TAILORING CRISPR/CAS SPECIFICITY USING CHEMICALLY MODIFIED GUIDE RNAS

by

Christopher Robert Cromwell

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Pharmacology

University of Alberta

Abstract

The ability to reliably target specific sequences of interest is a critical component of any gene editing technology. While the discovery of CRISPR-Cas systems have fundamentally simplified the way in which genomic material may be manipulated, the presence of off-target activity at undesired loci still remains a concern, especially with regards to human clinical applications. Conversely, the high levels of polymorphic sequences present within the natural population also present a unique challenge to the broader application of CRISPR-based therapeutics, as high levels of specificity may predispose some individuals to therapy failure.

Here, I describe how a variety of chemically modified nucleotides may be incorporated into the sequence-specific spacer region of the crRNA of either Cas9 or Cas12a, in order to modulate specificity. Initially, I demonstrated that chemically modified nucleotides (such as bridged nucleic acids and locked nucleic acids) are an effective strategy for broadly enhancing the specificity of Cas9 cleavage both *in vitro*, and in cells. In addition to providing design considerations for the effective use of BNA/LNA-modified crRNAs, I also identified a potential biochemical mechanism for the observed specificity improvements through smFRET experiments with collaborators. Building upon these discoveries, I expanded the scope of my work to include several additional chemically modified nucleic acids (or xeno nucleic acids; XNAs). Our lab hypothesized that while BNA/LNA incorporation lead to higher specificities, we may also be able to identify novel modifications with even greater fidelities. Among the XNA modified crRNAs (crXNAs) tested, I describe several crXNAs capable of imparting high levels of cleavage specificity to the Cas9 system. I also show that while different crXNAs may display similar levels of off-target

discrimination overall, each crXNA has a unique specificity profile regarding the off-target sequences they are able to effect.

Given our success with crXNAs, we also investigated the ability of chemically modified nucleotides to impart properties to CRISPR-based systems independent of enhancing specificity. With collaborators, I demonstrated that through the incorporation of universal bases into the spacer sequences of either Cas9 or Cas12a crRNAs, we are able to target multiple polymorphic sequences without loss of overall specificity. Alongside these collaborators, I helped to demonstrate the applicability of this technology to the targeting of genes with high levels of polymorphisms, as well as to the *in vitro* diagnostic platform known as DETECTR, for the detection of multiple HIV-1 variant sequences.

In summary, the work presented herein displays the versatility of chemically modified crRNAs for augmenting the capability of CRISPR-based technologies with novel properties. These findings outline a framework for the identification and characterization of chemically modified guide RNAs, introducing a versatile tool for future CRISPR-based research.

Preface and Co-Authorship

This thesis is an original body of work by Christopher Robert Cromwell.

Chapter 2 has been published: Cromwell, C.R., Sung, K., Park, J., Krysler, A.R., Jovel, J., Kim, S.K., Hubbard, B.P. Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat Commun* **9**, 1448 (2018). https://doi.org/10.1038/s41467-018-03927-0

Together, B.P.H and C.R.C conceived the study, as well as designed experiments in order to characterize the effects of BNA-modified and LNA-modified crRNAs on Cas9 cleavage specificity. Experiments pertaining to specificity of BNAs *in vitro* and in cells was performed by C.R.C, and assisted by A.R.K. Single-molecule FRET experiments were designed and performed by K.S. and J.P. under the supervision of S.K.K. Bioinformatic analysis of high-throughput sequencing was performed by J.J. All authors contributed to writing of the final manuscript.

Chapter 3 is under preparation for publication: Cromwell, C.R., Jovel, J., Kalmady, S.V., Greiner, R., Hubbard, B.P. High-throughput specificity profiling for the identification of Cas9 crXNAs with superior specificity.

B.P.H and C.R.C conceived the study, and designed the manner in which crXNAs would be modified. All experiments regarding activity and specificity of the crXNAs was performed by C.R.C. J.J. performed bioinformatic analysis of high-throughput specificity

profiling data. S.V.K and R.G. provided experimental advice on future machine learning applications, and designed the set of diverse Cas9 target sites tested.

Chapter 4 has been published: Krysler, A.R*., Cromwell, C.R*., Tu, T., Jovel, J., Hubbard, B.P. Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences. *Nat Commun* **13,** 1617 (2022). https://doi.org/10.1038/s41467-022-29202-x

The study was conceived by B.P.H, A.R.K and C.R.C. Experimental design for testing the incorporation of universal bases into crRNAs of Cas9 and Cas12a for activity and specificity was performed by B.P.H, A.R.K and C.R.C. *In vitro* activity, kinetics and thermodynamics of modified crRNAs was performed by A.R.K and C.R.C, and assisted by T.T. All experiments testing activity of Cas9 and Cas12a in cells was performed by C.R.C. J.J. performed bioinformatic analysis on high-throughput specificity profiling data. All authors contributed to writing of the manuscript. *Denotes equal contribution.

Acknowledgments

I would like to thank Dr Basil Hubbard, my supervisor and mentor for whom without, this thesis would not be possible. With your expertise and guidance, I've grown not only as a scientist and researcher, but as a person as well. I am incredibly grateful for the opportunity, and cannot wait to see the many accomplishments and discoveries this lab will continue to make.

To my friend and colleague Evan, thank you for all of the memories. We had quite a rollercoaster of a program together, but have had some amazing times that have seen us move across the country. To Karin, Angelica, Amanda, Robert and all of the other members of the Hubbard Lab (past and present), thank you for making it an incredible space to work. I know I've made some life-long friends in the process. I would also like to thank our long-term collaborator Dr Juan Jovel, whose expertise has been invaluable throughout my program. As well, a special thank you to Sudip, for your technical assistance, as well as the many great hours working together in TAGC. I also want to take this opportunity to thank my supervisory committee, Dr Robert Campbell and Dr Simonetta Sipione. With your feedback and advice, I truly believe I've not only become a better scientist in the lab, but also learned the invaluable skills necessary for presenting and sharing my work effectively.

I would also like to thank the people who have supported me throughout this journey outside of the lab. To Micheala, your support has done more than you will ever know and I am forever grateful. Thank you to my grandparents, for making my time as a student possible, and for always taking the time and interest to listen. Finally, I would like to thank my family for their unconditional love and support. I can never truly express the level of gratitude I have towards you all for making this accomplishment possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 History of Genome Engineering & CRISPR-Cas9

The ability to edit and manipulate genetic information has allowed for significant progress to be made in biology, biochemistry and medicine; facilitating discoveries from the molecular mechanisms of gene regulation[1-3], to the development of novel therapeutic strategies[4-6]. This has largely been accomplished through the use of prokaryotic-derived molecules, beginning in 1971 with the generation of specific DNA fragments through restriction enzyme digestion[7]. Since then, scientists have developed technologies which extend beyond the rigid binding requirements of restriction enzymes in the form of programmable nucleases. Such examples include: meganucleases[8], zinc-finger nucleases (ZFNs)[9], transcription activator-like effector nucleases (TALENs)[10] and RNA-guided engineered nucleases (RGENs) derived from the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas system[11]. Classically, the development of modular DNA-binding proteins which interact with custom DNA target sequences has been a complex process, often requiring extensive knowledge of protein engineering and molecular cloning[12-14]. However, the discovery and characterization of the CRISPR-Cas9 system over the past decade has functioned to absolve much of this complexity. In contrast to previous technologies which rely on DNA-protein interactions, the CRISPR-Cas9 system utilizes nucleic acid base pairing for sequence recognition, allowing for modular DNA targeting through simple nucleic acid sequence substitution, in place of engineering a new protein entirely.

Although the term CRISPR wasn't used until much later[15], the repeated sequences (which are a hallmark of all CRISPR systems) were initially observed in *Escherichia coli* by Yoshizumi Ishino

and his colleagues in 1987[16]. Contrasting to the tandem repeat sequences commonly found in genomes which occur directly adjacent to each other[17], the disparate repeats of CRISPR loci were found to be separated by non-repeating sequences, later known as spacers. Advancements in sequencing technology over the next decade allowed researchers to not only complete the Human Genome Project[18], but also allowed for the sequencing of many other organisms, including bacteria, archaea and bacteriophages. Computational analysis of these genomes led to the realization that the non-repeating spacer sequences of CRISPR loci belonged to viruses and other mobile genomic elements[19-21]. Additional analyses also resulted in several other key discoveries, including the identification of CRISPR sequences in 90% of sequenced archaea and 40% of sequenced bacteria[22]. Several well-conserved CRISPR-associated (*cas*) genes were also identified[15]. This presence of viral- and/or exogenously-derived sequences within the CRISPR arrays led many to hypothesize that this may be a form of adaptive immunity[19, 23], although the underlying mechanisms were yet to be elucidated.

The CRISPR-Cas system was first experimentally demonstrated to function as an adaptive immune system in 2007 by scientists at Danisco studying *Streptococcus thermophilus*, a bacterium widely used in the dairy industry for the production of yogurt and cheese[24]. Phillipe Horvath and Rodolphe Barrangou successfully demonstrated that through the addition or removal of spacer sequences from the genome of *S. thermophilus*, he could modulate the resulting resistance of the bacteria to phages bearing the same genomic sequences[25]. Following this discovery, researchers quickly began to uncover the mechanisms behind the CRISPR-Cas system, demonstrating that the activity of Cas enzymes are directly dictated by short CRISPR RNAs (crRNAs) transcribed from said spacer sequences[26]. Acquired spacer sequences were also observed to contain sequences highly

similar to each other at regions known as protospacer-adjacent motifs (PAMs)[27]. This was later demonstrated to be a critical component of DNA recognition for CRISPR systems, as well as being crucial for self, versus non-self recognition[28]. It was also demonstrated that of the numerous Cas proteins within *S. thermophilus*, Cas9 was the only protein with measurable nuclease activity against DNA[29]. In addition to a crRNA, a second auxiliary RNA, known as a trans-activating crRNA (tracrRNA) was also shown to be required for Cas9 activity[30]. Specifically, the tracrRNA base pairs with the repeat sequences found in the crRNA to form a dual-RNA hybrid structure[30]. This dual-RNA, known as a guide RNA (gRNA), directs Cas9 to degrade any DNA containing a complementary 20-nucleotide (nt) target sequence (or protospacer) and adjacent PAM[31, 32].

Perhaps one of the most critical discoveries in the timeline of CRISPR's adaptation as a genome editing tool, was the demonstration that CRISPR systems are self-contained units and can function heterologously in other species. This was performed by Virginijus Siksnys in 2011, when he was able to successfully express the entire CRISPR-Cas locus from *S. thermophilus* in *E. coli*[33]. This also demonstrated that all of the required components for this particular CRISPR-Cas system were now known. Immediately following this discovery, the individual components of the CRISPR-Cas system were biochemically characterized, which allowed for further simplification of the system[31, 32]. As mentioned above, the endogenous CRISPR-Cas9 system requires two non-coding RNA elements: a processed, mature crRNA (which contains the sequence-specific spacer region) and a tracrRNA. However it was demonstrated that these two RNAs may be covalently linked as a single, chimeric RNA known as a single guide RNA (sgRNA) facilitating a more simple approach to gRNA synthesis[32]. Shortly after, it was demonstrated that the CRISPR-Cas9 system may be reconstituted in eukaryotic cells for genome editing purposes[34-36]. For the first time, scientists now

had access to a highly customizable gene editing platform capable of targeting virtually any location within the genome through an interchangeable, short RNA oligonucleotide.

1.2 Biochemistry of CRISPR-Cas Systems

As new microbial genomes continue to be sequenced, the number of unique CRISPR-Cas systems identified continues to increase. While the nomenclature surrounding CRISPR classification remains fluid as new discoveries are made, CRISPR-Cas systems are generally categorized into one of two classes, each containing several types and sub-types (**Figure 1.1**). Class 1 is divided into types I, III and IV; while Class 2 contains types II, V and VI[37-39]. The main differentiator between classes is the main effector of nucleic acid degradation. Class 1 CRISPR-Cas systems utilize a multi-protein complex to degrade foreign genetic material, while Class 2 systems use only a single, generally larger Cas protein for interference[38]. These 6 types may be further divided into 19 sub-types based on the presence of a gene unique to each category[40].

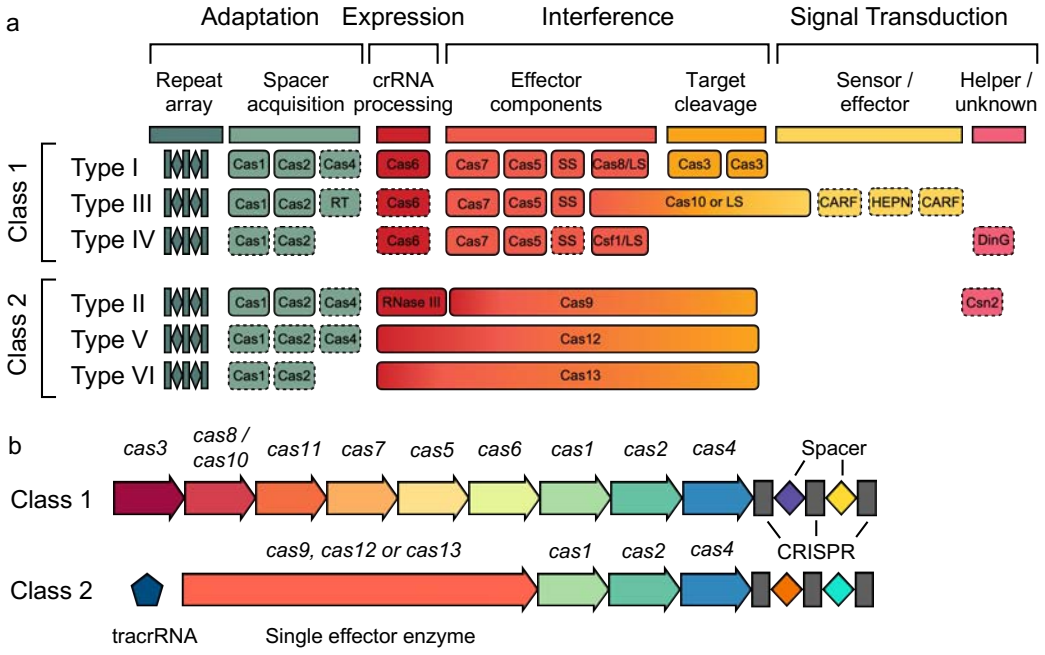


Figure 1.1. Classification of CRISPR systems
CRISPR-Cas systems are divided into two classes based on their effector modules. Class 2 systems employ a single, multi-domain effector protein, while Class 1 systems use a multi-protein effector complex to accomplish the same functions. (**a**) Diagrammatic representation of the functional components of CRISPR-Cas systems according to current nomenclature. Dashed outlines indicate components which are disposable and/or missing in some variants. LS = large subunit; SS = small subunit. (**b**) Representation of the genetic organization of both Class 1 and Class 2 CRISPR-Cas systems. Adapted from Makarova *et al*. 2020[39].

1.2.1 Cas9

While several CRISPR-Cas systems have been re-purposed for genome editing applications[41, 42], *Streptococcus pyogenes* Cas9 (SpyCas9), a Class 2, type II CRISPR system remains the most widely utilized. SpyCas9 (herein referred to as Cas9 unless otherwise noted) is a large, multi-domain, RNA-guided DNA endonuclease, 1,368 amino acids in length[28]. Two independent nuclease domains, an HNH-like nuclease domain and RuvC-like nuclease domain, function to cleave DNA, resulting in a blunt-ended, double-stranded DNA break (DSB) 3 base-pairs (bp) upstream of the PAM (5′-NGG-3′ for *S. pyogenes* Cas9)[31, 32, 43]. Cas9 also possess functional roles in the process of spacer acquisition and crRNA maturation[44], although this is beyond the scope of this thesis.

In its apo state (i.e., without a gRNA), Cas9 presents as two distinct lobes: the alpha helix rich recognition (REC) lobe, and the nuclease (NUC) lobe. The NUC lobe contains the well conserved HNH nuclease domain and split RuvC nuclease domain, as well as the C-terminal domain (CTD) (**Figure 1.2**)[45, 46]. The CTD contains several PAM-interacting sites, and is therefore required for DNA binding through PAM interrogation. However this region is largely disorganized in the apo state, indicating that prior to binding of a guide RNA (gRNA), the enzyme is kept in an inactive configuration which is unable to properly recognize target DNA[32]. Experiments involving DNA-

curtain assays demonstrated that apo-Cas9 is able to bind DNA non-specifically, although this DNA may be readily detached through the addition of competitor RNA (i.e., a gRNA)[47].

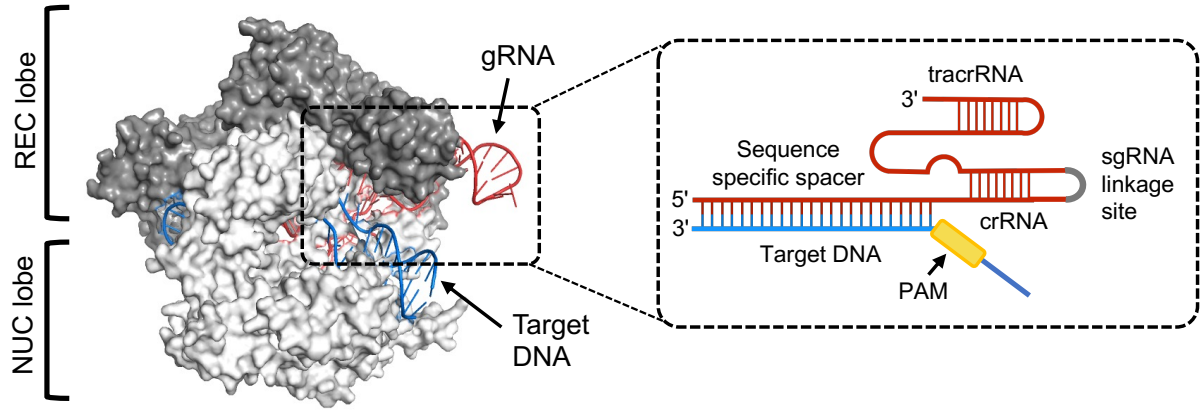


Figure 1.2. Crystal structure of Cas9 in complex with a gRNA and DNA target
Pymol model of Cas9 complexed with a gRNA (red) and target DNA (blue) (PDB: 4OO8)[46], with lobes individually colored (REC lobe = dark gray, NUC lobe = white) (left). A diagrammatic representation of the interactions between the gRNA and target DNA is also shown (right). The position in which the crRNA and tracrRNA may be covalently linked as a sgRNA is indicated in grey.

Structural studies which compared the nuclease domains of Cas9 with other homologous DNA-bound nucleases revealed similarities to the RNase H fold-containing members of the retroviral integrase superfamily[46]. As the RuvC nuclease domain of Cas9 contains an RNase H fold, this suggests that the domain is likely to employ a two-metal-ion biochemical mechanism for nicking of the non-target DNA strand (as is seen for other RNase H fold-containing nucleases within the retroviral integrase superfamily)[45, 46]. Common to other well conserved HNH nuclease domains, the HNH domain of Cas9 adopts a ββα-metal fold, and therefore likely utilizes a one-metal-ion dependent mechanism for nicking of the target strand of DNA[28]. Nucleases which employ one- or two-metal-ion catalysis as a mechanism for nucleic acid cleavage contain an often conserved

histidine, or an absolutely conserved aspartate residue respectively[48]. Cas9 mutagenesis studies demonstrated further support for these mechanisms of catalysis, as mutating either the HNH (H840A) or RuvC (D10A) domain (at these conserved residues) results in the conversion of Cas9 to a nickase. Mutation of both domains results in a catalytically deficient ("dead" or dCas9) version of Cas9 capable of RNA-guided DNA binding, but a complete loss of nuclease activity[32].

In order to form an active DNA surveillance complex and subsequently achieve site-specific DNA recognition and cleavage, Cas9 must be complexed with a gRNA (either as a crRNA-tracrRNA hybrid or sgRNA)[32, 45]. Within the crRNA, it is the 20 nt spacer sequence located at the 5′ end of the RNA oligonucleotide which confers the majority of the system's specificity[49, 50]. The tracrRNA has functions in Cas9 recruitment and facilitates binding of the crRNA to the Cas9 enzyme[51]. Of the 20 nts which make up the spacer region, the 10-12 nts proximal to the PAM are the most integral for cleavage specificity and are known as the seed region[32, 35, 47]. Mismatches present within the seed region often severely impeded or completely eliminate DNA binding and cleavage, while off-targets with sequences highly homologous to the seed region are often tolerated, even when mismatches are present elsewhere[52].

Extensive contacts between Cas9 and the ribose-phosphate backbone of the gRNA pre-order the 10-12 nt seed region in an A form conformation as a prerequisite to initial DNA interrogation[46]. It is hypothesized that this pre-ordering is thermodynamically favourable for target DNA binding[53]. In addition to the conformational rearrangements of the seed region, binding of the gRNA to Cas9 also pre-positions the PAM-interacting residues for PAM recognition (5′-NGG-3′ for *S. pyogenes* Cas9)[54]. In the apo state, these residues are disordered[55]. Contrary to the seed region, the 10 nt

PAM-distal region of the crRNA is completely disordered in the Cas9 sgRNA-bound crystal structure[46]. Although disordered, these nucleotides are protected from degradation in a cavity found between the HNH and RuvC nuclease domains, requiring additional conformational rearrangements for release[56].

DNA target recognition requires both complementarity between the 20 nt spacer sequence of the crRNA and DNA target, as well as the presence of a well-conserved PAM directly adjacent to said target[31, 32]. Recognition of a suitable DNA target occurs through three-dimensional diffusion in which the dwell time of the Cas9 gRNA complex is determined by the level of complementarity between the gRNA and target (but only when an appropriate PAM is present). In the absence of a PAM, the complex rapidly dissociates from the DNA[47, 57, 58]. PAM identification results in local DNA melting at a nucleation site directly adjacent, followed by RNA strand invasion (**Figure 1.3**)[47]. This results in an RNA-DNA hybrid structure, as well as a displaced strand of DNA known as an R-loop[46]. DNA unwinding proceeds in a 3′ to 5′ direction until all 20 nts of the spacer form Watson-Crick base pairs with the target DNA. The resulting heteroduplex exists in a distorted conformation that is largely A form and can be found between the REC and NUC lobes. This RNA-DNA heteroduplex is recognized by Cas9 through its geometry in a sequence-independent manner, rather than its nucleobase composition[54]. Target binding and R-loop formation result in conformational rearrangements within Cas9, which re-position the HNH domain for cleavage of the DNA target strand[59, 60]. This re-orientation of the HNH domain also causes rearrangements of two hinge regions which connect the HNH domain with the RuvC domain, demonstrating the allosteric control over the RuvC domain (via the HNH domain) for concerted DNA cleavage[56].

Following DNA cleavage, Cas9 remains tightly bound to the DNA until it is displaced through other cellular factors[47].

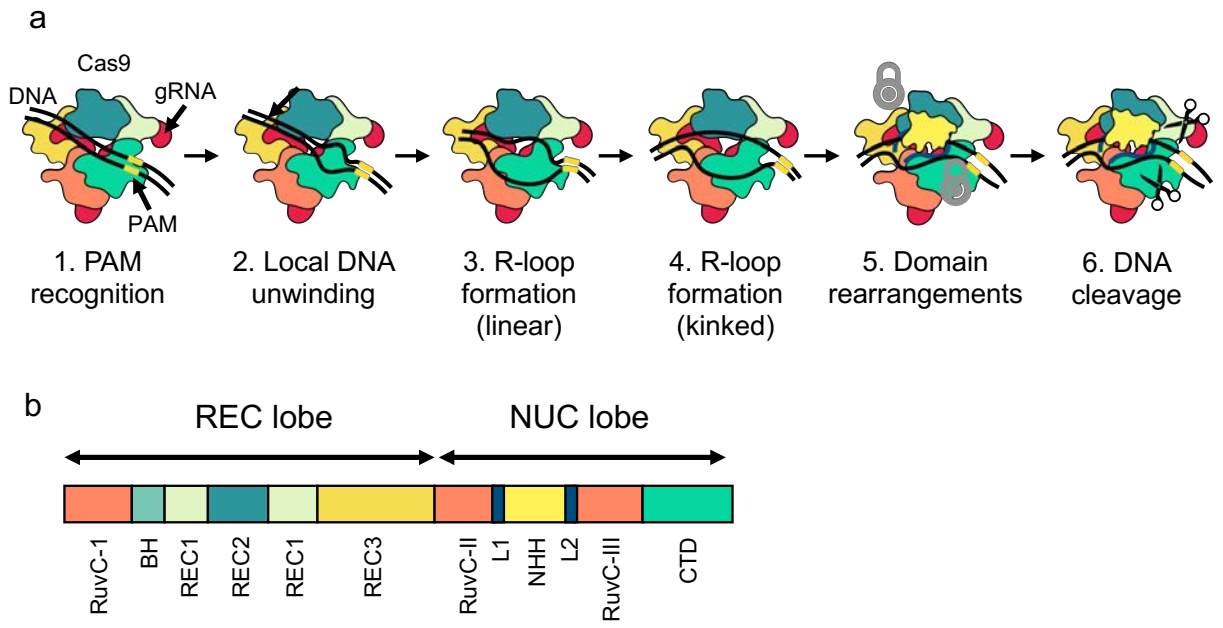


Figure 1.3. Mechanism of Cas9-mediated DNA cleavage
(**a**) Biochemical mechanism of Cas9 target recognition and cleavage. Initial binding of the Cas9 RNP complex to a target DNA is determined by the presence of a suitable PAM (yellow). Once bound, local DNA unwinding proceeds in a 3′ to 5′ direction, testing for complementarity between the sequence-specific spacer region of the crRNA, and the 20 nt DNA protospacer. This results in the formation of an R-loop, which is required for downstream conformational rearrangements. These R-loop mediated rearrangements allow for repositioning of the HNH and RuvC nuclease domains, followed by the introduction of a double-stranded DNA break. (**b**) Organization of the domains of *S. pyogenes* Cas9. BH = bridge helix, L1= linker 1, L2 = linker 2, CTD = C-terminal domain. Adapted from Bravo *et al*. 2020[60].

1.2.3 Cas12a

Cas12a (formerly known as Cpf1) is a Class 2, type V CRISPR-Cas system with functional similarities to Cas9 despite evolving through an independent pathway. Similarly to Cas9, Cas12a is a large, multi-domain effector nuclease containing both a REC and NUC lobe[32, 61]. However, unlike Cas9, the NUC lobe of Cas12a is devoid of an HNH nuclease domain; instead utilizing a

single RuvC nuclease domain to achieve dsDNA cleavage[61]. Cas12a also only requires a single crRNA for sequence-specific DNA recognition and cleavage, in contrast to the two RNA oligonucleotides required by Cas9[32, 61, 62]. Transcription of CRISPR arrays results in the biogenesis of long RNA oligonucleotides known as pre-crRNAs[63]. These pre-crRNAs are subsequently processed into mature crRNAs before binding to their respective endonucleases. For Cas9, this pre-crRNA hybridizes with the independently expressed tracrRNA; allowing for recognition by Cas9 followed by RNase III cleavage to form the mature crRNA[30]. Contrasting to this, Cas12a does not have the requirement for a tracrRNA or RNase III, as the enzyme is capable of processing its own crRNAs through the presence of a ribonuclease domain[64]. This provides some inherent advantage over Cas9, as Cas12a is therefore a more minimalistic system.

Cas12a recognizes a T-rich PAM (often 5′-TTTV-3′) located directly upstream of the protospacer sequence. Cas12a derived from *Francisella novicada* (the first CRISPR-Cas12a system to be described) recognizes a spacer sequence of 27-32 bp in length, although this varies between species[61, 65]. As seen with Cas9, the crRNA of Cas12a also displays a seed region integral for target binding accuracy, albeit shorter than that of Cas9. In Cas9, the seed sequence is often depicted as 10-12 nt in length, located at the 3′ end of the spacer[32]. For Cas12a, this region is only 5-6 nt in length[66]. However, the differences in seed region length do not appear to manifest as differences in specificity, as Cas9 and Cas12a have shown to have similar mismatch tolerances *in vitro*[67], with apparent higher fidelities of Cas12a *in vivo*[68, 69]. These apparent higher specificities may be due to the more stringent PAM restrictions of Cas12a when compared to Cas9. Complete hybridization of the RNA/DNA heteroduplex results in cleavage of both DNA strands via the single RuvC nuclease domain. As the target and non-target DNA strands follow independent biomechanical

pathways in order to reach the RuvC catalytic site, this results in a staggered DSB as opposed to the blunt end produced by Cas9[61, 62].

Cas12a also possesses a unique biochemical property in that following sequence-specific DNA cleavage, Cas12a is able to non-discriminately degrade surrounding ssDNA in a sequence-independent mechanism. This is known as collateral or *trans* cleavage, and is a property of all type V CRISPR-Cas systems[70]. Prior to formation of the RNA/DNA heteroduplex, a protein segment known as "the lid" sterically blocks a cleft within the NUC lobe containing the catalytic residues. After hybrid formation, conformational changes within the lid result in an α helix conformation which interacts with the crRNA[71]. This interaction dissociates specific polar contacts resulting in stabilization of the DNase site, making the catalytic cleft available for non-discriminate ssDNA degradation once the distal region of the dsDNA substrate is freed from the complex[71].

1.2.4 Cas13 and Cas7-11

While the work encompassed by this thesis focusses primarily on the manipulation of DNA through CRISPR-Cas systems, it is important to note that several CRISPR systems have been described which act on RNA as a substrate instead of DNA. Class 2, type VI CRISPR-Cas systems are characterized by the presence of an RNA-guided, RNA endonuclease Cas13 (formerly known as C2c2)[72]. Type VI CRISPR systems may be further classified based on the phylogeny of Cas13, including: Cas13a, Cas13b, Cas13c, Cas13d, Cas13x and Cas13y (although this continues to expand as new systems are identified)[73-77]. All type VI systems contain two conserved higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains which confer the ability to degrade single-stranded RNA (ssRNA)[72]. Unlike previously described CRISPR systems, Cas13

does not require a PAM for RNA recognition and binding, but instead a preference for specific nucleotides adjacent to the target (known as a protospacer flanking site (PFS) sequence). Cas13a derived from *Leptotrichia shahii* was the first to be described, and presents a PFS of a single, non-G motif immediately adjacent to the 3′ end of the protospacer sequence[73]. However, additional Cas13 enzymes from different species have shown to have more flexible targeting requirements; some without the need for a PFS altogether[78]. As with CRISPR-Cas12a systems, Cas13 also demonstrates collateral cleavage following sequence-specific RNA degradation (albeit against ssRNA instead of ssDNA). With the ability to target RNA over DNA, Cas13 has been successfully repurposed as a means of RNA knockdown in a variety of cellular contexts[72, 79-81].

