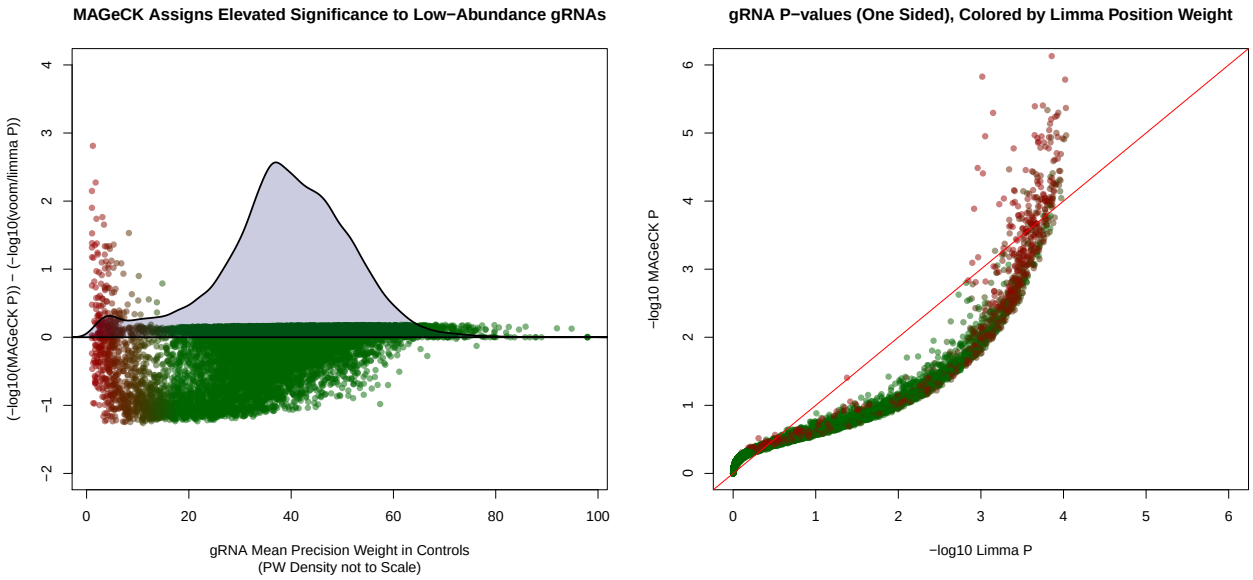

Figure 2: Abundance changes in gRNAs with low counts are assigned higher significance by MAGeCK relative to voom/limma. gRNAs assigned low precision weights by voom, indicating low numbers of counts, are drawn in red.

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 *MAGeCK*'s framework (Figure 3).

MAGeCK and *gCrisprTools* Prioritize the Same Targets

At the target level, *gCrisprTools* or *MAGeCK* largely return similar *P*-values when they are run fully independently (i.e., using the respective methods' preferred gRNA significance estimates and rankings, and using their respective implementations).

As shown below, there is a set of targets that both methods identify as equivalently optimal hits (Figure 4, right panel, dot in the lower left corner corresponding to best-ranked targets), and high-priority targets identified in one method are always among the best candidates of the other. *MAGeCK* does not prioritize any targets that *gCrisprTools* eliminates from consideration completely (Figure 4, vertical stripe in right panel). This stripe is due to differences in the treatment of targets with no associated gRNAs passing the α threshold (see above), and as expected this difference affects targets that are not highly prioritized by either algorithm.

Although the algorithms are largely consistent with each other, there are nevertheless a handful of targets that are assigned discordant scores by the two pipelines. *gCrisprTools* allows the user to specify an *MArrayLM* object containing comprehensive gRNA model estimates, and so to clarify the nature of these minor disagreements we ran both algorithms with the gRNA *P*-value estimates, rankings, and α settings defined by the *MAGeCK* algorithm.

