Journal of Biotechnology 211 (2015) 56–65

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity

Emily M. Anderson, Amanda Haupt, John A. Schiel, Eldon Chou, Hidevaldo B. Machado, Žaklina Strezoska, Steve Lenger, Shawn McClelland, Amanda Birmingham, Annaleen Vermeulen, Anja van Brabant Smith*

Dharmacon, Part of GE Healthcare, Lafayette, CO 80026, USA

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form 26 May 2015
Accepted 23 June 2015
Available online 17 July 2015

Keywords:
CRISPR-Cas9
Specificity
Gene editing
Genome engineering
Functional assay
crRNA

ABSTRACT

The discovery that the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) acquired immune system can be utilized to create double-strand breaks (DSBs) in eukaryotic genomes has resulted in the ability to create genomic changes more easily than with other genome engineering techniques. While there is significant potential for the CRISPR–Cas9 system to advance basic and applied research, several unknowns remain, including the specificity of the RNA-directed DNA cleavage by the small targeting RNA, the CRISPR RNA (crRNA). Here we describe a novel synthetic RNA approach that allows for high-throughput gene editing experiments. This was used with a functional assay for protein disruption to perform high-throughput analysis of crRNA activity and specificity. We performed a comprehensive test of target cleavage using crRNAs that contain one and two nucleotide mismatches to the DNA target in the 20mer targeting region of the crRNA, allowing for the evaluation of hundreds of potential mismatched target sites without the requirement for the off-target sequences and their adjacent PAMs to be present in the genome. Our results demonstrate that while many crRNAs are functional, less than 5% of crRNAs with two mismatches to their target are effective in gene editing; this suggests an overall high level of functionality but low level of off-targeting.

1. Introduction

The CRISPR–Cas9 system is a highly effective and simple technique for modifying the genome of higher eukaryotes, particularly mammalian cells that have not been readily amenable to gene editing. In the short period since the initial discovery of efficacy in mammalian cells (Cong et al., 2013; Mali et al., 2013; Jinek et al., 2013; Cho et al., 2013), there have been numerous studies that demonstrate the utility of the CRISPR–Cas9 system for genome engineering, from performing whole-genome-scale knockout screens elucidating gene function in cell culture (Shalem et al., 2014; Wang et al., 2014), to creating multiple mutations modeling tumors in the livers and lungs of adult mice (Xue et al., 2014; Platt et al., 2014) and creating mutations in the brains of adult mice (Swiech et al., 2015). This CRISPR system requires the Cas9 nuclease along with two short RNA molecules, the guiding CRISPR RNA (crRNA) and the trans-activating crRNA (tracrRNA), that hybridize to each other and direct Cas9 to the target location for cleavage based on sequence complementarity to the crRNA as well as proximity of the DNA target to a protospacer adjacent motif (PAM). In 2012, Jinek et al. (2012) demonstrated that the two small RNAs, crRNA and tracrRNA, can be combined into a single guide RNA (sgRNA) that can be expressed in mammalian cells. While the effectiveness and universality of CRISPR–Cas9 for genome engineering in mammalian systems has been clearly demonstrated, the activity and specificity of a particular crRNA or chimeric sgRNA is less clear.

The ability of any given crRNA to efficiently create a DSB in the target DNA can vary based on the guide RNA (gRNA; either crRNA or sgRNA) sequence and position in the targeted gene (Shalem et al., 2014; Wang et al., 2014; Koike-Yusa et al., 2014). A recent

Abbreviations: CRISPR, clustered regularly interspaced palindromic repeats; Cas9, CRISPR associated protein 9; crRNA, CRISPR RNA; tracrRNA, trans-activating CRISPR RNA; PAM, protospacer adjacent motif; sgRNA, single guide RNA; NGS, next generation sequencing; PSC, pluripotent stem cells; ACE, monoacetate orthoester; siRNA, small interfering RNA; T7E1, T7 endonuclease I; PCR, polymerase chain reaction; GFP, green fluorescent protein; EGFP, enhanced green fluorescent protein; Ubi-GFP, ubiquitin-conjugated green fluorescent protein; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; GUIDESeq, genome-wide unbiased identifications of DSBs evaluated by sequencing; DSB, double-strand break.

* Corresponding author.
E-mail address: anja.smith@ge.com (A.v.B. Smith).

http://dx.doi.org/10.1016/j.jbiotec.2015.06.427

study has identified sgRNA design rules that predict functional sgRNAs (Doench et al., 2014). Likewise, several publications have demonstrated variable levels of off-target cleavage for specific sgRNAs (Cong et al., 2013; Cho et al., 2014; Fu et al., 2013; Lin et al., 2014; Pattanayak et al., 2013; Hsu et al., 2013). For example, Fu et al. (2013) examined the activity of sgRNAs that contained single, double, and multiple mismatches with the target DNA. Their results demonstrated that, in general, single mismatches are well-tolerated at the 5′ end of the sgRNA, adjacent double mismatches can still result in cleavage for some sequences, and three or more adjacent mismatches at the 5′ end of the sgRNA eliminate activity. When testing potential off-target sites that were predicted bioinformatically, CRISPR–Cas9-dependent cleavage was only detected at a small subset of these endogenous potential off-target sites. In addition, off-targeting was demonstrated using sequences containing up to five nonadjacent mismatches to their target, suggesting that sequences with more mismatches to a potential off-target do not necessarily yield reduced off-targeting. Off-target sites including DNA or RNA bulges at certain positions relative to the sgRNA strand have also been shown to be cleaved by CRISPR–Cas9 (Lin et al., 2014). These results show context and sequence dependencies including a correlation with GC content. In contrast, whole-genome next-generation sequencing (NGS) of human pluripotent stem cell (PSC) clones or induced PSC clones targeted with CRISPR–Cas9 showed very few to no off-target mutations that could be attributed to CRISPR–Cas9 (Smith et al., 2014; Veres et al., 2014). In these studies, clonal artifacts derived from isolating CRISPR–Cas9-edited cells appeared to be a larger concern, suggesting that off-targeting is not a significant issue for using CRISPR–Cas9 in human PSCs. The rules governing which mismatches between the crRNA and target DNA are tolerated and therefore result in off-target effects remain unclear.

In consideration of the variety of reports outlined above, we performed a comprehensive evaluation of off-target potential of crRNAs containing mismatches with the target DNA. While Fu et al. evaluated transversion mutations at all possible single mismatches throughout 19 of the 20 nucleotides, their evaluation of double mismatches was limited to adjacent mismatches and a subset of variably spaced ones. Likewise, Hsu et al. evaluated varying combinations of mismatches, both interspersed and concatenated, but they did not assess all possible mismatch combinations, likely due to the technical infeasibility of generating and assaying large numbers of sgRNA designs. Here, we exploit a functional assay paired with synthetic crRNAs to assess the effects of mismatches in a high-throughput manner, including a comprehensive analysis of all possible crRNA positions containing two mismatches. This approach allows for a systematic and thorough investigation of positional preferences in off-targeting, since a very large number of combinations can be tested without prior knowledge of whether the off-target sites occur endogenously in the genome alongside a PAM. In a genomewide study design, for instance utilizing NGS, a much larger set of crRNAs would have to be assayed to represent every position in the crRNA targeting sequences with the assurance that all or a large number of the potential off-target sequences were represented in silico.