Cas7-11 was recently discovered, and is incredibly unique as it is a Class 1, type III CRISPR-Cas system which shares no identifiable similarities between domain architecture and function with previous CRISPR effector proteins. Like Cas13, Cas7-11 is also an RNA-guided, RNA endonuclease, but possesses no observable collateral cleavage following sequence-specific RNA degradation. Cas7-11 is also unique, as it is a single effector protein which originated from the fusion of multiple Cas7 subunits and an assumed Cas11 domain[82]. While significant work remains to be done regarding the biochemical mechanisms behind Cas7-11, it has already been demonstrated to function in *E. coli* for RNA interference applications, as well as in mammalian cells for RNA knockdown. Interestingly, Cas7-11 has also been shown to have no detectable off-target activity, or negative effects on cell viability[82].

1.3 Applications of CRISPR

CRISPR-Cas systems are powerful tools for genome editing, offering a wide range of potential applications (**Figure 1.4**). While the number of demonstrated uses far exceeds what is able to be discussed within this thesis, some of the most common applications are discussed below.

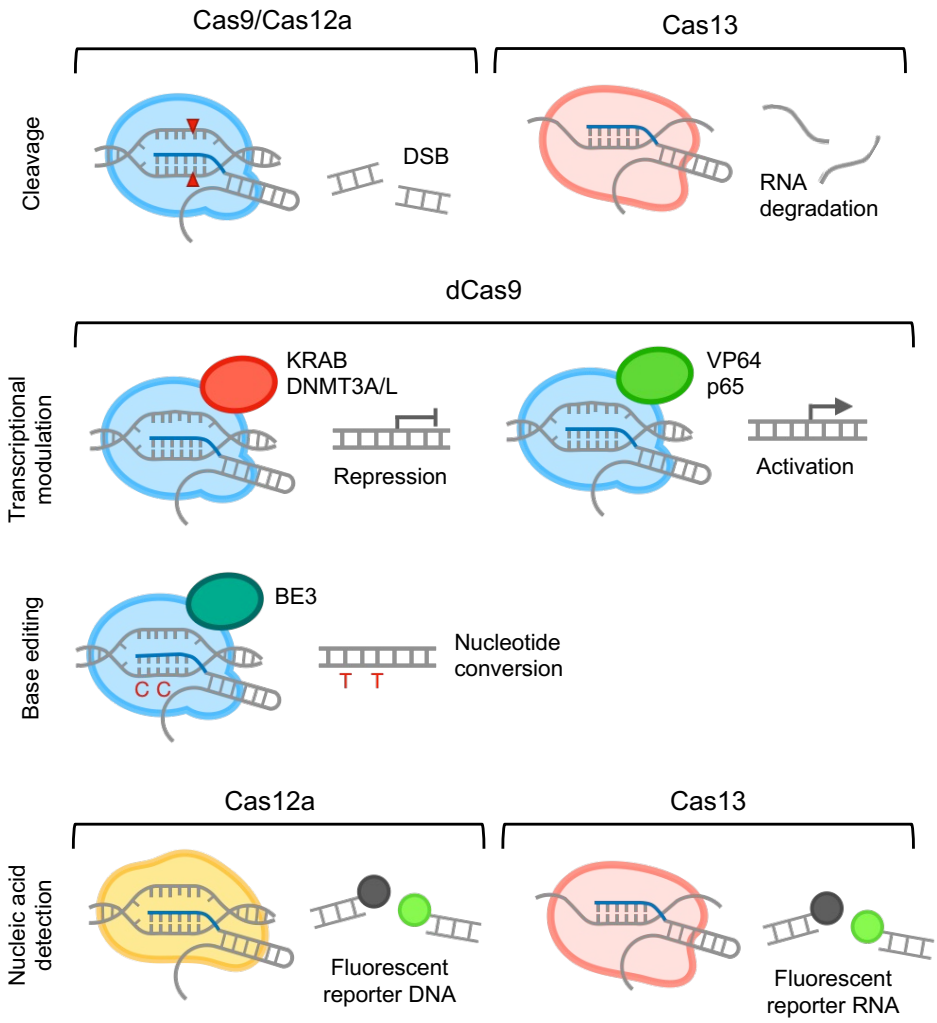


Figure 1.4. Common applications of CRISPR-Cas systems
Examples of DNA and RNA targeting CRISPR-Cas applications are shown. Catalytically active Cas enzymes may be used to bind and cleave DNA (Cas9 or Cas12a) or RNA (Cas13), while nuclease deficient Cas9 (dCas9) may function as a method for sequence-specific DNA targeting without cleavage. KRAB = Krüppel associated box; DNMT3A/L = DNA methyltransferase 3 alpha/like; VP64 = tetrameric VP16 transcription activator domain; p65 = NF-κB p65 subunit; BE3 = base editor.

1.3.1 Gene Editing Through Insertion or Deletion (Indels)

The first, and most prominent application of CRISPR-Cas9 as a technology, is the introduction of targeted DNA breaks at specific sites within the genome of an organism of interest. This was initially performed in mammalian cells using human codon-optimized Cas9 and gRNA, demonstrating the capability of the system for efficiency cellular activity[35, 36]. These DSBs are subsequently repaired through several endogenous DNA damage repair pathways which may be harnessed for the introduction, or removal of genetic material; this forms the foundation for gene editing as a technology (**Figure 1.5**)[83].

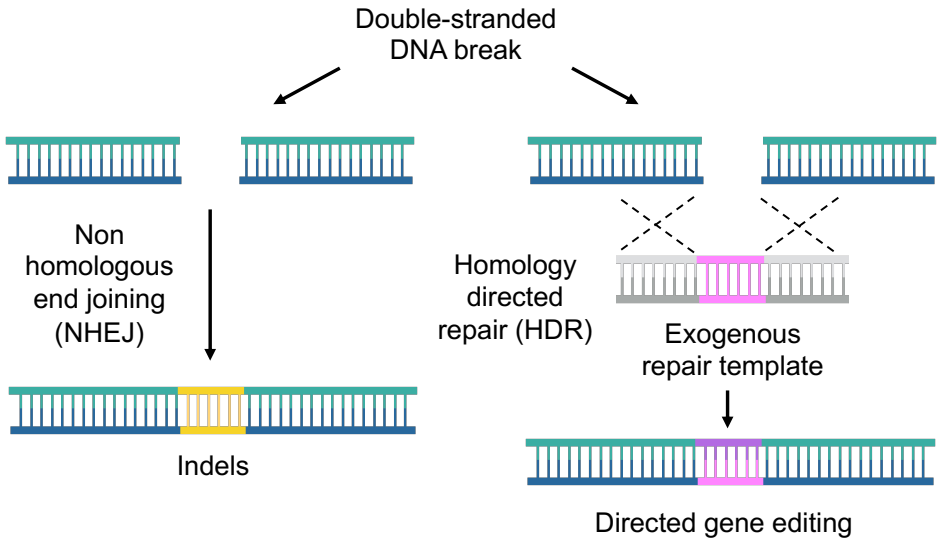


Figure 1.5. DNA repair pathways following Cas9-mediated DSBs
Following target cleavage, DSBs are most often repaired through either non-homologous end joining (NHEJ) or homology-directed repair (HDR) if a repair template is also provided.

Eukaryotic cells predominantly repair DSBs through the non-homologous end joining (NHEJ) repair pathway[84]. This pathway is highly error-prone and often leads to the accumulation of small insertions or deletions (indels)[85]. By targeting a coding exon with Cas9, the formation of indels

can be used to generate gene knockouts, as indel-mediated frameshift mutations often lead to premature stop codons or truncated gene products which disrupt functional protein expression[35, 86]. By simultaneously targeting two sites within a gene, it is also possible to introduce larger deletions, including deletions within the magnitude of megabases in length[87].

Alternatively to NHEJ, an exogenous repair template may be provided to stimulate the error-free, homology-directed repair (HDR) pathway (albeit at lower efficiencies)[88]. Donor repair templates containing epitope tags, or fluorescent protein-encoding genes have successfully been used in conjunction with HDR to monitor endogenously expressed proteins within their cellular environment[89, 90]. While NHEJ and HDR represent the predominant pathways in which DSBs are repaired in mammalian cells, there are several additional pathways (either independent or within NHEJ/HDR) such as microhomology-mediated end joining (MMEJ)[91] and the single-strand annealing (SSA) pathway[92] (although the intricacies of each are beyond the scope of this thesis).

1.3.2 Base Editing

Single nucleotide point mutations represent some of the most common genetic variants which cause disease in humans[93]. Therefore the ability to reliably introduce single nucleotide modifications at desired sequences of interest is critical for the generation of disease models and subsequent therapeutics. While it is possible to utilize Cas9-mediated HDR to generate such models, these strategies remain inefficient and incompatible with all cell types[94]. To circumvent this, single-base editing tools have been developed which utilize a Cas9 nickase (Cas9n) or nuclease-deficient dCas9 to achieve sequence-specific DNA targeting without the introduction of DSBs. Instead, the direct conversion of single nucleotides is accomplished through dCas9 or Cas9n

fusion proteins which are paired with a cytidine deaminase (either rat APOBEC1 or lamprey cytidine deaminase 1). These first generation base editors have been shown to be capable of C→T (or G→A) nucleotide conversion within a 5 bp window of activity located within the targeted protospacer sequence[95, 96].

Since their conception, several generations of base editors have been developed to facilitate improved base editing activity, specificity and targeting capability. For example, third generation base editors (BE3), which are fusion proteins of APOBEC1, a 16 amino acid XTEN linker, Cas9n as well as a uracil glycosylase inhibitor have been shown to achieve a nucleotide conversion rate of 15-75% in mammalian cells[95]. The addition of a uracil glycosylase inhibitor was incorporated in order to prevent base excision repair, therefore facilitating increased base editing efficiencies[95-97]. BE3 has also already been used in a variety of contexts, including: *in vivo* base editing within zebrafish and mice embryos through protein delivery[98, 99], AAV-mediated delivery to mice *in utero*[100], and direct injection of BE3- and corresponding gRNA-encoding mRNA into human embryos[101]. Notable base editor variations also include the use of high-fidelity Cas9 variants in place of wild-type Cas9[98], optimization of linker length between fusion proteins[102], and the use of human APOBEC3A for base editing in human cells[103, 104] and plants[105], as well as regions with high levels of DNA methylation[103].

1.3.3 Transcriptional Modulation

Looking beyond the introduction of Cas9-mediated DSBs, CRISPR-Cas9 has also been utilized as an effective tool for the sequence-specific regulation of genes[106]. This has largely been accomplished through the use of nuclease-deficient dCas9 (containing the mutations H840A and

D10A) as an effector of targeted DNA binding, rather than cleavage[32]. By targeting dCas9 to specific regulatory elements of a gene, dCas9 is able to repress transcription through steric hinderance of the RNA polymerase cellular machinery. This is known as CRISPR interference (CRISPRi)[3]. However, while CRISPRi has been demonstrated to be effective in prokaryotic cells, it is less effective in eukaryotic cells[2, 3, 107]. In order to function in the environment of a eukaryotic cell, dCas9 may be fused to the transcriptional repressor domain Krüppel-associated box (KRAB)[108]. Targeting of promoter regions, enhancer elements and 5′ untranslated regions (UTRs) by dCas9-KRAB have all been demonstrated to be effective techniques for the specific silencing of individual genes and non-coding RNAs[2, 109, 110]. Combining dCas9-KRAB with the catalytic domains of DNA methyltransferases such as DNMT3A and DNMT3L have also allowed for the heritable transcriptional silencing of genes[111].

Like the repressive capabilities of dCas9-KRAB, dCas9 may also be fused to transcriptional activators such as VP64, or the nuclear factor κB transactivating subunit (p65) for targeted gene activation (known as CRISPR activation; CRISPRa)[2, 112-114]. Gene activation has also been achieved through dCas9-mediated epigenetic modifications such as acetylation and demethylation[106]. dCas9 fused to the catalytic core of human p300 (a histone acetyltransferase) has been successfully used for the acetylation of histone H3 Lys27 when targeted to promoter or enhancer regions; leading to gene activation[115]. Fusion of the catalytic domain of methylcytosine dioxogenase TET1 to dCas9 has also been employed as a means of gene activation, specifically through DNA demethylation of promoter regions[116].

1.4 Clinical Applications

Advances in CRISPR-Cas technology has led to the development of powerful gene editing tools for both the manipulation of genetic information, as well as regulatory control over gene expression; both of which are now being applied to potential therapeutic applications.

1.4.1 Cas9

The first Phase I clinical trial using a CRISPR-Cas9 based therapeutic opened in 2018 in the United States. The goal of this trial was to utilize the *ex vivo* manipulation of autologous T cells from individuals suffering from relapsed tumours such as multiple myeloma, synovial sarcoma and melanoma for cancer immunotherapy treatment (NCT03399448)[117-119]. During this trial, T lymphocytes from patients were collected and engineered *ex vivo* via Cas9 to knockout the α and β chains of the endogenous T cell receptor (TCR) as well as the programmed cell death-1 (PD-1) protein. PD-1 has been shown to block the inhibitory signals that prevent recognition of tumour cells by circulating T cells[120]. Cas9-edited cells are then transduced with a lentiviral construct encoding an NY-ESO-1 antigen specific TCR, followed by infusion back into the patients. NY-ESO-1 is upregulated within relapsed tumours and is therefore a candidate for potential therapeutic targeting[121]. CRISPR-Cas9 has also been used *ex vivo* as a potential treatment for individuals with either sickle-cell anemia or β-thalassemia, both hemoglobin disorders (NCT03745287). Autologous hematopoietic stem cells and progenitor cells from peripheral blood were collected from patients, followed by Cas9-mediated disruption of the intronic erythroid-specific enhancer of the gene *BCL11A*. Disruption of *BCL11A* has been shown to increase fetal hemoglobin (HbF) levels which has therapeutic benefits for individuals suffering from major β-hemoglobin diseases[122, 123].

Although *ex vivo* strategies make up the majority of current CRISPR-based therapeutics, use of Cas9 *in vivo* is also being explored as a therapy for leber congenital amaurosis (LCA) (NCT03872479)[124]. LCA is a monogenic disease caused by a bi-allelic loss-of-function mutation within the gene *CEP290* resulting in blindness which manifests during childhood. The proposed therapy, termed EDIT-101, is specific to individuals bearing the intronic IVS26 mutation within *CEP290*. This mutation results in the creation of a strong splice donor site and therefore results in abnormal splicing of the gene. EDIT-101 utilizes the AAV5-mediated delivery of *Staphylococcus aureus* Cas9, and two *CEP290*-specific gRNAs whose targets flank the IVS26 mutation, directly into the retina of affected individuals. The resulting Cas9-mediated DSBs cause either an inversion, or deletion of the IVS26 region, subsequently circumventing the mutation-caused aberrant splice site[124].

1.4.2 Other Cas Enzyme Therapies

Independent of the DNA targeting capabilities of Cas9, other enzymes such as Cas13 are also being investigated for their potential use as therapeutic agents. Specifically, the RNA targeting ability of Cas13 is being explored as a potential antiviral therapy against ssRNA viruses including influenza A virus, vesicular stomatitis virus, and most recently, SARS-CoV-2 (albeit currently only in isolated human cells)[125, 126].

The prophylactic anti-viral CRISPR in huMAN cells (PAC-MAN) method is of particular interest, as it is currently being tested as a means for treatment of individuals with SARS-CoV-2[126]. The PAC-MAN method employs the utilization of Cas13d isolated from *Ruminococcus flavefaciens*, predominantly due to its small size and therefore amenability to viral packaging for downstream

delivery[75]. Cas13d-expressing A549 cells (a lung epithelial cell line) were transfected with pools of either plasmid or lentiviral constructs encoding crRNAs targeted to either the RNA-dependent RNA polymerase (*RdRP*) or nucleocapsid (*N*) gene. *RdRP* and *N* were chosen as therapeutic targets due to their high levels of observed conservation across SARS-CoV-2 and middle eastern respiratory syndrome (MERS; a viral respiratory infection caused by MERS-CoV) genomes isolated from patient samples. This pooled approach was able to achieve up to 90% repression of SARS-CoV-2 RNA production, with similar results also shown when using an H1N1 viral model[126]. Variants of the PAC-MAN method using Cas13a (Cas13-assisted restriction of viral expression readout; CARVER)[125] and Cas13b have also been described[127], illustrating the potential for CRISPR therapeutics beyond Cas9. While techniques such as PAC-MAN represents a novel approach to antiviral therapy, it is currently limited to a proof-of-concept; as delivery of Cas13d (and CRISPR-Cas systems in general) into human cells must be further developed before clinical use.

1.5 Diagnostics

Outside CRISPR's cellular gene editing applications, CRISPR-Cas systems have also been leveraged for use as diagnostic nucleic acid detection platforms. While the field of clinical diagnostics is expansive, the use of nucleic acid-based testing has become the industry standard for disease detection; making CRISPR-Cas technology a logical candidate for the development of novel methodologies.

1.5.1 Cas9-Based Detection

Examples of Cas9-based methods for nucleic acid detection include: nucleic acid sequence-based amplification (NASBA)-CRISPR cleavage (NASBACC)[128] and leveraging engineered tracrRNAs and on-target DNAs for parallel RNA detection (LEOPARD)[129]. NASBACC combines several previously described methods to achieve accurate nucleic acid detection, including: isothermal pre-amplification of target RNA, PAM-dependent Cas9-mediated DNA cleavage and a toehold sensor for colourimetric readout[128]. A toehold switch is a riboregulatory RNA containing a hairpin loop structure. This hairpin loop may be unfolded following binding of a trigger RNA, exposing a ribosome binding site and allowing for translation of a reporter protein (e.g., resulting in a colour change or other observable state)[130]. NASBACC utilizes a toehold trigger conjugated to a NASBA-amplified RNA sequence of interest, followed by reverse transcription to generate the DNA substrate for Cas9. If a suitable PAM is present within the DNA substrate, Cas9 cleaves the sequence, resulting in a truncated RNA product (after downstream T7 transcription). Absence of the PAM results in the full length trigger RNA being produced, which is then free to interact with the toehold switch to produce a colour change. By designing substrates which contain strain-specific PAM sequences, NASBACC has been successfully used to detect and differentiate between samples containing either American or African Zika virus variants in the low femtomolar range[128].

While NASBACC is able to detect a single substrate, LEOPARD has been demonstrated to function as a method for the multiplexed detection of several RNA sequences through Cas9-specific interactions. Specifically, LEOPARD utilizes the observation that Cas9 is able to hybridize with cellular RNAs to form non-canonical crRNAs (ncrRNAs). This was discovered

following sequencing of Cas9-bound RNAs within the natural context of the *Campylobacter jejuni* bacterium. Therefore, through engineering of a reprogrammed tracrRNA (known as a Rptr), it is possible to direct Cas9 to bind and cleave a transcript of interest. This effectively converts an individual transcript into a functioning gRNA. Fluorescent reporter DNAs can also be used as individual outputs for the detection of several unique transcripts simultaneously. Most recently, LEOPARD has also been shown to be capable as a detection method for SARS-CoV-2 in patients with COVID-19[129].

1.5.2 Cas12/Cas13-Based Diagnostics

In addition to Cas9-based *in vitro* diagnostic tools, type V (Cas12) and type VI (Cas13) CRISPR-Cas systems have also been utilized. Unlike type II (Cas9) systems, type V and VI systems are able to elicit non-specific collateral cleavage following sequence-specific recognition of either DNA (Cas12) or RNA (Cas13)[70, 72]. This unique characteristic has been exploited for the development of several *in vitro* diagnostic platforms, including the SHERLOCK[131] and DETECTR assays[70].

Specific high-sensitivity enzymatic reporter unlocking (SHERLOCK), is a Cas13-based assay which uses Cas13a derived from *Leptotrichia wadeii*[131]. Substrate RNA or DNA is first isothermally amplified through recombinase polymerase amplification (RPA; for DNA-based samples) or reverse-transcription RPA (RT-RPA; for RNA-based samples)[132]. As Cas13 recognizes RNA, a T7 promoter sequence is also added during amplification so that the sample may be transcribed prior to digestion. Sequence-specific recognition of the target by Cas13 complexed to a crRNA of interest results in *cis* cleavage of the on-target, followed by sequence-

independent *trans* collateral cleavage of surrounding RNA. Within the reaction, a reporter RNA consisting of a fluorophore and quencher moiety is also included. If a target of interest is present within the sample, the collateral cleavage following sequence recognition degrades the reporter RNA, freeing the fluorophore from the quencher and resulting in measurable fluorescence[131]. An improved version of the assay (SHERLOCKv2) has also been developed, demonstrating increased detection sensitivity (up to zeptomolar concentrations) and portability. Specifically, SHERLOCKv2 utilizes a laminar flow-based readout immobilized on a paper strip for the detection of a FAM-biotin reporter. Intact reporter molecules accumulate at the first line of the strip, while cleaved reporters bind to antibody-conjugated gold nanoparticles at the second line. This instrument-free detection system has successfully been applied to the detection of Zika virus, Dengue virus and SARS-CoV-2 virus RNA[131, 133, 134].

As Cas12 enzymes also display the ability for *trans* collateral cleavage, assays similar to the Cas13-based SHERLOCK system have also been developed. Examples include the DNA endonuclease-targeted CRISPR *trans* reporter (DETECTR) assay, which uses Cas12a derived from *Lachnospiraceae bacterium* for the detection of DNA substrates[70]. Like SHERLOCK, DETECTR also requires the addition of DNA reporters carrying a fluorophore and quencher for fluorescence-based detection. Additional Cas12-based techniques such as HOLMES (one-hour low-cost multipurpose highly efficient system) and HOLMESv2 have also been developed to utilize the *trans*-mediated collateral cleavage of reporter DNA[135, 136]. HOLMESv2 is of note, as it uses a thermostable Cas12b variant derived from *Alicylobacillus acidoterrestris* for single reaction amplification (using loop-mediated isothermal amplification; LAMP) and detection of DNA[136, 137]. Enzymes outside of Cas12 have also been demonstrated to have amenability for use in such assays

as SHERLOCK and DETECTR. Specifically, Cas14, a small Class 2 CRISPR-Cas system capable of recognition and cleavage of ssDNA has been implemented in DETECTR[138, 139]. Cas14-DETECTR has been shown to display better single nucleotide polymorphism (SNP) resolution than Cas12a, demonstrating the continuous expansion and improvement of such tools[140].

1.6 Specificity of CRISPR

A critical challenge for the utilization of CRISPR-Cas systems for both basic research and future therapeutics is the identification and minimization of off-target activity at sequences highly homologous to the target of interest. While an ideal gene editing tool would display singular targeting to a sequence of interest, CRISPR-Cas9 (and the larger CRISPR field as a whole) has been demonstrated to have non-negligible activity at off-target sites. As the specificity of CRISPR as a technology largely governs the validity of wherever it is implemented (including all of the applications discussed above), this challenge is of paramount concern as the field of genome engineering continues to grow.

1.6.1 Cas9 Specificity & Methods for Measuring Off-Target Activity

To probe the cellular specificities of Cas9, initial studies utilized *in silico* prediction of potential off-target sites based on their sequence similarity to a target of interest. This was followed by the experimental validation of Cas9 activity at each of the predicted sites through quantification of Cas9-mediated indels resulting from NHEJ[141-143]. These studies were able to definitively demonstrate that Cas9 is capable of high levels of nuclease activity at off-target sites distinct from the target of interest; with some experiments showing levels equal to, or higher than that of the on-target[141]. Generally, off-target activity was observed at sites with ≤5 mismatches within the spacer

sequence, although mutations within the PAM were also observed at lower frequencies (e.g., cleavage of 5′-NAG-3′ instead of the canonical 5′-NGG-3′)[142]. These early observations led to the development of the first off-target predicting programs such as E-CRISP[144] and the CRISPR Design Tool[142], which used sequence homology and mismatch position within the spacer to determine potential off-target sites for a given crRNA.

While these studies were the first to illuminate the propensity of Cas9-mediated off-target activity, they were limited in scope as they did not examine the genome as a whole, and were targeted in their approach. More specifically, any true off-target sites which did not fit the criteria outlined computationally were never examined. To rectify this, several genome-wide methods have since been described to more globally profile the specificity of Cas9 activity. These methods can be divided into one of two categories: cell-based and *in vitro* (cell-free) assays.

1.6.2 Cell-Based Off-Target Detection

The first genome-wide approach used to study Cas9 specificity was the integrase-defective lentiviral (IDLV) capture assay[145]. IDLV capture is based on the ability to measure integration of IDLVs (whose genome's are linear, dsDNA) into the sites of nuclease-mediated DSBs via NHEJ. Integrations are subsequently identified through linear amplification-mediated PCR (LAM-PCR) and aligned with a reference genome following next-generation sequencing (NGS)[146]. While unbiased and genome-wide, IDLV capture suffers from limited sensitivity due to the low integration rates of IDLVs and high levels of background integration independent of nuclease-induced DSBs[146].

Similar to IDLV capture, GUIDE-seq (genome-wide unbiased identification of DSBs enabled by sequencing) utilizes the integration of genetic material at DSB sites introduced by Cas9[147]. However, GUIDE-seq uses short, end-protected double-stranded oligodeoxynucleotide (dsODN) tags in place of lentiviral integration. This allows for a much more sensitive method, as integration of dsODNs is highly efficient, allowing for the detection of off-target events at frequencies of 0.1% or lower within a given population of cells. Following Cas9-mediated DNA cleavage and tag integration, sequences flanking the integration site are amplified through tag-specific primers and sequenced through NGS[147]. GUIDE-seq has been experimentally demonstrated to be able to identify bona fide off-target sites previously ignored by *in silico* prediction tools[142, 144]. While GUIDE-seq has been shown to be highly efficient, it is heavily dependent on the efficiency of transfection, and therefore precludes certain cell types and *in vivo* use[147]. However, a variation of GUIDE-seq known as GUIDE-tag has recently been described, which has demonstrated to be effective *in vivo* in both mouse lung and liver environments[148]. Distinct from GUIDE-seq, which utilizes wild-type Cas9, GUIDE-tag uses a monomeric streptavidin-conjugated Cas9 (Cas9-mSa) tethered to biotinylated dsDNA[149]. Using this tethered approach, GUIDE-tag has been shown to increase the efficiency of DNA integration both in cultured cells, and *in vivo* following the introduction of Cas9-mediated DSBs[148]. Several other variations to GUIDE-seq have also been described, including iGUIDE[150] and TTISS[151]. iGUIDE is a revised GUIDE-seq protocol which utilizes a lengthened dsODN (46 bp instead of 34 bp) to reduce mis-priming during library amplification and therefore subsequent false positive sites[150]. Tagmentation-based tag integration site sequencing (TTISS) is based on GUIDE-seq, but utilizes the previously published Tn5 transposase for tagmentation following cell lysis[152]. This streamlines the library preparation process, allowing for multiplexed analysis of GUIDE-seq samples for large-scale screening[151].

While IDLV capture and GUIDE-seq rely on integration-dependent methods for determination of Cas9 cellular specificity, an alternative method known as DISCOVER-seq (discovery of *in situ* Cas off-targets and verification by sequencing) uses localization of endogenous DNA repair proteins. Through chromatin immunoprecipitation sequencing (ChIP-seq) of repair proteins, it was identified that meiotic recombination 11 homolog 1 (MRE11) localizes to Cas9-mediated DSB sites with single nucleotide resolution[153]. MRE11 is a member of the MRE11-RAD50-NBS1 (MRN) complex which is involved in DSB repair and DNA damage responses in general[154-156]. DISCOVER-seq is therefore performed using traditional ChIP-seq methodology against MRE11 following transfection of cells with Cas9 and gRNA, and analyzed using custom software. As DISCOVER-seq uses endogenous cellular machinery for detection of DSBs, it is able to be used in both cell culture and *in vivo* contexts. It has also been shown that dsODNs required for GUIDE-seq are toxic to induced pluripotent stem cells (iPSCs), making DISCOVER-seq more advantageous in its amenability to multiple experimental formats. However, as DISCOVER-seq employs ChIP-seq protocols, it requires more starting material when compared to other methods, and has a slightly higher sensitivity threshold (a lower detection limit of 0.3% when compared to 0.1% of GUIDE-seq)[153].

1.6.3 In Vitro-Based Off-Target Detection

While cell-based techniques for off-target identification are able to provide broad insights into the promiscuity of engineered nucleases within a cellular context, they are limited to sequences which exist within the genome. While this is advantageous for the practical application

of CRISPR-Cas within cells, it does not provide the same level of detail as *in vitro*-based techniques which utilize more diverse starting materials.

The first depiction of Cas9 specificity using an *in vitro* high-throughput based approach was performed using an adapted assay initially developed for profiling the specificities of ZFNs[157] and TALENs[158]. This method utilizes a partially randomized oligonucleotide library designed to include a 10-fold coverage of all potential off-target sites with ≤8 mismatches (relative to the on-target); resulting in a binomially distributed population containing an average of 4.62 mutations per target[52]. A partially randomized library was chosen, as a fully randomized target sequence would be cumbersome and inefficient to sequence due to the required depth of coverage. Previous studies have also shown that Cas9 is unlikely to cleave off-target sites with >6 mismatches, making their inclusion in the starting library null[142]. As Cas9 digestion results in the production of blunt-ended DSBs containing free 5′ phosphate groups, digested library members may be selected for through the phosphate-dependent ligation of non-phosphorylated adaptor sequences. Post-selection library members may then be amplified through PCR and sequenced through NGS, followed by comparison with a pre-selection library to identify the relative Cas9-mediated enrichment of specific off-target sites[52]. While many of the sequenced library members have no real-world genomic locations, this technique can be used to provide additional insights into the manner in which Cas9 specificity is dictated, and may be applied to computational methods for increased predictive power of potential off-targets.

