RESEARCH ARTICLE

NANO · MICRO small
www.small-journal.com

Compartmentalized CRISPR Reactions (CCR) for High-Throughput Screening of Guide RNA Potency and Specificity

*Tinku Supakar, Ashley Herring-Nicholas, and Eric A. Josephs**

CRISPR ribonucleoproteins (RNPs) use a variable segment in their guide RNA (gRNA) called a spacer to determine the DNA sequence at which the effector protein will exhibit nuclease activity and generate target-specific genetic mutations. However, nuclease activity with different gRNAs can vary considerably in a spacer sequence-dependent manner that can be difficult to predict. While computational tools are helpful in predicting a CRISPR effector's activity and/or potential for off-target mutagenesis with different gRNAs, individual gRNAs must still be validated in vitro prior to their use. Here, the study presents compartmentalized CRISPR reactions (CCR) for screening large numbers of spacer/target/off-target combinations simultaneously in vitro for both CRISPR effector activity and specificity by confining the complete CRISPR reaction of gRNA transcription, RNP formation, and CRISPR target cleavage within individual water-in-oil microemulsions. With CCR, large numbers of the candidate gRNAs (output by computational design tools) can be immediately validated in parallel, and the study shows that CCR can be used to screen hundreds of thousands of extended gRNA (x-gRNAs) variants that can completely block cleavage at off-target sequences while maintaining high levels of on-target activity. It is expected that CCR can help to streamline the gRNA generation and validation processes for applications in biological and biomedical research.

1. Introduction

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) system (**Figure 1**) has emerged as a powerful tool for targeted gene editing,[1,2] allowing precise and efficient modifications of specific nucleic acid sequences within a genome. The process relies on a cofactor of the Cas9 protein known as a guide RNA (gRNAs) that is designed to be complementary to targeted DNA sequences in a modular segment of the RNA called their spacer.[3,4] Stable base-pairing between the spacer sequence and the target sequence activates the Cas9's nuclease domains to cleave the DNA.[3–7] If Cas9 cleavage occurs within a cell, cellular double-strand break repair (DSB) mechanisms can introduce permanent genetic changes into its genome at the site of the break. Therefore, by simply changing the sequence of the gRNA's spacer, Cas9 can be used to introduce specific mutations into a gene of interest[8–11] which often involves insertions or deletions during mutagenic DSB repair[8,9] that can cause a frameshift knockout of the gene, making it a very useful tool for biological studies and for gene therapies.[1,2]

However, the molecular process of target recognition and cleavage by Cas9 is complex and dynamic: the Cas9 enzyme recognizes a short motif known as a PAM[7] immediately next to the target sequence in the DNA and opens the double helix, allowing the gRNA's spacer sequence to displace the nontargeted strand nucleotide by nucleotide,[4] forming a structure known as an R-loop.[12,13] The formation of a stable R-loop structure drives a conformational change in the Cas9 that positions the nucleases appropriately to generate the DNA DSB. As a result of this dynamic process, predicting the ability of Cas9 to recognize and degrade different sequences is more difficult than, say, the design of PCR primers which is driven largely by the thermodynamics of base-pairing;[14,15] instead, this process, known as strand invasion by the gRNA, is believed to be a kinetic, nonequilibrium process, and the ability of Cas9 to recognize and introduce DSBs into DNA can vary significantly based on multiple properties of the gRNA's spacer sequence[16–26] and its ability to form a stable R-loop after strand invasion.[3–5,12,13] Cas9 RNPs will exhibit little or no nuclease activity for many gRNAs.[16–26] Significant efforts have gone into predicting the activity of Cas9 RNPs with different gRNAs,[16–26] and machine learning models[23–26] have emerged as useful tools for determining gRNA candidates

T. Supakar, A. Herring-Nicholas, E. A. Josephs
Department of Nanoscience
Joint School of Nanoscience and Nanoengineering
University of North Carolina at Greensboro
Greensboro, NC 27401, USA
E-mail: eajoseph@uncg.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.202403496

DOI: 10.1002/smll.202403496

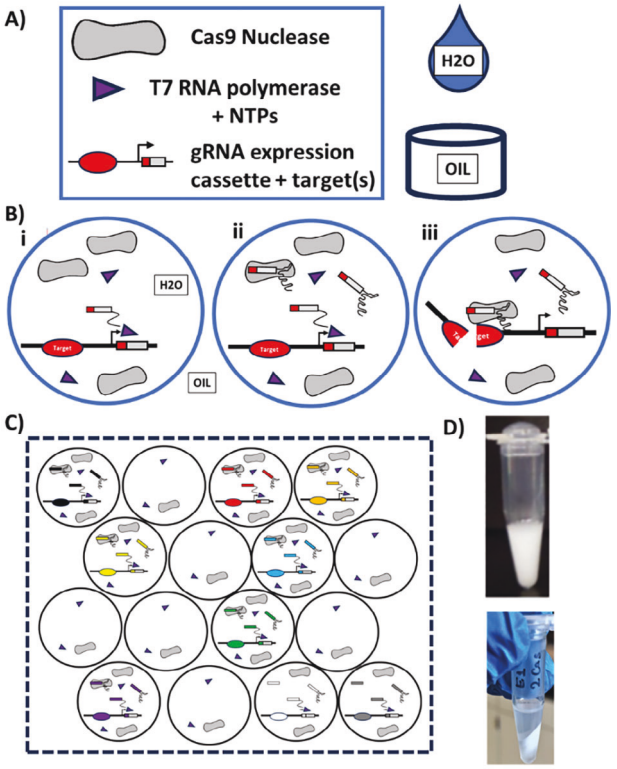

Figure 1. A) Compartmentalized CRISPR Reactions (CCR) are performed in water-in-oil emulsions containing high concentrations of Cas9 nucleases and T7 RNA polymerases with nucleotide triphosphates (NTPs) for RNA transcription, and low concentrations of DNA substrates so that individual DNA molecules are expected to be isolated within a single emulsion. B) Within each emulsion, unique gRNAs can be i) transcribed and ii) assembled into RNPs with Cas9, and then the RNPs can iii) either demonstrate or fail to demonstrate cleavage activity at an (off- or on-) target site of interest. C) If a library of different DNA molecules with different gRNA/target/off-target combinations (different colors) are used when generating the water-in-oil emulsions, large numbers of gRNA/target/off-target combinations can be evaluated simultaneously. D) (above) Photograph of water-in-oil emulsions in a microcentrifuge tube; (below) after the emulsions have been broken after the CCR reaction via centrifugation for subsequent analysis.

to knock out a gene of interest, however, these models often do not have excellent predictive power. Predictions of nuclease activity by these models have shown a correlation coefficient of only ≈0.4 with experimental results, and they depend strongly on the experimental system to which they are compared.[32]