2. Materials and methods

2.1. Generating a stable cell line expressing Cas9

2.1.1. Cas9 lentiviral particle production

A lentiviral vector containing a human codon-optimized *Streptococcus pyogenes* Cas9 gene under control of the hCMV promoter and a puromycin resistance gene was packaged into lentiviral particles using the Trans-Lentiviral Packaging Kit (Dharmacon, #TLP5919) (Kappes and Wu, 2001; Kappes et al., 2003).

2.1.2. Creation of Cas9 stable cell lines

Recombinant U2OS cells that stably express a mutant human ubiquitin fused to EGFP (Thermo Scientific BioImage Proteasome Redistribution Assay, #R0402102) were seeded at 100,000 cells per well in a 24-well plate (Costar, #3524). The following day, cells were transduced with 3 μL of Cas9 lentiviral particles (rough estimate of viral titer: 1×10^6 viral particles per mL). The cells were expanded into a 6-well plate (Thermo Scientific, #140675) 72 h post-transduction. After a 48-hour recovery and expansion period, the cells were exposed to 2 μg/mL of puromycin (Invivogen, #ant-pr-5). Cells that survived the selection were harvested and utilized in the following experiments and are referred to as Ubi-GFP U2OS–Cas9.

2.2. Design of crRNA and tracrRNA

Full-length crRNAs were designed by appending the *S. pyogenes* crRNA repeat sequence, 5′ GUUUUAGAGCUAUGCUGUUUUG 3′, to each target sequence.

crRNAs were designed to target VCP, PPIB, MYBB1, DNMT3B, CDKNIA, VEGFA and EMX1 (Supplementary Table 1). VEGFA and EMX1 crRNA sequences were obtained from Fu et al. (2013). For saturation studies, 266 crRNAs to most positions across the VCP were designed (Supplementary Table 2).

For specificity studies, two functional target sequences were selected (VCP_2: 5′ UCAUCAACAAUUAACCGAUU 3′ and PSMD7_1: 5′ AGCGGGGUGUGUCGCGAUGC 3′). For each one, a set of alternate sequences with disruptive 2-base mismatches across the first 20 targeting nucleotides was designed by replacement of each base with its Watson–Crick partner (A–U, U–A, G–C, and C–G) at two separate positions in all possible combinations of positions. This resulted in 190 2-base mismatch sequences for each of the two crRNAs (Supplementary Table 3).

The tracrRNA is a 74-nucleotide RNA (5′ AACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACU-UGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU 3′) based on the published *S. pyogenes* tracrRNA sequence (Jinek et al., 2012).

2.3. Synthesis of crRNA and tracrRNA

All crRNAs and tracrRNA were chemically synthesized on solid-phase support using 5′-silyl-2′-orthoester (ACE) chemistry (Scaringe, 2001; Wincott et al., 1995). crRNAs were deprotected, desalted, and used without further purification. The tracrRNA was additionally HPLC-purified.

2.4. Transfections

HEK293T, Ubi-GFP U2OS–Cas9, U2OS and ESD3 cells were seeded in a 96-well plate in a density appropriate for each specific assay. The following day crRNA and tracrRNA were individually resuspended in 10 mM Tris–HCl (pH7.5), 100 mM NaC, and 1 mM EDTA. 100 μM crRNA and tracrRNA were added together and the RNA was further diluted to 5 μM using sterile 1X siRNA Buffer (Dharmacon, B-002000-UB-100). A final concentration of 50 nM crRNA:tracrRNA (50 nM of each crRNA and tracrRNA) was used for transfection. The cells were transfected with DharmaFECT Duo Transfection Reagent (Dharmacon, # T-2010-03) using 100 ng of Cas9 plasmid (for co-transfections) and 50 nM crRNA:tracrRNA complex.

2.5. Genomic DNA isolation and mismatch-detection assay

Genomic DNA was isolated 72 h post-transfection by direct lysis of the cells in Phusion HF buffer (Thermo Scientific, #F-518L), proteinase K and RNase A for 1 h at 56 °C followed by heat inactivation at 96 °C for 5 min. PCR was performed with primers flanking the cleavage sites (Supplementary Table 4). 500 ng of PCR products was treated with T7 endonuclease I (T7EI; NEB, #M0302L) for 25 min at 37 °C and the samples were separated on a 2% agarose gel. Percent editing in each sample was calculated using ImageJ.

2.6. Sanger sequencing

Blunt-end PCR products generated for the mismatch-detection assay were cloned into pCR-Blunt using the Zero Blunt PCR Cloning Kit according to the manufacturer's protocol (Life Technologies, #K2700-20). Transformed colonies were picked into 96-well plates and grown overnight in NZYM PLUS broth (Teknova, #N1220) and 8% glycerol. Frozen 96-well plates were sent to Oblique Bio, Inc. (Huntsville, AL) for mini-prep and Sanger sequencing. Chromatograms were analyzed with Geneious Version 6.1.8, created by Biomatters (www.geneious.com).

2.7. Functional screening of gene editing

2.7.1. Proteasome assay

Ubi-GFP U2OS–Cas9 cells were plated at 3500 cells per well in Costar Black 96-well tissue culture treated plates (Costar, #3916) in DMEM high glucose supplemented with 10% fetal bovine serum. Cells were incubated at 37 °C, 5% CO_2 for 24 h before transfection. Cells were transfected in triplicate using 0.08 μL/well DharmaFECT #3 (Dharmacon, T-2003-01) diluted in reduced serum media (HyClone, SH30564.01). Cells were returned to the 37 °C, 5% CO_2 incubator for 72 h prior to analysis.

2.7.2. ArrayScan detection

Cells were washed with 1X DPBS and treated with 4% paraformaldehyde for 30 min. Cells were then washed two times with 1X DPBS and Hoechst stained for 5 min. Plates were analyzed on the Cellomics ArrayScan VTI HCS Reader (Thermo Scientific) at 5 × magnification, using a target activation method in the CellomicsScan Software (Thermo Scientific), where GFP-positive cells are identified as selected cells. A total of nine fields of view at 5× were imaged per well. Total number of cells and percent selected cells were reported and used for analysis. Individual images were analyzed for artifacts that could be erroneously calculated as GFP-positive cells by the analysis software; any wells containing these artifacts were discarded from the data set.

2.7.3. Data analysis

Percent selected (GFP positive) cells per well for each biological triplicate were normalized to average of percent selected cells for the positive control (perfect match sequence) on each plate. Average normalized percent GFP positive cells and standard deviation for each biological triplicate is reported. Robust z-score (x – median(sample)/MAD(sample) × 1.4826 where x = sample value and MAD is the median absolute deviation of the non-control samples) was also calculated. Any 2-base-mismatched crRNA to the targets VCP_2 or PSMD7_1 with a robust z-score of >100 was selected as a hit; this score threshold was chosen as it identifies all positive controls and no negative controls.

