



Modified sgRNA Design Improves Results of CRISPR Knockout Screens

Introduction

The CRISPR single-guide RNA (sgRNA) design published by Jinek, et al. in 2012 has become the gold standard for CRISPR-mediated gene knockout. It was developed by fusing two oligonucleotide components of the native *Streptococcus pyogenes* CRISPR system—the tracrRNA and crRNA—into a single molecule where the first 40 bases contains the 20 base variable targeting region and the first part of the initial stem-loop. This first 5' domain of the sgRNA corresponds to the crRNA in the native bacterial system. The rest of the sgRNA is derived from the tracr sequence in the bacterial system where it hybridizes with the crRNA. Thus, the sgRNA contains an initial variable region of about 20 bases followed by a constant sequence of about 80 nucleotides that contains all the key interactions with the Cas9 nuclease.

Several research groups have attempted to optimize the design of the initial variable region that defines the sequence the sgRNA targets, in order to ensure an effective knockout while minimizing off-target disruptions (Doench et al., Fu, et al.). However, as described above, the 20-base targeting sequence makes up only a small portion—about one-fifth of the sgRNA sequence. The other 80 bases that are downstream of this targeting sequence interact primarily with the Cas9 endonuclease to catalyze gene knockout. If modifications to this constant 3' region of the sgRNA molecule could improve knockout efficiency, any sgRNA could be made more effective and might offer a general approach to improve our pooled genome-wide sgRNA libraries.

One study (Chen, et al.) with an inactive Cas9 nuclease has shown that sequence modifications to the constant region improved Cas9 binding significantly. However, since an inactive Cas9 mutant was used in these studies, it wasn't clear if the changes would actually increase the rate or the knockout efficiency of the active CRISPR system, and further, if they did, the effect that they would have on results of CRISPR-based pooled genetic screens. We initiated a study to address these points.

Experiment and Results

To investigate if changes in the 3' Cas9-binding portion of the sgRNA did, in fact, increase the efficiency of CRISPR-mediated knockout, we first ran some initial experiments with a few sgRNA sequences targeting a GFP gene. We specifically looked at two modifications of the constant 3' region of the guide sequence mentioned in the citation noted previously (Chen, et al.). One modification swapped locations of adenine (A) and thymine (T) residues ("AT") to remove a transcription

terminator site. Another alteration adds a 5 nucleotide extension ("HE") to the stem of a stem-loop structure which should make it more stable and accessible to the Cas9 protein.

We found that these two substitutions did indeed increase the rate of target knockout—at least with the one target (Figure 1). Both the AT inversion and HE insertion significantly improved the knockout rate of GFP relative to the "wild type" (wt) sgRNA sequence. While the effects varied somewhat for each sequence, the overall positive impact of these changes on several targets in the GFP sequence was clear.

The knockout results of GFP with the modified sgRNAs indicated that including these changes to 3' guide sequences in a pooled library might increase the knockout rate and improve the overall results of loss-of-function genetic screens. To substantiate it further, we built four libraries each with a different design including the "wild type" (wt) sgRNA design control that is based on the sequences found in the bacterial CRISPR system as described by Jinek, et al. Two libraries incorporated the AT inversion and HE insertion described above, and the fourth library contained both the HE and AT modifications (HEAT).

With each library, we ran a "dropout viability screen" to identify the sgRNA structure performing better at identifying positive control essential genes present in the library. With this type of genetic screen, cells harboring sgRNA that knock

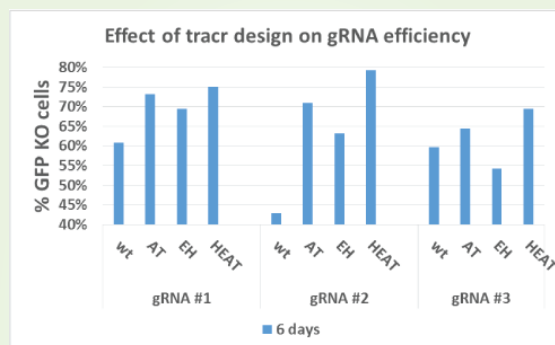


Figure 1. Lentiviral constructs with three different sgRNAs targeting a green fluorescent protein (GFP) sequence were transduced into cells stably expressing GFP. Four variants of each sgRNA were used: the "wild type" (wt) sequence, a variant with an AT inversion that eliminates a transcription termination site, a variant with an insertion (EH) that elongates and stabilizes a stem-loop structure, and a variant with both the HE and AT (HEAT) modifications. The change in GFP expression 6 days after transduction was assayed. In most cases, all three modified sgRNA designs reduced GFP fluorescence more quickly than the standard wt version.

out essential genes do not proliferate, so these sgRNA are underrepresented after several cell doublings. As you can see in Figure 2, the positive control sgRNAs—the ones that targeted known essential genes—are more consistently and strongly depleted in the screens with the libraries that contained the modified sgRNA.

To quantify the dropout improvement with the modified guides, we ran a Z score analysis of the depletion levels of 8 control sgRNAs to 20 different essential genes (Figure 3). The analysis showed a consistent increase in the magnitude of the signal for sgRNAs containing the modifications. Guides with the AT modification, in particular, were significantly less represented than the wt. However, guides containing both the AT and HE modifications consistently had lower depletion levels than guides with just the AT modification.