The nuclease domains of Cas9 can also be activated if, during strand invasion, the gRNA is able to stably bypass mispaired nucleotides at off-target sequences that are similar but not quite identical to the intended target sequence. Therefore, off-target effects are also considered when determining a gRNA for a target gene of interest to prevent unwanted mutagenesis.[16,18,25–31] In practice, trade-offs must often be made between the predicted on-target activity and the potential for off-target activities.[16,18,25–32] For CRISPR-based knockout of a gene of interest, computational tools will typically i) enumerate all possible target sequences within the exon sequences of that gene by the presence of a requisite PAM; ii) predict activity using one of many machine learning models; iii) identify the possible off-target sequences by sequence similarity to the target and for each gRNA predict nuclease activity at those sites; then iv) output the results.[32]

It can therefore be difficult to translate the output result of computational tools, as different models for on-target, off-target, and DSB repair outcomes can contradict one another,[32] and multiple gRNAs must be tested or validated in vitro by hand prior to use in order to efficiently generate the desired mutations in cells.[33–35] Conducting this validation for multiple gRNAs within cells one at a time can be cumbersome,[36–39] and often in vitro screening[34] is performed to assess nuclease activity using purified RNPs as a standard course. This too requires transcribing, purifying, and combining gRNAs individually with Cas9. Assessing large numbers of spacer/target/off-target combinations presents many practical difficulties[40,41] for the purpose of determining the most effective and most precise gRNA for a given application.

Here, we present a streamlined approach, inspired by the concept of compartmentalized self-replication (CSR)[41–50] to simultaneously assess CRISPR nuclease activity and specificity in vitro for multiple gRNA in parallel, in a process that we call a Compartmentalized CRISPR Reaction (CCR) screen (Figure 1). During a CCR screen, libraries of short (<500 bp), specially designed DNA molecules (**Figure 2**) are individually confined within micron-sized water droplets in water-in-oil emulsions along with high concentrations of T7 RNA polymerase and Cas9 nuclease. The emulsification process involves slowly adding the aqueous phase in a dropwise manner to a stirring mixture of mineral oil and surfactants. The DNA molecules contain the coding sequence for a gRNA downstream of a promoter for T7 RNA polymerase, as well as a targeted DNA sequence (or off-target) on the opposite end of the DNA molecule. Under buffer conditions, we have identified where both T7 RNA polymerase and Cas9 were found to be active

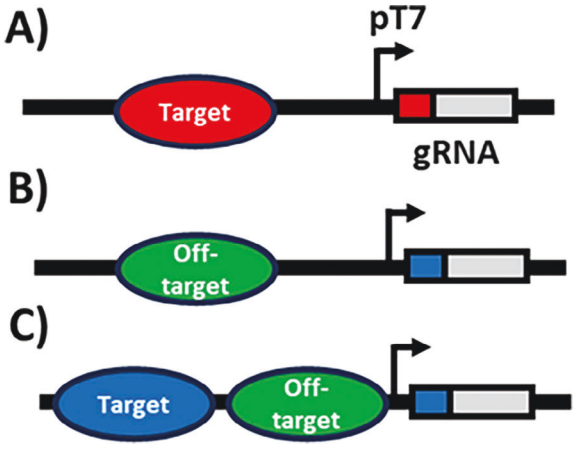

Figure 2. The DNA substrates are designed to contain a module for T7 RNA polymerase-based expression of a gRNA and A) a target for Cas9 with that gRNA, so that activity of the Cas9 within that emulsion with specific gRNA can be determined from the length of the resulting DNA substrates with that spacer sequence (red); B) a potential off-target site (green) for a gRNA of interest (blue); or C) a combination of both, where DNA molecules predominantly exhibiting cleavage patterns at the spacer-distal site would imply a gRNA with significantly higher activity on-target than at spacer-proximal off-target sites.

(Figure S1, Supporting Information), the gRNAs encoded on the isolated DNA molecules are transcribed and can associate with the Cas9 nucleases to form the active Cas9 RNP, which can then interact with the target sequence found on the same molecule within the emulsions (Figure 1B,C). After breaking the emulsions (Figure 1D), removing protein, and determining the length of the DNA molecules and the sequence of the spacer on the segment coding for the gRNA via run-off next-generation sequencing, we can determine the activity of many gRNAs simultaneously, in a single one-pot reaction. Here, this allows us to, for example, i) test the activities of *all* the gRNAs in the first exon of a gene of interest experimentally, ii) evaluate activities at the likely off-target sequences for a gRNA at the same time, and iii) screen hundreds of thousands of gRNAs variants for those with greater specificities and activities than a standard gRNA for a target of interest.

2. Result

Approximately 90% of the water-in-oil emulsions droplets were ≈850 nm in diameter on average (+/−218 nm, standard deviation (stdev.)) measured via dynamic light scattering (DLS), with ≈10% being larger (≈4.2 um in diameter, +/−1.042 um stdev.) (Figure S2, Supporting Information,). A back-of-the-envelope calculation, therefore, suggested that there were 3.5 × 10^14 water-in-oil emulsions per mL[40] and that adding ≈0.1 – 1 femtomoles of the DNA (≈600 million to ≈6 billion molecules) per mL would ensure a ratio of 1 DNA molecules for every 100–1000 emulsions, even within the large droplets. We hypothesized that these conditions would essentially ensure that each DNA molecule would be isolated within its own emulsion while still allowing us to test a large number of gRNA/target/off-targets independently and simultaneously (Figure 1C). To test whether these emulsions would successfully isolate the independent gRNA generation, RNP formation, and nuclease events, we designed a pair of DNA molecules each containing a gRNA expression cassette, its target on the distal end of the DNA molecule, and the target for the paired gRNA positioned proximally between its target and the promotor (**Figure 3**). This way, if each of the DNA molecules were isolated in their own emulsion, we might expect to see DNA with cuts only at the distal target sites, while if both species of the DNA pair were mixed (each in the same emulsion), we might expect to see DNA that was cut at the proximal target site (Figure 3A,B).

We performed this in vitro reaction either under standard in vitro reaction conditions (with no emulsions) (Figure 3A) or within the emulsions (Figure 3B) and determined the size of the resulting DNA molecules after an 8-h reaction period (Figure 3C). Those DNA molecules under standard conditions were essentially all cut at the proximal target site, as both gRNAs were generated together and CRISPR RNPs could easily access both respective targets on either molecule. In the emulsions, the overall fraction of cleaved molecules decreased, but those that were cleaved were highly cleaved at the distal sites (Figure 3C; Figure S3, Supporting information). This implied that most of the DNA molecules were isolated in individual emulsions that greatly reduced the potential for cross-reaction between RNPs with gRNAs generated from one DNA and the targets positioned on another DNA molecule.