3. Results

3.1. Efficient gene editing using Cas9 and synthetic crRNA and tracrRNA

Most published studies using CRISPR–Cas9 for gene editing have used a sgRNA cloned into and expressed from a plasmid to direct target cleavage. A wide range of gene editing efficiencies have been reported (<1–>50%), depending on the sgRNA and the target. While the use of a sgRNA is convenient for expression from a plasmid, this approach requires cloning and sequence verification of every construct; this approach is not amenable to high-throughput screening of crRNA sequences. We referred back to the original bacterial Type II CRISPR system and used a two-RNA approach, combining expression of the Cas9 protein with delivery of two synthetic RNA molecules endogenous to the bacterial CRISPR–Cas9 system: the processed (or mature) 42-mer crRNA and the 74-mer tracrRNA (nucleotides 15–89 of the primary transcript, derived from Supplementary Table 1 of Jinek et al. (2012)) (Fig. 1A). We synthesized these two RNA molecules using 2′ACE RNA chemistry, resulting in high yield and purity of full-length RNA.

We tested the efficiency of genome editing using synthetic crRNA and tracrRNA molecules on several gene targets, including those that have been published using sgRNAs. A mismatch-detection assay (for example, SURVEYOR or T7EI; Fig. 1B–E) is typically used for the detection and quantification of insertions and deletions (indel) mutations created by the CRISPR–Cas9 system (Vouillot et al., 2015). As shown in Fig. 1B, co-transfection of a plasmid expressing Cas9 in combination with synthetic crRNA and tracrRNA results in indels at the targeted site. The dual synthetic RNA approach is also effective in multiple cell lines (Supplementary Fig. 1A,B). To evaluate the off-target profile of Cas9 targeting using synthetic crRNAs and tracrRNA, we designed and synthesized crRNA molecules to three sequences in the human genome (two sites in VEGFA and one site in EMX1) that have previously been shown by a mismatch-detection assay to result not only in effective target cleavage but also in off-target editing at other sites in the genome (Fu et al., 2013). We detected off-target editing at the same locations reported for sgRNAs (Fig. 1C–E, Supplementary Fig, 2), and the efficiency of these events was consistent with efficiencies reported for sgRNAs (Mali et al., 2013), as detected by the T7EI mismatch-detection assay. These data demonstrate that the use of synthetic crRNA and tracrRNA molecules results in a similar pattern of on- and off-targeting in mammalian cells when compared to expressed sgRNAs.

3.2. Evaluation of a functional reporter assay to assess gene editing in a high-throughput manner

Because the sequence of the tracrRNA is constant regardless of the crRNA sequence, hundreds or even thousands of different crRNA targeting sequences can be evaluated in a high-throughput manner simply by synthesizing the 42-mer crRNA sequence and co-transfecting it with tracrRNA. We applied this approach to design a high-throughput assay which evaluates CRISPR–Cas9-mediated on- and off-targeting in mammalian cells. Although mismatch-detection assays (Fig. 1B–E) are fairly straight-forward, they are time-consuming, require identification and validation of PCR primers for each target site, and are not suitable for high-throughput screening of crRNAs. We therefore validated a reporter cell line stably expressing EGFP fused to a non-cleavable mutant ubiquitin moiety as a functional readout of CRISPR–Cas9 gene editing (Dantuma et al., 2000). These Ubi-GFP U2OS cells constitutively degrade the EGFP fusion through the proteasome pathway, resulting in little to no GFP fluorescence. However, if the proteasome complex or pathway is compromised or disrupted, the EGFP fusion

