Complete alignment identification of CRISPR-Cas9 genomic off-targets using Edit-R CRISPR specificity tool and a comprehensive analysis of positional mismatch tolerance

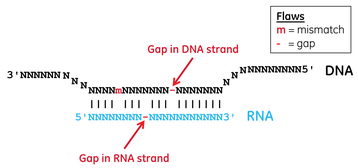
Emily M. Anderson, Shawn McClelland, Amanda Haupt, Eldon T. Chou, John Schiel, Annaleen Vermeulen, Žaklina Strezoska, Hidevaldo Machado, Steve Lenger, Amanda Birmingham, and Anja van Brabant Smith Dharmacon, 2650 Crescent Drive, Suite #100, Lafayette, CO 80026, USA

Abstract

The CRISPR-Cas9 system has the potential to advance basic and applied research; however, specificity of RNA-directed DNA cleavage events is not yet completely understood and can hamper its wider application. New findings on off-target effects and their determinants are being reported frequently. Recent work has demonstrated gene editing by CRISPR RNAs (crRNAs) containing bulges of up to 4 nucleotides, but existing design tools are unable to detect putative off-targets based on gapped alignments. We present the Dharmacon™ Edit-R™ CRISPR specificity tool, a simple web tool that leverages well-characterized alignment optimization techniques to perform rapid, customizable, and complete crRNA specificity checking including gap detection. The Edit-R CRISPR specificity tool is freely accessible dharmacon.gelifesciences.com/tools-and-calculators/crispr-specificity-tool.

In addition, we have comprehensively evaluated positional off-targeting propensities of the CRISPR system using a three-component platform of Cas9, synthetic crRNAs, and synthetic tracrRNA. We executed a systematic positional screen of crRNAs containing two nucleotide mismatches to the DNA target for two functional crRNAs. The application of a high-throughput reporter assay that directly measures functional activity of a central cellular process (ubiquitin-proteasome activity), allowed measurement of the relative cleavage activity of all disruptive two-mismatch combinations for each crRNA (190 combinations per sequence or 380 total), regardless of the need for the off-target target region to naturally occur in the genome adjacent to a PAM. Our results demonstrate that while off-targeting does occur in the presence of two mismatches between the crRNA and the target DNA, the overall levels of functional off-targeting are low relative to on-target activity. Analysis of the position of tolerated mismatches further clarifies the mechanism of CRISPR-Cas9 off-targeting as well as provides additional crRNA design rules for mammalian gene editing. These studies demonstrate the simplicity and high-throughput nature of this three-component system for elucidation of CRISPR-Cas9 function and mechanism.

Types of observed sequence flaws



Permutation	Description
AGGTCA	Query
AAGTCA	Mismatch at 2nd base
ACGTCA	
ATGTCA	
AGG-CA	Reference gap after 3rd base
AGGATCA	Query gap after 3rd base
AGGCTCA	
AGGGTCA	
AGGTTCA	

Figure 1. Examples of permutations corresponding to alignments containing a mismatch, a RNA gap, and a DNA gap.

Edit-R CRISPR specificity tool alignment is fast and complete

Tool	Seed		Time (s)	% Found by Flaw Count			
	Length	Interval		0	1	2	3
BLAST	N/A		24.8	100	100	80	23
Bowtie 2	6	6	23596	100	100	100	72
Bowtie 2	7	3	2498	100	100	95	68
Bowtie 2	7	6	544	100	100	95	53
Bowtie 2	8	3	286	100	100	95	56
Bowtie 2	9	5	6.27	100	100	70	29
Edit-R Tool	≤ 3 Flaws		2.30	100	100	100	100

Figure 2. Computation speed and percentage of flawed off-target sequences found in a genomic off-target search are compared between the Edit-R CRISPR specificity tool and commonly used mismatch search detection methods and settings.

Comparison of the Edit-R CRISPR specificity tool to other alignment methods

Target 1 - GGTCATCTGGGAGAAAAGCG CGG hg38; chr14:54748857-54748879				
# flaws:	0	1	2	3
Tool 1	1	0	0	16
Tool 2	1	0	0	--
Tool 3	1	0	0	7
Tool 4	1	0	0	7
Edit-R Tool	**1**	**0**	**4**	**169**

Figure 3. Number of human genome off-target alignments found for one target site by four published CRISPR-Cas9 design tools and the Edit-R CRISPR specificity tool. The rigor of the Edit-R tool is demonstrated by the complete identification of potential genomic off-targets due to both mismatches and gaps.