As anticipated, standardizing the gRNA-level statistics across both algorithms substantially improves the concordance of the resulting target estimates (Figure 5). 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 gRNAs but the 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

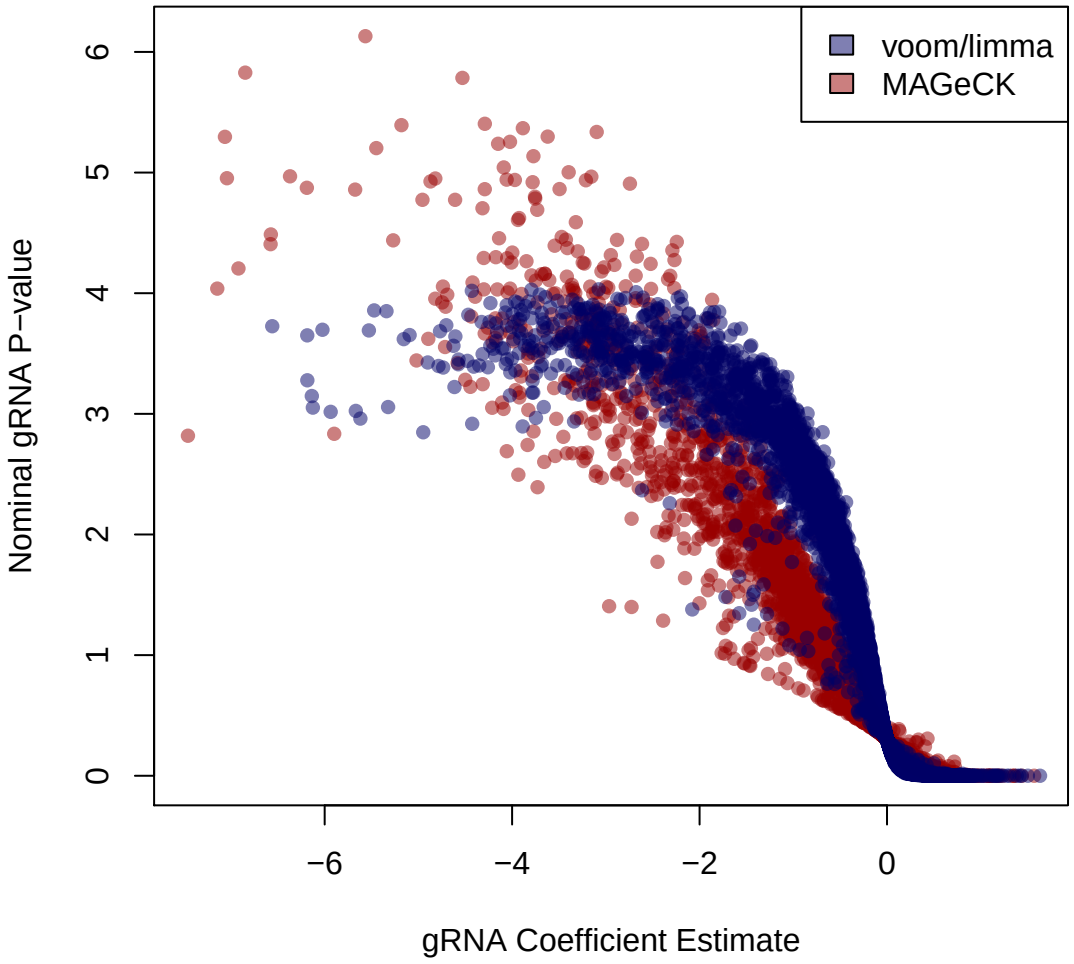

Figure 3: Fold changes in abundance (x-axis) are more closely associated with significance by MAGeCK relative to voom/limma.