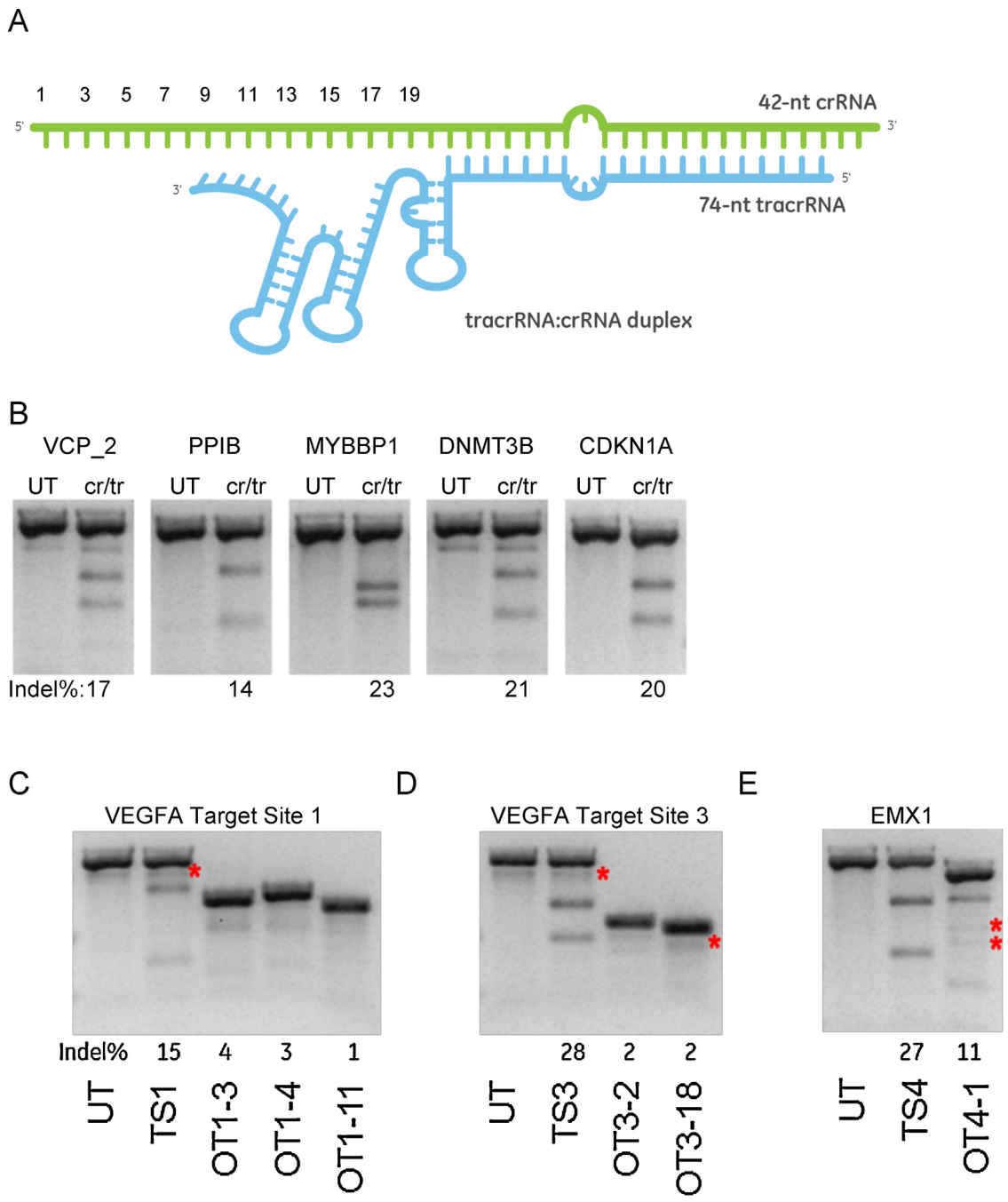
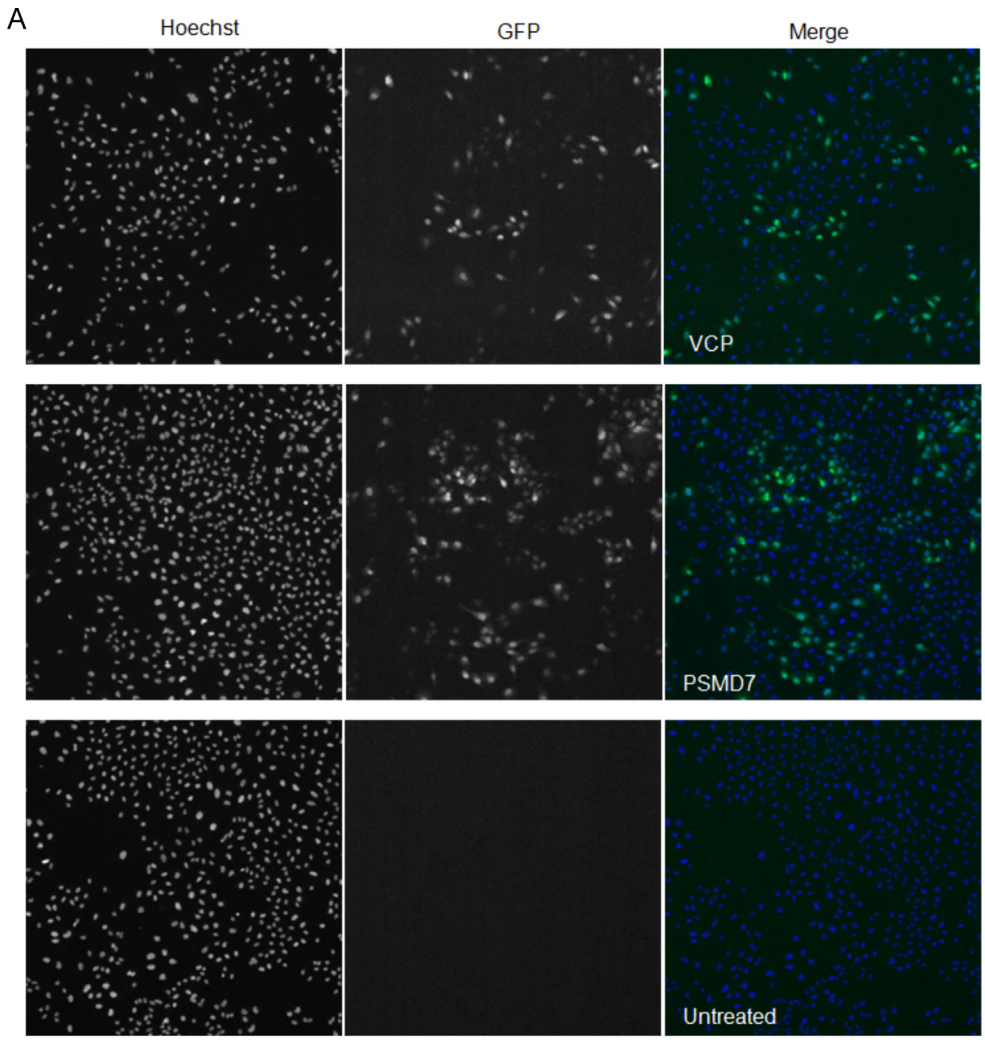


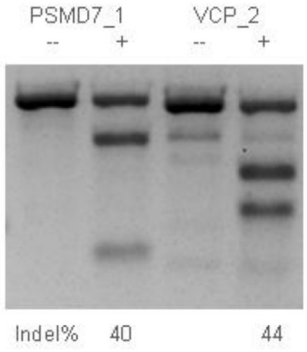
Fig. 1. CRISPR–Cas9-mediated gene editing using synthetic crRNAs and tracrRNA. (A) Diagram showing genome engineering using the CRISPR–Cas9 system consisting of the synthetic crRNA and tracrRNA. (B) Co-transfection of a plasmid expressing a humanized Cas9 and synthetic crRNA and synthetic tracrRNA results in targeted DNA cleavage in HEK293T cells. Cells were transfected with Cas9 plasmid, synthetic crRNA, and synthetic tracrRNA, and compared to untransfected cells (UT) using the T7 endonuclease I (T7EI) assay. The percent editing after 72 h is indicated at the bottom of each lane. (C–E) Transfections were performed as in (B) and off-target gene editing with synthetic crRNA and tracrRNA was determined for VEGFA (C–D) and EMX1 (E) as previously validated with sgRNA (Fu et al., 2013). Red asterisks indicate non-specific bands that are present in untransfected samples and were not included in the indel quantification (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

is no longer constitutively degraded and cells begin to accumulate Ubi-GFP.

We additionally stably integrated the Cas9 gene into this reporter cell line and performed CRISPR–Cas9 gene editing by transfecting synthetic crRNAs and tracrRNA targeting known components of the proteasome that we had previously validated using siRNAs (data not shown). As seen in Fig. 2A, crRNAs targeting VCP and PSMD7 resulted in detectable levels of EGFP, indicating that proteasome function had been disrupted in these cells, while there was no increase in EGFP fluorescence in untreated cells. The transfections were performed at a saturating concentration of 50 nM crRNA and tracrRNA to ensure maximal editing activity. The sensitivity of this reporter assay is high due to the very low levels of background in untreated cells (0–5 cells, where in an average well approximately 7700 cells are analyzed; 0.06%); we can therefore confidently detect positive functional activity with statistical significance down to 10% or less activity of the perfectly-matched crRNA. For instance, for a perfectly-matched target, 1500 cells are functionally disrupted in proteasome activity and therefore are EGFP-positive (22%). While 6 cells can theoretically be



B



C

Target Site Sanger Sequencing

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

Fig. 2. Gene editing in Ubi-GFP U2OS–Cas9 cells. (A) Ubi-GFP U2OS–Cas9 cells were transfected with synthetic crRNAs and tracrRNA targeting two genes involved in proteasome function, VCP and PSMD7 (using the VCP_2 crRNA or the PSMD7_1 crRNA), or were left untreated. GFP phenotype was assessed by high-content imaging 72 h post-transfection, and representative images are shown. The percent GFP-positive cells was quantified by calculating average values from biological triplicates: VCP = 10.3 ± 1.2%, PSMD7 = 10.6 ± 0.5%, Untreated = 0.0 ± 0.0%. (B) Assessment of gene editing using the T7EI mismatch-detection assay in untreated (−) 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 sequencing individually cloned PCR products from untreated (UT) or cells treated with the indicated crRNA and tracrRNA.