An advantage of CCR is the elimination of the need for cloning, and the libraries of individual DNA molecules needed for CCR can be synthesized inexpensively. These libraries can also be easily scaled to many members of the library as a pooled set of oligonucleotides, each containing both a unique target candidate and its paired gRNA candidate, that can be amplified using PCR to generate the double-stranded DNA molecules needed for CCR. Considering this, we then tested the utility of the CCR approach for larger screens by evaluating the on-target activity of all 18 possible spacers (**Figure 4A**)[16–25] that are capable of targeting sequences within the 126 bp exon 1 of the human gene *EMX1* (Chr2:72933825-72933951 in the hg38 assembly). Of those, one gRNA (gRNA 16) has been historically well-characterized for both on- and off-target activities.[51,52] CRISPOR, a computational tool for gRNA design, by default reports the efficiency scores or prediction of activity from two common machine learning approaches known as Doench '16 (sgRNA Designer)[16] and Moreno–Mateos (CrisprScan),[17] and it also reports prediction scores for eight other previously published methods.[18–25] As can be seen (Figure 4A), the different prediction methods are generally inconsistent with one another, with no one gRNA having high predicted efficacy scores across all methods, a situation which can complicate the interpretation of these outputs and decision-making regarding which gRNAs are most likely to result in an active RNP. According to the trend observed following the CCR screen of all 18 gRNAs simultaneously (Figures 1C and 4B), among the top performing RNPs, one includes the well-characterized gRNA (gRNA 16) which we have confirmed is active. Furthermore, the in vitro (using purified RNP complexes reacting individually with DNA molecules) validation of two gRNAs with demonstrated higher activity during CCR screen (gRNAs 12 and 10) revealed that they indeed had higher on-target activity compared to gRNA 16 (Figure 4B inset). These results show that CCR offers a straightforward approach for high-throughput screening and ranking of on-target activity, starting from a pooled oligonucleotide library.

In some circumstances, it may be desired or advantageous to simultaneously screen both on- and off-target propensities (Figures 2C and 5A), because even if a gRNA is highly active, it is not useful if it is also active at the potential off-target sites. To do so, we could use a DNA molecule similar in design to those used in Figures 2C and 3, with the on-target site located distally to the gRNA expression region and the off-target located proximally. This way, after screening, the identity of the spacer with the resulting length of the DNA molecule can be determined: if the DNA molecule cut only at the distal target, that is evidence that the RNP with that gRNA is active and specific since the off-target site remains uncut; but if it is cut at the proximal off-target, it indicates that the RNP with that gRNA is not specific enough, even if we do not know the activity on-target (if the molecule is uncut that is evidence of poor activity of the RNP with that gRNA). Following this design, we also screened four of the gRNAs from CRISPOR[32] with low predicted off-target activities, as well as gRNA 16 which is associated with several off-target sites with known off-target activity, totaling 18 predicted off-target sites (Figure 5B). The CCR method re-capitulated the relative off-target activity at four known off-targets for gRNA 16 and mirrored its in vitro activity at these sites[52] (Figure 5B inset), further

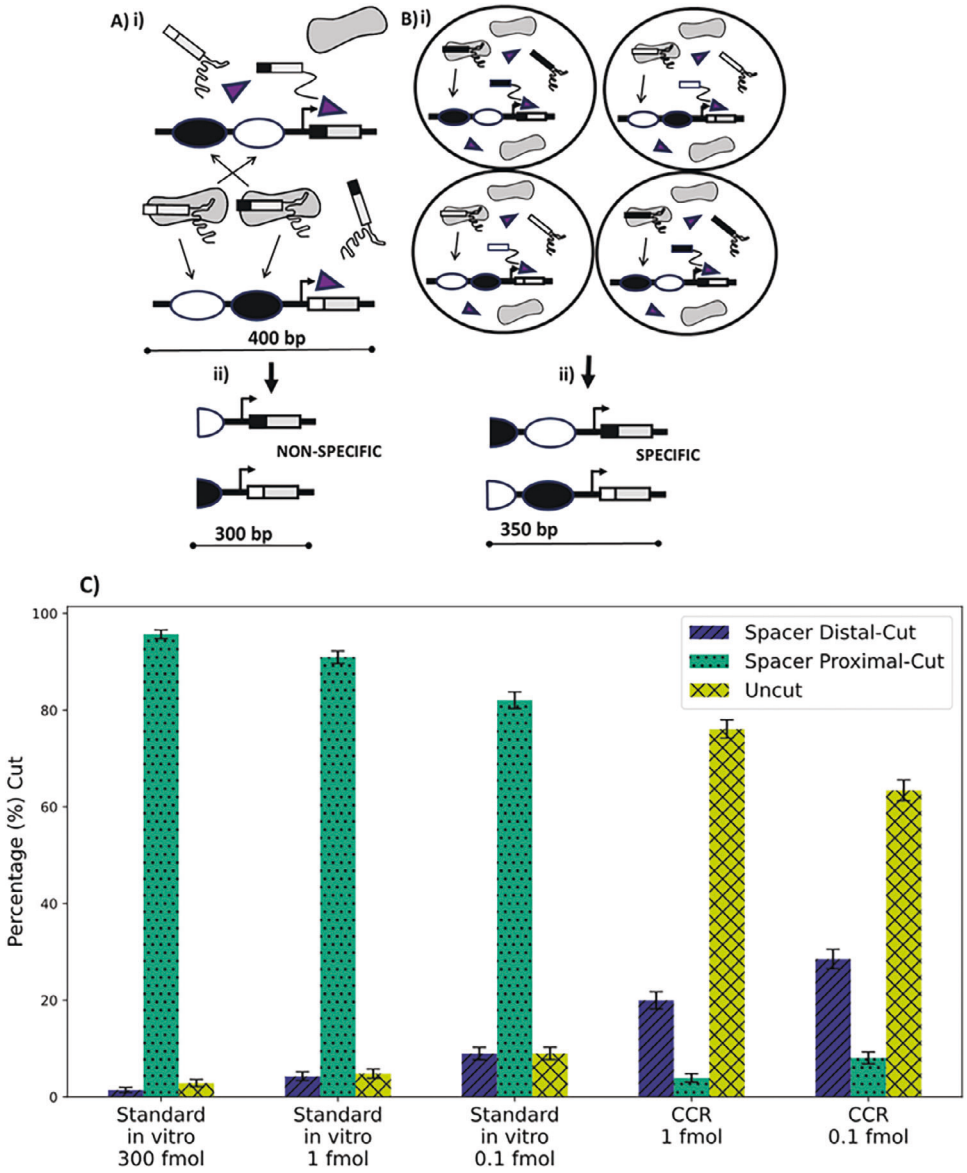

Figure 3. Confirmation that CCRs are performed with individual DNA molecules isolated within separate water-in-oil emulsions. A) Two DNA molecules are designed: each with a target (black or white) specific to the gRNA encoded on that molecule located on the distal end of the DNA (black or white, respectively) and the paired target for the other located proximal to the spacer (white or black, respectively). Under conditions where the two molecules were mixed, one would expect to see all of the DNA molecules cut at the proximal sites (300 bp), as the paired gRNAs could cause Cas9 to introduce DSBs at those sites, regardless of whether there were DSBs at the distal sites. B) Under conditions where the DNA molecules were isolated in emulsions, after CCR reactions one would expect mostly DNA molecules with DSBs at the distal site (≈350 bp). C) Under standard (with no emulsions) conditions where the molecules were introduced to both Cas9 and T7 RNA polymerase after the reaction most of the DNA molecules were determined to be cleaved at the spacer proximal site. However, those DNA molecules that reacted within water-in-oil emulsions were found to be significantly less likely to be cleaved at the spacer-proximal site, and the vast majority of those that had been cleaved were cleaved at the spacer-distal target, strongly implying that individual DNA molecules were isolated during the CCR reaction (See also Figure S3, Supporting Information).

demonstrating that CCR could be used to screen r a large array of off-targets as well. The off-target activity for the other gRNAs largely followed trends predicted by computational tools CFD and MIT off-target scores, but surprisingly gRNA 2, which computationally is not predicted to have a high amount of activity at off-target sites, exhibited the highest levels of activity during CCR at those off-target sites (Figure 5B). This highlights the importance of experimentally validating the computational predictions regarding gRNAs prior to use, which CCR makes significantly easier.

