

Expanding the Safety, Specificity, and Targeting Range of the CRISPR-Cas Genome Editing Toolbox

A THESIS PRESENTED

BY

VED V. TOPKAR

TO

THE DEPARTMENT OF BIOCHEMICAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

ARTIUM BACCALAUREUS (A.B.)

IN THE SUBJECT OF

CHEMICAL & PHYSICAL BIOLOGY

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

MAY 2016

Expanding the Safety, Specificity, and Targeting Range of the CRISPR-Cas Genome Editing Toolbox

ABSTRACT

The easy reprogrammability of the Cas9 RNA-guided endonuclease has made it one of the most versatile tools for genome editing. The protein's targeting activity occurs in three phases. First, it complexes with a guide RNA (gRNA), comprising of a spacer sequence and a double hairpin. Second, it scans DNA for a short Protospacer Adjacent Motif (PAM) sequence, which is specified by the protein itself. Third, when a PAM is found, Cas9 checks the gRNA spacer sequence for complementarity against the adjacent genomic DNA. If there is a match, Cas9 will cleave the DNA at that location. Though changing the spacer sequence on the gRNA reprograms the sequence specificity of Cas9, there is a natural limitation on its set of genomic targets due to the PAM sequence requirement. The canonical Cas9 used by most scientists, the *S. pyogenes* Cas9 (SpCas9), has a PAM sequence of NGG, limiting it to ≈ 1 in 8 loci in the human genome. We recently characterized the PAM specificity of evolved Cas9 variants, as well as the *S. aureus* Cas9 (Kleinstiver *et al.* 2015), showing that that Cas9 proteins from orthogonal bacterial strains can indeed be characterized. Here, we propose to attempt to characterize the PAM specificities of 5 more Cas9 proteins from a variety of bacterial strains.

Contents

1	INTRODUCTION	I
2	BACKGROUND	5
2.1	Selected Cas9 Proteins	8
2.2	Experimental Approach	8
2.2.1	Prediction of RNA Guides	8
2.2.2	Positive Selection	9
2.2.3	Negative PAM Depletion Assay	9
2.2.4	GUIDE-Seq Specificity Profiling	12
2.2.5	Expected Results and Potential Roadblocks	12
3	GUIDE-SEQ: AN UNBIASED ASSAY FOR GENOME-WIDE OFF-TARGET DETECTION	15
3.1	Exploratory Analysis of Potential Off-Target sites	15
3.2	Overview of the GUIDE-Seq Assay for Production of Raw Read Data	15
3.3	Bioinformatic Analysis of GUIDE-Seq data to infer off-target sites	15
3.4	Comparison of GUIDE-Seq results to ChIP-Seq and Computational	15
4	<i>GUIDESEQ</i> : A COMPUTATIONAL PIPELINE FOR ANALYZING GUIDE-SEQ DATA	16
4.1	Understanding the core steps of the <i>guideseq</i> pipeline	16
4.2	Example output of the software	16
5	EXPANDING THE TARGETING RANGE OF CRISPR-CAS WITH ORTHOGONALITY PROFILING AND PROTEIN ENGINEERING	17
5.1	Engineering the PAM-contacting residues of Cas9	18
5.2	Predicting the Activity of Orthogonal Cas9 Systems	18
5.3	Understanding the PAM Depletion Assay	18
5.4	PAM Depletion Assay Results	18

6	CHARACTERIZING THE SPECIFICITY OF CRISPR-CAS PROTEINS	19
6.1	Overview of the spacer subsequence randomization assay	19
6.2	Specificity Results of Randomization Assay	19
7	CasBLASTR: A TOOL FOR GENOME EDITING EXPERIMENT DESIGN AND ANALYSIS	20
7.1	Application Architecture	20
7.2	Results and Usage Statistics	20
8	CONCLUSION AND FUTURE STEPS	21
	REFERENCES	25

TO THE FAMILY, FRIENDS, AND FOOD THAT SUSTAINED ME THROUGH THIS PROCESS.

Acknowledgments

PEOPLE ARE AMAZING, I will acknowledge them here.

