GE Healthcare

Picking the best CRISPR-Cas9 targets for functional gene knockout: a machine learning algorithm based on both specificity and functionality

Shawn McClelland, Emily M. Anderson, Žaklina Strezoska, Elena Maksimova, Annaleen Vermeulen, Steve Lenger, Tyler Reed, and Anja van Brabant Smith
Dharmacon, now part of GE Healthcare, 2650 Crescent Drive, Suite #100, Lafayette, CO 80026, US

Introduction

The CRISPR-Cas9 system has the potential to significantly advance basic and applied research.

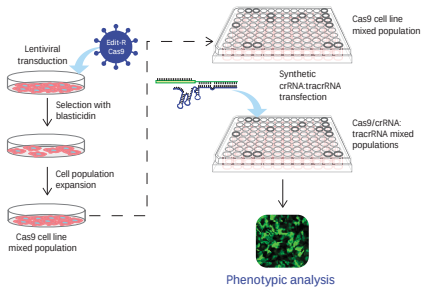
Functional gene knockout is an important tool for understanding a gene's role in a system or for specifically manipulating a known system to achieve a desired outcome.

Not all gene cleavage events result in functional knockout of the target protein.

Here we share important advancements that have helped to achieve the goal of picking the best crRNA targets for functional gene knockout, and not just formation of indels (insertions or the deletions of bases in the DNA).

FUNCTIONALITY

Improved experimentation



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

Rapid functional readout

Recombinant U2OS cell line stably expressing a mutant human ubiquitin fused to EGFP

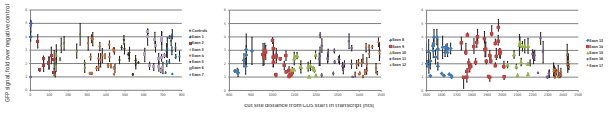
- Uncleavable ubiquitin moiety (Gly76Val)
- Constitutive degradation of the protein means very low EGFP fluorescence when the proteasome is functioning normally
- Disruption in proteasome-related components ***increases fluorescence***

Simple + Rapid = Large amounts of quality data

Machine learning

Trained functionality algorithm

An algorithm is important because crRNAs vary widely in their ability to cause functional gene disruption.



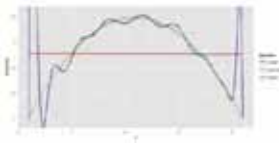
Dharmacon training set: 10 genes, 1115 crRNA target sites

Features examined include:

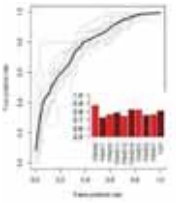
- nucleotide composition
- nearest neighbor effects
- PAM sequence
- position in exons
- distance from the start codon

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

Good measure of algorithm fit



Good fit of data while avoiding overfitting of data. Goal: To have good prediction of the training data set **AND** unrelated gene editing data



« ROC (Receiver Operating Characteristic) shows good fit of training set data. The ROC measures the area under the curve of True Positive Rate vs False Positive Rate.

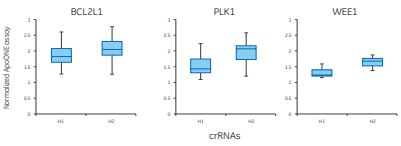
The ROC for our test set data is 0.78.

Validation of algorithm in other phenotypic assays

It is essential to verify that algorithm designs are tested in an assay with genes that are unrelated to the ones used to generate the training data. We used the ApoONE assay with algorithm-derived crRNA to new target genes known to further validate the functionality algorithm.

High scoring crRNAs show better function than low scoring crRNAs in an assay unrelated to the assay used to train the algorithm.

ApoONE assay box plots: crRNAs with the bottom half algorithm scores (H1) versus crRNAs with the top half of algoirthm scores (H2).

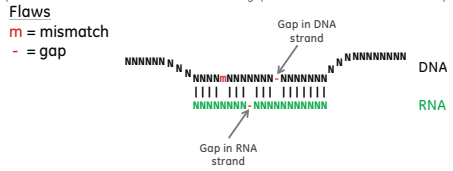


SPECIFICITY

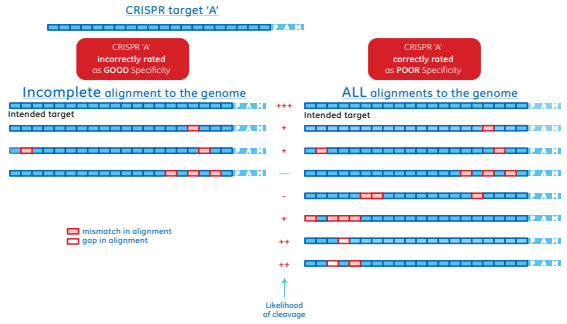
Rigorous analysis of potential off-targets

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. Most tools do not take into account gaps and most tools do not find all imperfect alignments.



The effect of incomplete alignment



Complete and fast alignment

Dharmacon's new alignment strategy finds all possible alignments (including alignments containing gaps) and allows us to design targets that are less likely to cause off-targets.

The Dharmacon CRISPR Specificity Analysis Tool provides comprehensive alignment.
Try it at: dharmacon.gelifesciences.com/tools-and-calculators/crispr-specificity-tool

Tool	% Alignments found by Flaw Count				Alignment Time (s)
	0	1	2	3	
BLAST	100	100	80	23	24.8
Bowtie 2	100	100	70	29	6.27
Dharmacon Alignment	100	100	100	100	2.30

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.

Target 1 - GGTCATCTGGGAGAAAAGCG CGG
hg38; chr14:54748857-54748879

	# Alignments found				
	0 flaws	1 flaws	2 flaws	3 flaws	4 flaws
Tool Z	1	0	0	16	227
Tool C	1	0	0	NR	NR
Tool W	1	0	0	7	142
Tool O	1	0	0	7	139
Dharmacon Alignment	1	0	4	169	3732

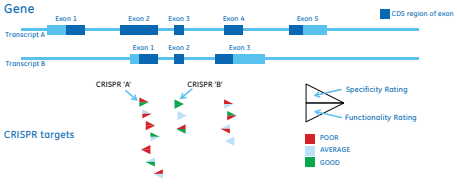
Target 2 - GCTCCACGTTTATAAATAGC TGG
hg38; chr14:54569552-54569574

	# Alignments found				
	0 flaws	1 flaws	2 flaws	3 flaws	4 flaws
Tool Z	1	0	1	7	81
Tool C	2	0	0	NR	NR
Tool W	NR	NR	NR	NR	NR
Tool O	2	0	1	4	58
Dharmacon Alignment	2	0	1	72	1915

Missing alignments
NR No Results reported for tool (error or inability)

Conclusion

Functionality & Specificity are used to select the optimal picks



CRISPR target 'A' – May function, but poor specificity
CRISPR target 'B' – OPTIMAL PICK, scores well in both functionality & specificity

gelifesciences.com/dharmacon