We then tested the effectiveness of CCR in identifying highly active and specific gRNA variants by screening a large library of gRNAs with 8 nucleotides (nt's) 5′- end extensions (x-gRNAs). As previously, reported,[51,52] gRNAs with short (8–12) nt extensions to the 5′- end immediately next to their spacer sequence are capable of eliminating the off-target activity of an RNP with that

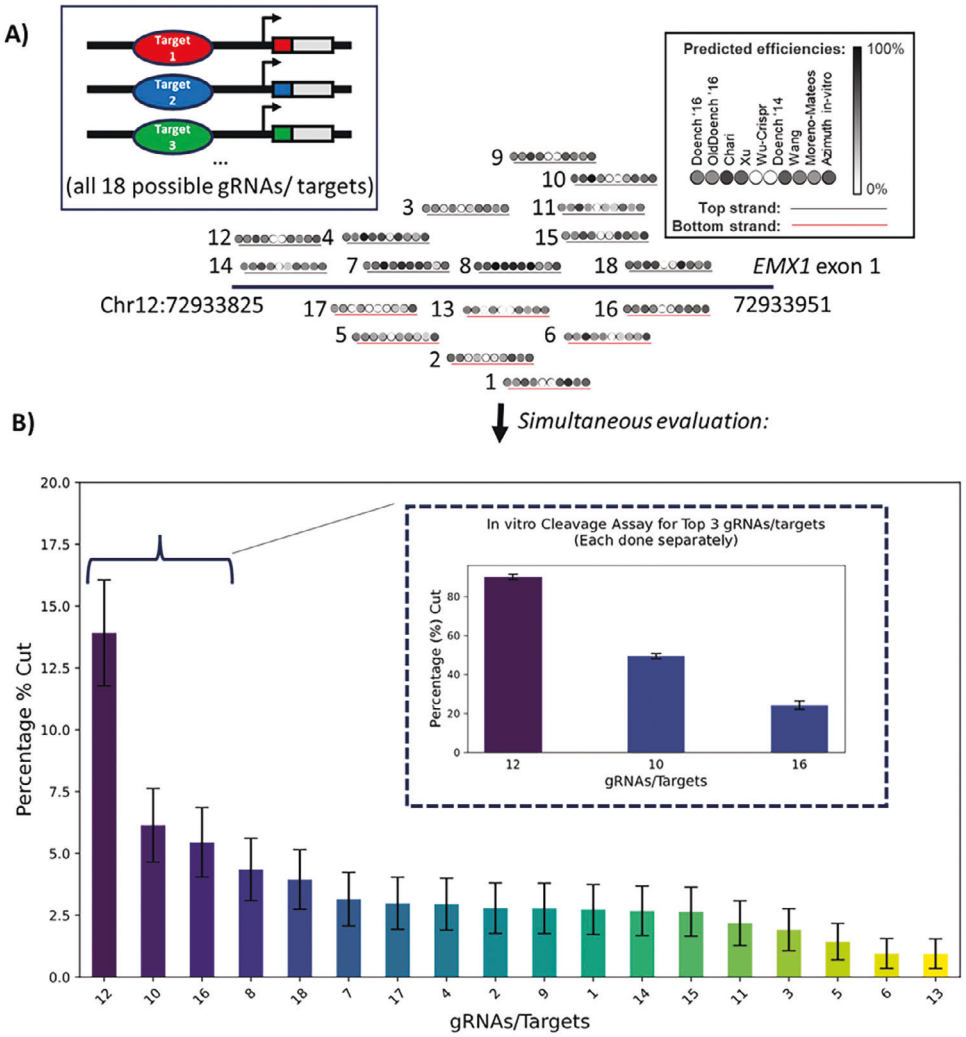

Figure 4. CCR to screen all gRNAs for a target gene in a one-pot reaction. A) The overlapping positions of all 18 possible gRNAs for exon 1 of *EMX1*, with computationally predicted activity for Cas9 at those sequences from ten different models[16–25] to predict gRNA activity. As can be seen, there is little agreement across models, which can make interpretation difficult. B) Observed cleavage rates for all 18 gRNAs reveal a gRNA known to be active (gRNA 16) to be among the best performing. Error bars are binomial confidence (95%) based on the number of next-generation sequencing reads for each gRNA/target DNA molecule. (inset) In vitro validation using purified RNPs shows the trend observed with CCR recapitulated the results from more labor-intensive screening for RNP activity with different gRNAs one at a time. Error bars are standard deviations across three identical replicates (see also Figure S4, Supporting Information).

x-gRNA while still maintaining its on-target activity as the original gRNA. However, the task of identifying the specific sequence of the 5′- nt's can be challenging, as many extensions either have no effect on specificity or also attenuate on-target activity. To screen the x-gRNAs library, we created a DNA library encompassing all possible 5′- extensions of 8 nt's (65536 possible sequences) for *EMX1* gRNA 16, which exhibits well-documented activity at several known off-target sites within the human genome, as recapitulated through CCR method, on a DNA molecule which had its target on the distal end and one each of the four most active off-target sites at the proximal end (**Figure 6A**): a total of 262144 separate gRNA/target/off-target combinations within a single CCR reaction. Following the CCR screen, the top five x-gRNAs were identified from the most common x-gRNA spacer sequences with spacer-distal cleavage patterns (Figure 5A; Figures S5 and S6, Supporting Information). Their efficacies were further validated in vitro using purified RNPs and it was found that they were able to completely block activity at the off-target site while maintaining on-target activity comparable to that of a high-specificity engineered variant of Cas9 known as eCas9 (Figure 6B). Therefore, CCR can be an effective tool to screen large numbers of gRNA variants with special attributes as well.