1.7 Strategies for Improving Specificity

Alongside the continual advancement of methodologies for the detection of Cas9-mediated off-target activity; numerous avenues have been explored regarding how to improve specificity of the system. These studies are largely interdisciplinary, ranging from computational and machine-learning based approaches, to the rational design of Cas9 enzyme variants for highly specific gene editing (**Figure 1.6**). Each method has inherent advantages and disadvantages, and as such, it will likely be a collaborative effort between strategies for the development and implementation of safe and accurate genome engineering in the future.

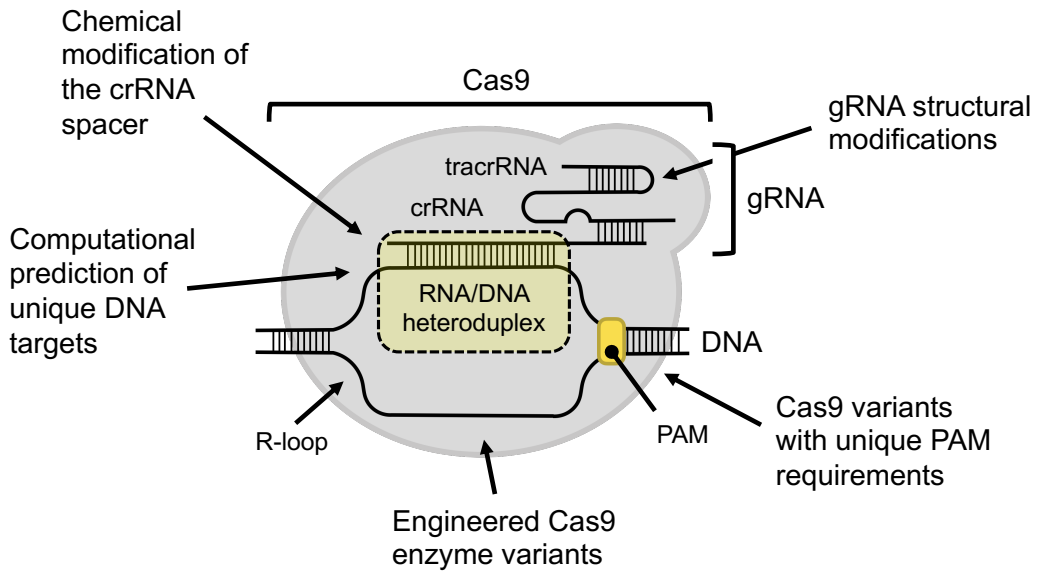


Figure 1.6. Strategies for modulating CRISPR-Cas9 specificity
Overview of the strategies which have currently been most commonly been utilized to improve, or otherwise modulate the specificity of Cas9 DNA cleavage.

1.7.1 Computational Design of gRNAs

Many computational tools have been developed for the prediction and ranking of Cas9 off-target sites, therefore providing an increase in specificity through the ability to design crRNAs

found to have fewer predicted off-target sites prior to their actual use. While the manner in which these programs were developed is beyond the scope of this thesis, it is important to give an overview of some of the available tools, and how they are being implemented. There are numerous examples of programs which have been developed that simply provide a list of potential off-target sites to the user. However these are of limited value, as there is no associated score linking the propensity for cleavage at each site[159-161]. More valuable examples include E-CRISP[144], CCTOP[162] and COSMID[163, 164], which rank predicted off-target sites through different means. E-CRISP ranks off-targets by their alignment scores[144], while CCTOP and COSMID use the position of mismatches within the crRNA to score each site[162, 163]. Off-targets with mismatches closest to the PAM are penalized for potential activity, as it has been shown that mismatches in this location often prevent DNA cleavage[47, 142]. Building upon the use of weighted scoring, CROP-IT divides the 20 nt protospacer into three regions, differentially weighting the number and position of mismatches within each three[165]. CROP-IT also incorporates cell-line-specific location accessibility scoring, using data obtained through DNase I-seq (a method which provides information on the availability of DNA for DNA binding proteins)[166]. Several machine-learning based algorithms have also recently been described[167-169]. Using publicly available, large-scale datasets, machine-learning models have been trained to predict Cas9 off-target activity through a number of biochemical properties, including sequence composition, the presence of bulges upon binding due to mismatched nucleotides, RNA/DNA heteroduplex thermodynamics and genome location. As the number of studies published using high-throughput approaches continues to increase, the availability of data to train predictive algorithms also increases, allowing for the continues improvement of predictive power.

1.7.2 Delivery of Cas9

The manner in which Cas9 and gRNA are delivered for gene editing purposes has also been shown to have significant impact on the specificity of the system. Initial studies used the delivery of plasmids encoding Cas9 and gRNAs for genome engineering. However, it has since been demonstrated that this continued expression leads to higher incidences of off-target activity. Specifically, Cas9 delivered by plasmid was shown to induce off-target mutagenesis at rates 13- to 20-fold higher than Cas9 delivered directly as a ribonucleoprotein (RNP) complex through electroporation[170] or cationic lipid delivery respectively[171]. This is likely due to the time spent within the cell, as Cas9 delivered by plasmid remains detectable for several days (through Western blotting), while RNP-delivered Cas9 is degraded within 24 hours[170].

Based on these observations, a number of inducible Cas9 variants have since been described as a means of further controlling the active time within a cell; including several split Cas9 systems, which only assemble following a particular stimulus. One particular example uses a split Cas9 fusion protein, where the two halves are fused to either the FK506 binding protein 12 (FKBP) or FKBP rapamycin (FRB) protein domain. These domains may be induced to dimerize following rapamycin treatment, allowing for the formation of active Cas9 and subsequent cleavage activity[172]. Another example of a drug-mediated Cas9 system uses Cas9 catalytically repressed via a drug-inducible intein[173]. Cas9 has also been shown to be photo-inducible through the fusion of photoactivatable dimerization domains known as Magnets (either positive or negative Magnet) to a split Cas9 protein (termed photoactivatable Cas9; paCas9)[174]. When compared to the above-mentioned strategies, paCas9 is advantageous as it represents a reversible system.

1.7.3 Engineered Cas9 Variants

All of the methods described so-far have employed the use of wild-type Cas9 for their respective specificity improvements. However, a large portion of the studies aimed at improving CRISPR-Cas9 off-target discrimination have centered on the generation of novel Cas9 variants with inherent increased specificity over their wild-type counterparts. This has been accomplished through several distinct methods, including rational engineering, random mutagenesis screens and directed evolution.

1.7.3.1 Rational Engineering of Cas9

The crystal structure of *S. pyogenes* Cas9 in complex with a gRNA and DNA target provided a strong foundation for the generation of several Cas9 variants through rational engineering[46, 54]; all of which display improved specificities over wild-type Cas9. Examples of these variants include: eSpCas9[175], Cas9-HF1[176], HypaCas9[177] and SuperFi-Cas9[60] (although more continue to be developed).

Both eSpCas9 and Cas9-HF1 are variants which employ similar strategies for improved DNA targeting specificities. Specifically, in both cases it was hypothesized that through the reduction of DNA binding energy it may be possible to disrupt off-target DNA binding while maintaining on-target activity[175, 176]. For eSpCas9, this was accomplished through rationally altering specific residues within the positively charged groove located between the HNH, RuvC and PAM interacting domains. This groove is positively charged so that it may stabilizing the non-target strand during DNA interrogation. By introducing neutralizing mutations, it was shown that the interaction with off-target DNA substrates were weakened, encouraging re-hybridization of the

target and non-target strands, and therefore requiring a higher stringency of Watson-Crick base pairing for cleavage[175]. Similarly, Cas9-HF1 was designed to reduce Cas9-mediated hydrogen bonding with the target DNA, as it was hypothesized that an excess of binding energy was present which facilitated cleavage of imperfect DNA targets. Mutations of four residues within Cas9 that make contact with the target DNA were shown to weaken Cas9-DNA interactions, preventing off-target activity through an increased thermodynamic threshold requirement (i.e., additional Watson-Crick base pairing)[176].

Independent of the DNA-binding characteristics of the Cas9 protein, HypaCas9[177] and SuperFi-Cas9[178] were rationally engineered based on observations made regarding how Cas9 (specifically the REC lobe) regulates proofreading and subsequent mismatch tolerance. Like eSpCas9 and Cas9-HF1, HypaCas9 was generated followed observations made possible by the crystal structure of Cas9[46]. Structural studies demonstrated that the REC3 domain of the REC lobe makes significant interactions with the RNA/DNA heteroduplex, undergoing conformational rearrangements following target binding[55, 56]. Experiments using single-molecule Förster resonance energy transfer (smFRET) also showed that the REC3 domain functions as an allosteric effector of HNH nuclease activation through recognition of the structure of the heteroduplex. HypaCas9 was developed through exploitation of these observations, specifically through directed mutagenesis of residues within REC3, resulting in an increased stringency of HNH activation; effectively locking the HNH nuclease in a restricted conformation when bound to mismatched DNA[177].

Similar to HypaCas9, SuperFi-Cas9 was also rationally engineered using exploitation of observations made regarding the biochemical mechanisms of mismatch tolerance. Building upon the observation that REC3 allosterically regulates HNH nuclease activation, cryo-electron microscopy (cryo-EM) was used to identify additional regulatory regions previously missing from crystal structures of Cas9[178]. The L1 and L2 linker regions, which function to tether the HNH domain to the rest of Cas9, are often absent from crystal structure data due to their intrinsic flexibility[46]. However, cryo-EM has shown that L1 forms numerous stabilizing interactions with the RNA/DNA heteroduplex, and is correlated with HNH conformational rearrangements required for HNH-mediated nuclease activity. Specifically, L1 functions to stabilize on-target DNA, as well as DNA bearing distally located mismatches (relative to the PAM). SuperFi-Cas9 was therefore rationally designed to prevent this unwanted DNA stabilization of off-target sequences through the mutation of 7 stabilizing residues in the L1 linker, resulting in an increase in observed specificity[60].

1.7.3.2 Random Mutagenesis Screens

While the rational design of Cas9 variants has successfully resulted in several highly specific enzymes, the random mutagenesis of Cas9 has also been used, producing variants including evoCas9[179], HiFi Cas9[180] and LZ3 Cas9[151]. While HiFi Cas9 was developed using a large library of randomly mutated Cas9-encoded genes[180], evoCas9 and LZ3 Cas9 were developed using a more stringent approach; only randomizing either the REC3 domain, or HNH and RuvC nuclease domains respectively[151, 179].

To generate evoCas9, error-prone PCR was used to construct a library of REC3 domain mutants, followed by cloning into a REC3-deficient Cas9 plasmid. The resulting plasmid pool was then expressed in a reporter yeast strain carrying both a genomic on-target, and genomic off-target sequence containing a single mismatch distal to the PAM. Yeast survival was linked to cleavage of the on-target gene (*TRP1*), while off-target activity was linked to a pigmentation change following cleavage of *ADE2*. Of the resulting variants, a Cas9 enzyme containing the mutations M495V, Y515N, K526E and R661L (denoted VNEL) was shown to be the most stringent for mismatch discrimination, albeit at a loss of ~20% on-target activity. Additional directed mutagenesis to mitigate this resulted in the generation of a second variant (containing R661Q; VNEQ) known as evoCas9[179]. EvoCas9 was experimentally demonstrated to possess comparable on-target activity to both eSpCas9 and wild-type Cas9, while being significantly more active than Cas9-HF1. GUIDE-seq also showed evoCas9 to be more specific than both previously described variants[179]. Similar to evoCas9, LZ3 Cas9 was generated through randomized mutagenesis concentrated within specific regions of Cas9; namely the HNH and RuvC nuclease domains. Using a self-targeting lentivirus pool (which functions to link a specific variant to a specific target), variant specificity and activity was measured through TTISS (described earlier) and repeated for four individual gRNA targets. LZ3 Cas9 was shown to have comparable specificity improvements with previously developed Cas9 variants, although with increased on-target activity[151].

Unbiased mutagenic screens in bacterial cells have resulted in the identification of additional Cas9 variants with high levels of observed specificity, including HiFi Cas9[180]. HiFi Cas9 was generated through low-fidelity-based PCR amplification of a pool of Cas9 mutants, followed by unbiased bacterial selection for reduction in off-target activity. Specifically, survival of the bacterium was

linked to on-target cleavage of a toxin-producing gene. Additionally, a known off-target sequence was incorporated into the plasmid which expressed Cas9, a gRNA and an antibiotic resistance gene. In this system, bacteria may only survive and proliferate if they cleave both on-target sequences (preventing toxin production) and do not cleavage off-target sequences (allowing for continued antibiotic resistance). Sequential screens using two independent Cas9 target sites resulted in identification of HiFi Cas9, a variant containing a single point mutation (R691A). Interestingly, the mutant N692A was found to be the second highest performing variant, with N692A also present in HypaCas9[177, 180]. Due to the proximity of these residues, it was hypothesized that HiFi Cas9 likely perpetuates an increased target sensitivity through a similar mechanism to HypaCas9[180].

1.7.3.3 Directed Evolution of Cas9

Directed evolution has also been used as a strategy for the development of novel Cas9 variants, including xCas9[181] and Sniper-Cas9[182], both of which employ bacterial-based methods, albeit in different capacities. xCas9 was developed using phage-assisted continuous evolution (PACE)[183], a technology previously applied to the evolution of proteins such as T7 RNA polymerase[183] and TEV protease[184], as well as other engineered nucleases such as TALENs[13]. In brief, PACE utilizes the biological linking of a desired activity of a protein of interest to the fitness of a bacteriophage which carries the corresponding gene. In this way, greater activity confers greater infectivity, allowing for the propagation of more advantageous mutations[183]. PACE was originally applied to Cas9 as a method for developing variants capable of recognizing a flexible PAM sequence (canonically 5′-NGG-3′)[181]. Nuclease-deficient dCas9 was fused to the bacterial polymerase subunit ω. In this way, binding and recognition of a PAM by dCas9-ω allows for

transcription of gene III, which is necessary for propagation of the phage[181, 185]. By including non-canonical PAM sequences, this would then allow for the selection of dCas9 variants capable of multi-PAM recognition. Multiple rounds of PACE resulted in the selection of a variant known as xCas9 3.7 (or xCas9). xCas9 was demonstrated to recognize several unique PAM sequences, including NG, GAT, GAA and CAA. Interestingly, when analyzed via GUIDE-seq, xCas9 was also shown to be markedly more specific than wild-type Cas9, an unintended but beneficial byproduct of the evolution process[181].

Similar to the yeast-based screen used in developing evoCas9, Sniper-Cas9 was identified using both positive and negative selection pressures for activity and specificity respectively[182]. Similar to previous methods, bacterial survival was linked to on-target mediated cleavage of the lethal *ccdB* gene[186], while specificity was linked to the absence of cleavage of an off-target sequence. Specifically, integration of several potential off-target sequences into the genome of the bacteria used allowed for off-target activity to function as negative selection pressure through the induction of lethal DSBs (in the presence of off-target cleavage). Several rounds of this selection resulted in identification of Sniper-Cas9, a variant with significantly higher on-target activity when compared to other similar variants, but lower relative specificity[151, 182].

1.8 Guide RNA Modification

While a significant amount of work has been put towards the engineering of Cas9 variants with higher levels of observed specificity, several studies have also utilized modification of the Cas9 gRNA structurally, or through incorporation of unique nucleotides to achieve similar outcomes. Contrasting to the challenges associated with protein manipulation, gRNAs may be

synthesized chemically or through biological expression *in vitro* or *in vivo*, forming the basis for the flexibility and diversity of gRNA-based engineering. These modified, synthetic gRNAs have been used to impart a number of properties to the CRISPR-Cas9 system, including: increased activity[187-191] and specificity[192-195], increased cellular stability[196], reduction in cytotoxicity and immunological responses[190, 197], as well as visualization of cellular components and genomic loci[198, 199]. However, to maintain the scope of this thesis, discussion will largely be limited to methods used for improving or otherwise modulating Cas9 specificity.

1.8.1 Structural Modifications

One of the first instances of gRNA modification as a means of reducing Cas9-mediated off-target activity was through the use of truncated gRNAs (tru-gRNAs) containing shortened spacers 17 nt in length (as opposed to the canonical 20 nt spacer)[193]. As it was previously demonstrated that mismatches within the PAM distal region of the crRNA were largely tolerated, it was reasoned that by reducing the length of the sequence-specific region, this would also reduce the presence of off-targets with corresponding mismatches. Originally, tru-gRNAs were shown to reduce off-target activity by up to 5,000-fold in human cells, however studies using GUIDE-seq later showed a more modest specificity improvement of 2- to 5-fold relative to Cas9 directed by full length gRNAs[147]. Additionally, it was also reported that in some instances, new off-target sequences not seen for full length gRNAs were also cleaved (most likely due to distal mismatches not tolerated by full length gRNAs). Along with the discovery of several Cas9 variants, it is now thought that tru-gRNAs function through reducing the available thermodynamic energy between the gRNA and DNA target, thereby reducing the tolerance for off-target binding[193]. In an inverse approach to tru-gRNAs, extended gRNAs containing 2 to 4 additional nucleotides have been

shown to have some specificity improvements. Specifically, gRNAs with two additional guanines located 5′ upstream of the sequence-specific spacer region have been demonstrated to display increased specificity[192]; while gRNAs with spacer up to 4 nt longer than wild-type sequences have been used mainly as a method for increasing base editor efficiency[200]. However these lengthened gRNAs have yet to be profiled for their genome-wide specificities, and therefore are not widely used.

Aside from simply changing the length of the spacer sequence, several studies have also shown that the addition of RNA secondary structures to the gRNA can modulate both activity as well as specificity. The addition of G-quadruplex (G4) structures to the 3′ end of sgRNAs has been demonstrated to boost *in vitro* serum stability and *in vivo* efficiencies[196], while the addition of a hairpin structure to the 5′ end of the sgRNA (hg-sgRNA) increased Cas9 cleavage specificity[201]. These hp-sgRNAs are thought to increase specificity through modulation of required thermodynamic energy. Specifically, during on-target binding the formation of the RNA/DNA heteroduplex is more energetically favourable, inducing unfolding of the hairpin. However, during off-target RNA/DNA base pairing, the hairpin is more energetically expensive to unfold, preventing complete R-loop formation and subsequent DNA cleavage. These hp-sgRNAs were also shown to be compatible with several Cas9 variants as well as Cas12a[201].

1.8.2 Chemical Modification

In addition to the modification of gRNA length and secondary structure, several studies have also demonstrated the use of chemically modified nucleotides within the gRNA of Cas9 as a means of improving Cas9-mediated cleavage specificity.

Although not as exotic as some other examples, DNA may be considered a form of chemically modified RNA, and has been shown to be well tolerated within the gRNA of Cas9, including both the crRNA[202-204] and tracrRNA[205]. The lower thermodynamic stability of a DNA/DNA duplex (when compared to an RNA/DNA heteroduplex) within the context of a partial DNA gRNA has been demonstrated to facilitate increased off-target discrimination both *in vitro* and in cells[202-204]. Interestingly, substitution of RNA sequences within the 3′ end of the crRNA external to the sequence-specific region results in higher specificities[204]. This is also true for the opposite, where replacement of RNA with DNA within the spacer region of the crRNA also results in reduced off-target activity[203]. However the crRNA is not universally tolerable of DNA substitutions, as replacement of more than 4 nt at the 3′ end, or 12 nt at the 5′ end results in abolishment of on-target activity[203]. Conflicting studies have also experimentally demonstrated that certain residues within the spacer region make key hydrogen bonds with the Cas9 enzyme, and therefore cannot be altered without loss of activity[202]. However this is somewhat debated, as additional studies have used DNA in said positions without any observed reduction in nuclease activity[203, 204]. Chimeric, partial DNA gRNAs have also been shown to be functional in other CRISPR-Cas systems such as Cas12[203].

Originally demonstrated as a means for improving cellular gRNA stability, the use of synthetic (and often non-natural) chemically modified nucleotides has been shown to be effective for the modulation of Cas9 specificity. While gRNA modification through incorporation of phosphorothioate (PS), 2′-fluoro (2′F) and 2′-OMe RNA were shown to improve stability, they did not affect measured specificity of Cas9 in cells or *in vivo*[190, 197]. Conversely, studies have

demonstrated that additional nucleotides such as 2′-*O*-methyl-3′-phosphonoacetate (PACE)[195], 2′,4′-BNA^{NC}[NMe] (BNA) and locked nucleic acids (LNAs)[194] are able to improve specificity through their incorporation into the crRNA of Cas9. The utilization of BNAs and LNAs within the crRNA of Cas9 will be discussed in **Chapter 2** of this thesis.

One of the first examples of chemically modified crRNAs for increased specificity was the incorporation of PACE-modified nucleotides within the 20 nt spacer of the Cas9 crRNA. However, to achieve these specificity improvements, PACE substitution was shown to be positionally-dependent, with incorporations at positions 5 and 11 to be the most effective[195]. This was hypothesized to be due to the proximity of position 11 to the 10 nt long seed region, as well as interactions made with the REC3 domain of Cas9. Moreover, position 11 of the gRNA interactions with a cluster of amino acids present within REC3[177]. Therefore, PACE substitution at this position may function to weaken the interactions with REC3, raising the energetic barrier for subsequent DNA cleavage and reducing off-target activity[195]. This reduction in available thermodynamic energy may also increase the off-rate of Cas9, also potentially contributing to increased specificity. This hypothesis also applies to PACE modification at position 5, as this position has also been demonstrated to make significant interactions with the REC3 domain, albeit through different amino acid residues. Cellular cleavage data of cells transfected with gRNAs containing PACE modifications at positions 5 and 11 showed no detectable levels of Cas9-mediated indels at 15 of 18 potential off-target sites, although the applicability of these modified gRNAs to a more diverse set of sequences must still be examined[195].

A number of additional chemically modified nucleotides have been demonstrated to be amenable to use within CRISPR-Cas systems, including Cas9[206], Cas12[207] and Cas13[208]. These include the aforementioned PS, 2′F and OMe RNAs in addition to: 2′-F-ANA (FANA), 2′, 5′-RNA, 2′-F-4′-Cα-OMe, 2′,4′-di-Cα-OMe, butane linkers and unlocked nucleic acids (UNAs) (among others)[206]. These chemically modified nucleotides have been shown to be tolerated within the Cas systems, allowing for on-target activity, but have yet to be comprehensively profiled for their abilities to improve specificity.

1.9 Statement of Hypothesis and Aims

The ability to reliably and accurately target a sequence of interest is a critical component for all genome editing applications, including CRISPR-Cas systems. Several strategies have been developed for the improvement of Cas9 specificity, such as: engineered Cas9 variants[151, 175-177, 179-182], computational design of gRNAs[162-165], structural alteration of the gRNA[192, 193, 201] and incorporation of chemically modified nucleotides into the sequence-specific region of the gRNA[194, 195, 203, 204]. While the former technologies have been extensively investigated, the use of chemically modified nucleotides for specificity enhancement remained largely underutilized. This is especially true when comparing the number of chemical modifications successfully used for specificity, with the number of modifications used in other contexts (i.e., activity and stability). Therefore, the overarching hypotheses tested in this thesis are as follows:

1. I hypothesized that bridged nucleic acids (BNAs) may be capable of improving the endonuclease specificity of CRISPR-Cas9 systems through their incorporation within the crRNA of Cas9 (**Chapter 2**).

2. Based on our findings detailed in **Chapter 2**, I hypothesized that additional chemically modified nucleotides may also be able to impart higher specificities to the CRISPR-Cas9 system (**Chapter 3**).
3. As the CRISPR-Cas9 system has been shown to be tolerable of a number of chemically modified nucleotides, I hypothesized that we may be able to modify the crRNA of Cas9 or other Cas enzymes with universal bases to allow for multi-sequence targeting of polymorphic sequences (**Chapter 4**).

The specific aims for each study are listed below by chapter:

Chapter 2: Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity

1. To determine where (and if) BNA and LNA modifications are most effective at improving Cas9 DNA cleavage specificity *in vitro*.
2. Identify a specificity profile through high-throughput, next-generation sequencing-based assays to determine characteristics of BNA-mediated specificity improvements.
3. Demonstrate *in vitro* specificity observations are able to be translated to a cellular context.
4. Determine a biochemical mechanism for BNA-mediated specificity enhancements.

Chapter 3: High-throughput specificity profiling for the identification of Cas9 crXNAs with superior specificity

1. Identify type, position and number of chemically modified nucleotides which result in the highest degrees of specificity improvement, while maintaining on-target activity, *in vitro*.
2. Determine whether or not observed specificity improvements are sequence-dependent, or function across a variety of biochemically diverse Cas9 targets.
3. Identify characteristics and design considerations which maximize the resulting specificity of chemically modified crRNAs

Chapter 4: Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences

1. Determine whether or not universal bases may be used within gRNAs to enable targeting of multiple target sequences *in vitro*.
2. Using a high-throughput specificity profiling assay, identify whether or not universal base incorporation results in decreased specificity globally.
3. Determine whether universal base modified Cas9 gRNAs retain their on-target activity in a cellular context.
4. Test the applicability of universal base modified gRNAs to CRISPR-Cas12-based *in vitro* diagnostic tools such as DETECTR.

Chapter 2

Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity

Christopher R. Cromwell[1], Keewon Sung[2], Jinho Park[2], Amanda R. Krysler[1], Juan Jovel[3], Seong Keun Kim[2] & Basil P. Hubbard[1]

[1]Department of Pharmacology, University of Alberta, Edmonton, Canada

[2]Department of Chemistry, Seoul National University, Seoul, Republic of Korea

[3]The Applied Genomics Core, Office of Research, University of Alberta, Edmonton, Canada

PUBLISHED:

Cromwell, C.R., Sung, K., Park, J. *et al.* Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat Commun* **9,** 1448 (2018).

https://doi.org/10.1038/s41467-018-03927-0

2.1 Abstract

Off-target DNA cleavage is a paramount concern when applying CRISPR-Cas9 gene-editing technology to functional genetics and human therapeutic applications. Here, we show that incorporation of next-generation bridged nucleic acids (2′,4′-BNA^{NC}[N-Me]) as well as locked nucleic acids (LNA) at specific locations in CRISPR-RNAs (crRNAs) broadly reduces off-target DNA cleavage by Cas9 *in vitro* and in cells by several orders of magnitude. Using single-molecule FRET experiments we show that BNA^{NC} incorporation slows Cas9 kinetics and improves specificity by inducing a highly dynamic crRNA–DNA duplex for off-target sequences, which shortens dwell time in the cleavage-competent, "zipped" conformation. In addition to describing a robust technique for improving the precision of CRISPR/Cas9-based gene editing, this study illuminates an application of synthetic nucleic acids.

2.2 Introduction

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)–Cas9 complex was originally characterized as a component of prokaryotic immune systems[209], but has now become a widely used tool for genome editing applications[28]. CRISPR-Cas9 has been applied to the generation of knockout and knock-in organisms ranging from yeast to mice[210], functional genomics[211] and epigenetic[212] screens, and proof-of-principle studies aimed at correcting genetic disease in mammals[213, 214]. Two non-coding RNA elements direct sequence-specific DNA cleavage by the Cas9 system[32]. The CRISPR-RNA (crRNA) contains a 20-nucleotide (nt) RNA sequence complementary to the target DNA sequence, while the transactivating crRNA (tracrRNA) acts as a bridge between the crRNA and Cas9 enzyme[35]. Hybridization of these RNA elements together constitute a guide RNA (gRNA); they can also be covalently linked to produce

a chimeric single-guide RNA (sgRNA)[35]. Recognition of a target sequence by Cas9 first involves the identification of an upstream protospacer adjacent motif (PAM) (5′-NGG-3′ in *S. pyogenes*) on the target DNA strand, followed by local DNA melting and hybridization of the first 10–12 bp of the 3′ end of the crRNA sequence (seed pairing), and formation an R-loop structure[28, 35, 55]. Complete hybridization between the guide segment and target DNA drives conformational changes in the HNH and RuvC nuclease domains on Cas9 that result in DNA cleavage 3-bp upstream of the PAM[31, 32, 59]. While mutations within the PAM sequence effectively abolish Cas9 cleavage activity[28, 215], mutations within the target sequence may be permitted[216], resulting in cleavage of off-target DNA sequences.

Cas9 DNA cleavage specificity is highly dependent on the crRNA sequence and correlates with target-crRNA folding stability[217]. A number of approaches have been deployed to improve Cas9 specificity[218], including the use of algorithms to computationally design gRNAs with minimal off-target activity[218], a paired Cas9 nickase system that employs two gRNAs for target recognition[219], and new delivery strategies displaying burst kinetics, such as Cas9 ribonucleoprotein (RNP) delivery[171]. In addition, several groups have engineered highly specific variants of Cas9, such as eSpCas9[175], SpCas9-HF1[176], and HypaCas9[177], by mutating residues on Cas9 involved in the formation of non-specific DNA interactions. Despite these advances, off-target cutting and generation of accessory mutations remains a significant barrier for Cas9-based gene editing[218].

While numerous studies have focused on engineering or otherwise modifying the Cas9 enzyme[95, 175-177, 219], few have investigated the possibility of altering the sequence or structure of its crRNA to improve specificity[218]. Reducing the number of nucleotides in the spacer sequence from 20 to

17–18 bp (tru-guides) improves Cas9 specificity, but reduces on-target cleavage efficiency[193]. Interestingly, bridged nucleic acids (BNAs) have previously been shown to improve mismatch discrimination in nucleic acid duplexes[220]. We hypothesized that incorporation of these synthetic nucleotides into crRNAs could improve Cas9 DNA cleavage specificity. Moreover, previous studies have demonstrated that chemical modification of crRNAs with 2′-fluoro-ribose[189], 2′-*O*-methyl 3′ phosphorothioate (MS)[188], and other moieties[202] increases gene editing efficiency by improving gRNA stability in cells and *in vivo*[191], suggesting another potential benefit of this technology.