1

Introduction

Much of the recent history of genome editing consists of building the site-specificity of a nuclease through engineered protein-DNA interactions. Though this technique has proven effective and useful, as seen in the examples of Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs), these platforms are difficult to use because of the complexity of programming protein-DNA specificity. In recent years, the CRISPR-Cas system, adapted from the CRISPR bacterial antiviral immune system, has been extensively used as a genome engineering alternative that relies on RNA-DNA complementarity as the homing interaction that guides the specificity of the Cas9 nuclease. The ease of using such a platform to cleave specific sites in the genome simply by choosing a

guide RNA sequence complementary to the target sequence has led to an explosion in its use, with enormous amounts of interest and capital being contributed to the further development and application of this technology (Doudna and Charpentier *et al.* 2014).

In what is perhaps the most important paper in the field, the Doudna (UC Berkeley) and Charpentier (Umea University, Sweden) labs teamed up in 2012 to fully characterize the homing mechanism of the Cas9 nuclease. Through *in vitro* cleavage assays, they found that Cas9, the tracrRNA, the crRNA with a spacer specifying the correct target, and magnesium ions, are all that is needed for cleavage of target DNA to occur. They further validated that complementarity between the spacer and the target is necessary for cleavage, and that one or two mismatches in this spacer::target complementarity is tolerated. Additionally, they showed that cleavage of a target depends on a “protospacer adjacent motif” (PAM) sequence, which is specified by the Cas9 protein, not the spacer (Jinek and Chylinski *et al.* 2012).

Finally, realizing that the CRISPR-Cas system could be an easily programmable nuclease, the authors constructed a synthetic platform that marked the first time that this system was engineered for easier biotechnological use. They fused the tracrRNA and crRNA sequences (as they would be after maturation by RNaseIII) together with a short linker loop and expressed the entire construct as a single sequence. This led to high cleavage activity, indicating that bypassing the maturation step increases the activity of CRISPR-Cas restriction. This fused RNA molecule concept, dubbed the guide RNA (gRNA) is now nearly universally used in all applications of the CRISPR-Cas platform.

Perhaps one of the largest questions in the CRISPR-Cas9 field is that of targetability. Since there is a PAM sequence requirement for all known Cas9 proteins, each can only tar-

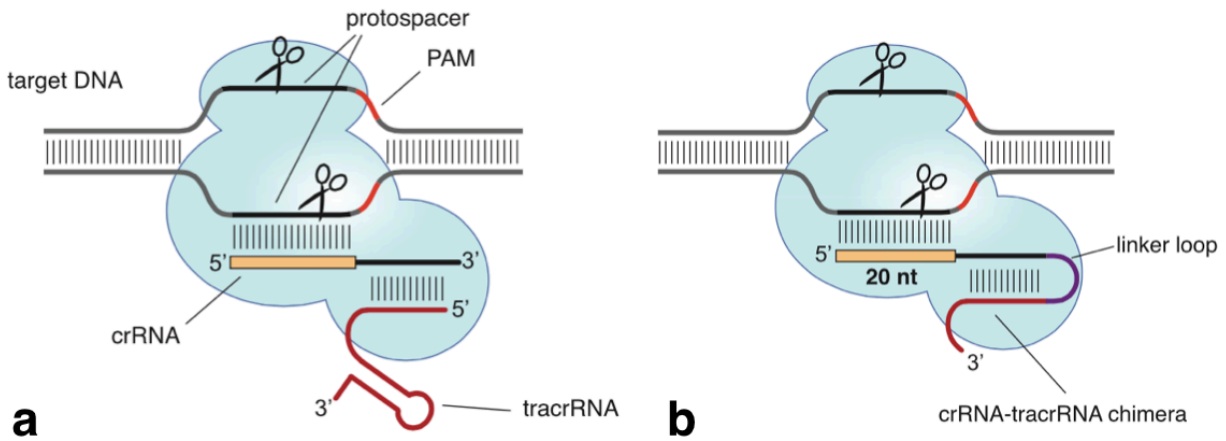


Figure 1.1: An RNA complex guides Cas9 to its target. **(a)** The matured crRNA and tracrRNA complex to guide Cas9. **(b)** A single gRNA comprising of the crRNA, a linker loop, and a crRNA guides Cas9 with the same efficiency as the dual-guide system. (Adapted from Jinek and Chylinski *et al.* 2012)

get a limited subset of all possible genomic targets. Interestingly, since the CRISPR-Cas system evolved as a bacterial immune system against viral DNA, and since different bacterial strains are attacked by vastly different phages, Cas9 proteins from different strains seem to have differing PAM specificities. Thus, the targeting range of the CRISPR-Cas platform can be expanded by adding more Cas9s to choose from (Kleinstiver *et al.* 2015). We explore this further in chapter 5. Additionally, we explore the ability of Cas9 to distinguish between a match and a mismatch between the spacer and target at each position through a spacer randomization assay, which is described in chapter 6.