3. Discussion

Water-in-oil emulsions technique have been useful for handling large sets of unique DNA molecules in parallel, such as in library preparation for next-generation sequencing (NGS)[41–43] and in directed evolution studies,[44,46,48] and here we show that the complete CRISPR reaction (gRNA transcription, RNP

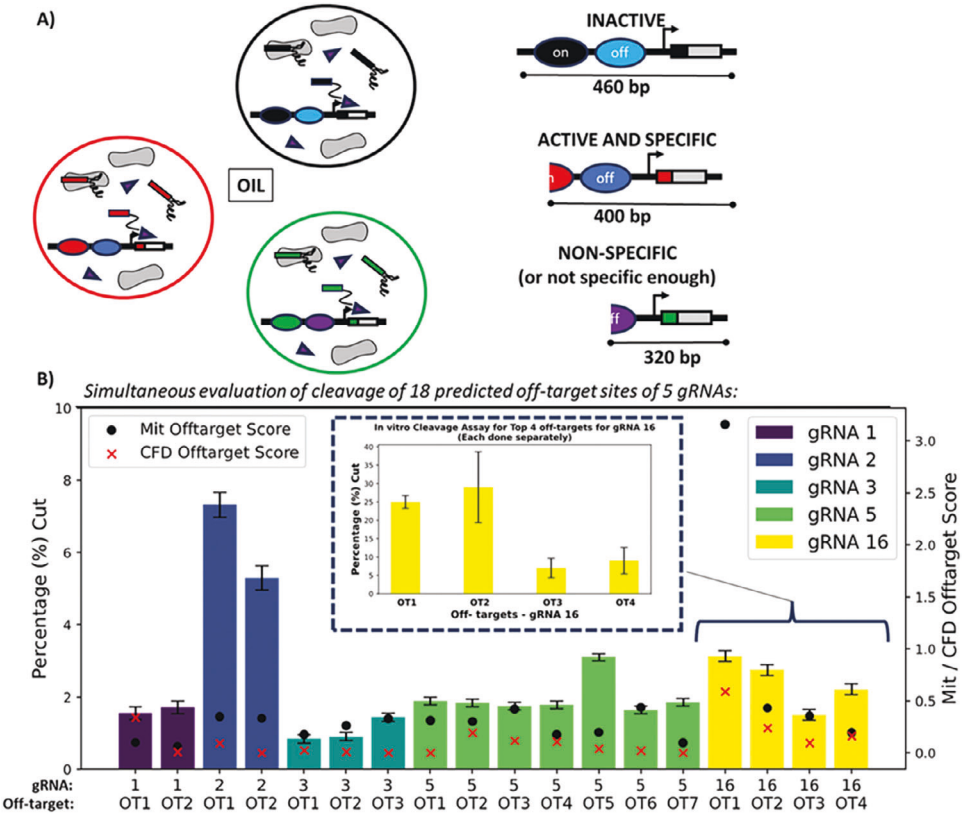

Figure 5. Screening different gRNAs for off-target activity simultaneously in parallel using CCR. A) A DNA molecule designed similarly to the one used in Figure 2C, with the on-target site located distal to the gRNA expression region and the off-target located proximally. The recorded length of the DNA molecules can be used to determine the off-target cleavage efficiency for the gRNAs. If the DNA molecule is only cleaved at the distal target, it indicates that the RNP with that particular gRNA is active and specific in relation to that off-target site, as the off-target site remains uncut; whereas, if cleavage occurs at the proximal off-target, it suggests that the RNP with that gRNA is not sufficiently specific in relation to that off-target site. B) Cleavage rates at the off-target sites of the different gRNAs, predicted to have very low off-target activity (1, 2, 3, and 5) and one with known off-targets (gRNA16), with predicted off-target scores (right, arrows, and dots). Error bars are binomial confidence (95%) based on the number of next-generation sequencing reads for each gRNA/target DNA molecule. (inset) Off-target activity for gRNA 16 using CCR follows that pattern observed when activity at those sites is validated one at a time using purified RNPs.[52]

formation, and target cleavage) can be reconstituted within individual emulsions to screen large libraries of gRNAs/targets/off-targets in a streamlined and experimentally simplified manner. Whenever considering screens starting from oligonucleotide libraries, there is always a concern with library generation; an advantage here is that in the specific case of CCR experiments, those variations can be normalized as a result of whether the sequencing reveals each target as "cut" versus "un-cut," divided by the total number of reads related to that oligonucleotide. This means that even if subtle variations in the library exist as a result of synthesis, the relative activity of an RNP with a specific activity can still be determined from the results of a CCR experiment. We expect CCR can be applied to screen large sets of gRNAs for other CRISPR effectors as well, particularly considering that the available computational tools for gRNA design are currently advanced for only a select few effectors, while new effectors with diverse properties are increasingly discovered.[53]

CCR represents a powerful yet accessible method to easily screen many candidate gRNAs (provided from the outputs of computational tools for gRNA design) for both activity and specificity and as a way of rapidly identifying highly active and specific gRNA variants from large sets of gRNA/target/off-target combinations. Furthermore, CRISPR biotechnologies have not only been used for applications of gene editing but also diagnostics and other specific applications like RNA knockdown.[54,55] We have no reason to believe that, with small modifications, the CCR technique cannot be used for rapid screening of gRNAs for these other applications. Diagnostic applications, for example, can also require examining large numbers of potential candidates for nuclease activities, and CCR can help to play a role in reducing the burden of labor in this regard. Beyond the application of identifying efficient gRNAs in a parallelized manner, the CCR technique can be utilized for conducting large-scale studies on CRISPR biochemical kinetics from an intriguing biophysical chemistry perspective. Additionally, exploring the effects of various factors such as temperature and salt concentrations across numerous sequences simultaneously could provide valuable insights, particularly considering that sequence diversity is often a confounding variable under each condition.

Lastly, and perhaps most importantly, is to understand how the results of CCR might predict the activity of a CRISPR RNP within a cell of interest. The activity of an RNP with a specific