Conclusion

In screens with complex pooled CRISPR libraries, hundreds of cells pick up the same sgRNA in a background of millions of cells with other sgRNAs. Detecting a phenotype associated with one particular sgRNA requires that the cells harboring this sgRNA respond in a similar manner in the timeframe of the assay. Thus, the more quickly and consistently the sgRNA acts on the target, the better the response will be.

The findings from this study clearly show that it is possible to substantially improve the quality of sgRNA libraries with just a few changes in the constant 3' region of the sgRNA where the Cas9 protein binds. Constructs utilizing the modified sgRNA structure knocked out target genes more quickly and effectively than those with the standard wt sgRNA. Further, libraries expressing sgRNA with the HE and AT modifications generated stronger and more robust results than those with the standard wt structure, thereby decreasing the chance of missing essential genes during screening.

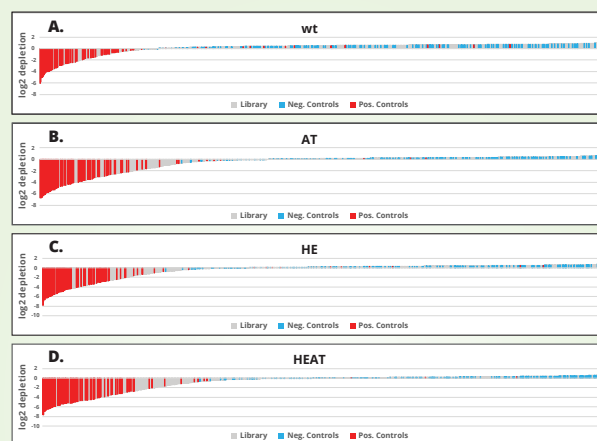


Figure 2. Four libraries, each with 1,000 constructs, were constructed with 8 sgRNA targeting each of 100 genes, 100 negative controls targeting non-coding genomic DNA, and 100 non-targeting negative controls. The sgRNAs in each library were designed with a different 3' domain: the wt sequence, the AT inversion variant, the HE insertion variant, and the variant with both modifications (HEAT). The libraries were transduced into Cas9-expressing CML cell lines. After 3 weeks of growth, the representation of each sgRNA construct in the surviving cell population was quantified by NGS and compared with its representation in the original library. Changes in representation of each sgRNA are shown. The positive controls that target essential genes (lines in red) appear to be more strongly depleted in the libraries with the modified sgRNA structures.

References

- Chen B, et al, Cell. 2013 Dec 19;155(7):1479-91. doi: 10.1016/j.cell.2013.12.001.
- Doench JG, et al., Nat Biotechnol. 2014 Dec;32(12):1262-7. doi: 10.1038/nbt.3026. Epub 2014 Sep 3.
- Fu, et al (Nat. Biotechnol. 2014, 23:279-84).
- Jinek M, et al., Science. 2012 Aug 17;337(6096):816-21. doi: 10.1126/science.1225829

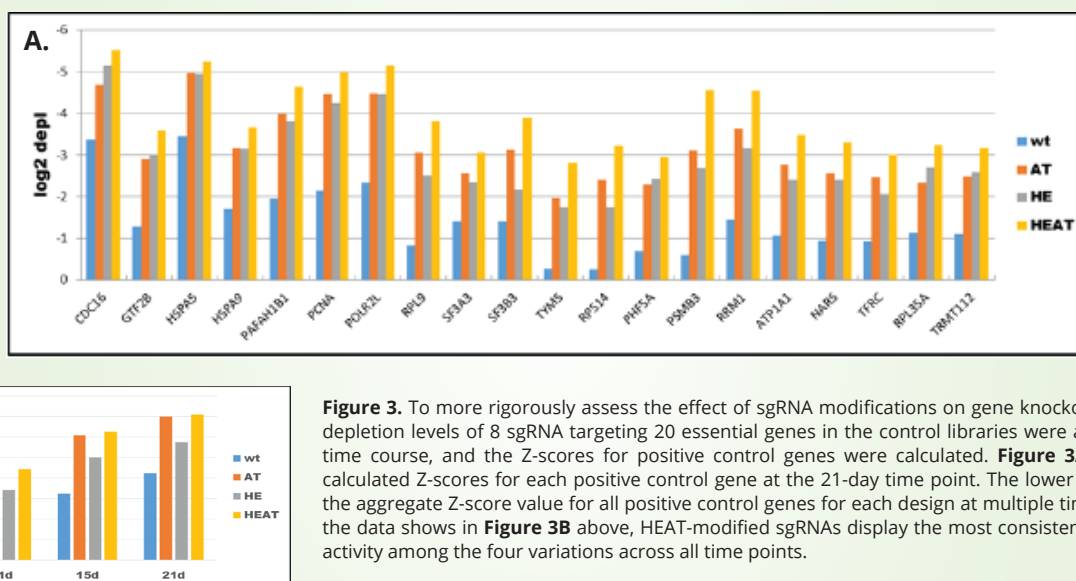


Figure 3. To more rigorously assess the effect of sgRNA modifications on gene knockout rates, the depletion levels of 8 sgRNA targeting 20 essential genes in the control libraries were analyzed in a time course, and the Z-scores for positive control genes were calculated. **Figure 3A** shows the calculated Z-scores for each positive control gene at the 21-day time point. The lower panel shows the aggregate Z-score value for all positive control genes for each design at multiple time points. As the data shows in **Figure 3B** above, HEAT-modified sgRNAs display the most consistent and robust activity among the four variations across all time points.

For more information and to subscribe to updates, email us at info@collecta.com.