Of similar import is specificity. Because RNA and DNA can form complementary complexes even in the presence of a small number of mismatches between the two molecules' sequences, CRISPR-Cas9 often cleaves at off-target locations in a somewhat unpredictable manner. This has huge implications in applications of CRISPR-Cas9, as off-target muta-

genesis can possibly have catastrophic effects on a cell. We have recently described an assay for detecting off-target cleavage events by Cas9 in an unbiased manner, which can be used to better understand the specificities of different Cas9 proteins (Tsai *et al.* 2015). We explore the preliminary results of the application of the GUIDE-seq assay in chapters 3 and 4.

2

Background

Much of the recent history of genome editing consists of building the site-specificity of a nuclease through engineered protein-DNA interactions. Though this technique has proven effective and useful, as seen in the examples of Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs), these platforms are difficult to use because of the complexity of programming protein-DNA specificity. In recent years, the CRISPR-Cas bacterial immune system has been extensively used as a genome engineering alternative that relies on RNA-DNA complementarity as the homing interaction that guides the specificity of the Cas9 nuclease. The ease of using such a platform to cleave specific sites in the genome simply by choosing a guide RNA sequence complementary to the

target sequence has led to an explosion in its use, with enormous amounts of interest and capital being contributed to the investigation of further development and application of this technology (Doudna and Charpentier *et al.* 2014).

The molecular basis of Cas9 specificity comes from two independent RNA sequences: the CRISPR RNA (crRNA) and the transactivating CRISPR RNA (tracrRNA). In the wild, the tracrRNA and crRNA complex at a region of complementarity, are processed by endogenous nonspecific exonucleases, and then associate with Cas9 for a final crRNA-tracrRNA-Cas9 complex. When using the system in the lab, the crRNA and the tracrRNA are expressed as a single guide RNA (gRNA) strand to simplify the formation of the final RNA-Protein complex. Cas9 then scans genomic DNA for a short Protospacer-Adjacent Motif (PAM) sequence, which is hard-coded into the protein. Once a match is found, Cas9 then alters its conformation to check for complementarity between the 20 nucleotide gRNA spacer sequence and the adjacent genomic DNA. If a match is found, Cas9 then makes a double stranded break at that locus (Jinek *et al.* 2012, Anders *et al.* 2014).

Once a double stranded break occurs in a eukaryotic cell, cellular machinery quickly repairs the break through either non-homologous end joining (NHEJ) or homology directed repair (HDR). In the case of NHEJ, random blunt end resections or small nucleotide insertions can occur, causing a deletion or insertion of one or more nucleotides in the final repaired site. This phenomenon is often used to induce frame-shift mutations into a gene to knock it out. In the case of HDR, a donor strand of DNA is used to template a precise insertion at the site of repair (Doudna and Charpentier 2014).

Perhaps one of the largest questions in the CRISPR-Cas9 field is that of targetability. Since there is a PAM sequence requirement for all known Cas9 proteins, each can only tar-

get a limited subset of all possible genomic targets. Interestingly, since the CRISPR-Cas system evolved as a bacterial immune system against viral DNA, and since different bacterial strains are attacked by vastly different phages, the Cas9 proteins from different strains seem to have differing PAM specificities. Thus, the targeting range of the CRISPR-Cas platform can be expanded by adding more Cas9s to choose from (Kleinstiver *et al.* 2015).

Of similar import is specificity. Because RNA and DNA can form complementary complexes even in the presence of a small number of mismatches between the two molecules' sequences, CRISPR-Cas9 often cleaves at off-target locations in a generally unpredictable manner. This has huge implications in applications of CRISPR-Cas9, as off-target activity can possibly have catastrophic effects on a cell. We have recently described an assay for detecting off-target cleavage events by Cas9 in an unbiased manner, which can be used to better understand the specificities of different Cas9 proteins (Tsai *et al.* 2015).