detected above background (0.08%), we conservatively picked a statistical threshold for significance corresponding to approximately 150 cells per well (~2%), resulting from a Robust z-score calculation, as described in Section 2. We also evaluated treated cells for the presence of indels using the T7EI assay (Fig. 2B). We consistently observed that the indel mutation percentages seen using a mismatch-detection assay are higher in the Ubi-GFP U2OS–Cas9 cell line than those seen in co-transfection experiments (Supplementary Fig. 1A compared to Fig. 2B); possible explanations for this discrepancy include transfection efficiency of the Cas9-expressing

plasmid or levels of Cas9 protein required for efficient gene editing. To further analyze the indels in the PSMD7 and VCP genes after transfection with PSMD7- and VCP-targeting crRNAs (PSMD7_1 and VCP_2, respectively), we cloned the PCR products containing the target sites from transfected cell populations into *Escherichia coli* and sequenced individual colonies. As shown in Fig. 2C and Supplementary Table 5, sequencing results showed a variety of indels around the crRNA target site. To rigorously control for detection of indels by sequencing, untransfected cells were sequenced at the editing target site to detect the background levels of indels that are likely due to inherent errors in the PCR, sequencing method and possibly mutations in the population. As expected, the frequency of mutations in the treated population was significantly higher than untransfected background and consistent with the frequency determined by T7EI. This background level is not due to Cas9 expression alone in this cell line and is consistent with reported error rates of PCR and Sanger sequencing (Ewing et al., 1998). The EGFP fluorescence results (Fig. 2A) show lower levels of gene editing than the sequencing results (Fig. 2C), but can be explained by the fact that the EGFP fluorescence detected in this assay will only reflect indels that result in functional disruption of the target protein and is therefore likely to be an underestimate of the total number of indels. In addition, functional disruptions will require disruptive indels in all (or most) alleles of the gene; thus, in U2OS cells that are known to be polyploid (Janssen and Medema, 2013), indels may occur at one or more alleles but not be sufficient to disrupt the proteasome to the extent that results in the GFP-positive phenotype in this assay. Nonetheless, significant levels of functional mutations are detected in this assay using the CRISPR–Cas9 system, confirming that the Ubi-GFP U2OS–Cas9 cell line can be used as a high-throughput assay for determining crRNA functionality. We used the assay to test 266 crRNAs (each proximal to an NGG PAM) across all coding exons of the VCP gene. Our results indicate that many crRNAs (244 of 266 or 92%) exhibit some level of functionality, as defined by >1.1-fold increase in GFP fluorescence, or disruption of proteasome function leading to the stability and detection of Ubi-GFP (Supplementary Table 2 and Supplementary Fig. 3), highlighting the fact that filtering of crRNAs based on specificity predictions will likely still allow for many functional crRNAs to a given gene.

3.3. *Functionality of crRNAs containing single, double, or triple mismatches to the intended targets*

To evaluate the potential of crRNAs to target and cleave off-target sequences with imperfect complementarity, we synthesized crRNAs with mismatches to the VCP and PSMD7 genes as a proxy for off-targeting in the Ubi-GFP U2OS–Cas9 cell line. First we synthesized crRNAs containing a single disruptive transversion mismatch to the DNA target (A to U, C to G, U to A, G to C) throughout the 20-nucleotide length of the crRNA targeting VCP (VCP_2 crRNA) or PSMD7 (PSMD7_1 crRNA) (Supplementary Table 3). Transversion mismatches were selected to narrow down the possible sequence space from a mismatch matrix that rapidly becomes intractable when considering combinations of two or more mismatches, while still sampling comprehensively across every position of the targeting region of the crRNA. As previously described (Cho et al., 2014; Fu et al., 2013; Hsu et al., 2013) using both a reporter mismatch assay and a sampling of predicted endogenous off-targets (Fu et al., 2013), single-base mismatches are tolerated (resulting in detectable and significant mismatch targeting relative to the on-target activity), particularly if the mismatches occur in the PAM-distal end of the crRNA, away from the seed region, defined here as nucleotides 1–10 of the crRNA (Fig. 3A).

We then synthesized crRNAs with all possible combinations of two-nucleotide disruptive mismatches to the DNA target for the VCP_2 crRNA and the PSMD7_1 crRNA, yielding 190 mismatch-containing crRNAs per gene (Supplementary Table 3). We evaluated the ability of these 380 crRNAs containing two mismatches to their targets to create indels in their respective gene (thereby resulting in EGFP fluorescence), despite having imperfect complementarity to the target. As shown in Fig. 3B and C (light blue data points in inset graphs), fewer than 5% of the crRNA sequences that contain two disruptive mismatches to their target are able to effectively cleave and create mutations in their target, as detected by EGFP fluorescence. For the crRNA targeting VCP (VCP_2), only 7 of the 190 crRNAs (3.7%) with two mismatches remain able to target the VCP gene as assessed by a statistical cut-off of robust z score > 100 (Fig. 3B and Table 1), which is roughly 10% relative activity compared to the on-target effect. As described earlier, a few mismatched sequences exhibited signal in this assay even below this threshold, but they would not be robustly reproducible enough to produce a functional effect. Therefore we deem them statistically insignificant in terms of representing what might truly be detected in a cell population as proxy for an endogenous off-target effect. For the crRNA targeting PSMD7 (PSMD7_1), only 8 of the 190 crRNAs (4.2%) with two mismatches remain able to target the PSMD7 gene using the same statistic (Fig. 3C and Table 1). Using T7EI, we confirmed that the 15 mismatch-containing crRNAs targeting VCP and PSMD7 resulting in EGFP fluorescence do indeed trigger indel formation and for these functional target sites, the percent editing from the mismatch-detection assay correlates with the magnitude of phenotypic readout (Fig. 4A and B). Of the 15 sequences that are able to cleave the target DNA site despite the presence of two mismatches, 11 of these contain a mismatch at position 1, or the most PAM-distal position. This result is consistent with previous data demonstrating that the identity of position 1 is not important for effective gene editing; indeed, many studies using expressed sgRNAs replace position 1 with a G, which is a requirement of the U6 promoter driving sgRNA expression (Cong et al., 2013; Koike-Yusa et al., 2014; Fu et al., 2013). Our results show that both positions 1 and 2 are tolerant of mismatches between the RNA and target DNA for both crRNAs that we tested, suggesting that mismatch tolerance may extend beyond position 1. Consistent with the single-mismatch data (Fig. 3A), those mismatches closer to the seed region (at positions 9, 10 and 13 from the 5′ PAM-distal end) are also tolerated in spite of this proximity.