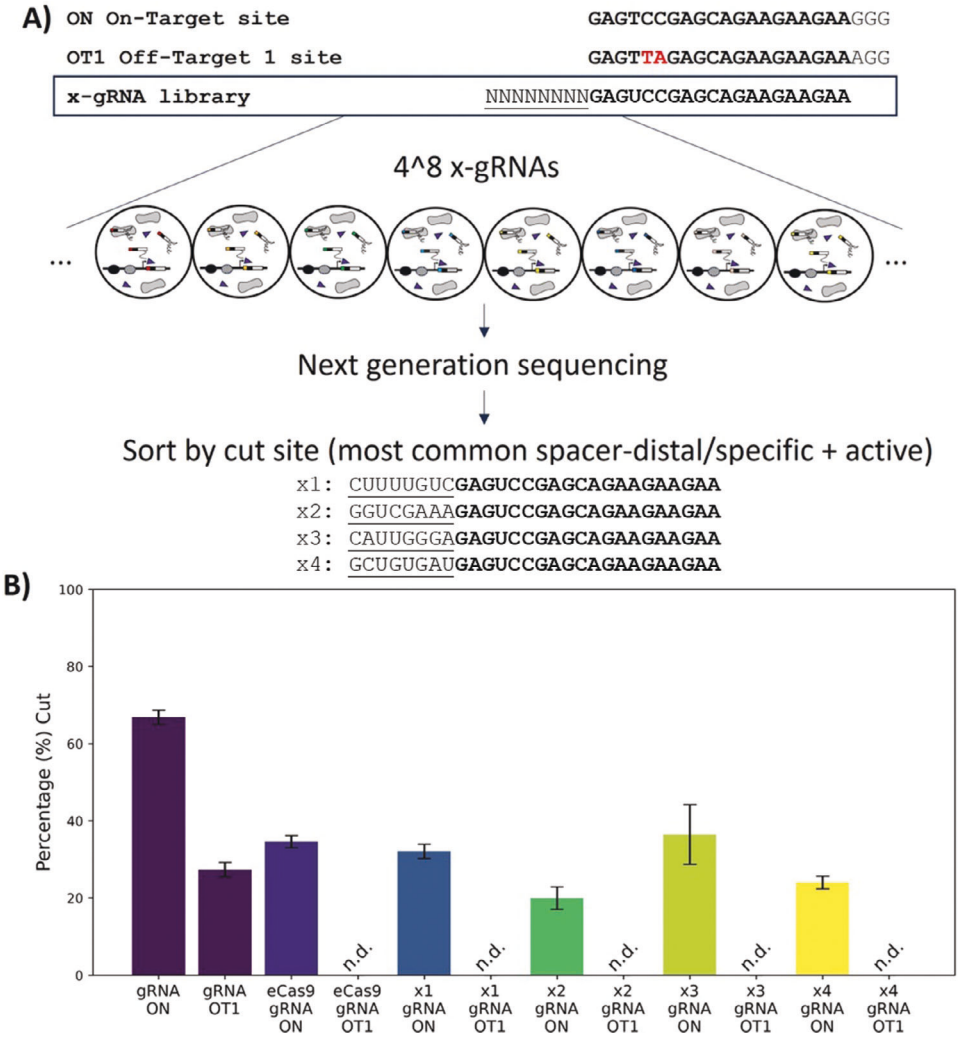

Figure 6. CCR to screen hundreds of thousands of gRNA variants for high activity and specificity. A) gRNAs with 5′ extensions of 8 nucleotides (x-gRNAs) are known to be capable of increasing the specificity of the gRNA, although identifying the precise sequence of the x-gRNA can be difficult. With CCR, all 4^8 x-gRNA sequences can be evaluated simultaneously. B) Validated in vitro one at a time, x-gRNAs identified by having the most common specific and active (spacer-distal) cleavage pattern all exhibited on-target (ON) activities comparable to an engineered high-specificity Cas9 (eCas9) while not having any detectable nuclease activity at their off-target site (OT1).

gRNA is very much depends on the specific cell line and its delivery method.[56,57] For example, differences in chromatin state can significantly influence CRISPR activity[58] at that site. However, without RNP catalytic activity, those experiment-specific factors are rendered moot, and so a rapid and simple evaluation of that activity can help to streamline the more complex and expensive experiments downstream.

4. Experimental Section

Protocol for Single Pot in Vitro CRISPR-Induced Cleavage Reactions: DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) and then resuspended to a stock concentration of 100 µM. DNA oligonucleotides were further diluted to a concentration of 300 fmol which is visible in agrarose gel stained with SYBR Safe dye. All the reagents were then mixed in the following order to set up CRISPR-mediated cleavage reaction: 7 µL nuclease-free water, 1 µL DNA Oligonucleotides (300 fmol), 2 µL NEB r 3.1 Buffer (10X), 1 µL RNAse Inhibitor (New England Biolabs #M0314S), 0.4 µL (2.5 mM) ribonucleotide triphosphate mix (NEB: N0466S), 3 µL (1000 nM) Cas9 nuclease, *S. pyogenes* ((NEB; M0386T), 2 µL T7 RNA polymerase (New England BioLabs). The reaction mixture was then incubated for a duration of 3 h at 37 °C. This was followed by proteinase K digestion, where 1 µL of proteinase K (ThermoFisher kit #EO0491) was added, and the mixture was incubated at 56 °C for 10 min. Subsequently, 2 µL of RNAse A was added to degrade the sgRNA molecules. The resulting products were then separated on a 3% agarose gel, stained with SYBR Safe, and analyzed using ImageJ.

Protocol for Single Pot CRISPR-Induced Cleavage Reactions in Emulsion: Water-oil emulsions were produced using the bulk mixing of water and oil phase following the procedure outlined by Williams et al.[41] The oil phase was prepared by thoroughly mixing the following components, 2.25 mL -Span 80 4.5% (vol/vol), 200 µL -Tween 80 0.4% (vol/vol), 25 µL- Triton X-100 0.05% (vol/vol) and Mineral oil to 50 mL. A modification in the protocol was introduced for the aqueous phase, which comprised the following reagents for the CRISPR-induced cleavage reaction: **Table 1**.

Table 1. Composition of the Aqueous phase.

Reagents	Vol [µL] added in solution
DNA oligonucleotides	0.1 femtomoles
NEB r3.1 Buffer (10x)	26 µL
RNAse Inhibitor (40,000 u/mL) (New England Biolabs #M0314S)	5 µL
Ribonucleotide triphosphate mix (NEB: N0466S)	5.2 µL
Cas9 nuclease, S. pyogenes (1000 nM) (NEB; M0386T)	1 µL
T7 RNA polymerase (50,000 u/mL) (NEB; M0251S)	5.2 µL
Nuclease-free water	Up to a Total Vol. of 260 µL

In a cryovial with a magnetic stir bar on a magnetic plate at an intermediate setting (1150 rpm), 400 µL of the oil phase was placed. Subsequently, 200 µL of the aqueous reaction mixture was slowly added to the oil phase over a 5-min duration. It was important to note that the entire setup was maintained in a cold room at 4 °C during the emulsification process. After this step, a stable water-oil emulsion was formed, followed by an 8-h incubation at 37 °C. The emulsion was then disrupted by centrifugation (15 000 rpm, 10 min) to recover the water phase, while the oil phase was pipetted out. Next, 1 µL of proteinase K (~20 mg/mL; ThermoFisher kit #EO0491) was added to degrade the RNP, and the mixture was incubated at 56 °C for 10 min. Following this, 2 µL of RNAse A was added to degrade the sgRNA molecules. The DNA library from the aqueous phase was then extracted through a cleanup process with magnetic beads added in a 0.9X ratio to the aqueous phase. The purified DNA library was subsequently enriched using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB# E7645S), followed by PCR with NEBNext Multiplex Oligos for Illumina (NEB# E7730S). The schematic of the process is depicted in Figure S7 (Supporting Information). The resulting products were also separated on a 3% agarose gel, stained with SYBR Safe, and then submitted for NGS sequencing.