First-generation BNAs, or locked nucleic acids (LNAs) (**Figure 2.1a**), are conformationally restricted RNA nucleotides in which the 2′ oxygen in the ribose forms a covalent bond to the 4′ carbon, inducing N-type (C3′-endo) sugar puckering and a preference for an A-form helix[220]. LNAs display improved base stacking and thermal stability compared to RNA, resulting in highly efficient binding to complementary nucleic acids and improved mismatch discrimination[220, 221]. They also display enhanced nuclease resistance[221]. LNAs have been successfully used in numerous applications ranging from single-nucleotide polymorphism (SNP) detection assays[221] to siRNA[222]. However, LNAs do have several limitations[223]. For example, oligonucleotides with multiple consecutive LNAs are unable to form nucleic acid triplex structures due to their rigidity[224, 225]. Furthermore, LNA-modified antisense oligonucleotides induce hepatotoxicity in mice[226]. Next-generation N-methyl substituted BNAs (2′,4′-BNA^{NC}[N-Me]) (**Figure 2.1a**) were designed to circumvent some of these issues. The six-membered bridged structure in BNA^{NC}s provides more conformational flexibility for nucleic acid binding and greater nuclease resistance due to steric bulk[223]. Recent work also suggests that BNA^{NC} nucleotides are less toxic than LNA nucleotides

when delivered to cultured cells[227]. Here, we show that incorporation of BNA^{NC}s and LNAs at specific positions within crRNAs broadly improves Cas9 DNA cleavage specificity.

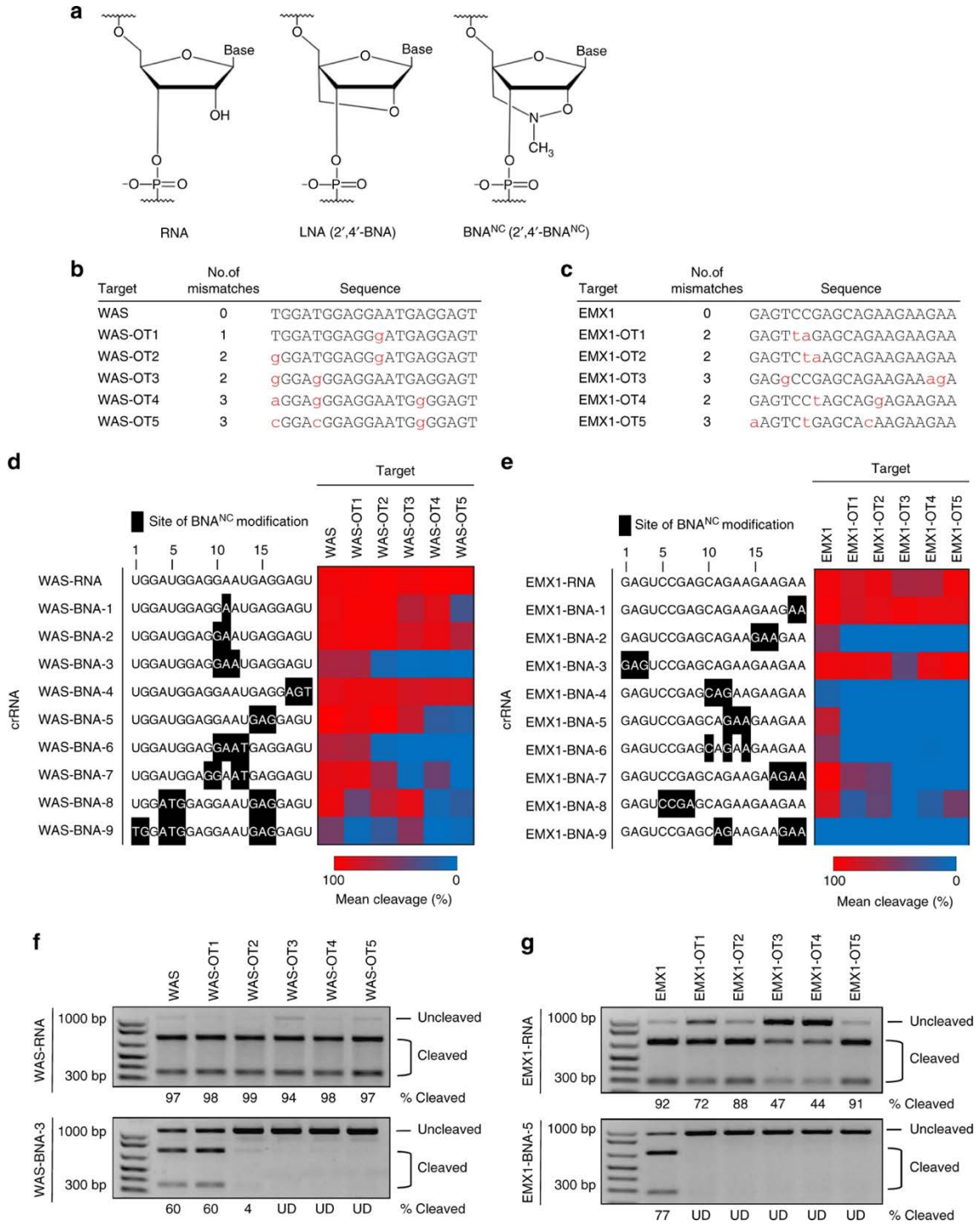


Figure 2.1. BNA^{NC} incorporation reduces off-target cleavage *in vitro*
(**a**) Chemical structures of RNA, LNA (2′,4′-BNA), and BNA^{NC} (2′,4′-BNA^{NC} [NMe]) nucleotides. (**b**) WAS and (**c**) EMX1 on-target and off-target sequences used for in vitro and cellular cleavage assays. Mismatches are indicated by red lowercase lettering. Heat map showing in vitro cleavage specificity for the unmodified crRNA and 9 BNA^{NC}-modified crRNAs toward either (**d**) WAS or (**e**) EMX1 on-target and off-target sequences (as listed in Figure 2.1b,c); mean shown ($n = 2$). crRNA and BNA^{NC}-modified sequences are shown to the left of the corresponding heat map. BNA^{NC} modifications are indicated in black. Targets that were highly cleaved in vitro are indicated in red, while targets that were not cleaved are indicated in blue. Gel showing relative cleavage efficiencies of the unmodified and most specific BNA^{NC}-modified crRNAs on a 1-kb linear DNA fragment containing either the (**f**) WAS or (**g**) EMX1 on-target and off-target sequences. The two bottom bands are cleavage products, while the top band is full-length substrate. The molar ratio of Cas9 RNP complex to target DNA was 30:1 for these experiments. Quantification of cleavage percentages was determined using densitometry (ImageJ), and are shown below each lane. Lanes in which no cleavage products were observed are marked as undetected (UD). Values used to generate heatmaps are presented in **Table 2.1**

Table 2.1. *In vitro* cleavage assay data used to generate Figure 2.1

	Target					
crRNA	**WAS**	**WAS-OT1**	**WAS-OT2**	**WAS-OT3**	**WAS-OT4**	**WAS-OT5**
WAS-RNA	81 ± 16	76 ± 2.8	80 ± 1.5	89 ± 0.72	73 ± 4.0	75 ± 8.2
WAS-BNA-1	69 ± 2.0	72 ± 2.4	95 ± 1.9	54 ± 3.1	66 ± 1.1	10 ± 1.2
WAS-BNA-2	67 ± 0.82	70 ± 0.17	77 ± 7.5	59 ± 0.70	58 ± 2.0	39 ± 1.7
WAS-BNA-3	63 ± 8.8	67 ± 10	9.3 ± 1.2	5.6 ± 0.08	7.5 ± 2.9	0.0 ± 0.0
WAS-BNA-4	68 ± 13	64 ± 4.0	85 ± 4.6	91 ± 0.15	59 ± 3.1	66 ± 0.48
WAS-BNA-5	65 ± 7.3	16 ± 4.1	34 ± 5.0	23 ± 4.1	0.0 ± 0.0	0.0 ± 0.0
WAS-BNA-6	55 ± 8.2	53 ± 8.9	2.9 ± 2.9	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
WAS-BNA-7	66 ± 6.6	64 ± 1.4	43 ± 2.5	2.2 ± 2.2	26 ± 6.4	0.0 ± 0.0
WAS-BNA-8	47 ± 3.0	31 ± 1.9	25 ± 1.2	66 ± 0.05	9.7 ± 1.2	23 ± 1.3
WAS-BNA-9	25 ± 0.90	10 ± 4.4	22 ± 0.34	39 ± 3.0	5.7 ± 5.7	4.4 ± 4.4

	Target					
crRNA	**EMX1**	**EMX1-OT1**	**EMX1-OT2**	**EMX1-OT3**	**EMX1-OT4**	**EMX1-OT5**
EMX1-RNA	65 ± 1.8	92 ± 3.0	46 ± 2.6	9.6 ± 0.71	8.7 ± 0.13	48 ± 17
EMX1-BNA-1	63 ± 6.2	73 ± 9.6	50 ± 4.2	17 ± 4.1	16 ± 4.3	44 ± 13
EMX1-BNA-2	4.9 ± 2.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-3	63 ± 1.8	61 ± 0.65	47 ± 2.2	4.1 ± 4.1	18 ± 0.20	59 ± 1.2
EMX1-BNA-4	7.4 ± 1.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-5	55 ± 4.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-6	38 ± 1.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-7	69 ± 12	8.6 ± 0.32	3.2 ± 3.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-8	50 ± 2.1	6.0 ± 1.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-9	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Values showing *in vitro* cleavage specificity for unmodified crRNA and 9 BNA^{NC}-modified crRNAs towards either WAS or EMX1 on- and off-target sequences (as listed in **Figure 2.1b**, **c**); Mean ± SE (n = 2). Experiments were performed using 150 nM Cas9 RNP and 5 nM DNA.

2.3 Results

2.3.1 BNA^{NC} incorporation reduces off-target cleavage *in vitro*

To test the hypothesis that incorporation of BNA^{NC}-modified nucleotides into crRNAs improves Cas9 cleavage specificity, we selected two previously characterized crRNAs directed toward the *WAS*[145] and *EMX1*[228] genes, for which *in vitro* and cellular off-target sites had previously been identified (**Figure 2.1b**, **c**), and designed variants with BNA^{NC} substitutions. Previous work has demonstrated that local mismatch discrimination can be improved in DNA–DNA hybrids when LNAs are incorporated in the vicinity of mismatched bases, with an LNA triplet centered on the mismatch yielding the best results[220]. Therefore, we generated a series of 9 crRNAs with sequential or alternating substitutions of 1, 2, 3 or 4 BNA^{NC}s (or in pairs) corresponding to the key mismatch positions of the five most abundant cellular off-target sites of the WAS and EMX1 crRNAs, respectively (**Figure 2.1b-e**). Using an *in vitro* cleavage assay with linear DNA substrates, we screened the ability of Cas9 to cleave the on-target sequence and off-target sequences using the original crRNAs and these BNA^{NC}-substituted crRNAs (WAS/EMX1-BNA-1-9) at two different concentrations (**Figure 2.1d-g**, **Figure 2.2**, **Tables 2.1** and **2.2**). Surprisingly, we found that substitution of the centrally located GAA triplet in positions 10–12 of the WAS crRNA with BNA^{NC} (WAS-BNA-3) abolished off-target cleavage on all but one of the off-target sequences (**Figure 2.1d**, **f**, **Figure 2.2a**, **c**). Interestingly, the off-target sequence that was unaffected contained a single A–G mismatch located within the substituted triplet. crRNAs containing substitutions of only 1 or 2 BNA^{NC}s, and substitutions in flanking positions within the crRNA demonstrated more heterogeneous effects, but still generally improved Cas9 specificity (**Figure 2.1d**, **Figure 2.2a**). A similar trend was observed for the EMX1 crRNAs, at both low and high concentrations of RNP complex (**Figure 2.1e**, **g**, **Figure 2b**, **d**, **Tables 2.1** and **2.2**). Remarkably,

BNA^{NC} substitutions at positions 12–14 of the EMX1 crRNA (EMX1-BNA-5) abolished cleavage on all five off-target sites tested, while only marginally decreasing on-target cleavage activity (**Figure 2.1e**, **g**, **Figure 2.2b**, **d**). Discontinuous BNA^{NC} substitutions at positions 10, 12, and 14 (EMX1-BNA-6) yielded similar improvements but with lower on-target activity, while EMX1-BNA-4 and EMX1-BNA-9 showed very low activity, possibly due to mis-folding of the crRNA (**Figure 2.1e**).

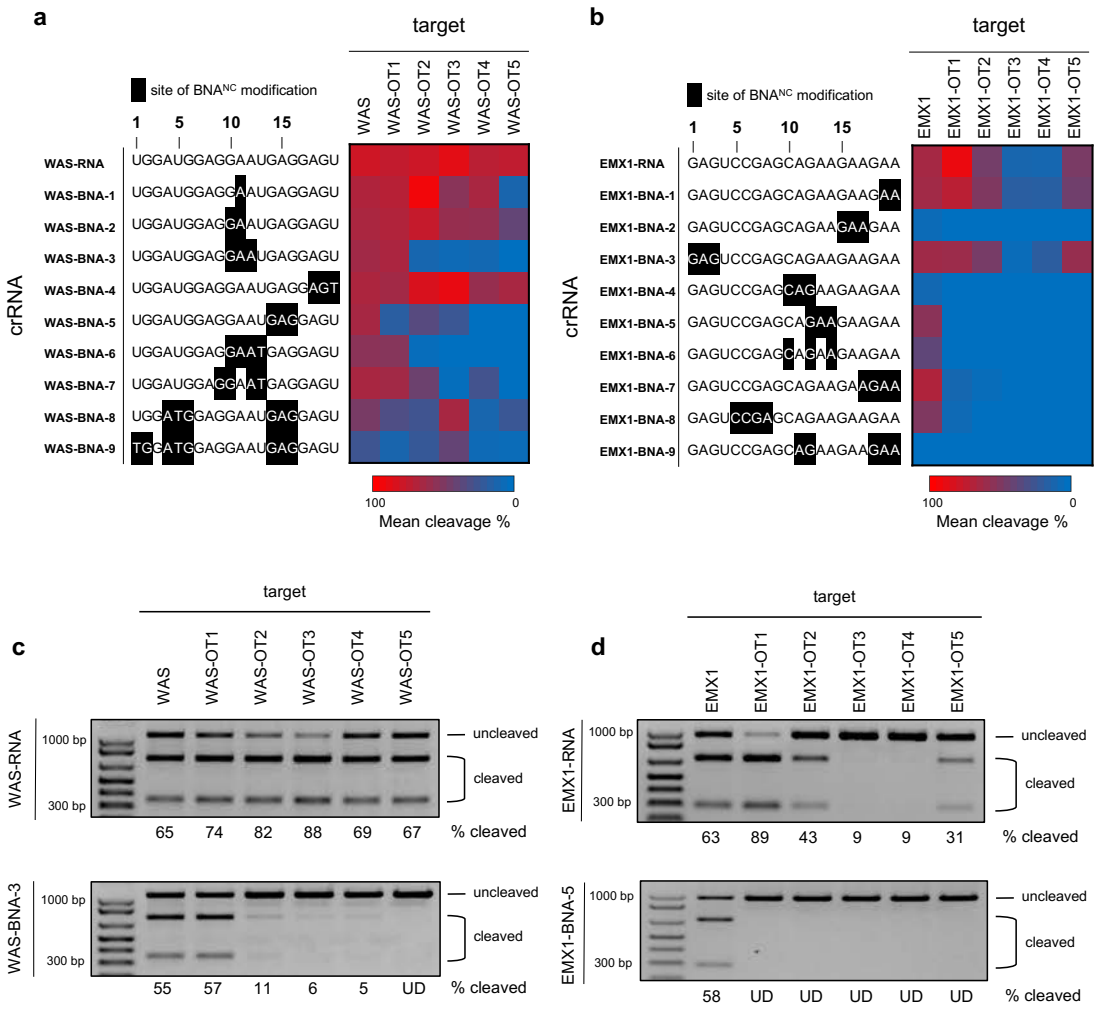


Figure 2.2. Incorporation of BNA^{NC} into crRNAs improves Cas9 cleavage specificity *in vitro*

Heat map showing *in vitro* cleavage specificity for unmodified crRNA and 9 BNA^{NC}-modified crRNAs towards either (**a**) WAS or (**b**) EMX1 on- and off-target sequences (as listed **Figure 2.1b**, **c**); Mean shown ($n = 2$). crRNA and BNA^{NC}-modified sequences are shown to the left of the corresponding heat map. BNA^{NC} modifications are indicated in black. Targets that were highly cleaved *in vitro* are indicated in red, while targets that were not cleaved are indicated in blue. Gel showing relative cleavage efficiencies of the unmodified and most specific BNA^{NC}-modified crRNAs on a linear 1-kb DNA fragment containing either the (**c**) WAS or (**d**) EMX1 on- and off-target sequences. The two bottom bands are cleavage products, while the top band is full-length substrate. The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments. Quantification of cleavage percentages was determined using densitometry (ImageJ), and are shown below each lane. Lanes in which no cleavage products were observed are marked as undetected (UD). Values used to generate heatmaps are shown in **Table 2.2**.

Table 2.2. *In vitro* cleavage assay data used to generate Figure 2.2

	Target					
crRNA	**WAS**	**WAS-OT1**	**WAS-OT2**	**WAS-OT3**	**WAS-OT4**	**WAS-OT5**
WAS-RNA	81 ± 16	76 ± 2.8	80 ± 1.5	89 ± 0.72	73 ± 4.0	75 ± 8.2
WAS-BNA-1	69 ± 2.0	72 ± 2.4	95 ± 1.9	54 ± 3.1	66 ± 1.1	10 ± 1.2
WAS-BNA-2	67 ± 0.82	70 ± 0.17	77 ± 7.5	59 ± 0.70	58 ± 2.0	39 ± 1.7
WAS-BNA-3	63 ± 8.8	67 ± 10	9.3 ± 1.2	5.6 ± 0.08	7.5 ± 2.9	0.0 ± 0.0
WAS-BNA-4	68 ± 13	64 ± 4.0	85 ± 4.6	91 ± 0.15	59 ± 3.1	66 ± 0.48
WAS-BNA-5	65 ± 7.3	16 ± 4.1	34 ± 5.0	23 ± 4.1	0.0 ± 0.0	0.0 ± 0.0
WAS-BNA-6	55 ± 8.2	53 ± 8.9	2.9 ± 2.9	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
WAS-BNA-7	66 ± 6.6	64 ± 1.4	43 ± 2.5	2.2 ± 2.2	26 ± 6.4	0.0 ± 0.0
WAS-BNA-8	47 ± 3.0	31 ± 1.9	25 ± 1.2	66 ± 0.05	9.7 ± 1.2	23 ± 1.3
WAS-BNA-9	25 ± 0.90	10 ± 4.4	22 ± 0.34	39 ± 3.0	5.7 ± 5.7	4.4 ± 4.4

	Target					
crRNA	**EMX1**	**EMX1-OT1**	**EMX1-OT2**	**EMX1-OT3**	**EMX1-OT4**	**EMX1-OT5**
EMX1-RNA	65 ± 1.8	92 ± 3.0	46 ± 2.6	9.6 ± 0.71	8.7 ± 0.13	48 ± 17
EMX1-BNA-1	63 ± 6.2	73 ± 9.6	50 ± 4.2	17 ± 4.1	16 ± 4.3	44 ± 13
EMX1-BNA-2	4.9 ± 2.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-3	63 ± 1.8	61 ± 0.65	47 ± 2.2	4.1 ± 4.1	18 ± 0.20	59 ± 1.2
EMX1-BNA-4	7.4 ± 1.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-5	55 ± 4.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-6	38 ± 1.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-7	69 ± 12	8.6 ± 0.32	3.2 ± 3.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-8	50 ± 2.1	6.0 ± 1.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
EMX1-BNA-9	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Values showing *in vitro* cleavage specificity for unmodified crRNA and 9 BNANC-modified crRNAs towards either WAS or EMX1 on- and off-target sequences (as listed **Figure 2.1b**, **c**); Mean ± SE (n = 2). Experiments were performed using 15 nM Cas9 RNP and 5 nM DNA.

2.3.2 BNA^{NC} incorporation broadly improves specificity *in vitro*

CRISPR-Cas9 is typically not effective in resolving SNPs[229] or single-nucleotide mismatches residing within the PAM-distal portion of the guide sequence[230]. Based on our finding that BNA^{NC}-substituted crRNAs improve specificity, we speculated that they might improve discrimination of single mismatch off-target sequences. To test this hypothesis, we generated a series of target sequences corresponding to the WAS and EMX1 sites bearing individual mutations at 2 bp

intervals and assayed their ability to be cleaved *in vitro* by Cas9 using either the unmodified crRNAs, or their most specific BNA^{NC}-modified counterparts. For the WAS sequence, we found that WAS-BNA-3 dramatically improved single mismatch discrimination at both PAM-proximal and PAM-distal regions of the target sequence, relative to the control, but had little effect on mismatches located in regions directly overlapping BNA^{NC} substitutions (**Figure 2.3a**, **Table 2.3**). Similarly, for the EMX1 sequence, EMX1-BNA-5 displayed improved specificity toward off-target sequences bearing single mutations in the PAM-proximal and PAM–distal regions, compared to the unmodified crRNA (**Figure 2.3b**, **Table 2.3**).

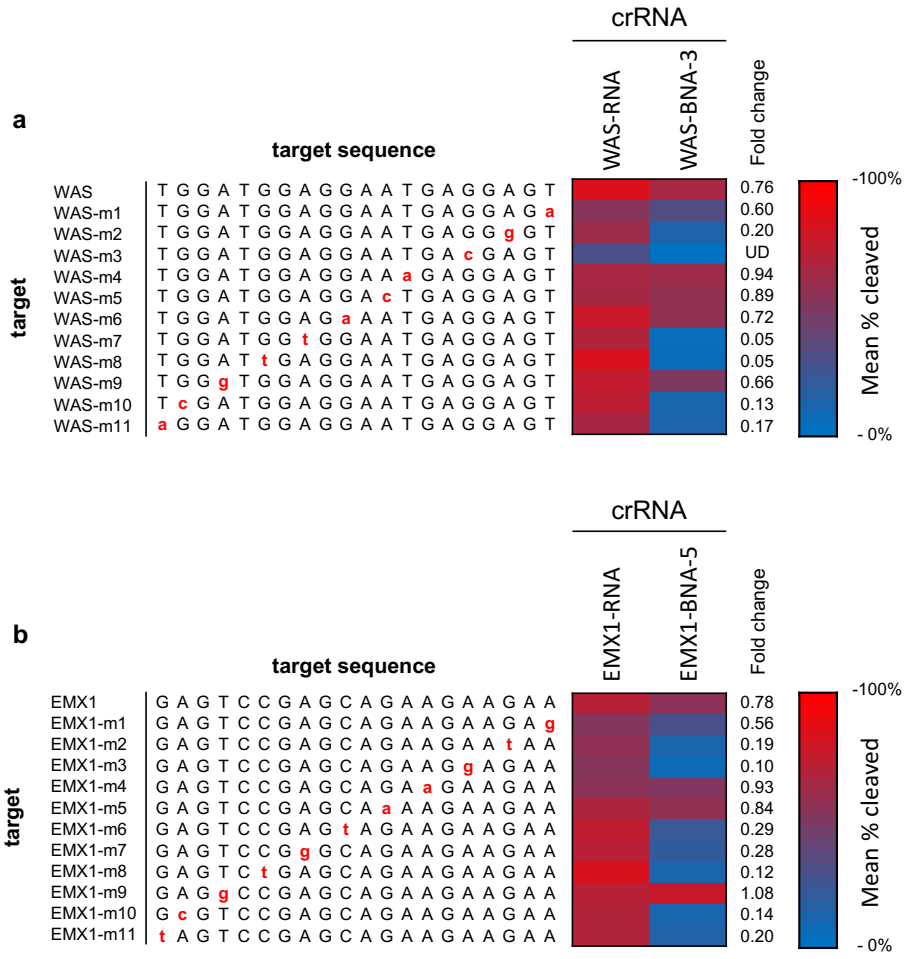


Figure 2.3. Incorporation of BNA^{NC} into crRNAs improves Cas9 cleavage specificity towards off-target sequences with single-nucleotide mismatches *in vitro*

(**a**) Heat map showing cleavage specificity of WAS-RNA and WAS-BNA-3 crRNAs towards on- and off-target sequences containing single-nucleotide mismatches; Mean shown (n = 2). Fold change indicates percent cleavage of WAS-BNA-3/WAS-RNA for each target. (**b**) Heat map showing cleavage specificity of EMX1-RNA and EMX1-BNA-5 crRNAs towards on- and off-target sequences with single-nucleotide mismatches; Mean shown (n = 2). Nucleotide mismatches are indicated by red lowercase lettering. Targets which were highly cleaved *in vitro* are indicated in red, while targets which were not cleaved are indicated in blue. Fold change indicates percent cleavage of EMX1-BNA-5/EMX1-RNA for each target. A designation of UD indicates no detectable cleavage of the target *in vitro*. Values used to generate the above heatmaps are shown in **Table 2.3**.

Table 2.3. *In vitro* cleavage assay data used to generate Figure 2.3

	crRNA	
Target	**WAS-RNA**	**WAS-BNA-3**
WAS	86 ± 7.5	66 ± 9.3
WAS-m1	54 ± 2.7	32 ± 3.5
WAS-m2	61 ± 8.6	12 ± 2.1
WAS-m3	30 ± 1.0	0.0 ± 0.0
WAS-m4	66 ± 1.5	62 ± 4.7
WAS-m5	65 ± 5.8	57 ± 3.4
WAS-m6	79 ± 2.8	57 ± 1.1
WAS-m7	70 ± 5.3	3.2 ± 3.2
WAS-m8	85 ± 1.9	4.0 ± 0.70
WAS-m9	76 ± 3.1	50 ± 0.50
WAS-m10	74 ± 2.2	9.7 ± 1.5
WAS-m11	65 ± 0.22	11 ± 0.07

	crRNA	
Target	**EMX1-RNA**	**EMX1-BNA-5**
EMX1	71 ± 9.1	56 ± 7.3
EMX1-m1	52 ± 9.8	29 ± 1.2
EMX1-m2	57 ± 5.3	11 ± 0.26
EMX1-m3	53 ± 8.3	5.4 ± 1.1
EMX1-m4	55 ± 7.5	51 ± 0.23
EMX1-m5	68 ± 4.3	56 ± 1.5
EMX1-m6	75 ± 2.0	22 ± 3.3
EMX1-m7	72 ± 8.7	20 ± 0.93
EMX1-m8	84 ± 4.0	11 ± 0.14
EMX1-m9	72 ± 1.1	78 ± 6.3
EMX1-m10	70 ± 0.52	9.9 ± 1.4
EMX1-m11	69 ± 0.06	14 ± 0.31

Table 2.3. Values showing *in vitro* cleavage specificity for unmodified crRNA and 9 BNA^{NC}-modified crRNAs towards either WAS (left) or EMX1 (right) on- and off-target sequences containing single-nucleotide mismatches (as listed in **Figure 2.3**); Mean ± SD (n = 2). Experiments were performed using 15 nM Cas9 RNP and 5 nM DNA.

To globally assess how BNA^{NC} modification of crRNAs influences Cas9 specificity, we employed a previously described *in vitro* high-throughput specificity profiling assay[13, 52, 158]. Briefly, this technique selects DNA sequences that have undergone cleavage from a library of $>10^{12}$ off-target sequences which contains a 10-fold coverage of all sequences with ≤8 mutations relative to the on-target sequence[13, 52, 158]. We performed this cleavage assay at both high and low concentrations of Cas9 RNP complex. We observed that WAS-BNA-3 and EMX1-BNA-5 broadly reduced the frequency of off-target cleavage on sequences containing 3–8 mutations relative to the on-target sequence, compared to the unmodified crRNAs (**Figure 2.4**). Moreover, we saw a significant reduction in the mean number of mutations in each selected sequence for the BNA^{NC}-modified

crRNAs, relative to their unmodified counterparts (**Table 2.4**). Using this dataset, we next calculated enrichment scores for each base at each position within the target sequence for the crRNAs (**Figure 2.5a**, **b**). We found that specificity was dramatically improved for both the WAS-BNA-3 and EMX1-BNA-5 crRNAs at nearly all positions in their respective target sequences (relative to the unmodified crRNAs), except those in the vicinity of the BNA^{NC} substitutions (**Figure 2.5a**-**d**, Figure **2.6**). Collectively, these results establish that incorporation of BNA^{NC} nucleotides into central positions of crRNAs broadly improves Cas9 specificity *in vitro*.

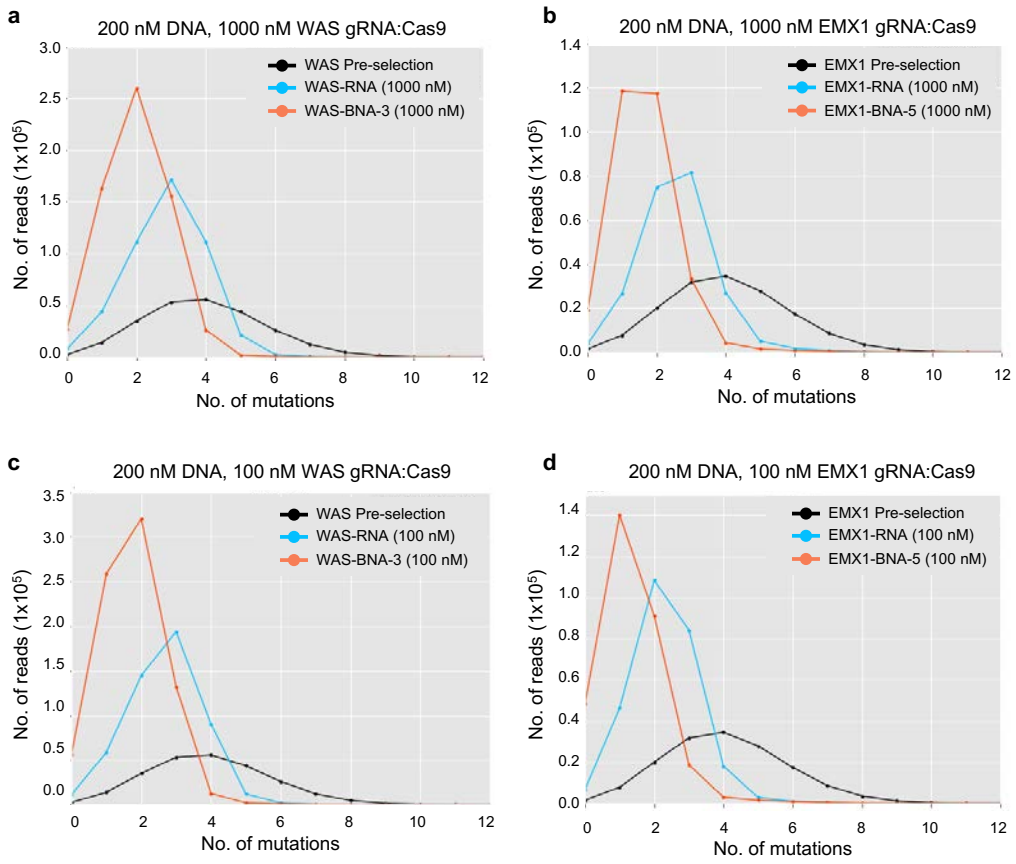


Figure 2.4. Distribution of mutations in pre- and post-selection libraries following *in vitro* library selection with unmodified or BNA^{NC}-modified crRNAs

In vitro selections were performed using 200 nM pre-selection library with 1000 nM Cas9 RNP complex targeting either (**a**) WAS or (**b**) EMX1. *In vitro* selections were performed using 200 nM pre-selection library with 100 nM Cas9 RNP complex targeting either (**c**) WAS or (**d**) EMX1. Cas9 RNP complexes were pre-assembled using unmodified or BNA^{NC}-modified crRNAs prior to DNA digestion. Distributions of mutations within pre-selection (black) and post-selection (colored) libraries are shown. Mutations are counted from the 20-base-pairs comprising the Cas9 target site, as well as the 2-base-pair PAM (N of NGG is excluded). Additional statistics are shown in **Table 2.4**.