For every crRNA, there are 1140 possible sequences containing three mismatches to the target sequence, even if only disruptive transversion mismatches as described above are considered. Given that most of the two-mismatch sequences tested did not result in off-targeting and the three mismatch set is large, we chose to evaluate only a subset of three-mismatch sequences that we anticipated were the most likely to off-target, rather than the entire set (Supplementary Table 3). Two approaches for choosing three-mismatch sequences were used: first, to those two-mismatch sequences that resulted in the most off-targeting, as shown above, we added a disruptive mismatch at each remaining position of the targeting 20-mer. Second, positions in the two-mismatch study that showed any off-targeting were combined and tested as three-mismatch combinations. We found that most of the three-mismatch crRNA sequences showed no functional targeting (Fig. 3D, Supplementary Fig. 4) with the exception of the sequence with mismatches at positions 1, 2, and 3 of the PSMD7_1 crRNA, which was able to target at low levels. This result is consistent with the single- and double-mismatch data, suggesting that nucleotide identity and RNA:DNA complementarity and binding at the 5′ PAM-distal end is not required for nuclease activity and that mismatches located at these positions can be sources of off-targeting.

Although the overall number of functional two-mismatch sequences was low, some sequences resulted in a relatively high percentage of gene editing at the target despite the presence of

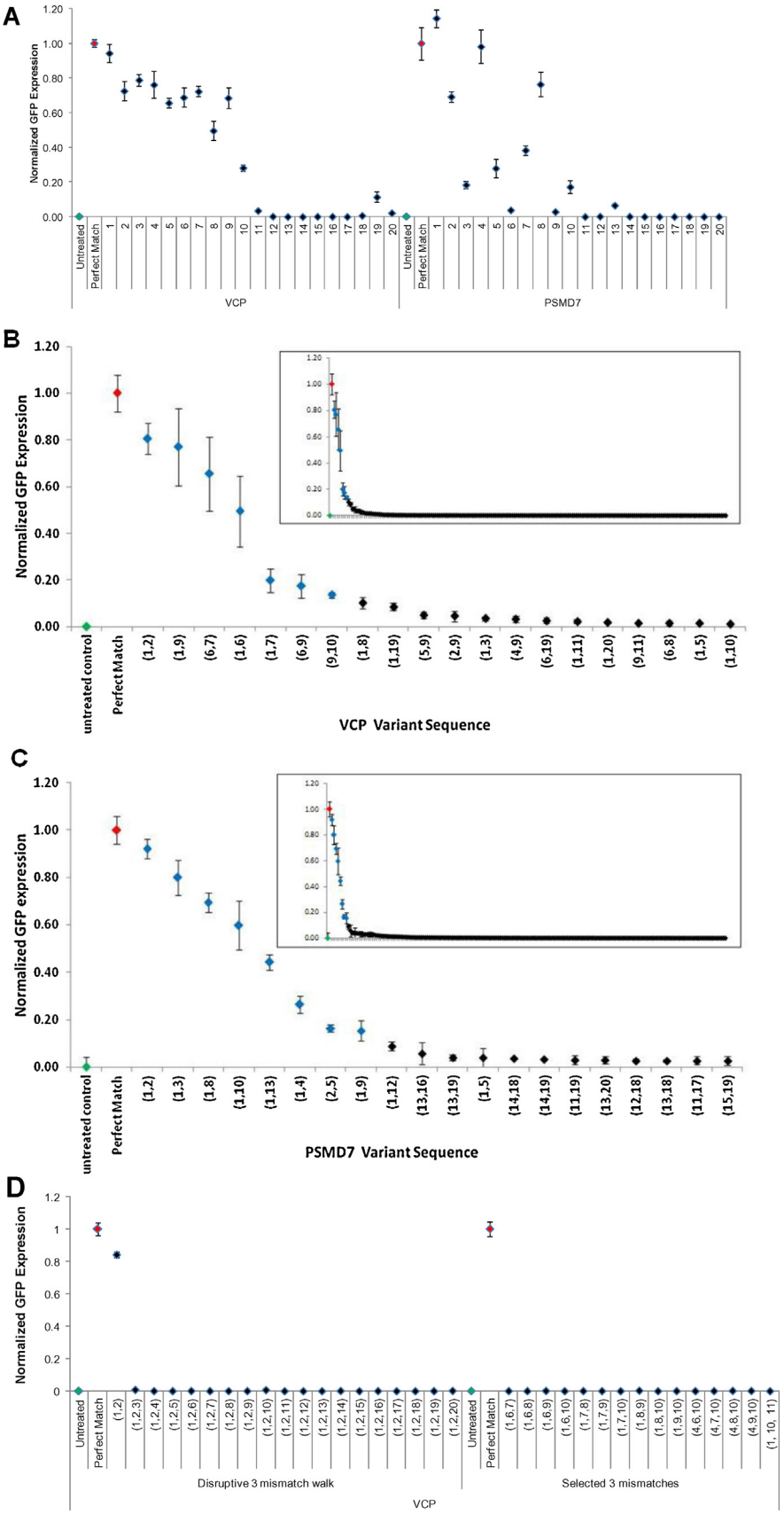


Fig. 3. Gene targeting and relative off-targeting in a phenotypic assay. (A) Functionality of crRNAs containing single-base mismatches at every position of the 20-nt crRNA spacer sequence using the Ubi-GFP U2OS–Cas9 EGFP fluorescence assay. (B) A ranked plot of 2-base-mismatched crRNA variants of VCP_2. The 20 highest-signal mismatched crRNAs are plotted in the main graphic and the full set of 190 crRNAs is shown in the inset. (C) A ranked plot of 2-base-mismatched crRNA variants of PSMD7_1. The 20 highest-signal mismatched crRNAs are plotted in the main graphic and the set of 190 crRNAs is shown in the inset. Untreated sample control (green), perfect match crRNA control (red), 2-base-mismatched crRNAs exhibiting a phenotype with robust z-score > 100 (blue), and crRNAs with insignificant signal (black). (D) Functionality of crRNAs containing three mismatches to the target site in VCP. Untreated sample control (green), perfect match crRNA control (red) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 1
Significantly off-targeting crRNA guide sequences.

Name	crRNA guide sequence																				Avg%	Std Dev	Robust z-score
VCP_2 (1,2)	a	g	A	U	C	A	A	C	A	A	U	U	A	A	C	C	G	A	U	U	0.81	0.07	785
VCP_2 (1,6)	a	C	A	U	C	u	A	C	A	A	U	U	A	A	C	C	G	A	U	U	0.50	0.15	482
VCP_2 (1,7)	a	C	A	U	C	A	u	C	A	A	U	U	A	A	C	C	G	A	U	U	0.20	0.05	193
VCP_2 (1,9)	a	C	A	U	C	A	A	C	u	A	U	U	A	A	C	C	G	A	U	U	0.77	0.17	750
VCP_2 (6,7)	U	C	A	U	C	u	u	C	A	A	U	U	A	A	C	C	G	A	U	U	0.66	0.16	638
VCP_2 (6,9)	U	C	A	U	C	u	A	C	u	A	U	U	A	A	C	C	G	A	U	U	0.17	0.05	169
VCP_2 (9,10)	U	C	A	U	C	A	A	C	u	u	U	U	A	A	C	C	G	A	U	U	0.14	0.01	132
PSMD7_1 (1,2)	u	c	C	G	G	G	G	U	G	U	G	U	C	G	C	G	A	U	G	C	0.92	0.06	896
PSMD7_1 (1,3)	u	G	g	G	G	G	G	U	G	U	G	U	C	G	C	G	A	U	G	C	0.80	0.04	779
PSMD7_1 (1,4)	u	G	C	c	G	G	G	U	G	U	G	U	C	G	C	G	A	U	G	C	0.26	0.03	257
PSMD7_1 (1,8)	u	G	C	G	G	G	G	a	G	U	G	U	C	G	C	G	A	U	G	C	0.69	0.07	676
PSMD7_1 (1,9)	u	G	C	G	G	G	G	U	c	U	G	U	C	G	C	G	A	U	G	C	0.15	0.02	150
PSMD7_1 (1,10)	u	G	C	G	G	G	G	U	G	a	G	U	C	G	C	G	A	U	G	C	0.60	0.04	582
PSMD7_1 (1,13)	u	G	C	G	G	G	G	U	G	U	G	U	g	G	C	G	A	U	G	C	0.44	0.10	431
PSMD7_1 (2,5)	A	c	C	G	c	G	G	U	G	U	G	U	C	G	C	G	A	U	G	C	0.16	0.04	159