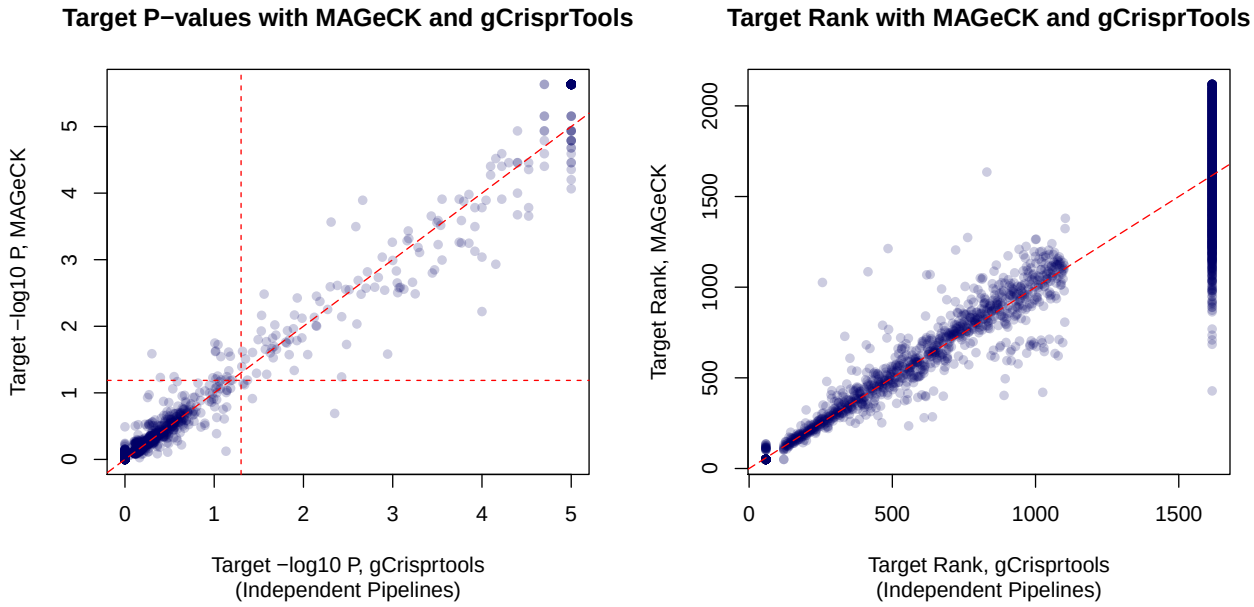

Figure 4: MAGeCK and gCrisprTools similarly estimate the significance of target-level depletion. In the left panel, dotted lines indicate an FDR cutoff of 0.5 for each algorithm.

with a gene are disqualified (see above), but targets in this category are deprioritized by both methods.

As mentioned above, *gCrisprTools* uses combined target scoring by default. When applied to the gRNA data generated by *MAGeCK* (i.e., comparing identical gRNA P-values and effect estimates but not ranking methods), the scoring methods broadly agree both in terms of the respective target rankings and P-value estimates (Figure 6). Notably, a handful of targets assigned weak significance by *MAGeCK*'s P-value scoring method do not achieve nominal significance by *gCrisprTools*' Combined approach, and vice versa (see below).

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*'s P-value scoring more than *gCrisprTools*' combined approach, and vice versa. It is important to note 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, α cutoffs, and fold changes estimated by *MAGeCK*), the handful of targets substantially preferred by *MAGeCK*'s target-level P-value scoring relative to *gCrisprTools*' Combined scoring method typically are associated with a single particularly strong gRNA signal (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 small number of targets whose significance estimates (P-values) change by more than 0.4 on the log10 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).