Table 2.4. Statistics of sequences from *in vitro* high-throughput library selection experiment

Selection	Sequence count	Mean mutations	SD mutations	*P*-value
WAS-RNA (1000 nM)	470856	2.880	1.119	$< 2.2 \times 10^{-16}$
WAS-BNA-3 (1000 nM)	635675	1.999	0.948	$< 2.2 \times 10^{-16}$
WAS-LNA-3 (1000 nM)	758713	2.372	1.026	$< 2.2 \times 10^{-16}$
WAS-RNA (100 nM)	513723	2.660	1.076	$< 2.2 \times 10^{-16}$
WAS-BNA-3 (100 nM)	783753	1.747	0.924	$< 2.2 \times 10^{-16}$
WAS-LNA-3 (100 nM)	865363	2.070	0.949	$< 2.2 \times 10^{-16}$
WAS pre-selection	249602	3.945	1.735	NA
EMX1-RNA (1000 nM)	222094	2.585	1.097	$< 2.2 \times 10^{-16}$
EMX1-BNA-5 (1000 nM)	295408	1.646	0.928	$< 2.2 \times 10^{-16}$
EMX1-LNA-5 (1000 nM)	338344	1.893	0.922	$< 2.2 \times 10^{-16}$
EMX1-RNA (100 nM)	268000	2.289	1.003	$< 2.2 \times 10^{-16}$
EMX1-BNA-5 (100 nM)	304230	1.342	0.956	$< 2.2 \times 10^{-16}$
EMX1-LNA-5 (100 nM)	390684	1.652	0.873	$< 2.2 \times 10^{-16}$
EMX1 pre-selection	154510	4.063	1.763	NA

Statistics are shown for pre-selection libraries, as well as post-selection libraries following digestion with unmodified, BNA^{NC}- or LNA-modified crRNA. Total counts are shown for sequences which passed computational filtering, as well as had Phred scores >30 at each position along the Cas9 target site and PAM. Mean mutations were calculated using the 20 nucleotides of the target site, as well as 2 of the 3 comprising the PAM (N of NGG was excluded). *P*-values were determined through comparison of 150 000 randomly sampled sequences of post-selection and corresponding pre-selection libraries using a Mann-Whitney test.

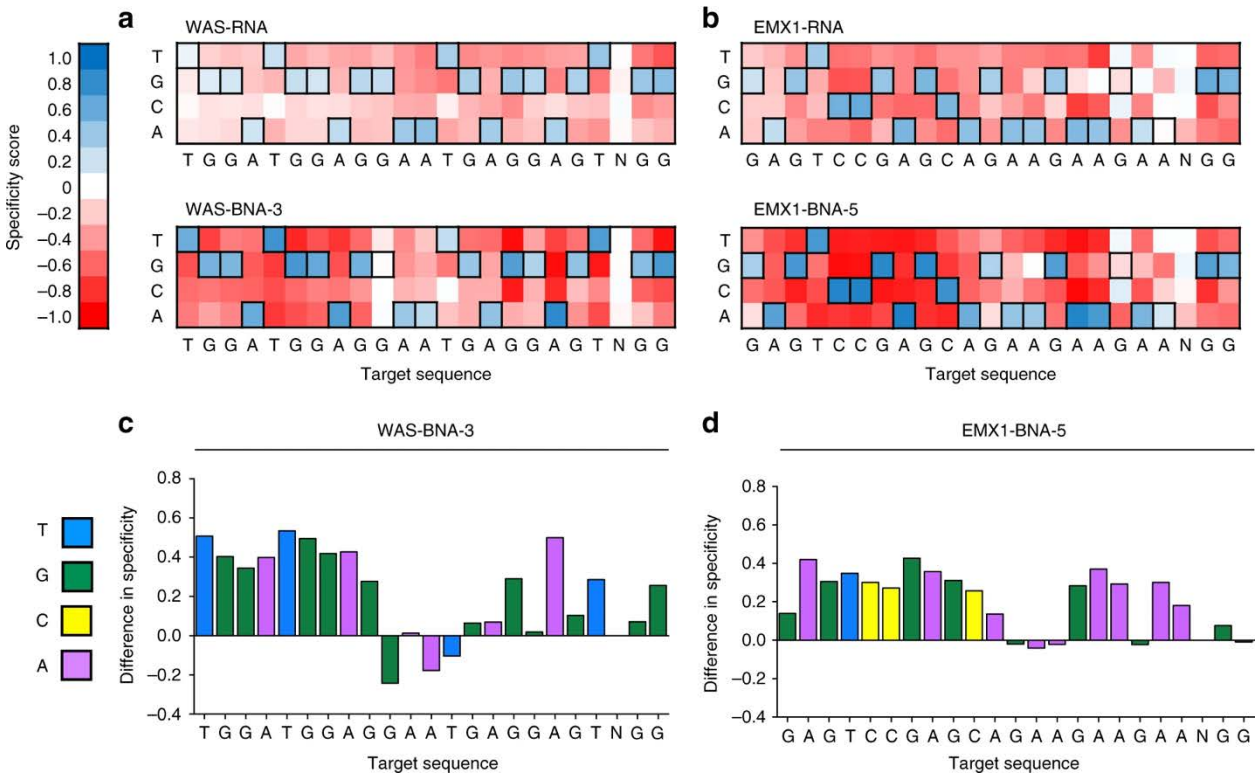


Figure 2.5. BNA^{NC} incorporation broadly improves specificity *in vitro*
Heat maps showing DNA cleavage specificity scores across >10^{12} off-target sequences for either unmodified (top) or BNA^{NC}-modified crRNAs (bottom) targeting (**a**) WAS or (**b**) EMX1. Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of −1.0 (dark red) correspond to 100% enrichment against a specific base-pair at a specific position. Black boxes denote the intended target nucleotides. Bar graph showing the quantitative difference in specificity score at each position in the 20 base-pair target site and 2 base-pair PAM (N of NGG excluded), between the unmodified and BNA^{NC}-modified crRNA for (**c**) WAS or (**d**) EMX1 target sequences. A score of zero indicates no change in specificity. Difference in specificity was calculated as specificity score$_{BNA^{NC}}$−specificity score$_{RNA}$. Experiments were performed with 200 nM pre-selection library and 1000 nM Cas9 RNP

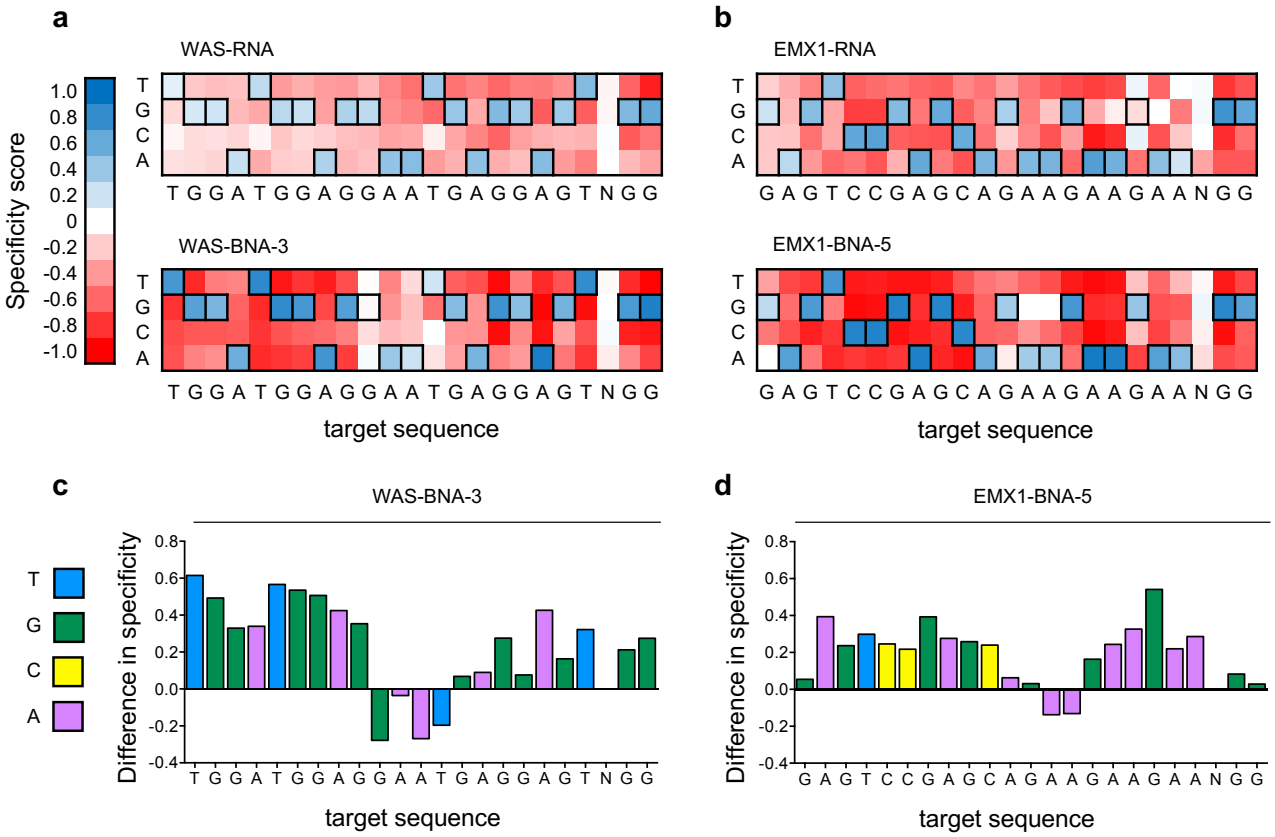


Figure 2.6. *In vitro* specificity profiling results for unmodified and BNA^{NC}-modified crRNAs

Heat maps showing DNA cleavage specificity scores across >10^{12} off-target sequences for either unmodified (top) or BNA^{NC}-modified (bottom) crRNAs targeting (**a**) WAS or (**b**) EMX1. Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of -1.0 (dark red) correspond to 100% enrichment against a specific base-pair at a specific position. Black boxes denote the intended target nucleotides. Bar graph showing the quantitative difference in specificity score at each position in the 20 base-pair target site and 2 base-pair PAM (N of NGG excluded), between the unmodified and BNA^{NC}-modified crRNA for (**c**) WAS or (**d**) EMX1 target sequences. A score of zero indicates no change in specificity. Difference in specificity was calculated as, specificity score$_{BNA^{NC}}$ – specificity score$_{RNA}$. Experiments were performed with 200 nM pre-selection library and 100 nM Cas9 RNP complex.

2.3.3 BNA^{NC}-modified crRNAs are compatible with Cas9 variants

Recently, variants of Cas9 with improved specificity have been engineered[175-177]. To examine if BNA^{NC}-substituted crRNAs could be used in conjunction with these variants to further boost specificity, we tested the activity of eSpCas9, a Cas9 variant with substitutions that reduce non-specific interactions with the non-complementary DNA strand[175], *in vitro* using our WAS and EMX1 on-target and off-target sequences and the corresponding unmodified crRNAs. For the WAS sequence, we found that eSpCas9 reduced off-target cleavage of WAS-OT3 and WAS-OT5, but had little effect on WAS-OT1, WAS-OT2, and WAS-OT4, compared to SpCas9 (**Figure 2.7**). Next, we repeated the cleavage assay of the WAS on-target and off-target sequences using eSpCas9 in combination with WAS-BNA-5. We selected this modified crRNA to avoid redundancy in specificity improvements, since it was able to reduce cleavage of off-target sites distinct from those affected by eSpCas9. Strikingly, we found that the combination was additive, resulting in elimination of nearly all off-target activity (**Figure 2.7a**, **c**). These results demonstrate that BNA^{NC}-modified crRNAs can complement the specificity enhancements of next-generation Cas9 variants.

a

	SpCas9		eSpCas9	
	WAS-RNA	WAS-BNA-5	WAS-RNA	WAS-BNA-5
WAS	81 ± 15	66 ± 10	64 ± 22	61 ± 3
WAS-OT1	77 ± 5	17 ± 6	88 ± 13	8 ± 2
WAS-OT2	81 ± 5	34 ± 7	94 ± 2	UD
WAS-OT3	89 ± 10	23 ± 6	5 ± 2	UD
WAS-OT4	73 ± 6	UD	42 ± 15	UD
WAS-OT5	76 ± 11	UD	7 ± 3	UD

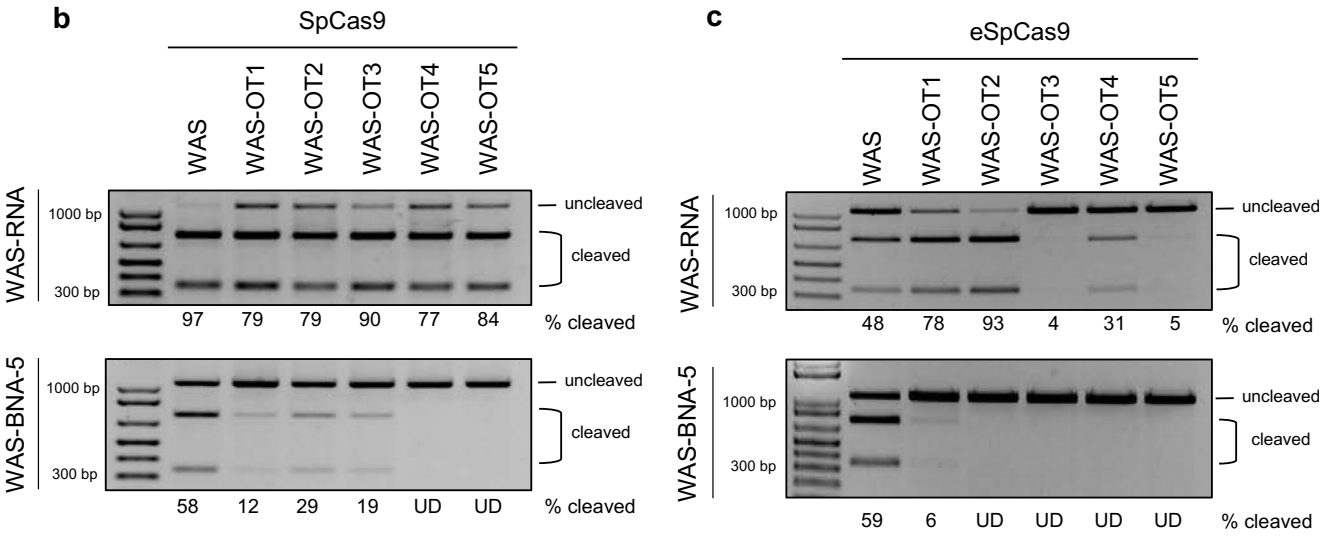


Figure 2.7. BNA^{NC}-modified crRNAs show additive specificity when used in combination with eSpCas9 *in vitro*

(**a**) Table showing cleavage specificities of WAS-RNA and WAS-BNA-5 crRNA towards WAS on- and off-target sequences (shown in **Figure 2.1b**) when complexed with either SpCas9 or eSpCas9; Mean ± SD shown (n = 2). Reactions in which no cleavage products were observed are marked as undetected (UD). Gel showing cleavage assay results for WAS-RNA and WAS-BNA-5 crRNA complexed with (**b**) SpCas9 or (**c**) eSpCas9 using WAS on- and off-target containing sequences. For all cleavage gels, the top band is full-length DNA substrate, while the two bottom bands are cleavage products. The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments. Quantification of cleavage products was determined by densitometry (ImageJ), and are shown below each lane. Lanes in which no cleavage products

2.3.4 BNA^{NC} incorporation improves Cas9 specificity in cells

To test if the improved specificity of the BNA^{NC}-modified crRNAs observed *in vitro* translates into reduced off-target cleavage in cells, we transfected U2OS and HeLa cells stably expressing

Cas9 with either unmodified or each of the 9 BNA^{NC}-modified gRNAs corresponding to either the *WAS* or *EMX1* loci. Initially, we examined the activity of these gRNAs in cells by measuring cleavage of the on-target site and one off-target site using the T7 endonuclease I assay[231]. Consistent with our *in vitro* findings, we found that BNA^{NC}-modified crRNAs generally induced lower cleavage rates at the off-target site (**Figure 2.8**). We also observed that on-target activity was reduced in several instances. To quantitatively measure the on-target and off-target cleavage rates of WAS-BNA-3 and EMX1-BNA-5 and their unmodified counterparts in cells, we performed high-throughput sequencing. As shown in **Table 2.5**, while cleavage of the on-target site and OT1 was reduced by a factor of ~2–3-fold in HeLa cells, cleavage of OT2, OT3, OT4, and OT5 was reduced by >17,000-fold, >24,000-fold, >11,000-fold, and >24,000-fold, respectively (estimating the lower limit of detection to be 0.003%) using WAS-BNA-3. We observed similar specificity improvements using EMX1-BNA-5, compared to the unmodified EMX1 crRNA, in both cell types (**Table 2.6**, statistics and *P*-values in **Table 2.7**). Finally, to determine if BNA^{NC}-modified crRNAs alter how Cas9 cuts DNA, or its subsequent repair, we compared both the size and location of insertions/deletions generated using WAS-BNA-3 or EMX1-BNA-5 to those generated using their unmodified crRNA counterparts. We found that the pattern of indel formation was highly similar in both cases (**Figures 2.9**-**2.11**). These data establish BNA^{NC}-modification of crRNAs as a new strategy to improve Cas9 DNA cleavage specificity in cells.

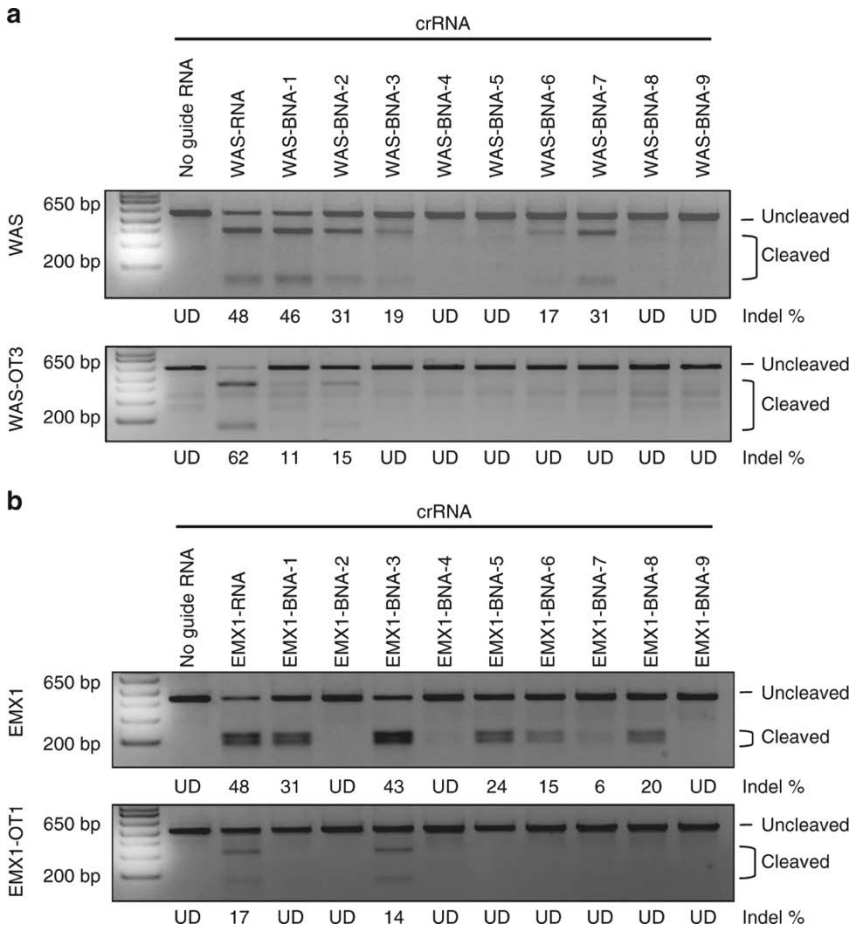


Figure 2.8. BNA^{NC} incorporation increases Cas9 specificity in cells
Gel showing relative cellular cleavage efficiencies of the unmodified, and 9 BNA^{NC}-modified crRNAs in U2OS-Cas9 cells targeting (**a**) *WAS* or (**b**) *EMX1* on-target (top) or off-target (bottom) sequences, as determined by T7 endonuclease I digestion. Mock transfections lacking guide RNA were used as controls. Modification frequencies were determined using densitometry (ImageJ) and are indicated below each lane. Lanes in which no cleavage products were observed are marked as undetected (UD)

Table 2.5. Cellular modification rates using *WAS*-targeting crRNAs

Cell Type	U2OS-Cas9	U2OS-Cas9	U2OS-Cas9	HeLa-Cas9	HeLa-Cas9	HeLa-Cas9
Treatment	No gRNA	WAS-RNA	WAS-BNA-3	No gRNA	WAS-RNA	WAS-BNA-3
WAS	< 0.003	58.392	14.572	< 0.003	56.020	19.475
WAS-OT1	0.003	17.548	9.400	0.004	18.135	8.726
WAS-OT2	0.092	59.553	0.034	0.120	53.400	< 0.003
WAS-OT3	< 0.003	73.142	< 0.003	< 0.003	70.859	< 0.003
WAS-OT4	< 0.003	36.506	< 0.003	< 0.003	33.151	< 0.003
WAS-OT5	< 0.003	75.239	< 0.003	< 0.003	71.119	< 0.003

Table summarizing modification frequencies of on-target and off-target sequences in U2OS-Cas9 and HeLa-Cas9 cells using either unmodified or BNA^{NC}-modified crRNAs targeting *WAS*, as determined by high-throughput sequencing. Modification frequencies were calculated by dividing the number of sequences bearing insertions or deletions (indels) in the target site by the total number of sequences and multiplied by 100 to give a percent. Mock transfections lacking gRNA were used as controls (see **Table 2.7** for additional data)

Table 2.6. Cellular modification rates using *EMX1*-targeting crRNAs

Cell Type	U2OS-Cas9	U2OS-Cas9	U2OS-Cas9	HeLa-Cas9	HeLa-Cas9	HeLa-Cas9
Treatment	No gRNA	EMX1-RNA	EMX1-BNA-5	No gRNA	EMX1-RNA	EMX1-BNA-5
EMX1	0.004	69.181	27.393	< 0.003	69.118	25.649
EMX1-OT1	0.005	12.233	< 0.003	0.009	11.677	< 0.003
EMX1-OT2	0.031	1.007	0.020	0.025	0.607	0.020
EMX1-OT3	0.068	< 0.003	0.047	0.005	< 0.003	< 0.003
EMX1-OT4	0.073	< 0.003	< 0.003	0.042	0.016	0.051
EMX1-OT5	< 0.003	0.113	0.002	0.060	0.141	< 0.003

Table summarizing modification frequencies of on-target and off-target sequences in U2OS-Cas9 and HeLa-Cas9 cells using either unmodified or BNA^{NC}-modified crRNAs targeting *EMX1*, as determined by high-throughput sequencing. Modification frequencies were calculated by dividing the number of sequences bearing insertions or deletions (indels) in the target site by the total number of sequences and multiplied by 100 to give a percent. Mock transfections lacking gRNA were used as controls (see **Table 2.7** for additional data)

Table 2.7. Statistics of cellular modification frequencies, sample size and *P*-values for high-throughput sequencing of Cas9:gRNA cleavage in U2OS-Cas9 and HeLa-Cas9 cells using no guide RNA, unmodified, BNA^{NC}- or LNA-modified crRNAs

Cell Line	Site	crRNA	Indel	Total Counts	*P* Value
U2OS-Cas9	WAS	Control	7	135063	
U2OS-Cas9	WAS	WAS-RNA	74404	127421	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS	WAS-BNA-3	19041	130670	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS	WAS-LNA-3	66896	186551	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT1	Control	6	265507	
U2OS-Cas9	WAS-OT1	WAS-RNA	36671	208973	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT1	WAS-BNA-3	27173	289132	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT1	WAS-LNA-3	55509	200033	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT2	Control	0	151841	
U2OS-Cas9	WAS-OT2	WAS-RNA	84656	142152	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT2	WAS-BNA-3	77	226039	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT2	WAS-LNA-3	79930	226778	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT3	Control	0	120982	
U2OS-Cas9	WAS-OT3	WAS-RNA	89731	122681	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT3	WAS-BNA-3	1	196946	1
U2OS-Cas9	WAS-OT3	WAS-LNA-3	281	313339	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT4	Control	1	244429	
U2OS-Cas9	WAS-OT4	WAS-RNA	66337	181716	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT4	WAS-BNA-3	0	189146	1
U2OS-Cas9	WAS-OT4	WAS-LNA-3	560	154888	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT5	Control	2	216660	
U2OS-Cas9	WAS-OT5	WAS-RNA	218854	290880	$< 2.2 \times 10^{-16}$
U2OS-Cas9	WAS-OT5	WAS-BNA-3	0	232344	0.2328
U2OS-Cas9	WAS-OT5	WAS-LNA-3	58	177406	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS	Control	4	175815	
HeLa-Cas9	WAS	WAS-RNA	130684	233282	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS	WAS-BNA-3	35276	181138	$< 2.2 \times 10^{-16}$

Cell Line	Site	crRNA	Indel	Total Counts	*P* Value
HeLa-Cas9	WAS	WAS-LNA-3	52499	150807	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT1	Control	6	162391	
HeLa-Cas9	WAS-OT1	WAS-RNA	25450	140379	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT1	WAS-BNA-3	12276	140675	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT1	WAS-LNA-3	66605	246307	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT2	Control	1	200567	
HeLa-Cas9	WAS-OT2	WAS-RNA	88936	166572	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT2	WAS-BNA-3	0	196263	1
HeLa-Cas9	WAS-OT2	WAS-LNA-3	61926	205026	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT3	Control	4	225024	
HeLa-Cas9	WAS-OT3	WAS-RNA	113162	159701	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT3	WAS-BNA-3	7	149701	0.1292
HeLa-Cas9	WAS-OT3	WAS-LNA-3	266	172325	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT4	Control	0	220588	
HeLa-Cas9	WAS-OT4	WAS-RNA	73969	223128	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT4	WAS-BNA-3	2	196905	0.2224
HeLa-Cas9	WAS-OT4	WAS-LNA-3	114	291185	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT5	Control	2	201738	
HeLa-Cas9	WAS-OT5	WAS-RNA	207868	292282	$< 2.2 \times 10^{-16}$
HeLa-Cas9	WAS-OT5	WAS-BNA-3	4	212485	0.6881
HeLa-Cas9	WAS-OT5	WAS-LNA-3	5	330680	0.7167
U2OS-Cas9	EMX1	Control	6	146532	
U2OS-Cas9	EMX1	EMX1-RNA	112844	163114	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1	EMX1-BNA-5	78223	285556	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1	EMX1-LNA-5	86812	185446	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT1	Control	6	121079	
U2OS-Cas9	EMX1-OT1	EMX1-RNA	16929	138393	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT1	EMX1-BNA-5	3	167545	0.1789
U2OS-Cas9	EMX1-OT1	EMX1-LNA-5	8	158349	1
U2OS-Cas9	EMX1-OT2	Control	41	133703	
U2OS-Cas9	EMX1-OT2	EMX1-RNA	494	119068	$< 2.2 \times 10^{-16}$

Cell Line	Site	crRNA	Indel	Total Counts	*P* Value
U2OS-Cas9	EMX1-OT2	EMX1-BNA-5	45	199083	0.1865
U2OS-Cas9	EMX1-OT2	EMX1-LNA-5	24	180737	0.00099
U2OS-Cas9	EMX1-OT3	Control	50	73931	
U2OS-Cas9	EMX1-OT3	EMX1-RNA	0	98752	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT3	EMX1-BNA-5	71	151126	0.05264
U2OS-Cas9	EMX1-OT3	EMX1-LNA-5	20	133801	1.558×10^{-9}
U2OS-Cas9	EMX1-OT4	Control	92	126743	
U2OS-Cas9	EMX1-OT4	EMX1-RNA	2	126081	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT4	EMX1-BNA-5	4	159035	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT4	EMX1-LNA-5	0	194305	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT5	Control	71	166087	
U2OS-Cas9	EMX1-OT5	EMX1-RNA	56	163340	$< 2.2 \times 10^{-16}$
U2OS-Cas9	EMX1-OT5	EMX1-BNA-5	40	173485	1
U2OS-Cas9	EMX1-OT5	EMX1-LNA-5	72	151924	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1	Control	64	203514	
HeLa-Cas9	EMX1	EMX1-RNA	126491	183008	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1	EMX1-BNA-5	46803	182476	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1	EMX1-LNA-5	96279	199447	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1-OT1	Control	12	135311	
HeLa-Cas9	EMX1-OT1	EMX1-RNA	16999	145571	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1-OT1	EMX1-BNA-5	7	148785	0.2503
HeLa-Cas9	EMX1-OT1	EMX1-LNA-5	2	106278	0.02969
HeLa-Cas9	EMX1-OT2	Control	42	167364	
HeLa-Cas9	EMX1-OT2	EMX1-RNA	328	144454	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1-OT2	EMX1-BNA-5	29	145830	0.3442
HeLa-Cas9	EMX1-OT2	EMX1-LNA-5	37	165871	0.6532
HeLa-Cas9	EMX1-OT3	Control	60	111185	
HeLa-Cas9	EMX1-OT3	EMX1-RNA	2	138824	0.1505
HeLa-Cas9	EMX1-OT3	EMX1-BNA-5	0	127723	0.01016
HeLa-Cas9	EMX1-OT3	EMX1-LNA-5	0	123049	0.01144
HeLa-Cas9	EMX1-OT4	Control	50	119940	

Cell Line	Site	crRNA	Indel	Total Counts	*P* Value
HeLa-Cas9	EMX1-OT4	EMX1-RNA	20	129032	0.0001025
HeLa-Cas9	EMX1-OT4	EMX1-BNA-5	82	160850	0.2911
HeLa-Cas9	EMX1-OT4	EMX1-LNA-5	55	153363	0.4914
HeLa-Cas9	EMX1-OT5	Control	90	151920	
HeLa-Cas9	EMX1-OT5	EMX1-RNA	26	137555	4.024×10^{-11}
HeLa-Cas9	EMX1-OT5	EMX1-BNA-5	21	138296	$< 2.2 \times 10^{-16}$
HeLa-Cas9	EMX1-OT5	EMX1-LNA-5	82	148466	$< 2.2 \times 10^{-16}$

Statistics are shown for cellular modification frequencies following transfection with unmodified, BNA^{NC}- or LNA-modified crRNAs into stably expressing Cas9 cells. Controls were performed using no guide RNA during transfection. Indels are defined as the number of observed sequences containing insertions or deletions which are consistent with Cas9-mediated cleavage. Total number of sequences for each library is reported as the number of reads following high-throughput sequencing which were successfully mapped to a reference amplicon and having a mean Phred score >30. *P*-values were calculated using a Fisher's exact test between each Cas9-treated sample and the corresponding untreated control. The sensitivity of high-throughput sequencing for the detection of genomic off-target cleavage is limited by the amount of genomic DNA (gDNA) used as input. Each sample was run with 100 ng of gDNA, which is equivalent to ~33 000 genomes. Therefore, the theoretical detection limit of this technique is ~1 in 33 000, which we have indicated as < 0.003%.