two mismatches. In order to further characterize the functionality of these mismatched crRNAs, we performed a dose-curve analysis of the 3 strongest double-mismatch sequences for the PSMD7_1 and VCP_2 crRNAs (mismatches at positions 1, 2; 1, 3; and 1, 8 for PSMD7_1 and mismatches at positions 1, 2; 1, 9; and 6, 7 for VCP_2). Results from the phenotypic assay for both PSMD7_1 and VCP_2 crRNAs indicate that while there is a reduction in editing activity of crRNAs containing two mismatches, there is a concurrent reduction in editing efficiency of the perfect match crRNA (Fig. 4C and D). This suggests that a reduction in concentration of crRNA cannot effectively reduce off-targeting without also reducing on-target editing. The data also show that gene editing efficiency begins to plateau at a concentration of approximately 10 nM crRNA:tracrRNA complex, consistent with Cas9, rather than crRNA or tracrRNA, being the limiting component for editing.

4. Discussion

Specificity of all gene editing technologies is a concern for researchers since genetic background is paramount for experimental interpretation. CRISPR–Cas9 gene editing technology off-targeting is of particular concern because this technology relies on sequence binding, and imperfect complementarity between the crRNA target sequence and genomic DNA can potentially direct unintended editing.

While the effect of mismatches to target (when generated in an sgRNA targeting sequence) has been considered previously (Cho et al., 2014; Fu et al., 2013; Pattanayak et al., 2013; Hsu et al., 2013), a comprehensive study of the effects of all combinations of two mismatches between the crRNA and target DNA has not been published to date. We used a novel high-throughput approach

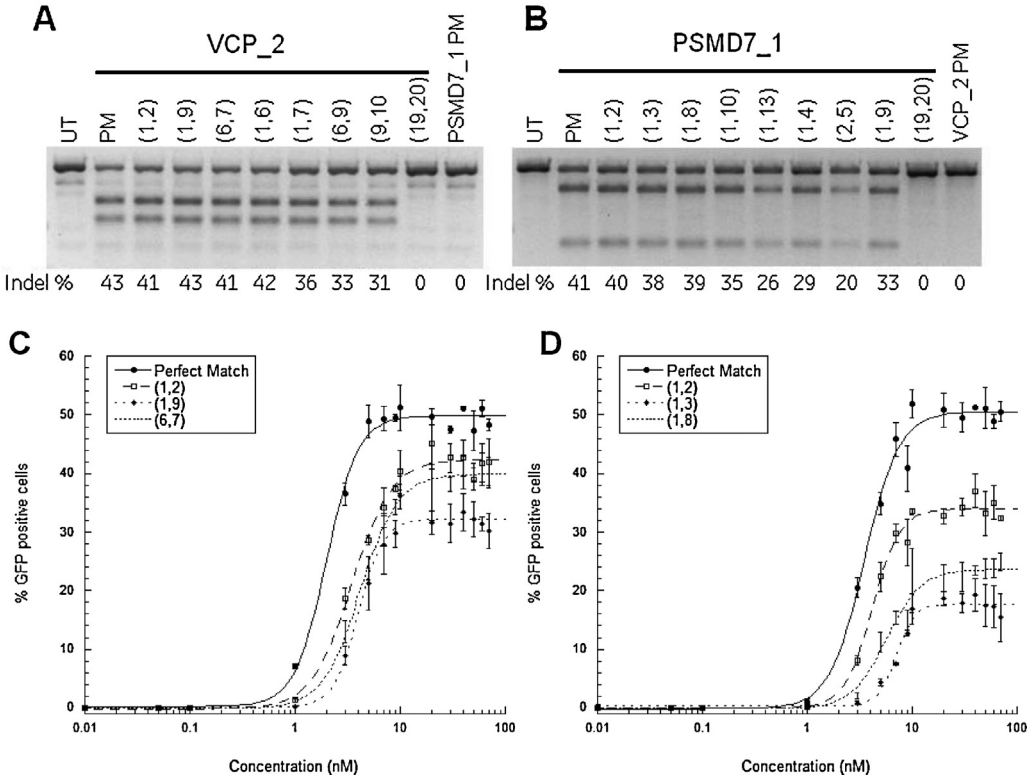


Fig. 4. Characterization of off-targeting crRNAs: T7EI mismatch-detection assay of crRNAs that are either perfectly matched to their target or contain two mismatches in the crRNA. (A) VCP_2. (B) PSMD7_1. Dose curve of crRNAs targeting (C) VCP and (D) PSMD7, both with complete complementarity as well as with the indicated double mismatches, assessed using the EGFP phenotypic assay.

involving synthesis of synthetic crRNAs along with a sensitive, functional reporter assay that identifies crRNA target sites that result in functional gene knockout. This assay will only identify crRNA sites that result in a functional knockout and will not be useful for the study of crRNA sites that result in indels but do not result in functional knockout likely due to the editing position in the gene. Therefore, only crRNA sites that result in functional knockout and phenotype with a large dynamic range were selected for the systematic evaluation of functional gene editing of all positional combinations of double mismatches in the context of two native crRNA sequences. The main advantage of our approach is that the contribution of each position in the targeting 20-mer could be evaluated on a well-by-well basis without concern for whether the mismatched sequence occurs endogenously in the genome, and without the additional confounding effects such as variable PAMs, differing genomic context of targeted sequence, and so forth for each endogenous off-target pairing.

Consistent with previous results, we observed that single mismatches are fairly well-tolerated, especially in the 5′ PAM-distal end of crRNAs. crRNAs containing mismatches in the seed region, defined here as nucleotides 11–20 of the crRNA, did not exhibit activity, with two minor exceptions: positions 13 and 19 showed detectable activity < 10% relative to the perfect match control. Conclusions from the comprehensive examination of two-base mismatches also agree with published data on selected endogenous off-targets. Namely, consecutive two-base mismatches are generally not active, with the exception of positions 1 and 2. Two-base mismatches exhibiting function are well-separated (in the non-seed region) and most often involve one mismatch at the most distal end from the PAM, indicating that the identity of this base is dispensable for activity and specificity of Cas9. Tolerance of mismatches distal to the PAM is consistent with the PAM surveillance mechanism and conformational changes required for Cas9 cleavage that have been proposed in biochemical studies (Sternberg et al., 2014; Szczelkun et al., 2014).