2.1 SELECTED CAS9 PROTEINS

Strain	CRISPR System Type	Putative PAM
<i>Lactobacillus buchneri</i>	Type II-A	5'-AAAA-3'
<i>Neisseria meningitidis</i>	Type II-A	5'-NNNNGATT-3'
<i>Francisella novicida</i>	Type II-B	5'-NG'3'
<i>Campylobacter jejuni</i>	Type II-C	5'-NNNNACA-3'
<i>Pasteurella multocida</i>	Type II-C	5'-GNNNCNNA-3'

Table 2.1: Selected Cas9 proteins and their strains of origins and putative (predicted) PAM sequences.

The five orthogonal Cas9 proteins selected for characterization are from a variety of families and CRISPR system subtypes, and have very disparate putative PAM sequences. All orthologs were chosen from the phylogenetic tree of Cas9 orthologues presented in the supplemental documents in Fonfara *et al* (2013). Note the inclusion of *N. meningitidis* Cas9, which has been profiled previously in the literature, and will be used in part as a positive control. It should also be noted that all selected Cas9s have been shown to have *in vitro* activity with a dual-RNA (non-chimeric) guide in the supplemental information of Fonfara *et al* (2013).

2.2 EXPERIMENTAL APPROACH

2.2.1 PREDICTION OF RNA GUIDES

The first step in using any Cas9 is to determine its crRNA and tracrRNA to construct a chimeric gRNA to guide the Cas9 activity. In general (for type II CRISPR systems), the Cas9 open reading frame, CRISPR repeats, and tracrRNA are all very near one another, generally within a single $\approx 5kb$ stretch. The location of these loci can easily be found through a simple BLAST search of SpCas9 against the bacterial genome to find the orthogonal Cas9

gene. Then, a repeat masking algorithm is applied to find fixed-length repeats in the region $\pm 2kb$ from the Cas9 protein. These repeats are the CRISPR repeats, which are spaced by fixed-length spacer sequences. Finally, the crRNA repeats are aligned against the same $\pm 2kb$ region to find the tracrRNA, which starts at the region of complementarity and ends at a poly-T transcriptional termination signal. Once these two sequences are found, they are fused together via a GAAA linker loop to form a single gRNA. This method was used computationally implemented to predict (with success) the gRNA for SaCas9 in Kleinstiver *et al.* 2015. The same pipeline will be used to predict the crRNA, tracrRNA, and gRNA of the orthogonal Cas9 proteins in this study.

2.2.2 POSITIVE SELECTION

To test if the predicted gRNA can indeed transactivate and guide the orthogonal Cas9 protein, a positive selection assay will be used. In this assay, a toxic gene is placed in front of a glucose-inducible promoter on a plasmid containing a set Cas9 target site containing the predicted PAM sequence. The plasmid is co-transformed with a plasmid containing both the orthogonal Cas9 gene and the gRNA gene and the bacteria are grown in glucose-containing media. If there is bacterial growth, then the gRNA activated and guided Cas9 to cleave the target site on the toxic plasmid-containing gene, keeping the bacterium alive. Thus, this positive selection experiment can be used as an early screen for gRNA efficiency.

2.2.3 NEGATIVE PAM DEPLETION ASSAY

Once the gRNA is confirmed to work through the positive selection assay, a negative selection assay will be used to fully characterize the PAM specificity of the protein. In this assay, an antibiotic resistance gene is placed on the same plasmid as a fixed target site followed

by a randomized 8bp PAM region (i.e. NNNNNNNN). This plasmid is co-transformed with a plasmid containing the orthogonal Cas9 and gRNA, and the bacteria are grown in antibiotic-containing media. If the randomized PAM matches the Cas9 PAM, then the resistance plasmid is cleaved, and the bacterium dies. At the end of the assay, the region containing the spacer and PAM is sequenced using high-throughput sequencing. The relative abundance of each PAM sequence is analyzed computationally. PAM sequences that are the most depleted are the most active PAM sequences for the Cas9 protein. An example of the results from this assay, as found in Kleinstiver *et al.* 2015 can be found in Figure 1. The code used to determine relative PAM frequencies can be found in the supplemental information of that same text.