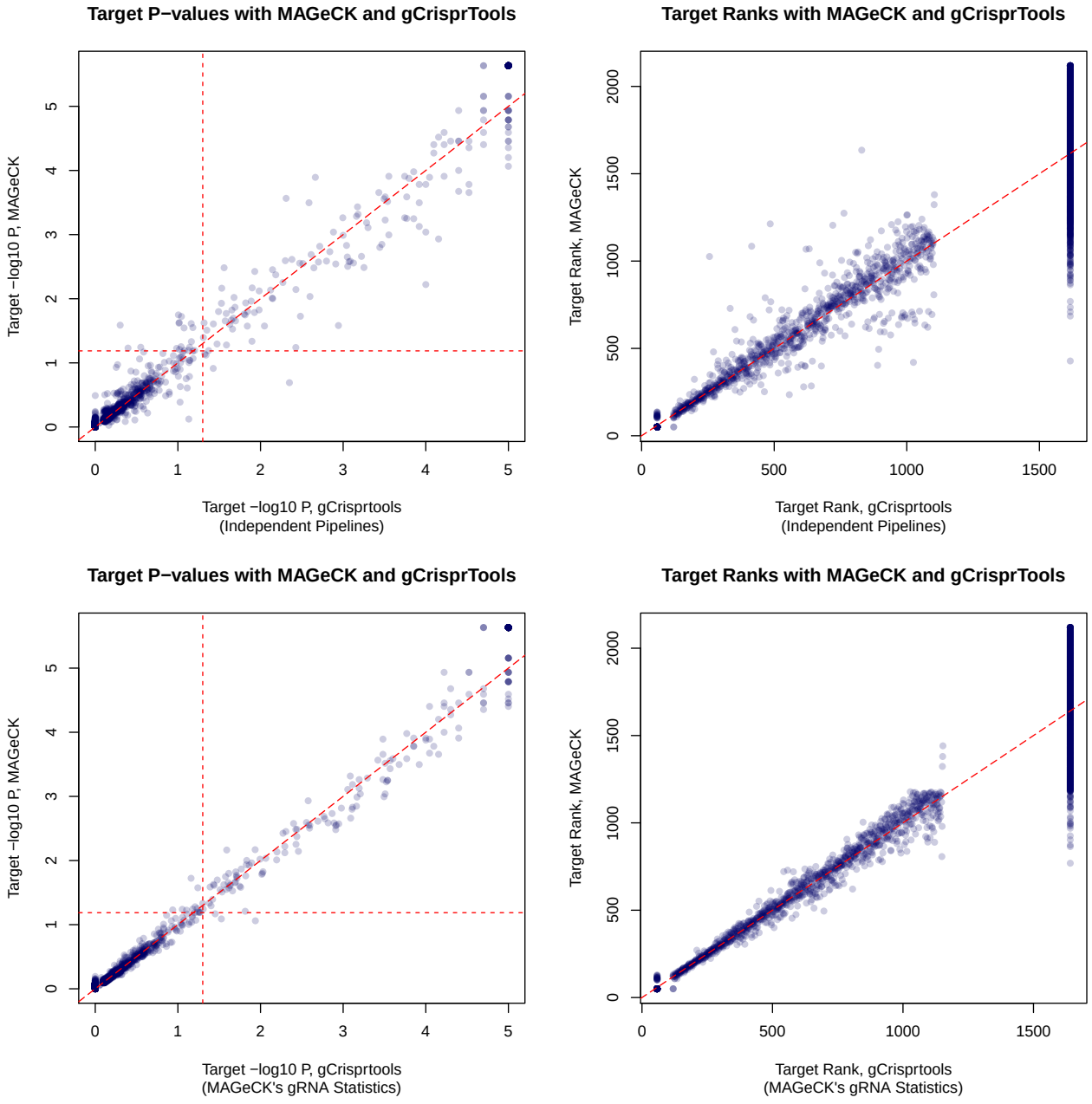

Figure 5: The consistency of target P-value estimates by MAGeCK and gCrisprTools is improved when both algorithms are run using identical gRNA-level statistics (see text for details). In the left panels, dotted lines indicate an FDR cutoff of 0.5 for each algorithm.