Surprisingly, the overall number of active two-base-mismatched sequences was lower than expected based on prior studies (Fu et al., 2013; Hsu et al., 2013) that found a hit rate of 6–26% of endogenous 2- and 3-base mismatched off-target sites (depending on initial targeting sequence) to exhibit fractional cleavage activity. Previous publications concluded that highly GC-rich sequences (60–80%) yield more off-targets containing any number of mismatches between one and three. However we used sequences containing 30% (VCP_2) and 70% GC content (PSMD7_1), respectively, and found an overall hit rate of only 15/380, or 4%. Our results most closely agree with the conclusion of Cho et al. (2014), who stated “RGENs cannot efficiently cleave chromosomal DNA with mismatches of two or more nucleotides, at least one of which occurs in the seed region”. Considering that either on-target or off-target activity will strictly require a PAM, (thus limiting the vast number of possible 1-, 2-, and 3-base-mismatched off-targets to only the subset which occur in the genome adjacent to a PAM) and that only a small fraction of these possible sites, in context of the current study, show activity, our results support the conclusion that off-targeting is not as much of an impediment to the use of CRISPR–Cas9 as was initially suggested. Several methods have recently emerged that enable profiling of genome-wide off-target cleavage by CRISPR–Cas9 (Tsai et al., 2015; Frock et al., 2015; Wang et al., 2015). While GUIDE-Seq may provide an approach to detect genome-wide off-target cleavage for single sgRNAs in a low throughput manner, our method allows detection of functional genome editing and serves as a proxy method for off-targeting through the use of numerous, unbiased mismatches to a crRNA with similar sensitivity to these published methods but also in a high-throughput manner with respect to both targets and off-targets detected.

Bioinformatic approaches have been used to avoid designing crRNAs that have potential off-target sites in the genome with a small number of mismatches (1 or possibly 2). However, when filtering out all crRNAs with any 2- and 3-base-mismatched sites against the entire human genome, the number of crRNAs remaining for any given target rapidly decreases. Since crRNA functionality differs, and several crRNAs should be tested per target to ensure effective gene editing, a solely bioinformatic approach is not a viable strategy to identify highly specific crRNAs. In addition, for creating gene knockins through homology-directed repair, the location of the CRISPR–Cas9-mediated DSB is even more important than for knockout. Therefore, understanding the off-target potential of a given crRNA that may have one or more mismatches to potential targets in the genome remains important. Several papers have proposed modifications of the standard CRISPR–Cas9 system to enhance specificity and reduce off-target cleavage (Ran et al., 2013; Fu et al., 2014; Tsai et al., 2014). While these techniques have been shown to increase the overall specificity of CRISPR–Cas9, some off-targeting will remain, and in some cases enhanced specificity using these modified CRISPR–Cas9 systems may come at the expense of decreased on-targeting efficiency. In cases where the mismatched off-target has mismatches only in the PAM-distal region, off-targets containing mismatches at position 1 are likely to be prevalent. This information is consistent with a recent publication (Tsai et al., 2015) and further supports the need to consider position 1 in crRNA design rules. More target/off-target sets will need to be tested to predict precise bioinformatic rules for off-targeting; however, the rapid ability to test many synthetic crRNAs per gene target and some understanding of mismatch tolerance are essential to triage designs for highly specific and functional crRNAs.

Conflict of interest

All authors are employed or were formerly employed by Dharmacon. Some of the materials used in this study are products sold by Dharmacon. This does not alter our adherence to all the Elsevier Publishing Ethics.

Acknowledgement

This work was funded by Dharmacon, part of GE Healthcare. We acknowledge Dharmacon Manufacturing for producing synthetic RNA molecules.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2015.06.427

References

Cho, S.W., Kim, S., Kim, J.M., Kim, J.S., 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 31, 230–232.
Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., Kim, J.S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132–141.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.
Dantuma, N.P., Lindsten, K., Glas, R., Jellne, M., Masucci, M.G., 2000. Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. Nat. Biotechnol. 18, 538–543.
Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., Root, D.E., 2014. Rational design of highly active sgRNAs for CRISPR–Cas9-mediated gene inactivation. Nat. Biotechnol. 32, 1262–1267.
Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175–185.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., Alt, F.W., 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186.
Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826.
Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., Joung, J.K., 2014. Improving CRISPR–Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32, 279–284.
Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832.
Janssen, A., Medema, R.H., 2013. Genetic instability: tipping the balance. Oncogene 32, 4459–4470.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.
Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., Doudna, J., 2013. RNA-programmed genome editing in human cells. Elife 2, e00471.
Kappes, J.C., Wu, X., 2001. Safety considerations in vector development. Somat. Cell Mol. Genet. 26, 147–158.
Kappes, J.C., Wu, X., Wakefield, J.K., 2003. Production of trans-lentiviral vector with predictable safety. Methods Mol. Med. 76, 449–465.
Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C., Yusa, K., 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. 32, 267–273.
Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., Bao, G., 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 42, 7473–7485.
Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-guided human genome engineering via Cas9. Science 339, 823–826.
Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., Liu, D.R., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839–843.
Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al., 2014. CRISPR–Cas9 knockin mice for genome editing and cancer modeling. Cell 159, 440–455.
Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389.
Scaringe, S.A., 2001. RNA oligonucleotide synthesis via 5′-silyl-2′-orthoester chemistry. Methods 23, 206–217.
Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al., 2014. Genome-scale CRISPR–Cas9 knockout screening in human cells. Science 343, 84–87.
Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., et al., 2014. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell stem cell 15, 12–13.
Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67.
Swiech, L., Heidenreich, M., Banerjee, A., Habib, N., Li, Y., Trombetta, J., Sur, M., Zhang, F., 2015. In vivo interrogation of gene function in the mammalian brain using CRISPR–Cas9. Nat. Biotechnol. 33, 102–106.
Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., Seidel, R., 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. U. S. A. 111, 9798–9803.
Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., Joung, J.K., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. 32, 569–576.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. Nat. Biotechnol. 33, 187–197.
Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Talkowski, M.E., Musunuru, K., 2014. Low incidence of off-target mutations in individual CRISPR–Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell 15, 27–30.
Vouillot, L., Thelie, A., Pollet, N., 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. G3 (Bethesda) 5, 407–415.
Wang, T., Wei, J.J., Sabatini, D.M., Lander, E.S., 2014. Genetic screens in human cells using the CRISPR–Cas9 system. Science 343, 80–84.
Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., Yee, J.K., 2015. Unbiased detection of off-target cleavage by CRISPR–Cas9 and TALENs using integrase-defective lentiviral vectors. Nat. Biotechnol. 33, 175–178.
Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., Usman, N., 1995. Synthesis, deprotection, analysis and purification of RNA and ribozymes. Nucleic Acids Res. 23, 2677–2684.
Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., et al., 2014. CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature 514, 380–384.