In Vitro Digestion Reactions: DNA targets containing the target sequences were synthesized by Twist Bioscience. Subsequently, they were PCR amplified using the provided universal primers, purified, and resuspended in nuclease-free water to a concentration of 100 nM. For each reaction, three technical replicates were prepared in the following order: 7 µL of nuclease-free water, 1 µL of the target DNA substrate (100 nM), 1 µL of 10x Cas9 Nuclease Reaction buffer (composed of 200 mM HEPES, 1 M NaCl, 50 mM MgCl2, and 1 mM EDTA with a pH of 6.5 at 25 °C), and 1 µL of Cas9-RNP (1 mM). The assembled reactions were then incubated for 1 h at 37 °C, followed by digestion with proteinase K (ThermoFisher enzyme #EO0491) (1 µL) at 56 °C for 10 min. The resulting products were resolved on a 3% agarose gel stained with SYBR Gold and analyzed using ImageJ. For analysis of the gel image in Image J, the fluorescence intensity was normalized by the length of the DNA fragments, and the fraction cleaved was calculated using Equation (1):

$$\text{Fraction cleaved} = \left(\frac{\text{Normalised cleaved band intensity}}{\sum \text{Normalised Cleaved and Uncleaved band Intensities}}\right) \times 100 \quad (1)$$

Dynamic Light Scattering (DLS) Measurement: The size of water droplets in water-oil emulsions was measured using a dynamic light scattering (DLS) instrument, Zetasizer Nano ZS90 (Malvern Instruments, Malvern, UK). The average value of 3 runs was taken as a final result. The stability of the water–oil emulsion was tested by measuring the water droplets immediately after emulsion formation and after an interval of 8 h.

Analysis of the Next-Generation Sequencing Data: The NGS sequencing data had sequences with pre-trimmed adaptors and barcode sequences. The paired-end was then merged into longer, single reads using FLASH (Fast Length Adjustment of SHort reads)[59] bioinformatic tool. A script was written in Python that used the Bio-python[60] package for reading the FASTQ files. Each dataset here is a combination of three identical repeat experiments. The algorithm begins by checking the sequence length of each read from the input data to ensure it meets a certain criterion. Next, the sequences that pass the length check undergo alignment using the Smith–Waterman algorithm[61] with reference sequences, which include a complete list of different targets and off-targets present in the DNA library. Once the alignment scores are obtained, the algorithm evaluates the alignment score threshold and selects sequences that exceed the threshold. This step ensures that only sequences with high alignment scores, indicating a good match, are considered for further analysis. Following this, based on their alignment scores the sequences are assigned to categories (Spacer–Distal /Spacer-Proximal/ Uncut) with respect to different targets or off-targets present. The algorithm then checks the occurrence of sequences in each category by counting how many times each sequence appears within its assigned category. To calculate the percentage cut, the ratio of the number of sequences in the cut state (spacer-proximal or distant) to the number of uncut sequences in each category was taken. Finally, the algorithm selects the top hits, which were the sequences with the highest percentage cut or most frequent occurrences. For the analysis of results from the extended guideRNAs (x-gRNAs) screening experiment, a Python script was written to extract 8-nucleotide sequences (extended gRNA extensions) adjacent to the (T7 Promoter + off-target) sequence from the filtered large fragments list and stored them as a CSV file.

The reference code used for the analysis of the NGS data is provided at https://github.com/neelarka/CCR-Analysis. The complete workflow for the analysis of the NGS data is shown in Figure S8 (Supporting Information):

Statistical Analysis: For the in vitro CRISPR cleavage assay using the standard protocol three technical replicates were used as listed in their respective figure captions. Statistical analysis such as the generation of confidence intervals was performed using a Python Script. Each Next-Generation Sequencing (NGS) dataset comprises three identical repeated experiments. The error bars depicted were binomial confidence intervals (95%) determined by the number of next-generation sequencing reads for each gRNA/target DNA molecule.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors wish to acknowledge financial support from NIH National Institute of General Medical Sciences (NIGMS) (R35GM133483), NIH National Institute of Biomedical Imaging and Bioengineering (NIBIB) (R21EB033595), as well as NSF (award #2027738) which provided support for T.S and DoD (Contract #W911QY2220006) which provided support for A.H.N. This work was performed in part at the Joint School of Nanoscience and Nanoengineering, a member of the Southeastern Nanotechnology Infrastructure Corridor (SENIC) and National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the NSF (Grant ECCS-1542174).

Conflict of Interest

EAJ is an inventor on patents and provision patents associated with CRISPR-related technologies, and EAJ and AHN are inventors on a provisional patent related to x-gRNAs.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

CRISPR, gRNA design, guide RNA, high-throughput screen, microemulsions

Received: April 30, 2024
Revised: May 22, 2024
Published online:

[1] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, F. Zhang, *Nat. Protoc.* **2013**, *8*, 2281.
[2] P. D. Hsu, E. S. Lander, F. Zhang, *Cell* **2014**, *157*, 1262.
[3] F. Jiang, J. A. Doudna, *Ann. Rev. Biophys.* **2017**, *46*, 505.
[4] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, *Nature* **2014**, *7490*, 62.
[5] M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, *Science* **2014**, *343*, 1247997.
[6] H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, *Cell* **2014**, *156*, 935.
[7] C. Anders, O. Niewoehner, A. Duerst, M. Jinek, *Nature* **2014**, *513*, 569.
[8] M. Kosicki, K. Tomberg, A. Bradley, *Nat. Biotechnol.* **2018**, *36*, 765.
[9] E. K. Brinkman, T. Chen, M. de Haas, H. A. Holland, W. Akhtar, B. van Steensel, *Mol. Cell* **2018**, *70*, 801.
[10] F. Allen, L. Crepaldi, C. Alsinet, A. J. Strong, V. Kleshchevnikov, P. De Angeli, P. Páleníková, A. Khodak, V. Kiselev, M. Kosicki, A. R. Bassett, *Nat. Biotechnol.* **2019**, *37*, 64.
[11] J. A. Gagnon, E. Valen, S. B. Thyme, P. Huang, L. Ahkmetova, A. Pauli, T. G. Montague, S. Zimmerman, C. Richter, A. F. Schier, *PLoS One* **2014**, *9*, e98186.
[12] F. Jiang, D. W. Taylor, J. S. Chen, J. E. Kornfeld, K. Zhou, A. J. Thompson, E. Nogales, J. A. Doudna, *Science* **2016**, *351*, 867.
[13] G. Mullally, K. Van Aelst, M. M. Naqvi, F. M. Diffin, T. Karvelis, G. Gasiunas, V. Siksnys, M. D. Szczelkun, *Nucleic Acids Res.* **2020**, *48*, 6811.
[14] W. Rychlik, *PCR Protocols: Current Methods and Applications*, Springer, Berlin, Heidelberg **1993**, pp. 31–40.
[15] T. Mann, R. Humbert, M. Dorschner, J. Stamatoyannopoulos, W. S. Noble, *Nucleic Acids Res.* **2009**, *37*, e95.
[16] J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, *Nat. Biotechnol.* **2016**, *34*, 184.
[17] M. A. Moreno-Mateos, C. E. Vejnar, J. D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha, A. J. Giraldez, *Nat. Methods* **2015**, *12*, 982.
[18] M. Stemmer, T. Thumberger, M. del Sol Keyer, J. Wittbrodt, J. L. Mateo, *PLoS One* **2015**, *10*, e0124633.
[19] W. Chen, A. McKenna, J. Schreiber, M. Haeussler, Y. Yin, V. Agarwal, W. S. Noble, J. Shendure, *Nucleic Acids Res.* **2019**, *47*, 7989.
[20] H. Xu, T. Xiao, C. H. Chen, W. Li, C. A. Meyer, Q. Wu, X. S. Liu, *Genome Res.* **2015**, *25*, 1147.
[21] N. Wong, W. Liu, X. Wang, *Genome Biol.* **2015**, *16*, 218.
[22] J. G. Doench, E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier, D. E. Root, *Nat. Biotechnol.* **2014**, *32*, 1262.
[23] D. Wang, C. Zhang, B. Wang, B. Li, Q. Wang, D. Liu, H. Wang, Y. Zhou, L. Shi, F. Lan, Y. Wang, *Nat. Commun.* **2019**, *10*, 4284.
[24] N. Fusi, I. Smith, J. Doench, J. Listgarten, *BioRxiv* **2015**, 021568.
[25] R. Chari, N. C. Yeo, A. Chavez, G. M. Church, *ACS Synth. Biol.* **2017**, *6*, 902.
[26] G. Chuai, H. Ma, J. Yan, M. Chen, N. Hong, D. Xue, Q. Liu, *Genome Biol.* **2018**, *19*, 80.
[27] J. Listgarten, M. Weinstein, B. P. Kleinstiver, A. A. Sousa, J. K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J. G. Doench, N. Fusi, *Nat. Biomed. Eng.* **2018**, *2*, 38.
[28] S. Bae, J. Park, J. S. Kim, *Bioinformatics* **2014**, *30*, 1473.
[29] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, *Nat. Biotechnol.* **2013**, *31*, 827.
[30] V. Pliatsika, I. Rigoutsos, *Biol. Direct* **2015**, *10*, 4.
[31] S. Xie, B. Shen, C. Zhang, X. Huang, Y. Zhang, *PLoS One* **2014**, *9*, e100448.
[32] M. Haeussler, K. Schönig, H. Eckert, A. Eschstruth, J. Mianné, J. B. Renaud, S. Schneider-Maunoury, A. Shkumatava, L. Teboul, J. Kent, J. S. Joly, *Genome Biol.* **2016**, *17*, 148.
[33] M. F. Sentmanat, S. T. Peters, C. P. Florian, J. P. Connelly, S. M. Pruett-Miller, *Sci. Rep.* **2018**, *8*, 888.
[34] S. Q. Tsai, N. T. Nguyen, J. Malagon-Lopez, V. V. Topkar, M. J. Aryee, J. K. Joung, *Nat. Methods* **2017**, *14*, 607.
[35] T. Arndell, N. Sharma, P. Langridge, U. Baumann, N. S. Watson-Haigh, R. Whitford, *BMC Biotechnol.* **2019**, *19*, 71.
[36] Y. Zhou, S. Zhu, C. Cai, P. Yuan, C. Li, Y. Huang, W. Wei, *Nature* **2014**, *509*, 487.
[37] S. L. Jason, K. Yusa, *Methods* **2019**, *164*, 29.
[38] C. Bock, P. Datlinger, F. Chardon, M. A. Coelho, M. B. Dong, K. A. Lawson, T. Lu, L. Maroc, T. M. Norman, B. Song, G. Stanley, *Nat. Rev. Methods Prim.* **2022**, *2*, 8.
[39] T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, *Science* **2014**, *343*, 80.
[40] S. S. Terekhov, I. E. Eliseev, L. A. Ovchinnikova, M. R. Kabilov, A. D. Prjibelski, A. E. Tupikin, I. V. Smirnov, A. A. Belogurov Jr., K. V. Severinov, Y. A. Lomakin, S. Altman, A. G. Gabibov, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 27300.
[41] R. Williams, S. G. Peisajovich, O. J. Miller, S. Magdassi, D. S. Tawfik, A. D. Griffiths, *Nat. Methods* **2006**, *3*, 545.
[42] A. D. Griffiths, D. S. Tawfik, *Trends Biotechnol.* **2006**, *24*, 395.
[43] K. Shao, W. Ding, F. Wang, H. Li, D. Ma, H. Wang, *PLoS One* **2011**, *6*, e24910.
[44] B. M. Paegel, G. F. Joyce, *Chem. Biol.* **2010**, *17*, 717.
[45] B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, T. K. Kitano, *Anal. Chem.* **2011**, *83*, 8604.
[46] D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652.
[47] K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, D. S. Tawfik, *Anal. Biochem.* **2004**, *325*, 151.
[48] O. J. Miller, K. Bernath, J. J. Agresti, G. Amitai, B. T. Kelly, E. Mastrobattista, V. Taly, S. Magdassi, D. S. Tawfik, A. D. Griffiths, *Nat. Methods* **2006**, *3*, 561.
[49] Y. Schaerli, F. Hollfelder, *Mol. BioSyst.* **2009**, *5*, 1392.
[50] V. Taly, B. T. Kelly, A. D. Griffiths, *ChemBioChem* **2007**, *8*, 263.
[51] D. D. Kocak, E. A. Josephs, V. Bhandarkar, S. S. Adkar, J. B. Kwon, C. A. Gersbach, *Nat. Biotechnol.* **2019**, *37*, 657.
[52] A. Herring-Nicholas, H. Dimig, M. R. Roesing, E. A. Josephs, *Commun. Biol.* **2024**, *7*, 86.
[53] E. V. Koonin, J. S. Gootenberg, O. O. Abudayyeh, *Biochemistry* **2023**, *62*, 3465.
[54] J. M. Lesinski, T. Moragues, P. Mathur, Y. Shen, C. Paganini, L. Bezinge, B. Verberckmoes, B. Van Eenooghe, S. Stavrakis, A. J. deMello, D. A. Richards, *ChemRxiv* **2024**, https://doi.org/10.26434/chemrxiv-2024-82q4c.
[55] D. C. Swarts, M. Jinek, *Mol. Cell* **2019**, *73*, 589.
[56] K. Murugan, A. S. Seetharam, A. J. Severin, D. G. Sashital, *bioRxiv* **2020**, 2020.

[57] M. S. Liu, S. Gong, H. H. Yu, K. Jung, K. A. Johnson, D. W. Taylor, *Nat. Commun.* **2020**, *11*, 3576.
[58] C. Kuscu, S. Arslan, R. Singh, J. Thorpe, M. Adli, *Nat. Biotechnol.* **2014**, *32*, 677.
[59] T. Magoč, S. L. Salzberg, *Bioinformatics* **2011**, *27*, 2957.
[60] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. De Hoon, *Bioinformatics* **2009**, *25*, 1422.
[61] T. F. Smith, M. S. Waterman, *J. Mol. Biol.* **1981**, *147*, 195.