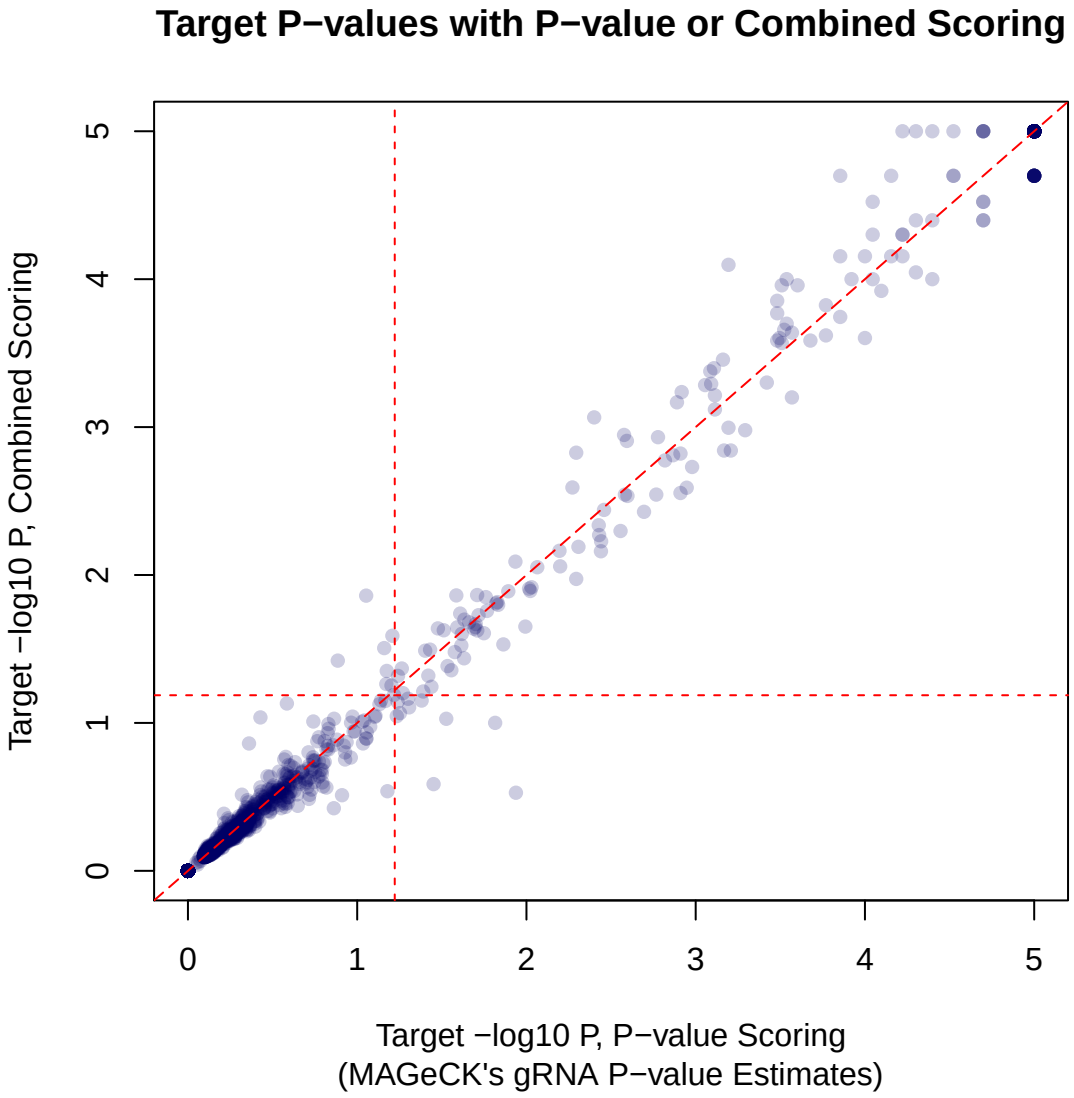

Figure 6: Combined Scoring substantially influences the target-level significance estimates of a small number of targets.