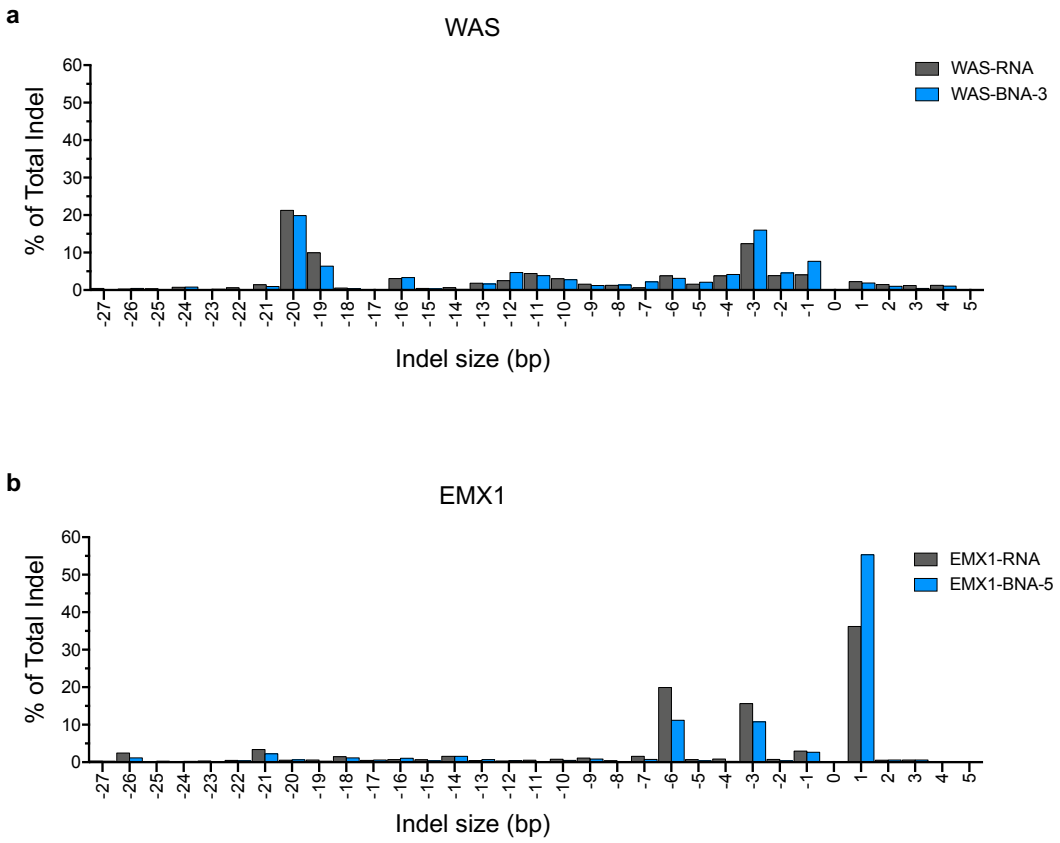


Figure 2.9. Comparison of indel size resulting from Cas9 DNA cleavage with unmodified or BNA^{NC}-modified crRNAs

Distributions of indel sizes are shown for either (**a**) WAS or (**b**) EMX1 using unmodified (grey) or BNA^{NC}-modified (blue) crRNAs in U2OS-Cas9 cells. Distribution of indel sizes is based on sequences obtained following high-throughput sequencing of transfected cells. The number of reads for each indel size was normalized to the total number of indel reads. Only indels ranging from -27 bp (27 bp deletion) to +5 bp (5 bp insertion) are shown.

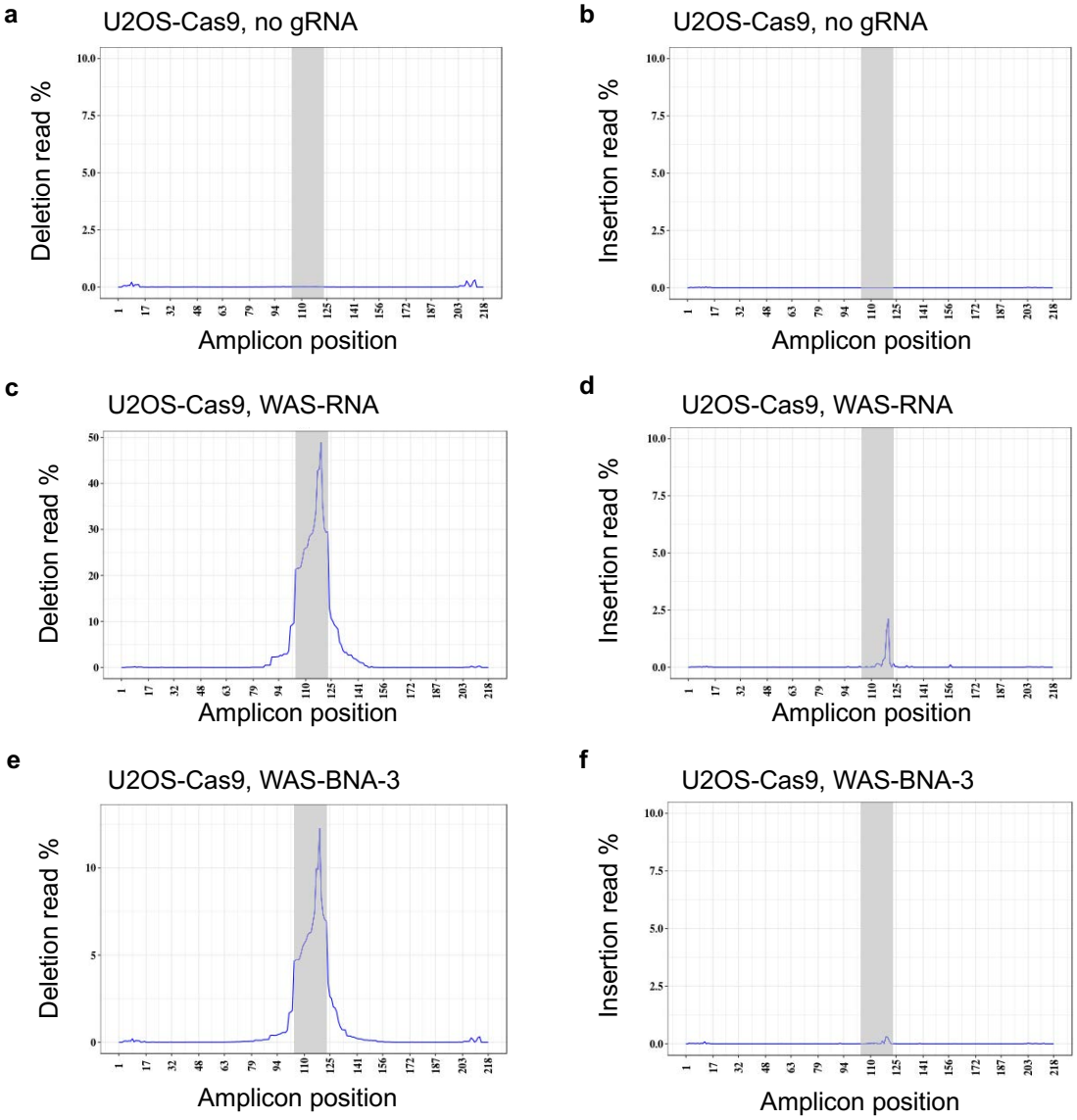


Figure 2.10. Distribution of indel location using unmodified or BNA^{NC}-modified crRNAs on WAS target

Distribution frequencies of insertions (right) and deletions (left) across the *WAS* target site in U2OS-Cas9 cells transfected with no gRNA (**a**, **b**), WAS-RNA (**c**, **d**) or WAS-BNA-3 (**e**, **f**) are shown. Grey coloring indicates the location of the genomic CRISPR target site along the amplicon. Distribution of indel positions is based on sequences obtained following high-throughput sequencing of transfected cells.

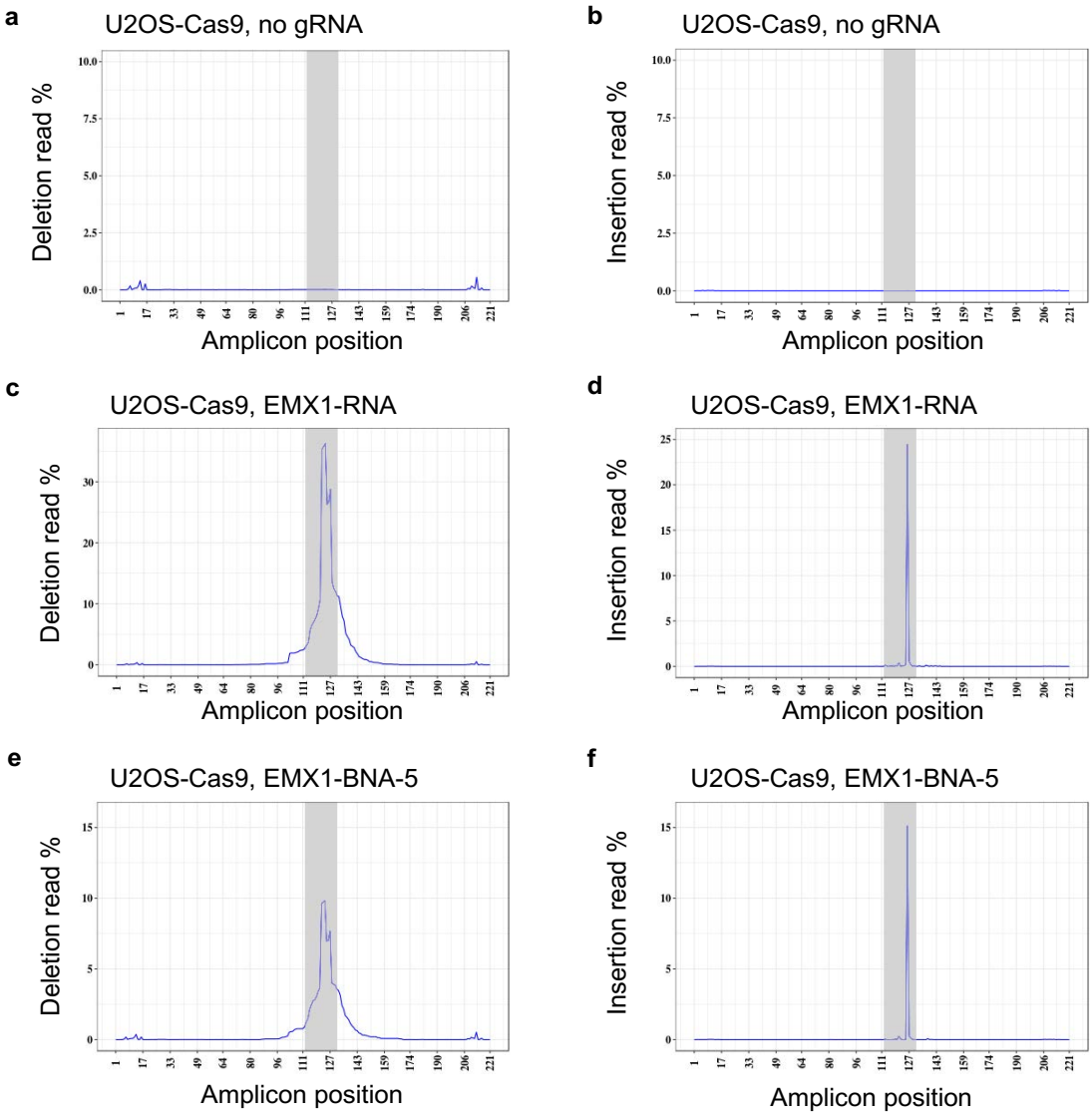


Figure 2.11. Distribution of indel location using unmodified or BNA^{NC}-modified crRNAs on EMX1 target

Distribution frequencies of insertions (right) and deletions (left) across the *EMX1* target site in U2OS-Cas9 cells transfected with no gRNA (**a**, **b**), EMX1-RNA (**c**, **d**) or EMX1-BNA-5 (**e**, **f**) are shown. Grey coloring indicates the location of the genomic CRISPR target site along the amplicon. Distribution of indel positions is based on sequences obtained following high-throughput sequencing of transfected cells.

2.3.5 BNA^{NC}s improve specificity via a conformational mechanism

We next investigated the mechanism underlying the specificity improvements of BNA^{NC}-modified crRNAs. At least five distinct stages in the Cas9 cleavage reaction have been resolved: tracrRNA/crRNA loading onto Cas9, binding of the RNP complex to target DNA, DNA melting and PAM-proximal hybridization (open conformation), complete R-loop formation (zipped conformation), and structural rearrangement of the nuclease domains leading to cleavage[232]. To identify which of these stages is altered by BNA^{NC}-modified crRNAs, we performed a variety of biochemical and biophysical experiments. We found that altering the tracrRNA:crRNA ratio in on-target and off-target cleavage reactions using either WAS-RNA or WAS-BNA-3 yielded similar effects (WAS-OT3 was not cleaved when WAS-BNA-3 crRNA was used), suggesting that BNA^{NC} incorporation does not alter tracrRNA hybridization (**Figure 2.12a**, **b**). In addition, titrations using different concentrations of annealed gRNA, whole RNP complex, and target DNA produced similar results between WAS-RNA and WAS-BNA-3 cleavage reactions with the on-target WAS sequence, while cleavage of WAS-OT3 was abolished in all cases when WAS-BNA-3 crRNA was used (**Figure 2.12c**-**h**). Using an electrophoretic mobility shift assay (EMSA), we observed that BNA^{NC} incorporation did not alter the ability of nuclease-deficient Cas9 (dCas9) to bind to DNA containing both on-target and off-target sequences (**Figure 2.13**).

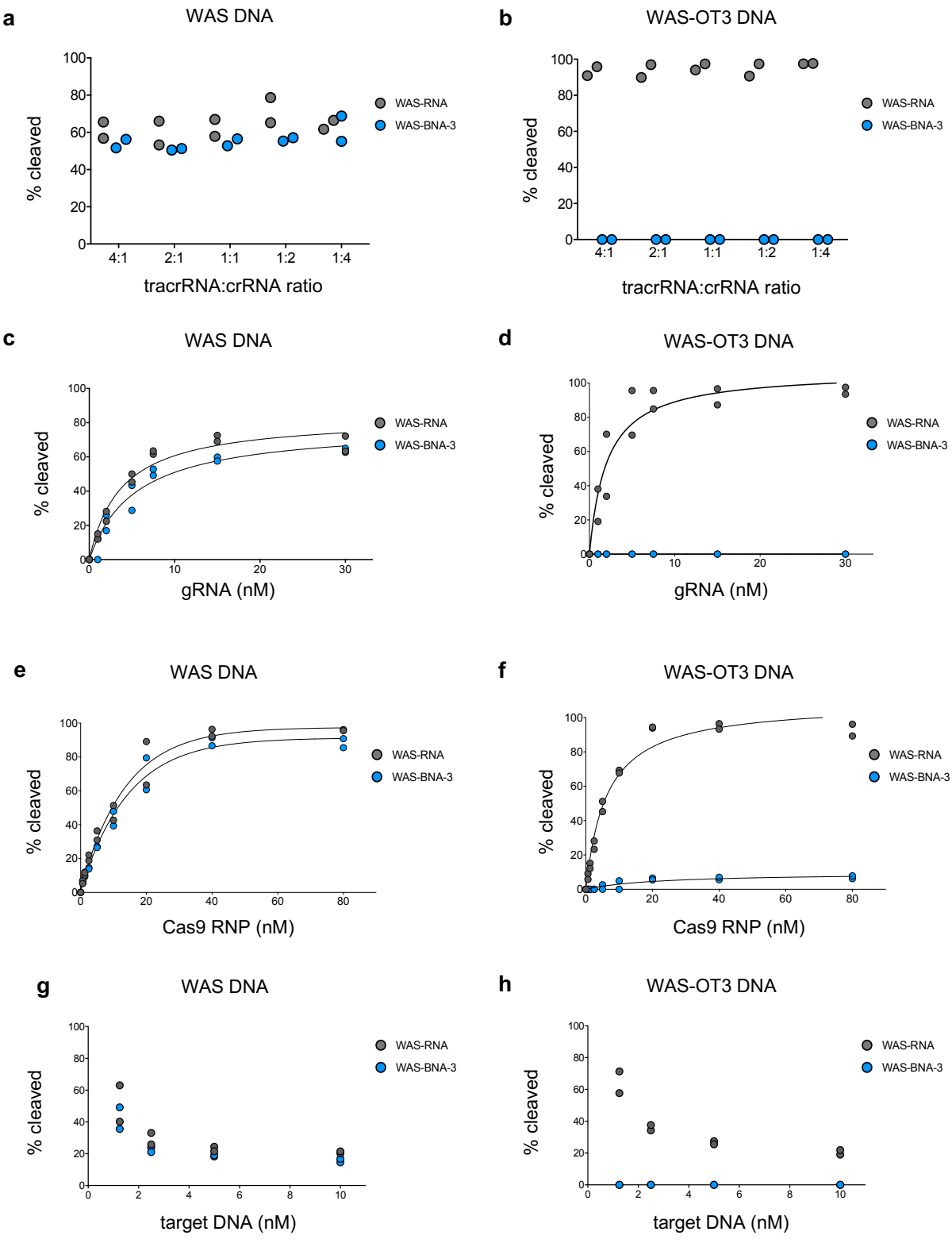

a
WAS DNA
% cleaved
tracrRNA:crRNA ratio
4:1
2:1
1:1
1:2
1:4
WAS-RNA
WAS-BNA-3
b
WAS-OT3 DNA
c
gRNA (nM)
d
e
Cas9 RNP (nM)
f
g
target DNA (nM)
h

Figure 2.12. Effect of BNA^{NC}-modified crRNAs on affinity, activity, and kinetics of Cas9 *in vitro*

Graph showing the effect of differential tracrRNA/crRNA annealing ratios on (**a**) WAS or (**b**) WAS-OT3 DNA cleavage *in vitro* using WAS-RNA or WAS-BNA-3 crRNA and fixed Cas9 concentrations (15 nM); Individual data points shown (n = 2). gRNAs were annealed to a final concentration of 100 nM using excess tracrRNA or crRNA prior to Cas9 RNP complex formation. Graph showing the effect of increasing gRNA concentrations on (**c**) WAS or (**d**) WAS-OT3 DNA cleavage *in vitro* using WAS-RNA or WAS-BNA-3 crRNA and fixed Cas9 concentrations (15 nM); Individual data points shown (n = 2). Graph showing the effect of increasing Cas9 RNP complex concentration on (**e**) WAS or (**f**) WAS-OT3 DNA cleavage *in vitro* using WAS-RNA or WAS-BNA-3 crRNA; Individual data points shown (n = 2). Graph showing effect of increasing (**g**) WAS or (**h**) WAS-OT3 DNA target on cleavage *in vitro* using WAS-RNA or WAS-BNA-3 crRNA and Cas9 at fixed concentrations; Individual data points shown (n = 2).

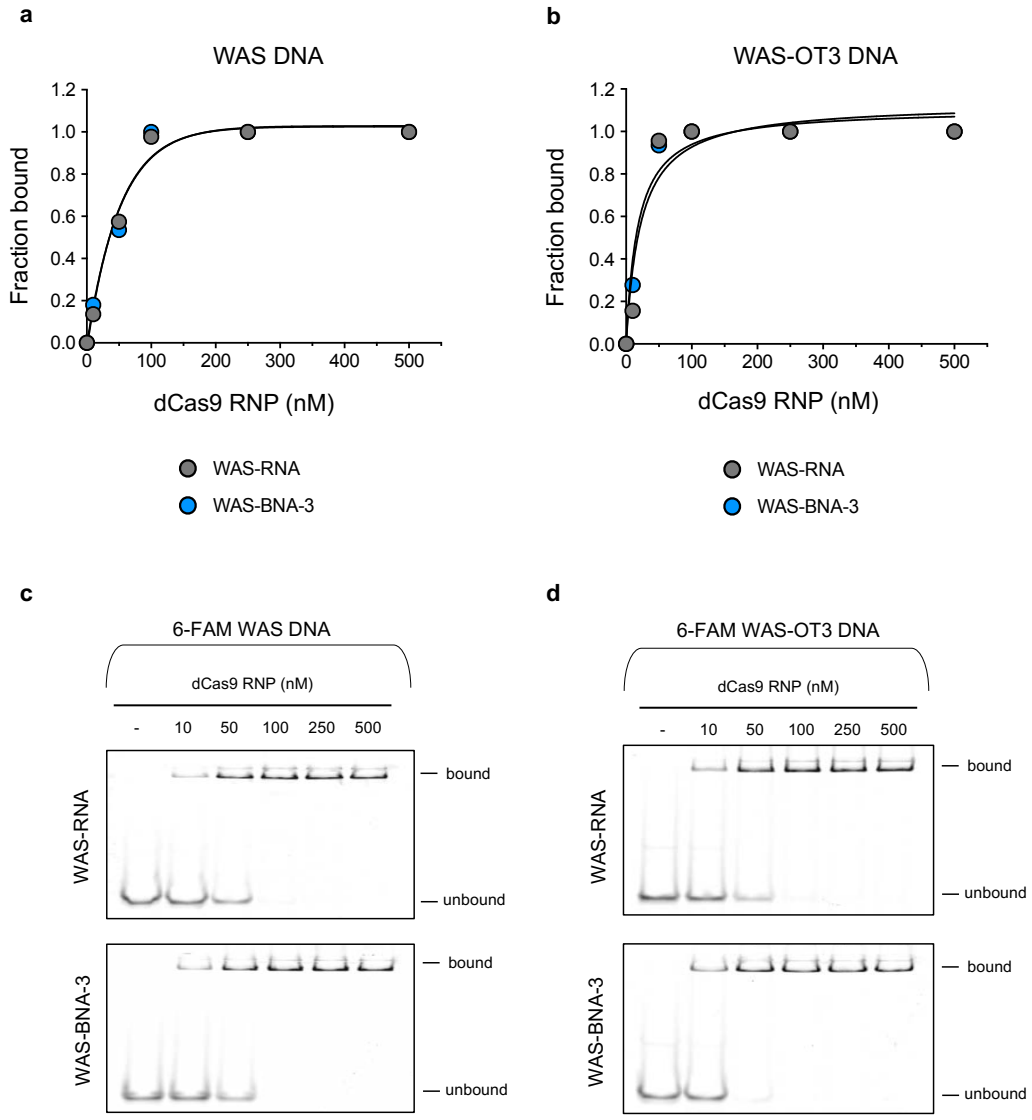


Figure 2.13. Effect of BNA^{NC}-modified crRNAs on the ability of nuclease-deficient Cas9 (dCas9) to bind DNA target sequences

Graph showing the effect of increasing dCas9 RNP concentration on the ability to bind (**a**) WAS or (**b**) WAS-OT3 DNA containing sequences using WAS-RNA (grey) or WAS-BNA-3 (blue) crRNA (n = 1). Corresponding gel showing the ability of dCas9 to bind (**c**) WAS or (**d**) WAS-OT3 DNA containing sequences using WAS-RNA (top) or WAS-BNA-3 (bottom) crRNAs at increasing concentrations. Substrate DNA concentration remained constant at 50 nM. For all binding experiments, the top band is dCas9-bound DNA, while the bottom band is unbound DNA. Quantification of bound percentages was determined using densitometry (ImageJ).

To study changes in crRNA-DNA hybridization, we first performed a melting temperature (T_m) analysis of WAS-RNA and WAS-BNA-3 annealed to single-strand on-target and off-target DNA templates, in the absence of Cas9. This experiment revealed slight differences in duplex stability, as T_m values for WAS-BNA-3 duplexes were higher than those of WAS-RNA (**Figure 2.14**).

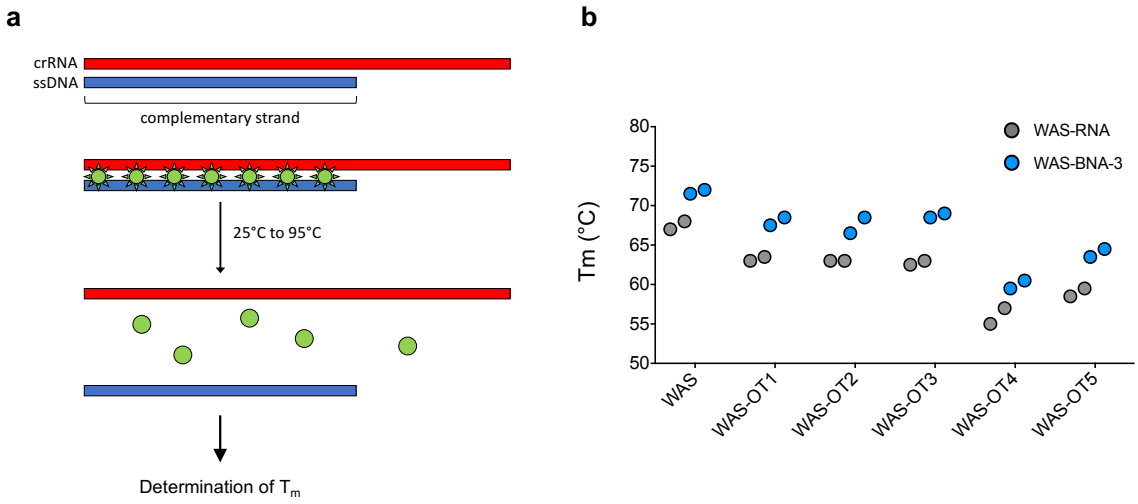


Figure 2.14. Effect of BNANC modifications on crRNA/DNA melting temperature
(**a**) Schematic diagram of assay used to determine melting temperature. crRNA (red) interacts with complementary ssDNA (blue) to form an RNA / DNA heteroduplex. SYBR Green I (green circle) intercalates and fluoresces during duplex formation. (**b**) Graph showing the experimentally determined melting temperatures for WAS on- and off-targets (listed in **Figure 2.1b**) using WAS-RNA (grey) or WAS-BNA-3 (blue) crRNA; Individual data points shown (n = 2).

Since the differences in T_m values between WAS-RNA and WAS-BNA-3 crRNA-DNA heteroduplexes were nearly equivalent for on-target and off-target sequences, we employed a previously described single-molecule fluorescence resonance energy transfer (smFRET) assay to monitor changes in hybridization between the crRNA and target DNA in the presence of Cas9. This assay employs a Cy5-labeled crRNA complexed to Cas9 and a Cy3-labeled DNA substrate immobilized on a quartz surface[232]. Briefly, changes in FRET occurring between the Cy3-Cy5 dye pair during R-loop formation represent the transitional dynamics between the partially zipped

("open"; low-FRET) and the high fully zipped ("zipped"; high-FRET) sub-conformations of the Cas9 complex, which correspond to an intermediate and cleavage-competent state, respectively (**Figure 2.15a**)[232]. We used this technology to measure the relative ratio of Cas9 complexes in each state using either WAS-RNA or WAS-BNA-3 in conjunction with the on-target sequence (WAS), or an off-target sequence (WAS-OT4). For the on-target sequence, we found that most molecules were populated in the zipped conformation for both WAS-RNA and WAS-BNA-3 crRNAs, while a drastic decrease in the proportion of molecules in the zipped conformation was observed for WAS-BNA-3, relative to WAS-RNA, when the off-target sequence was used (**Figure 2.15b**). This result is consistent with the specificity improvements observed using WAS-BNA-3 *in vitro* and in cells. To examine differences in population kinetics between complexes containing either WAS-RNA or WAS-BNA-3, we analyzed single-molecule time trajectories (**Figure 2.15c**). In accordance with the FRET histograms, trajectories for the on-target substrate were mainly docked in the zipped conformation regardless of BNA^{NC} incorporation (**Figure 2.16a**). For the off-target DNA, however, the majority of molecules exhibited repetitive transitions between the two sub-conformations (**Figure 2.16c**). We measured the dwell time in each state and found that the time spent in both the open and the zipped conformation was reduced by half using WAS-BNA-3 crRNA, compared to the unmodified crRNA (**Figure 2.15d**). This implies that BNA^{NC}-modified crRNAs accelerate the dynamic transitions on the off-target substrate, leading to a shortened dwell-time in the cleavage-competent zipped conformation and a reduction of off-target cleavage. Because these findings suggest that BNA^{NC}-modified crRNAs may alter Cas9 enzyme kinetics, we performed several time course experiments using an in vitro assay to compare rates of cleavage between WAS-BNA-3, WAS-BNA-2, and EMX1-BNA-5, and their unmodified crRNA counterparts. We found that on-target DNA cleavage was slower using BNA^{NC}-modified crRNAs

compared to the unmodified crRNAs in all instances where specificity was improved (**Figures 2.17-2.19**). Taken together, these results suggest that the specificity improvements imparted by BNA^{NC}-modified crRNAs likely stem from delayed reaction kinetics coupled to an impaired ability to form a productive zipped conformation, a prerequisite for DNA cleavage[232], on off-target sequences.