Cas9-mediated gene editing using synthetic crRNA:tracrRNA

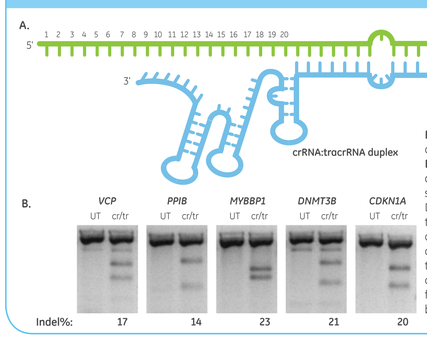


Figure 4. A. Schematic of the crRNA:tracrRNA synthetic RNA molecules. **B.** Co-transfection of a plasmid expressing a humanized Cas9, synthetic crRNA, and synthetic tracrRNA results in targeted DNA cleavage in HEK293T cells. Cells were transfected with Cas9 plasmid, synthetic crRNA, and synthetic tracrRNAs and compared to untransfected cells (UT) using the T7 endonuclease I (T7EI) assay. The calculated percent indel (definition of indel) formation after 72 hours is indicated at the bottom of each lane.

A high-throughput functional assay for proteasome function

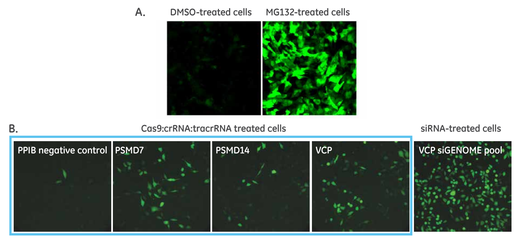
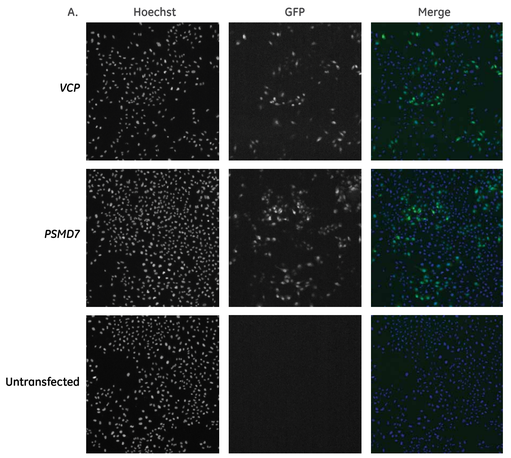


Figure 5. Recombinant U2OS cell line stably expressing a mutant human ubiquitin fused to EGFP. A Gly76Val mutation results in an uncleavable ubiquitin moiety fused to EGFP, which results in constitutive degradation of the protein and little detectable EGFP fluorescence. **A.** Inhibition of proteasome activity with chemical treatment (MG132) results in EGFP fluorescence. **B.** Inhibition of proteasome activity with CRISPR-Cas9 or siRNAs targeting proteasome components results in EGFP fluorescence.

Editing in Ubiquitin-GFP U2OS-Cas9 cells



C.

	VCP_2	
	Untransfected	+ crRNA:tracrRNA
Total sequences	143	114
Insertions	2 (1.4%)	18 (15.8%)
Deletions	1 (0.7%)	37 (32.5%)
Insertions and Deletions	3 (2.1%)	55 (48.2%)
	PSMD7_1	
	Untransfected	+ crRNA:tracrRNA
Total sequences	140	113
Insertions	1 (0.7%)	14 (12.4%)
Deletions	1 (0.7%)	31 (27.4%)
Insertions and Deletions	2 (1.4%)	45 (39.8%)

Figure 6. A. Ubiquitin (Ubi)-GFP U2OS-Cas9 cells were transfected with synthetic crRNA and tracrRNAs targeting two genes involved in proteasome function, *VCP* and *PSMD7* (VCP_2 crRNA or PSMD7_1 crRNA, respectively) or untreated. The GFP phenotype was assessed by high content imaging 72 hours post-transfection and representative images are shown. The percent GFP-positive cells was quantified by calculating average values from biological triplicates: VCP_2 = 10.3 ± 1.2%, *PSMD7* = 10.6 ± 0.5%, Untransfected = 0 ± 0%. **B.** Assessment of gene editing using the T7EI mismatch detection assay in untransfected (-) or cells treated with the indicated crRNA and tracrRNA (+). The percent gene editing for treated samples is indicated below each lane. **C.** Summary of results from Sanger sequencing of individually cloned PCR products from untransfected or cells treated with the indicated crRNA and tracrRNA.