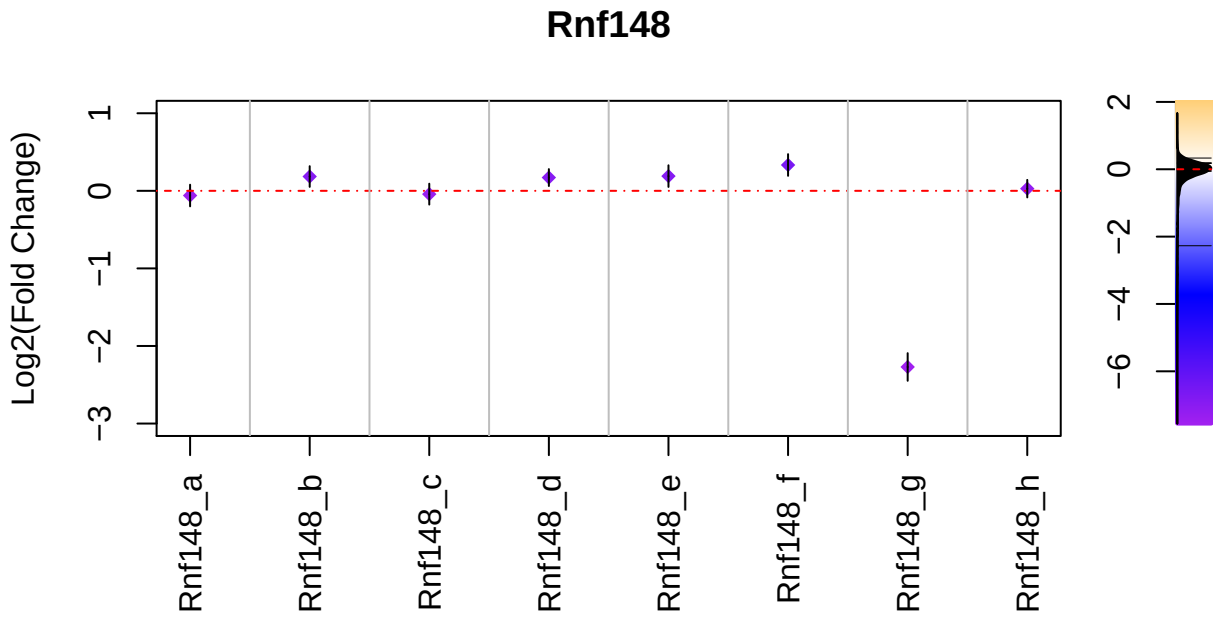

Figure 7: Example of a target whose depletion is assigned greater significance with MAGeCK's P-value scoring.