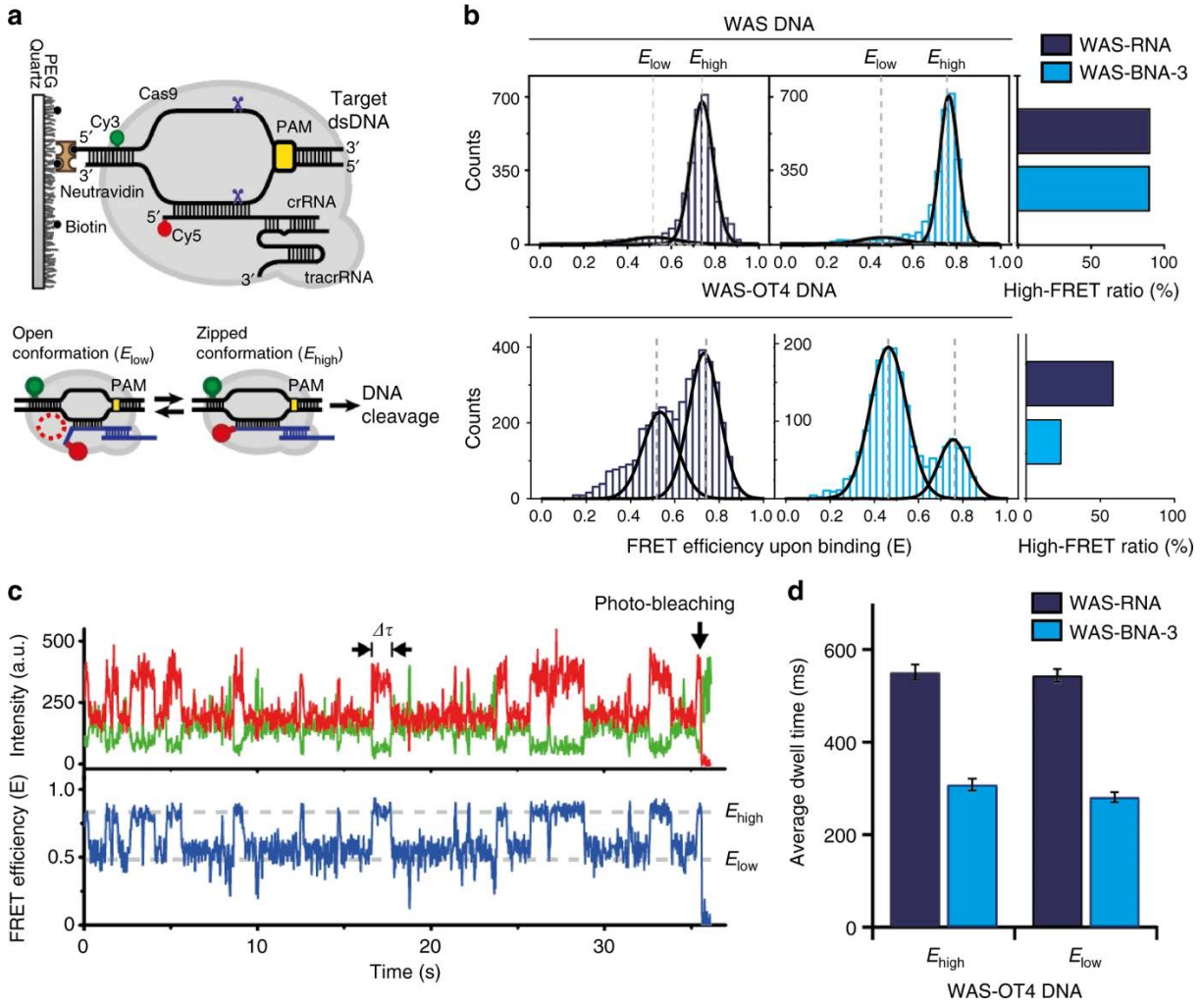


Figure 2.15. BNA^{NC} incorporation influences conformational transitions
(**a**) Schematic diagram for smFRET experiments showing a Cas9 RNP complex consisting of Cy5-labeled crRNA, tracrRNA, and Cas9, bound to a Cy3-labeled dsDNA immobilized on a quartz surface. (**b**) Histograms showing FRET efficiency after equilibration for the WAS DNA (upper) or WAS-OT4 DNA (lower) target sequences using WAS-RNA (dark blue) or WAS-BNA-3 (light blue) crRNAs; black curves represent Gaussian fits. (**c**) Time trace for Cas9 on WAS-OT4 DNA using WAS-BNA-3 crRNA indicating repetitive transitions between the open and zipped conformations. Dwell time in each conformation is indicated as $\Delta\tau$. (**d**) Comparison of Cas9 dwell times between WAS-RNA and WAS-BNA-3 crRNA using the WAS-OT4 DNA template; mean ± SD shown

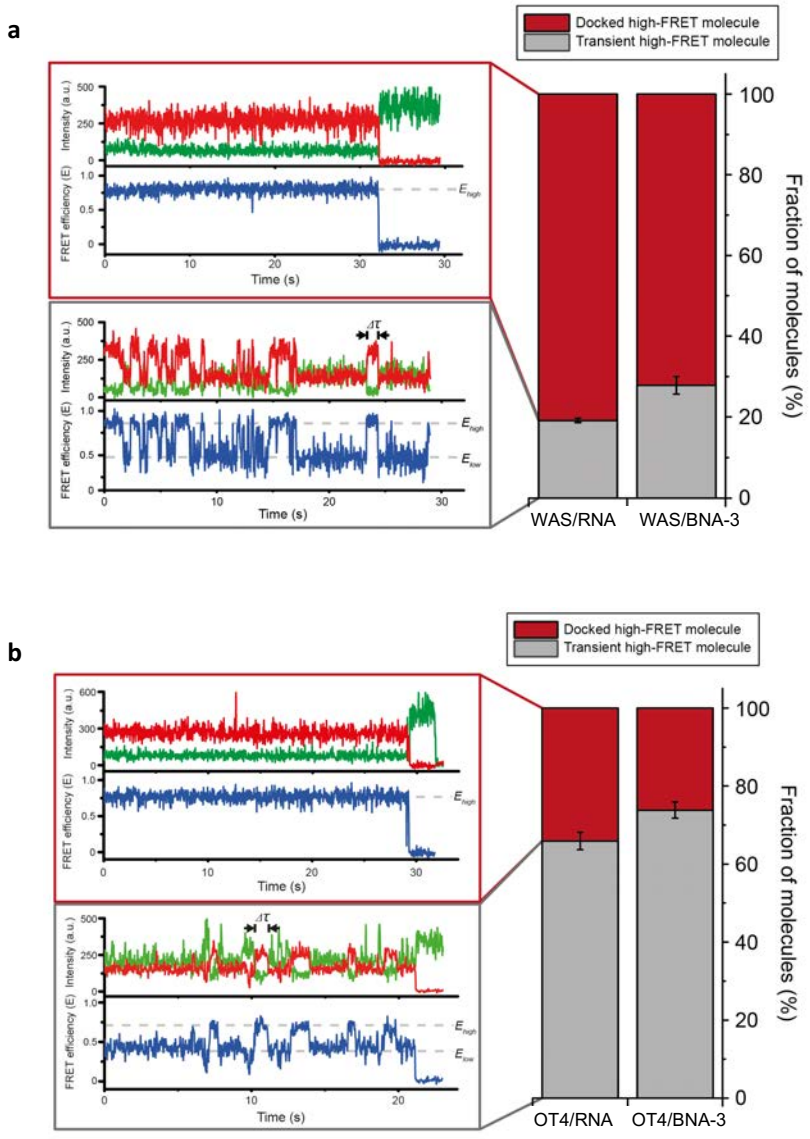


Figure 2.16. Kinetic analysis of time spent in open to zipped transition and docked states using single-molecule FRET

Fraction of molecules observed in either docked high-, or transient high-FRET states for unmodified and BNA^{NC}-modified crRNAs using (**a**) WAS or (**b**) WAS-OT4 DNA containing sequences. Representative time-trajectories reveal two distinct modes within the high-FRET state; docked mode (red) and transient mode (grey).

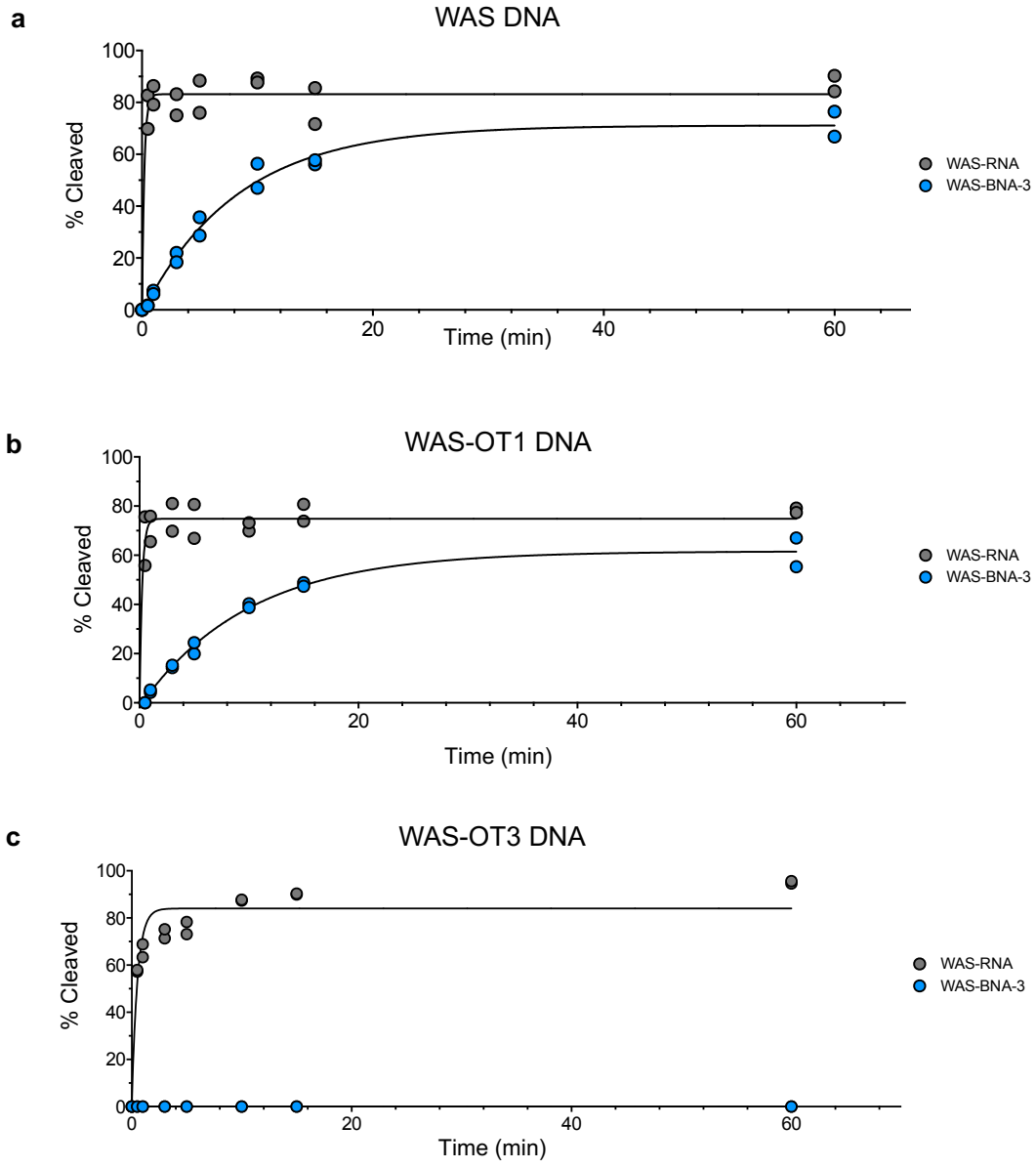


Figure 2.17. Cas9 *in vitro* cleavage kinetics using either WAS-RNA or WAS-BNA-3 crRNAs on several target sequences

Graph showing the *in vitro* cleavage of (**a**) WAS, (**b**) WAS-OT1 or (**c**) WAS-OT3 DNA containing sequences using either WAS-RNA (grey) or WAS-BNA-3 (blue) crRNA over time; Individual data points shown ($n = 2$). The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments.

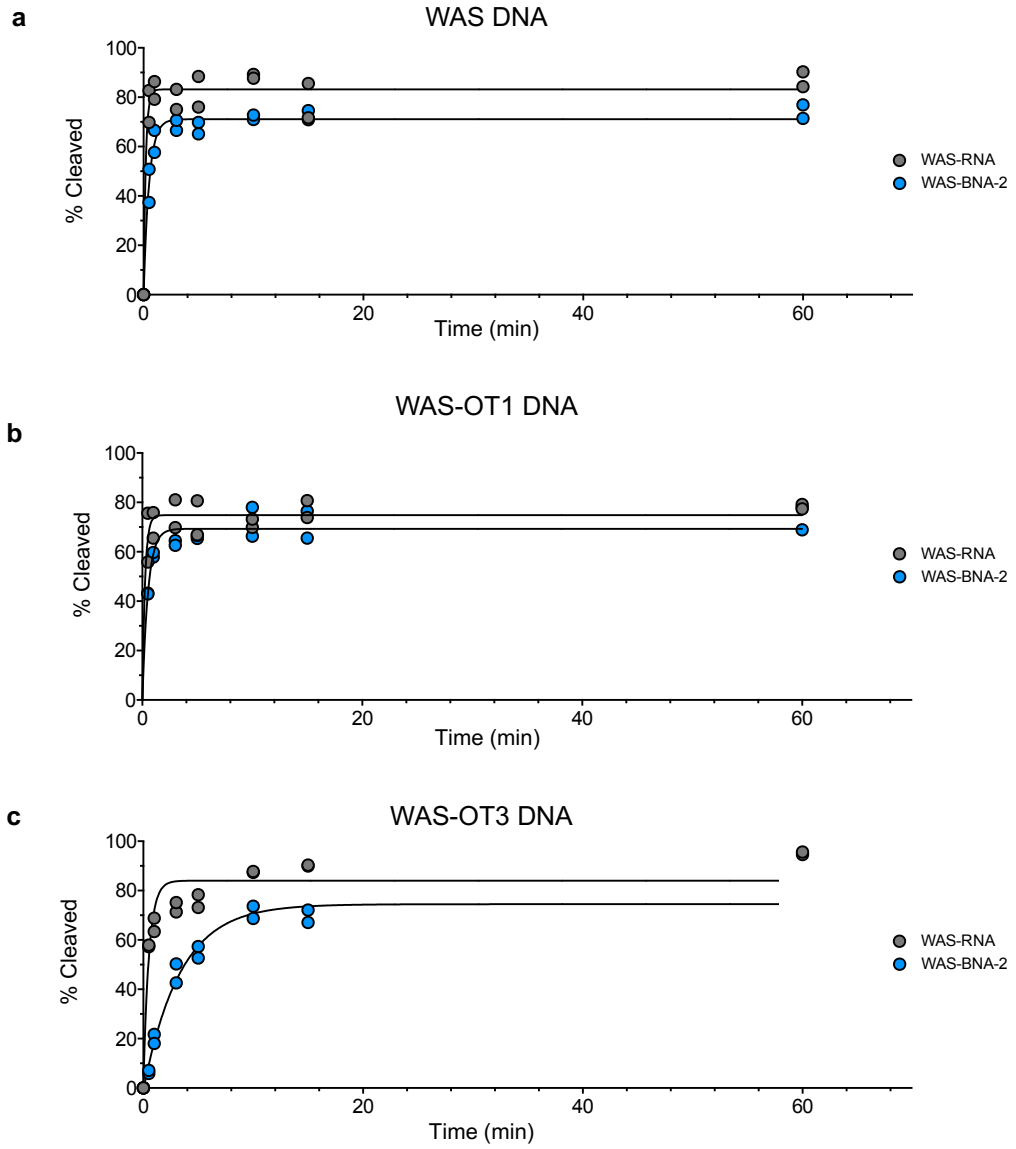


Figure 2.18. Cas9 *in vitro* cleavage kinetics using either WAS-RNA or WAS-BNA-2 crRNAs on several target sequences

Graph showing the *in vitro* cleavage of (**a**) WAS, (**b**) WAS-OT1 or (**c**) WAS-OT3 DNA containing sequences using either WAS-RNA (grey) or WAS-BNA-2 (blue) crRNA over time; Individual data points shown ($n = 2$). The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments. WAS-RNA values from **Figure 2.17** are shown again as reference.

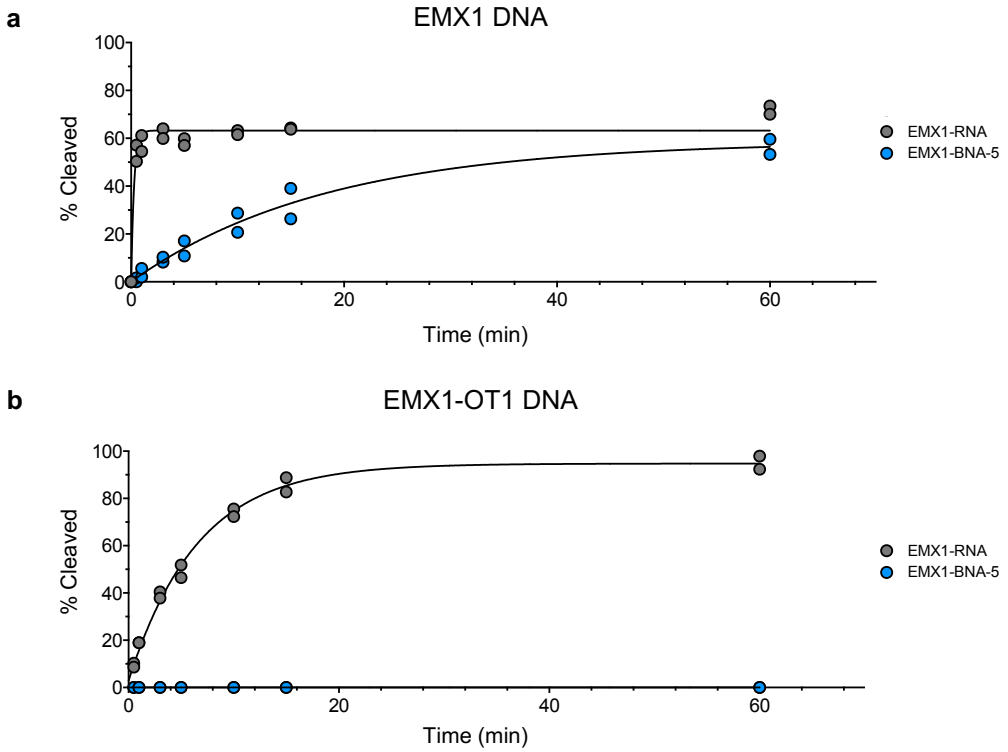


Figure 2.19. Cas9 *in vitro* cleavage kinetics using either EMX1-RNA or EMX1-BNA-5 crRNAs on several target sequences
Graph showing the *in vitro* cleavage of (**a**) EMX1 or (**b**) EMX1-OT1 DNA containing sequences using either EMX1-RNA (grey) or EMX1-BNA-5 (blue) crRNA over time; Individual data points shown ($n = 2$). The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments.

2.3.6 Incorporation of LNAs into crRNAs improves Cas9 specificity

To examine if the Cas9 specificity improvements imparted by BNA^{NC}-modified crRNAs are conferred by a general property of BNAs, or if they are unique to these particular nucleotides, we substituted BNA^{NC} modifications in three of the WAS and EMX1 crRNAs (WAS-BNA-3/5/6 and EMX1-BNA-5/6/7) with LNA modifications. Like BNA^{NC}-modified crRNAs, WAS, and EMX1-directed LNA-modified crRNAs (WAS-LNA-3/5/6 and EMX1-LNA-5/6/7) induced less off-target cleavage by Cas9 compared to unmodified crRNAs *in vitro* (**Figure 2.20**, **Table 2.8**). In addition, high-throughput specificity profiling revealed that Cas9 specificity is broadly improved using WAS-LNA-3 and EMX1-LNA-5, compared to the corresponding unmodified crRNAs

(**Figure 2.21**, **2.22** and **Table 2.4**). However, while specificity enhancements between EMX1-BNA-5 and EMX1-LNA-5 were comparable, *in vitro* DNA cleavage by Cas9 was substantially more promiscuous with WAS-LNA-3 than with WAS-BNA-3 (**Figure 2.20**-**2.22**).

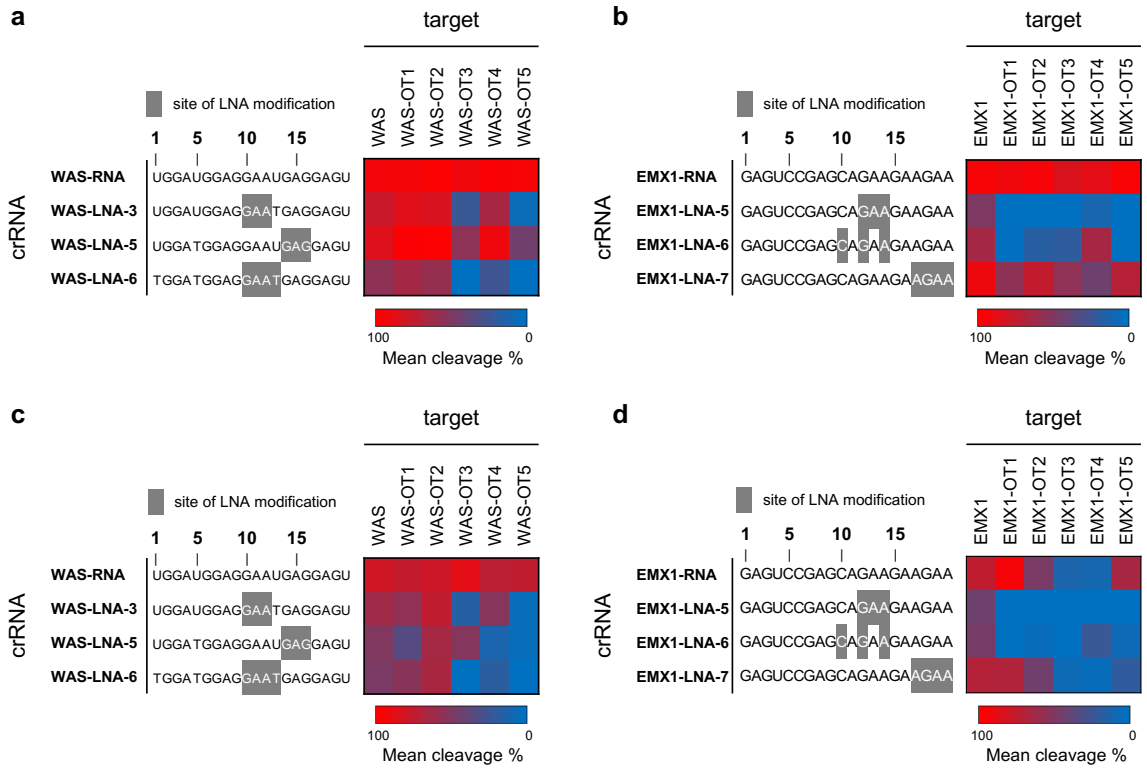


Figure 2.20. Incorporation of LNA into crRNAs improves Cas9 cleavage specificity *in vitro*

Heat map showing cleavage specificity for unmodified and 3 LNA-modified crRNAs towards either (**a**) WAS or (**b**) EMX1 on- and off-target sequences (as listed in **Figure 2.1b, c**) Mean shown (n = 2). LNA modifications are indicated in grey. Targets which were highly cleaved *in vitro* are indicated by red, while targets which were not cleaved are indicated by blue. The molar ratio of Cas9 RNP complex to target DNA was 30:1 for these experiments. Values used to generate heatmaps are shown in **Table 2.8**. Heatmap showing cleavage specificity for unmodified and 3 LNA-modified crRNAs towards either (**c**) WAS or (**d**) EMX1 on- and off-target sequences; Mean shown (n = 2). The molar ratio of Cas9 RNP complex to target DNA was 3:1 for these experiments. Values used to generate heatmaps are shown in **Table 2.8**. WAS-RNA and EMX1-RNA values from **Figure 2.1** and **Figure 2.2** are shown again as reference.

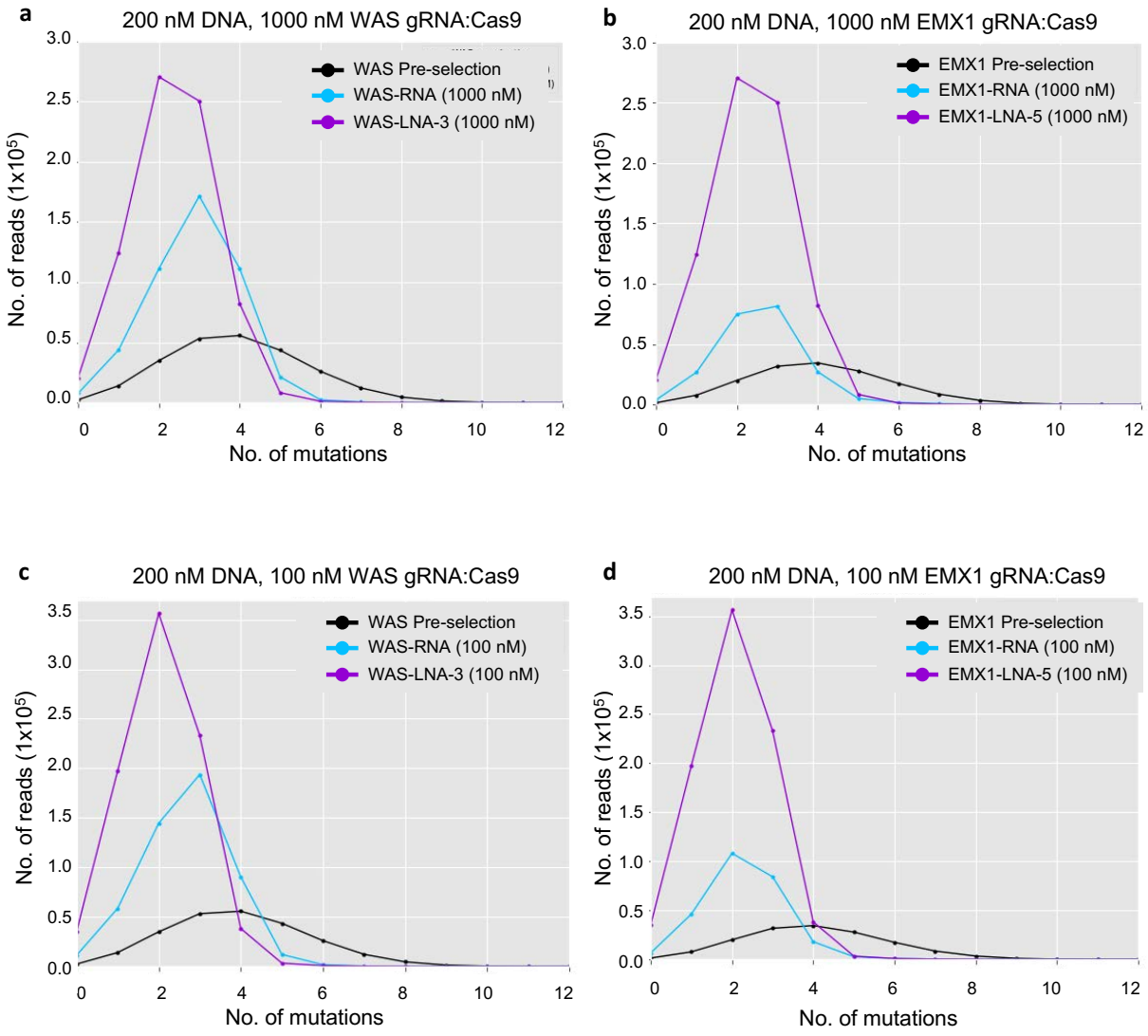


Figure 2.21. Distribution of mutations in pre- and post-selection libraries following in vitro library selection with unmodified or LNA-modified crRNAs

In vitro selections were performed using 200 nM pre-selection library with 1000 nM Cas9 RNP complex targeting either (**a**) WAS or (**b**) EMX1. *In vitro* selections were performed using 200 nM pre-selection library with 100 nM Cas9 RNP complex targeting either (**c**) WAS or (**d**) EMX1. Cas9 RNP complexes were pre-assembled using unmodified or LNA-modified crRNAs prior to DNA digestion. Distributions of mutations within pre-selection (black) and post-selection (colored) libraries are shown. Mutations are counted from the 20-base-pairs comprising the Cas9 target site, as well as the 2-base-pair PAM (N of NGG is excluded). Pre-selection, WAS-RNA and EMX1-RNA values from **Figure 2.4** are shown again as reference.