PAM	Count	Frequency
GTT	27360	3.00391958805898E-2
TGT	26445	2.9034595579758699E-2
TTT	24614	2.70242970542704E-2
GGT	23149	2.5415838649114501E-2
TAT	22106	2.42707040985496E-2
TGC	21063	2.3125569547984798E-2
AGT	20840	2.2880732534776699E-2
GCT	20368	2.2362512488883501E-2
:	:	:
TAG	6433	7.0629439729471598E-3
CAG	4531	4.9746928558096603E-3
AAG	866	9.5080203335492597E-4
GAG	457	5.0175118850254196E-4
GGG	386	4.2379859685335002E-4
TGG	322	3.5353147198647402E-4
AGG	309	3.39258462247889E-4
CGG	278	3.0522282364049599E-4

Figure 2.1: Table Excerpt of PAM Depletion Assay analysis for EGFP site 1 for wild-type *S. pyogenes* Cas9. Cells are in descending frequency order. Note that the most depleted PAM sequences are those in the format NGG, and are depleted by two orders of magnitude. The second most depleted set are those in the format NAG, which are less depleted but still somewhat depleted compared to the rest of the PAM sequences.

2.2.4 GUIDE-SEQ SPECIFICITY PROFILING

GUIDE-Seq is a method for unbiased detection of off-target nuclease activity that we recently described in Tsai *et al.* 2014. In this assay, a short, end-stabilized DNA strand is captured into double stranded breaks through NHEJ capture, wherein the blunt ends of the oligo is captured into the blunt double stranded break by endogenous DNA repair machinery. These regions are then linearly amplified off of the known oligo sequence and sequenced with high-throughput sequencing. By mapping these sequences back to the genome, the sites of off-target cleavage events are elucidated.

In this study, GUIDE-seq will be performed for each orthogonal Cas9 for a number of targetsites to determine the number of off-target cleavage events that each protein is generally responsible for. This will provide a set of data that will be valuable in determining which Cas9 is best used in a given application (e.g. a Cas9 with very low off-target effects may be better for therapeutic applications).

2.2.5 EXPECTED RESULTS AND POTENTIAL ROADBLOCKS

The methods outlined above are, in a way, a “Hail Mary” approach to test the activity of the selected Cas9s. For the positive selection to work, three things must simultaneously be successful. First, both the chimeric guide and the Cas9 must both be successfully expressed in high numbers, and be non-toxic to the cell. Second, the predicted crRNA and tracrRNAs must be correct, and the chimeric guide constructed from them must activate and guide the Cas9. Third, the Cas9 must induce a double stranded break in the targeted DNA sequence. Only then will the plasmid containing the toxic gene be degraded, and the cell will live. To ensure that the system should work in its totality, the *N. meningitidis*

Cas9 (NmCas9) will be used as a positive control, as it has been profiled in a number of published studies, and so its activity in the positive selection can be expected. In the case of further complications, individual steps of the expression and cleavage process will be evaluated. For example, the expression of the protein and the gRNA could be evaluated via RT-qPCR. One other possible concern is whether or not the tracrRNA secondary structure in the gRNA will accurately reflect the endogenously processed tracrRNA that would ordinarily transactivate the Cas9 protein. For this reason, multiple different truncations of the tracrRNA will be predicted and tested. Though we expect at least some nonempty subset of the 5 selected Cas9s to work in this assay, it is possible that only the NmCas9 will work successfully, and that the other four Cas9s simply do not work *in vivo* outside of their host organisms.

The negative PAM depletion assay also carries a few assumptions that must be taken into consideration. In particular, the random 8bp PAM is assumed to be fully random, with every sequence having an equal probability of appearing. However, due to inherent biases in the oligonucleotide synthesis process, this is not always indeed the case. To account for this, the PAM depletion assay will be carried out with the same randomized PAM library with *S. pyogenes* Cas9, which has the most well-characterized PAM sequence. In doing so, we can make sure that there are not significant biases in the PAM library that may skew our PAM characterization results.