Systematic positional two-base mismatch design

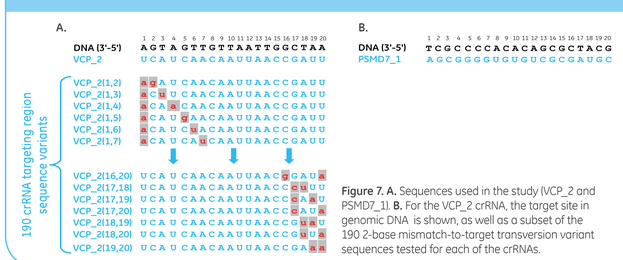


Figure 7. A. Sequences used in the study (VCP_2 and PSMD7_1). **B.** For the VCP_2 crRNA, the target site in genomic DNA is shown, as well as a subset of the 190 2-base mismatch-to-target transversion variant sequences tested for each of the crRNAs.

Gene targeting and relative off-targeting in Ubiquitin-GFP U2OS-Cas9 cells

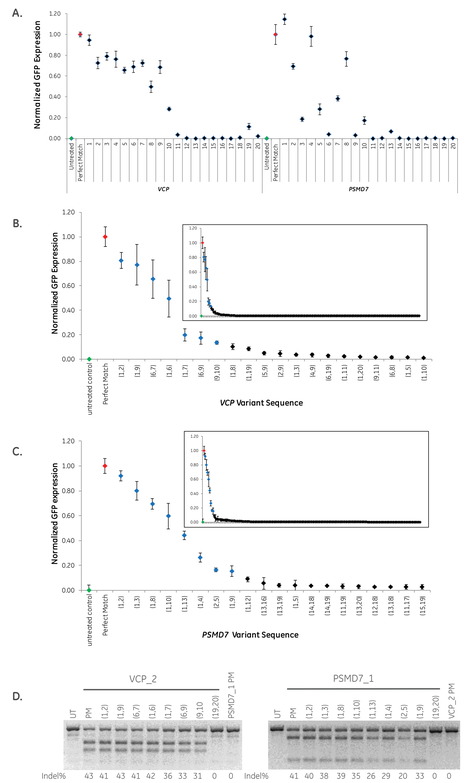


Figure 8. A. Functionality of crRNAs containing single base transversions that result in disruptive mismatches at every position of the 20 nt crRNA targeting sequence in the Ubi-GFP U2OS-Cas9 EGFP cells. **B.** A ranked plot of 2-base mismatched crRNA variants of VCP_2 and PSMD7_1. The 20 highest signal mismatched crRNAs are plotted along with the full sets of 190 crRNAs in the inset. **C.** A ranked plot of 2-base mismatched crRNA variants of PSMD7_1. The 20 highest signal mismatched crRNAs are plotted along with the full sets of 190 crRNAs in the inset. **B-C.** Untreated control (green), perfect match crRNA (red), 2-base mismatched crRNAs exhibiting a phenotype with robust Z-score > 100 (blue), and crRNAs with insignificant signal (black). **D.** T7EI mismatch detection assay of VCP and PSMD7 crRNAs that are either perfectly matched to their target or contain two mismatches in the crRNA. All crRNAs containing two mismatches resulted in off-targeting (EGFP fluorescence) with the exception of VCP_2 (19,20) and PSMD7_1 (19,20) which also did not show a phenotype. **B-D.** Position of mismatch as indicated (x,y) where position 1 is distal and position 20 is proximal to the PAM.

Conclusions

- The Edit-R CRISPR specificity tool is both fast and exhaustive at predicting crRNA genomic off-target sites with up to three flaws
- Single mismatches are fairly well-tolerated, especially in the 5' PAM-distal end of crRNAs
- crRNAs with mismatches in the seed region, position 11-20 of the crRNA, did not exhibit significant activity
- Consecutive two-base mismatches are generally not active
- Two-base mismatches exhibiting function tend to be nonconsecutive (in the non-seed region) and most often involve one mismatch at the most distal end from the PAM

horizondiscovery.com/dharmacon

For more information:
If you have any questions directly related to this document, contact our experts:
UK +44 (0) 1223 976 000
USA +1 800 235 9880, +1 303 604 9499
Website horizondiscovery.com/contact-us