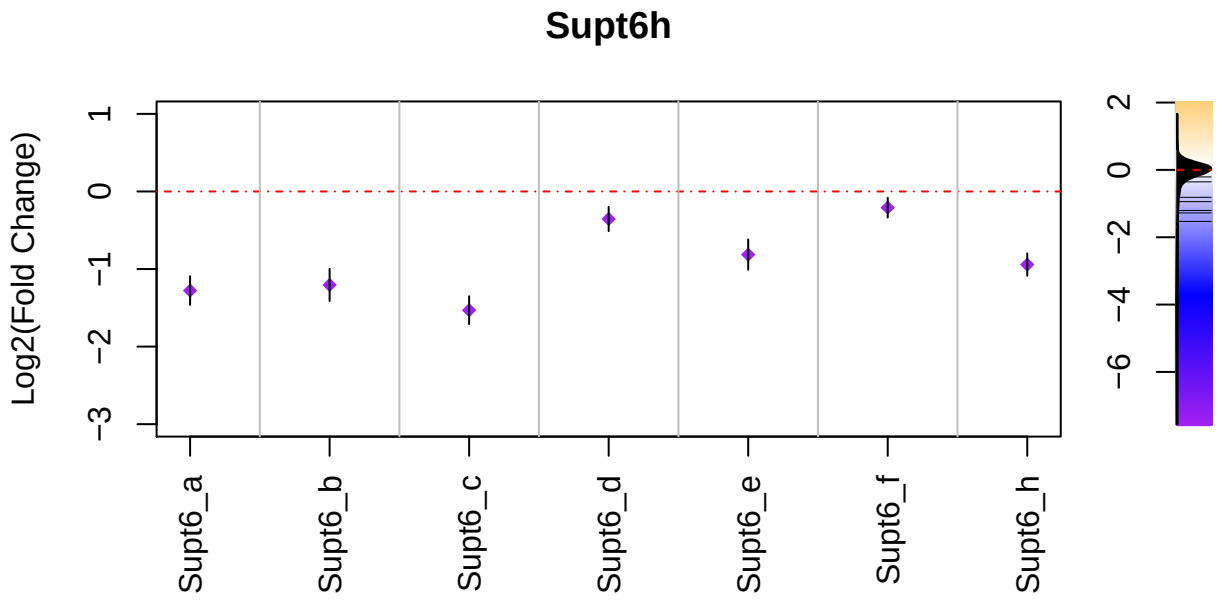

Figure 8: Example of a target whose depletion is assigned greater significance with gCrisprTools' Combined scoring.

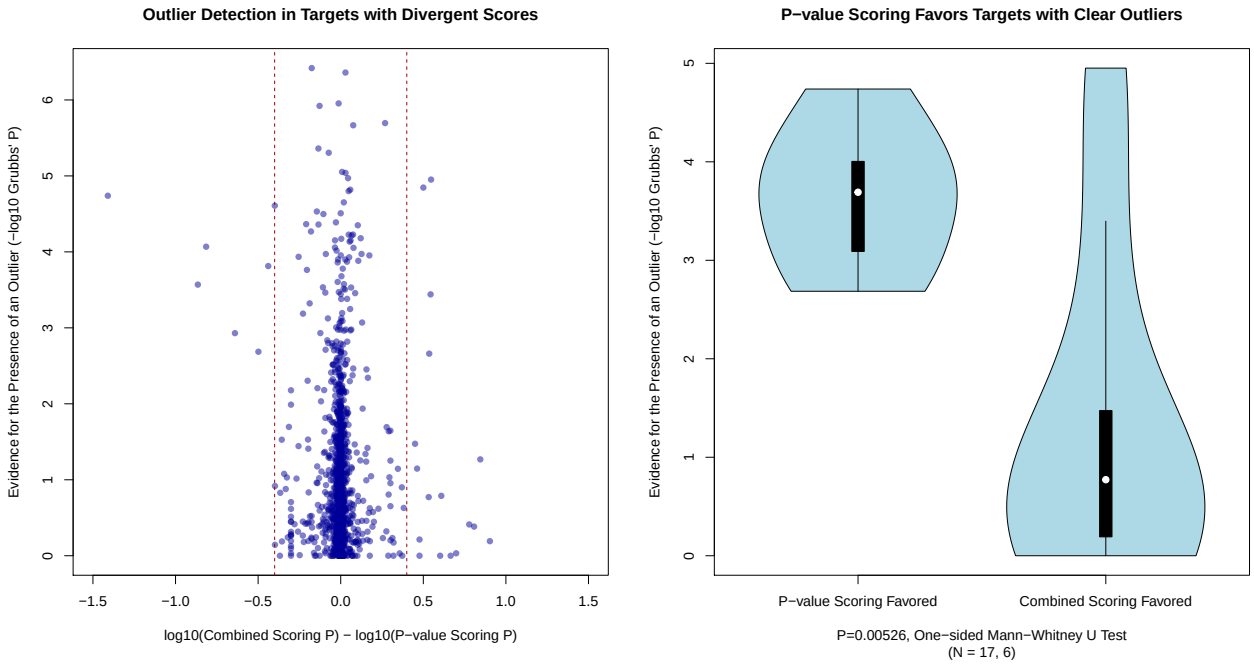

Figure 9: P-value scoring increases detection of targets with outlier signals.

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.