Abstract
Although the literature shows most CRISPR-Cas9 guide RNAs will produce significant indel formation, not all guide RNAs produce a functional gene knockout, which is the desired result for the majority of these experiments. To better understand the parameters affecting the efficiency for a functional gene knockout, we utilized synthetic crRNA and tracrRNA, which can be chemically synthesized rapidly without the need for cloning and sequencing. We systematically transfected >1100 synthetic crRNA:tracrRNA targeting components of the proteasome into a reporter cell line in which knockout of proteasome function results in fluorescence of a ubiquitin-EGFP fusion protein that is normally degraded by the proteasome pathway. Using the results from the functional assay, we developed and trained a machine-learning algorithm to score crRNAs based on how likely they were to produce functional knockout of targeted genes (functionality score). To minimize potential off-targets, we developed a rigorous specificity tool that is able to detect and score mismatches as well as gapped alignments that are typically missed using most existing specificity tools (specificity scores). We combined this comprehensive specificity check with our functionality algorithm to select and score highly specific and functional crRNAs for any given gene target and also generated a whole-genome arrayed crRNA library for screening applications.

Functionality

Improved experimentation

- Lentiviral transduction
- Selection with Marsidin
- Cell population expansion
- Synthetic crRNA:tracrRNA transfection
- Cris-cell line mixed populations
- Phenotypic analysis

Establishment of a Cas9-expressing cell line greatly improved phenotypic consistency and ease-of-use of Edit-R synthetic crRNA:tracrRNA to assess >1100 target regions across multiple genes in a high-throughput manner.

Rapid functional readout

Cas9 stable cells transfected with crRNA:tracrRNA + RNA treated

Simple + Rapid = Large amounts of quality data

- Recombinant LUC cell line stably expressing a mutant human ubiquitin fused to EGFP
- Undetectable ubiquitin moiety (Gly76Val)
- Constitutive degradation of the protein means very low EGFP fluorescence when the proteasome is functioning normally
- Disruption in proteasome-related components increases fluorescence

Machine learning

Trained functionality algorithm

Dharmacon training set: 50 genes, 1135 crRNA target sites

Features examined include:
- nucleotide composition
- position in exons
- nearest neighbor effects
- distance from the start codon
- rRNA sequence

Data was used to select features and multi-dimensional features that were highest predictors

crRNAs vary widely in their ability to cause functional gene disruption

Good measure of algorithm fit

Good fit of data while avoiding overfitting of the data, since the goal is to have a good prediction of the training data set AND unrelated gene editing data.

Validation of algorithm in other phenotypic assays

- ROC (Receiver Operating Characteristic) shows good fit of training data set. The ROC measures the area under the curve of True Positive Rate vs False Positive Rate.
- The ROC for our test set data is 0.78

Specificity

Rigorous analysis of potential off-targets

A perfect alignment is not required for off-targeting

Chromosomal sites with flaws in their alignment can create indels and are potential off-targets. Flaws include not just mismatches, but gaps as well. Indels at chromosomal sites can lead to potential off-targets, so they should be considered when checking specificity. Most tools do not take into account gaps, therefore most tools do not find all imperfect alignments.

The effect of incomplete alignment

The Dharmacon CRISPR Specificity Analysis Tool provides comprehensive alignment. Try it at dharmacon.com/Tools-and-Calculators/CRISPR-specificity-tool

Complete and fast alignment

The Dharmacon CRISPR Specificity Analysis Tool provides comprehensive alignment. Try it at dharmacon.com/Tools-and-Calculators/CRISPR-specificity-tool

Comparison of number of possible off-target alignments

Using two targets, we used multiple publicly available alignment tools to look at the predicted off-targets and sorted them by the total number of flaws. Results below show that the Dharmacon alignment strategy can identify potential off-targets with much more rigor than other tools.

Conclusion

We have developed a crRNA design algorithm that finds optimal picks using a functionality and specificity score. As a result, higher scoring crRNAs showed better functionality than lower scoring crRNAs.

Dharmacon and Edit-R are trademarks of Horizon Discovery Group. ©2018 Horizon Discovery—All rights reserved.