If we show that one or more of these Cas9s have activity *in vivo*, then the overall sequence specificity of the Cas9 can be profiled in an unbiased fashion through GUIDE-Seq. Though we have shown previously that off-target profiling cannot be theoretically or computationally predicted (Tsai *et al.* 2015), we hypothesize that Cas9s with shorter putative PAM se-

quences (e.g. *F. novicida*, which has a putative PAM of 5'-NG-3') will have less stringent sequence specificity.

3

GUIDE-Seq: An unbiased assay for genome-wide off-target detection

- 3.1 EXPLORATORY ANALYSIS OF POTENTIAL OFF-TARGET SITES
- 3.2 OVERVIEW OF THE GUIDE-SEQ ASSAY FOR PRODUCTION OF RAW READ DATA
- 3.3 BIOINFORMATIC ANALYSIS OF GUIDE-SEQ DATA TO INFER OFF-TARGET SITES
- 3.4 COMPARISON OF GUIDE-SEQ RESULTS TO CHIP-SEQ AND COMPUTATIONAL

4

guideseq: A computational pipeline for analyzing GUIDE-Seq data

4.1 UNDERSTANDING THE CORE STEPS OF THE *GUIDESEQ* PIPELINE

4.2 EXAMPLE OUTPUT OF THE SOFTWARE

5

Expanding the Targeting Range of CRISPR-Cas with Orthogonality Profiling and Protein Engineering

- 5.1 ENGINEERING THE PAM-CONTACTING RESIDUES OF CAS9
- 5.2 PREDICTING THE ACTIVITY OF ORTHOGONAL CAS9 SYSTEMS
- 5.3 UNDERSTANDING THE PAM DEPLETION ASSAY
- 5.4 PAM DEPLETION ASSAY RESULTS

6

Characterizing the Specificity of CRISPR-Cas Proteins

6.1 OVERVIEW OF THE SPACER SUBSEQUENCE RANDOMIZATION ASSAY

6.2 SPECIFICITY RESULTS OF RANDOMIZATION ASSAY

7

CasBLASTR: A tool for Genome Editing Experiment Design and Analysis

7.1 APPLICATION ARCHITECTURE

7.2 RESULTS AND USAGE STATISTICS

8

Conclusion and Future Steps

References

- [1] Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519), 569–573.
- [2] Baltimore, D., Berg, P., Botchan, M., Carroll, D., Charo, R. A., Church, G., Corn, J. E., Daley, G. Q., Doudna, J. A., Fenner, M., Greely, H. T., Jinek, M., Martin, G. S., Penhoet, E., Puck, J., Sternberg, S. H., Weissman, J. S., & Yamamoto, K. R. (2015). Biotechnology. A prudent path forward for genomic engineering and germline gene modification. *Science (New York, N.Y.)*, 348(6230), 36–38.
- [3] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), 1709–1712.
- [4] Chylinski, K., Le Rhun, A., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology*, 10(5), 726–737.
- [5] Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602–607.

- [6] Doudna, J. A. & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213), 1258096–1258096.
- [7] Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K. S., Lécivain, A.-L., Bzdrenga, J., Koonin, E. V., & Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research*, 42(4), 2577–2590.
- [8] Jansen, R., Embden, J. D. A. v., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, 43(6), 1565–1575.
- [9] 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 (New York, N.Y.)*, 337(6096), 816–821.
- [10] Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*.
- [11] Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., Sun, Y., Bai, Y., Songyang, Z., Ma, W., Zhou, C., & Huang, J. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & cell*.

- [12] Marraffini, L. A. & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463(7280), 568–571.
- [13] Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), 174–182.
- [14] Nuñez, J. K., Lee, A. S. Y., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, 519(7542), 193–198.
- [15] Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, 151(Pt 3), 653–663.
- [16] Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520(7546), 186–191.
- [17] Staals, R. H. J., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., Koehorst, J. J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P. J., Urlaub, H., Heck, A. J. R., Nogales, E., Doudna, J. A., Shinkai, A., & van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Molecular cell*, 56(4), 518–530.

- [18] Taylor, D. W., Zhu, Y., Staals, R. H. J., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E., & Doudna, J. A. (2015). Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science (New York, N.Y.)*, (pp. aaa4535).
- [19] 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., Aryee, M. J., & Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, 33(2), 187–197.