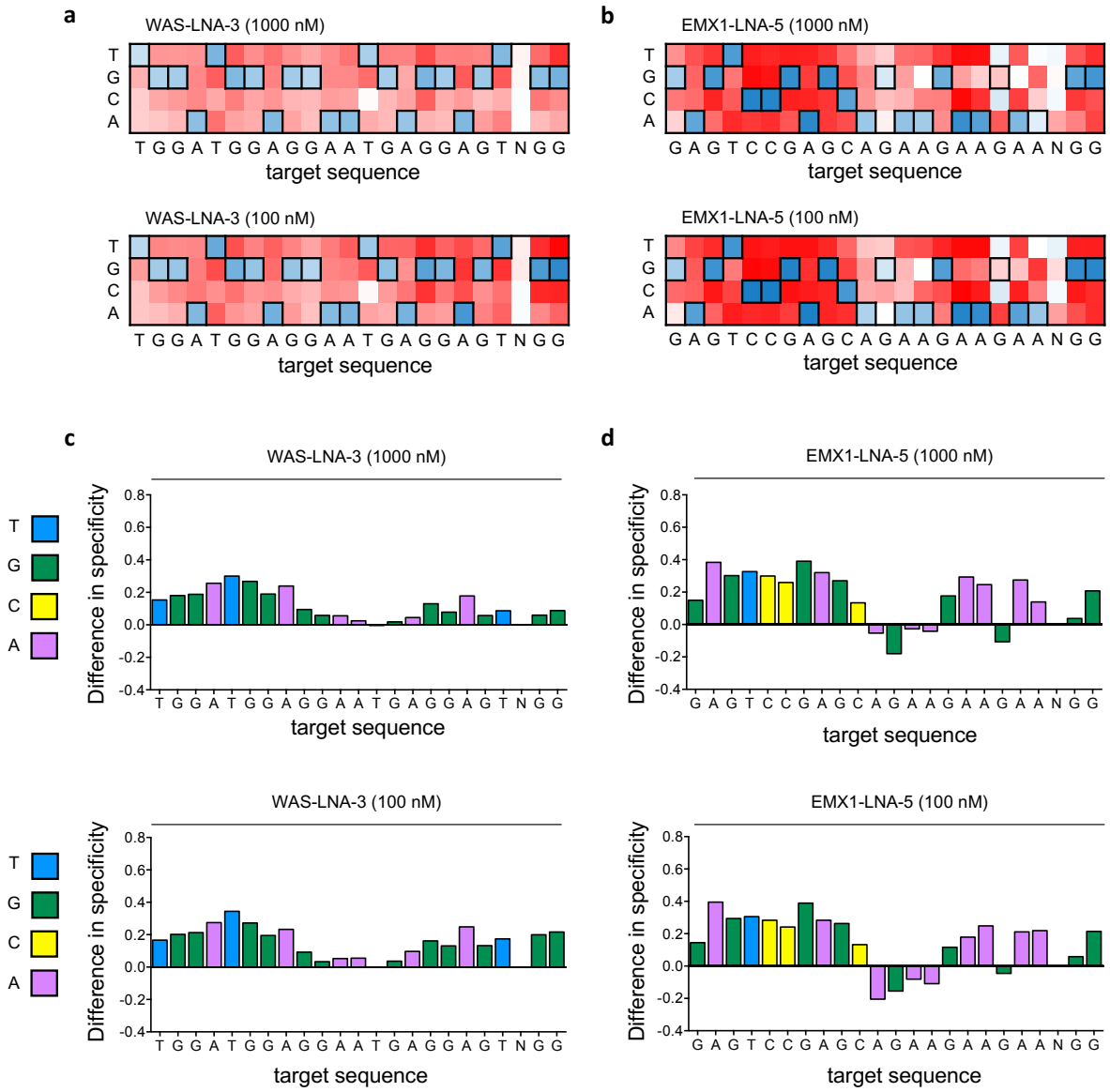


Figure 2.22. ***In vitro*** **specificity profiling results for LNA-modified crRNAs**
Heat maps showing DNA cleavage specificity scores across $>10^{12}$ off-target sequences for LNA-modified crRNAs targeting (**a**) WAS or (**b**) EMX1 using 1000 nM Cas9 RNP complex (top) or 100 nM Cas9 RNP complex (bottom). Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of -1.0 (dark red) correspond to 100% enrichment against a specific base-pair at a specific position. Black boxes denote the intended target nucleotides. Bar graph showing the quantitative difference in specificity score at each position in the 20 base-pair target site and 2 base-pair PAM (N of NGG excluded), between the unmodified (shown in **Figure 2.5** and **Figure 2.6**) and LNA-modified crRNA for (**c**) WAS or (**d**) EMX1 target sequences using 1000 and 100 nM Cas9 RNP complex. A score of zero indicates no change in specificity. Difference in specificity was calculated as, specificity $\text{score}_{\text{BNA}}^{\text{NC}}$ – specificity $\text{score}_{\text{RNA}}$.

Table 2.8. *In vitro* cleavage assay data used to generate Figure 2.20

		Target					
	crRNA	**WAS**	**WAS-OT1**	**WAS-OT2**	**WAS-OT3**	**WAS-OT4**	**WAS-OT5**
150 nM RNP	**WAS-LNA-3**	79 ± 1.5	87 ± 0.92	85 ± 0.04	22 ± 1.2	66 ± 1.4	4.5 ± 0.61
	WAS-LNA-5	87 ± 7.3	98 ± 0.08	99 ± 0.03	55 ± 9.3	93 ± 3.8	45 ± 7.3
	WAS-LNA-6	56 ± 6.4	65 ± 8.8	59 ± 8.0	0.0 ± 0.0	24 ± 4.6	0.0 ± 0.0
15 nM RNP	**WAS-LNA-3**	62 ± 0.17	57 ± 11	75 ± 0.62	15 ± 1.0	53 ± 3.3	2.5 ± 0.70
	WAS-LNA-5	51 ± 5.1	34 ± 7.4	65 ± 17	52 ± 23	10 ± 10	3.2 ± 3.2
	WAS-LNA-6	48 ± 7.9	54 ± 8.5	67 ± 5.3	0.0 ± 0.0	17 ± 7.6	0.0 ± 0.0

		Target					
	crRNA	**EMX1**	**EMX1-OT1**	**EMX1-OT2**	**EMX1-OT3**	**EMX1-OT4**	**EMX1-OT5**
150 nM RNP	**EMX1-LNA-5**	**49 ± 13**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**9.7 ± 9.7**	**0.0 ± 0.0**
	EMX1-LNA-6	**64 ± 0.20**	**0.5 ± 0.5**	**16 ± 0.61**	**18 ± 1.1**	**65 ± 1.0**	**0.0 ± 0.0**
	EMX1-LNA-7	**93 ± 0.37**	**56 ± 4.7**	**76 ± 1.3**	**56 ± 0.91**	**43 ± 1.2**	**71 ± 3.7**
15 nM RNP	**EMX1-LNA-5**	**43 ± 6.7**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**
	EMX1-LNA-6	**47 ± 1.5**	**0.0 ± 0.0**	**2.9 ± 2.9**	**0.0 ± 0.0**	**22 ± 0.44**	**5.7 ± 0.11**
	EMX1-LNA-7	**70 ± 15**	**70 ± 3.4**	**44 ± 1.1**	**5.3 ± 0.07**	**3.0 ± 0.71**	**19 ± 1.8**

Values showing *in vitro* cleavage specificity for unmodified crRNA and 3 LNA-modified crRNAs towards either WAS (top) or EMX1 (bottom) on- and off-target sequences (as listed **Figure 2.1b**, **c**); Mean ± SE (n = 2). Experiments were performed using 150 nM or 15 nM Cas9 RNP and 5 nM DNA.

These results were mirrored in HeLa and U2OS cells. For example, while WAS-LNA-3 dramatically reduced cleavage of WAS-OT5 by >23,000-fold vs. WAS-RNA, it had lesser or no effect on WAS-OT2, WAS-OT3, and WAS-OT4, in stark contrast to WAS-BNA-3 (**Table 2.5**, **Table 2.7**, and **Table 2.9**). The distribution of indel size and location produced by Cas9 DNA cleavage was unaffected by LNA incorporation (Figures **2.23**-**2.25**). Moreover, both Cas9 activity and enzyme kinetics were reasonably similar (activity was slightly reduced) using either WAS-LNA-3 or WAS-RNA (**Figure 2.26**). Overall, these data establish that LNA substitutions at

specific locations in crRNAs improve Cas9 specificity *in vitro* and in cells, although to a lesser extent than the corresponding BNA^{NC} substitutions.

Table 2.9. Cellular modification rates induced by unmodified or LNA-modified crRNAs targeting *WAS* or *EMX1*

Cell Type	U2OS-Cas9	U2OS-Cas9	U2OS-Cas9	HeLa-Cas9	HeLa-Cas9	HeLa-Cas9
Treatment	**No gRNA**	**WAS-RNA**	**WAS-LNA-3**	**No gRNA**	**WAS-RNA**	**WAS-LNA-3**
WAS	< 0.003	58.392	35.859	< 0.003	56.020	34.812
WAS-OT1	0.003	17.548	27.750	0.004	18.135	27.041
WAS-OT2	0.092	59.553	35.246	0.120	53.400	30.204
WAS-OT3	< 0.003	73.142	0.090	< 0.003	70.859	0.154
WAS-OT4	< 0.003	36.506	0.362	< 0.003	33.151	0.383
WAS-OT5	< 0.003	75.239	0.032	< 0.003	71.119	< 0.003

Cell Type	U2OS-Cas9	U2OS-Cas9	U2OS-Cas9	HeLa-Cas9	HeLa-Cas9	HeLa-Cas9
Treatment	**No gRNA**	**EMX1-RNA**	**EMX1-LNA-5**	**No gRNA**	**EMX1-RNA**	**EMX1-LNA-5**
EMX1	0.004	69.181	46.813	< 0.003	69.118	48.273
EMX1-OT1	0.005	12.233	< 0.003	0.009	11.677	< 0.003
EMX1-OT2	0.031	1.007	0.013	0.025	0.607	0.022
EMX1-OT3	0.068	< 0.003	0.015	0.005	< 0.003	< 0.003
EMX1-OT4	0.073	< 0.003	< 0.003	0.042	0.016	0.036
EMX1-OT5	< 0.003	0.113	0.059	0.060	0.141	< 0.003

Table summarizing modification frequencies of on- and off-target sequences in U2OS-Cas9 and HeLa-Cas9 cells using either unmodified or LNA-modified crRNAs targeting *WAS* or *EMX1*, as determined by high-throughput sequencing. Modification frequencies were calculated by dividing the number of sequences bearing insertions or deletions (indels) in the target site by the total number of sequences. Mock transfections lacking gRNA were used as controls (See **Table 2.7** for additional data). No gRNA, WAS-RNA and EMX1-RNA values from **Tables 2.1** and **2.2** are shown again as reference.

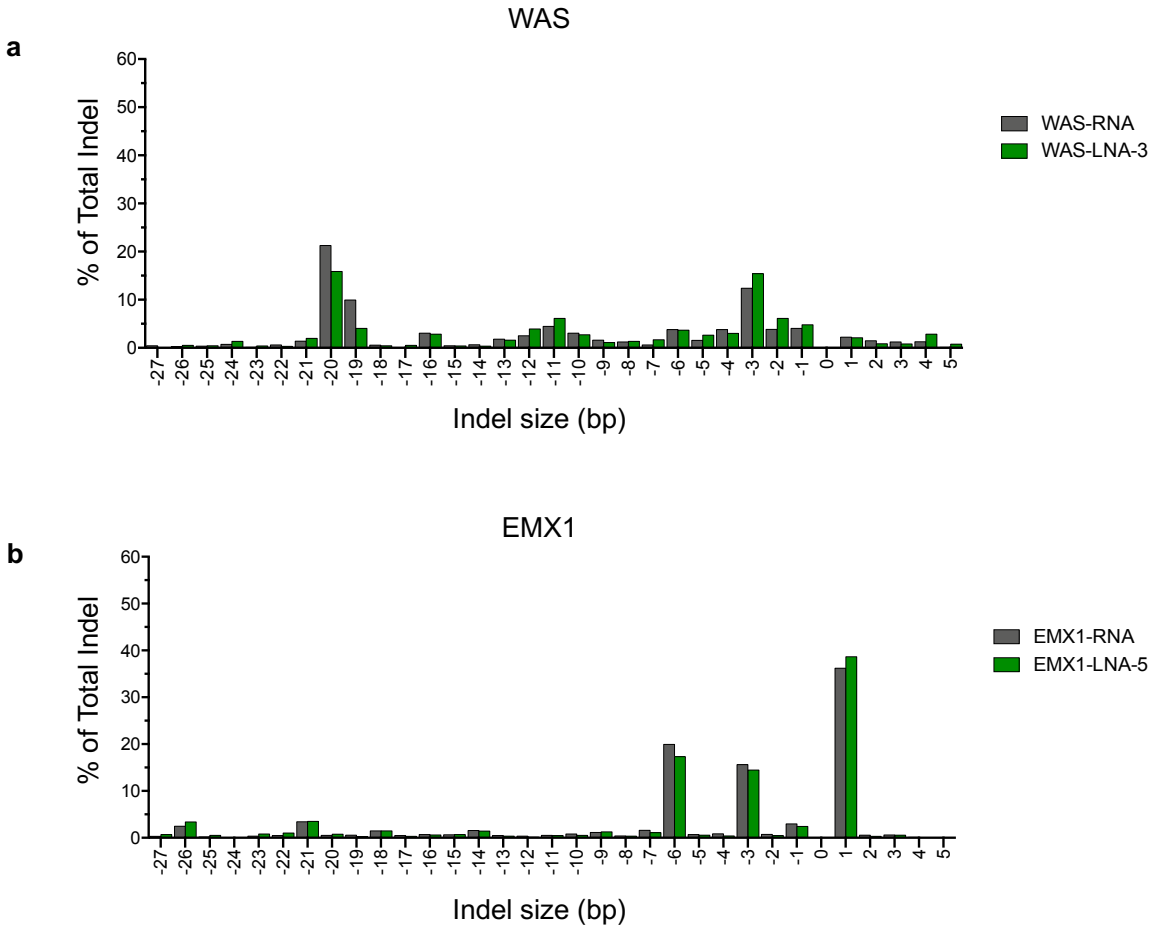


Figure 2.23. Comparison of indel size resulting from Cas9 DNA cleavage with unmodified or LNA-modified crRNAs

Distributions of indel sizes are shown for either (**a**) WAS or (**b**) EMX1 using unmodified or LNA-modified crRNAs in U2OS-Cas9 cells. Distribution of indel sizes is based on sequences obtained following high-throughput sequencing of transfected cells. The number of reads for each indel size was normalized to the total number of indel reads. Only indels ranging from -27 bp (27 bp deletion) to +5 bp (5 bp insertion) are shown. WAS-RNA and EMX1-RNA values from **Figure 2.9** are shown again as reference.

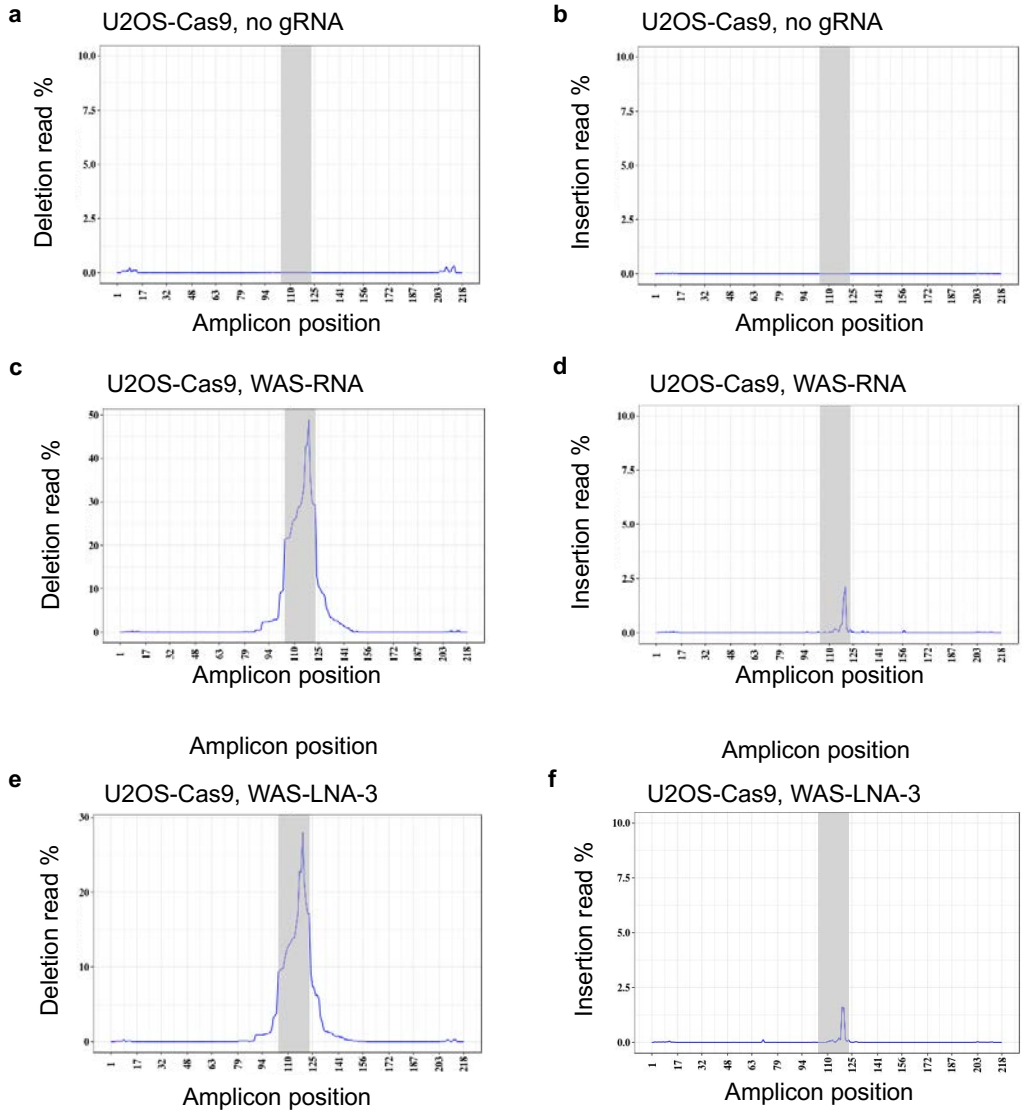


Figure 2.24. Distribution of indel location using unmodified or LNA-modified crRNAs on WAS target

Distribution frequencies of insertions (right) and deletions (left) across the *WAS* target site in U2OS-Cas9 cells transfected with no gRNA (**a**, **b**), WAS-RNA (**c**, **d**) or WAS-LNA-3 (**e**, **f**) are shown. Grey coloring indicates the location of the genomic CRISPR target site along the amplicon. Distribution of indel positions is based on sequences obtained following high-throughput sequencing of transfected cells. No gRNA and WAS-RNA values from **Figure 2.10** are shown again as reference.

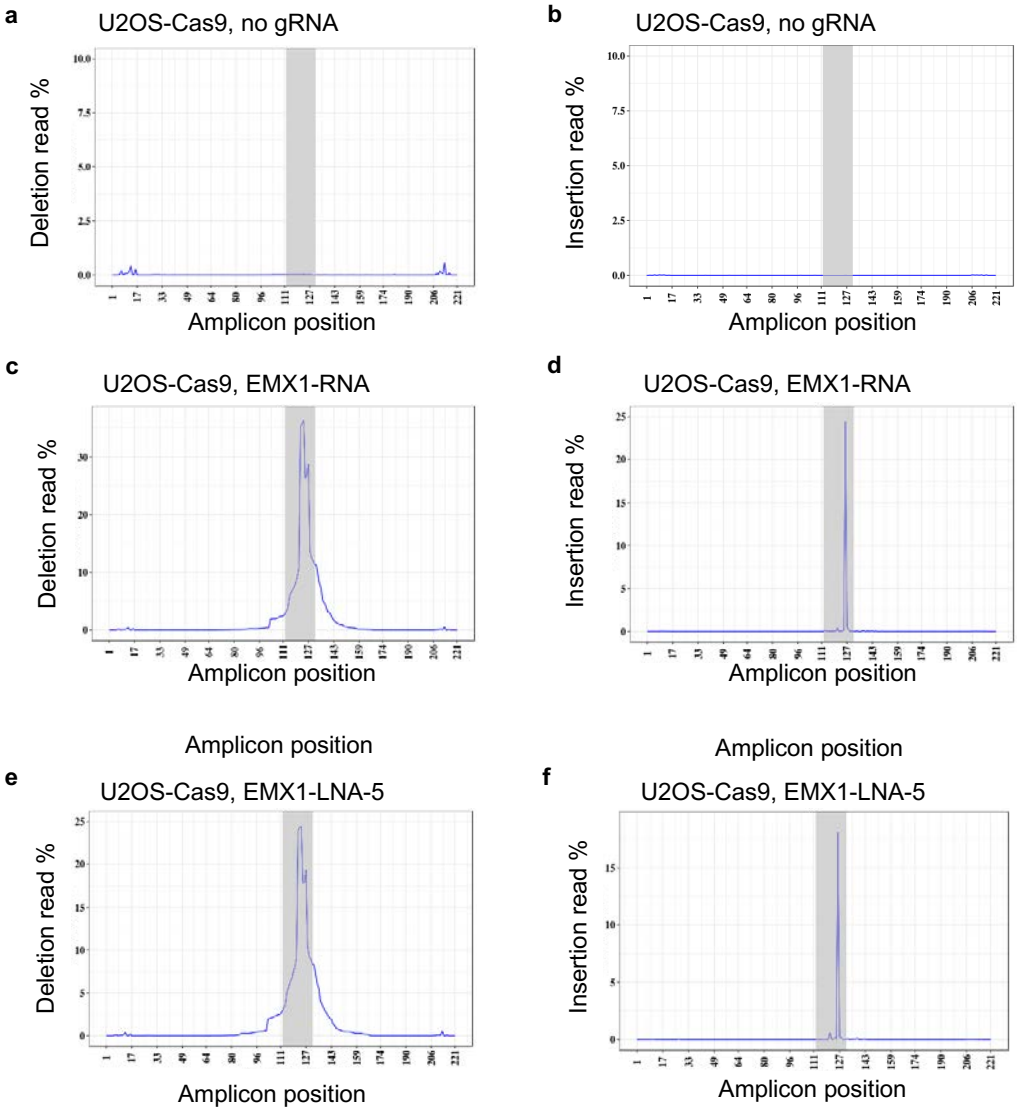


Figure 2.25. Distribution of indel location using unmodified or LNA-modified crRNAs on EMX1 target

Distribution frequencies of insertions (right) and deletions (left) across the *EMX1* target site in U2OS-Cas9 cells transfected with no gRNA (**a, b**), EMX1-RNA (**c**, **d**) or EMX1-LNA-5 (**e**, **f**) are shown. Grey coloring indicates the location of the genomic CRISPR target site along the amplicon. Distribution of indel positions is based on sequences obtained following high-throughput sequencing of transfected cells. No gRNA and EMX1-RNA values from **Figure 2.11** are shown again as reference.

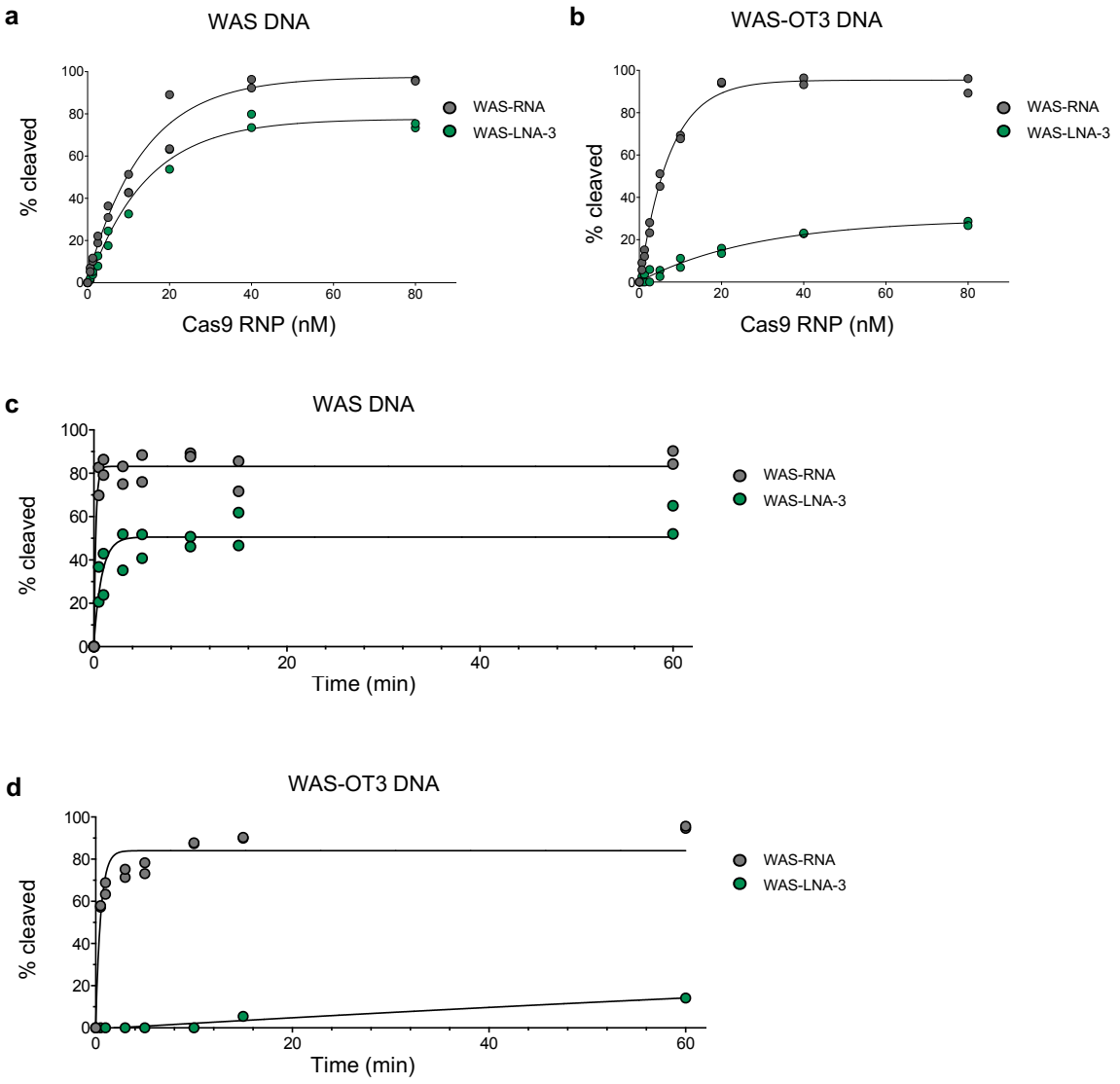


Figure 2.26. Effect of LNA-modified crRNAs on activity and kinetics of Cas9 *in vitro*
Graph showing the effect of increasing Cas9 RNP complex concentration on WAS (**a**) or WAS-OT3 (**b**) DNA cleavage using WAS-RNA or WAS-LNA-3 crRNA; Individual data points are shown (n = 2). Graph showing the *in vitro* cleavage of (**c**) WAS or (**d**) WAS-OT3 DNA containing sequences using either WAS-RNA or WAS-LNA-3 crRNAs over time; Individual data points are shown (n =2). The molar ratio of Cas9 RNP complex to target DNA was 3:1 for **c** and **d**. WAS-RNA values from **Figure 2.12** and **Figure 2.17** are shown again as reference.

2.4 Discussion

Crystal structures have revealed that Cas9 forms seven hydrogen bonds with ribose 2′ hydroxyl groups in crRNA nucleotides, four of which occur within the guide sequence[46, 202]. Remarkably, RNA nucleotides at all locations within the guide segment can be recoded as DNA nucleotides, with the exception of position 16[202]. Moreover, up to 70% of ribose sugars in sgRNAs can be chemically modified with 2′ methoxy or 2′ fluoro moieties without disruption of Cas9 activity[191]. Consistent with these findings, our results demonstrate that Cas9 is highly tolerant toward incorporation of small stretches of 1–4 BNA^{NC} and LNA modifications throughout the crRNA guide segment. While a few of the modified crRNAs we generated showed low activity *in vitro* and in cells, such as EMX1-BNA-2 and EMX1-BNA-9 (**Figure 2.1**, **Figure 2.2** and **Figure 2.8**), we speculate that this was caused by mis-folding due to stabilization of secondary structures by BNA^{NC} modifications[220].

Surprisingly, our results indicate that incorporation of BNA^{NC} or LNA nucleotides within the central region (positions 10–14) of crRNAs substantially increases specificity in the PAM-proximal and PAM-distal regions (**Figures 2.1**-**2.6** and **2.20**-**2.22**). This is distinct from previous work showing that inclusion of LNAs in isolated DNA–DNA duplexes enhances mismatch discrimination locally[220]. Examination of these central nucleotides in the Cas9 complex crystal structure reveals that they do not make robust contacts with the Cas9 enzyme (**Figure 2.27**)[54]. Therefore, we speculate that BNA^{NC} incorporation alters how crRNAs hybridize with the target DNA sequence. A conformational mechanism is supported by our biophysical data showing that BNA^{NC} incorporation facilitates repetitive transitions of the Cas9 complex between the open and zipped conformations on off-target sequences (**Figures 2.15** and **2.16**) possible due to the

destabilization of either states. Based on our T_m experiments (**Figure 2.14**), we theorize that BNA^{NC} or LNA substitutions stabilize stacking to their corresponding DNA bases (increased relative to RNA), but destabilize hybridization in adjacent positions, in the context of Cas9. Structural studies have shown that the crRNA "seed" region, positions ~9–20 in the crRNA[55], is bound by Cas9 as a pre-ordered A-form helix, while the PAM-distal portion of the crRNA is maintained in a disordered state[55]. Since our most specific BNA^{NC} and LNA-modified crRNAs contain substitutions in positions 10–14, which overlap with the terminal bases in the seed sequence, it is possible that they promote an extended A-form structure throughout the entirety of the crRNA that alters DNA target hybridization. This is consistent with our observations that BNA^{NC} or LNA substitutions occurring early in the seed sequence (near the PAM), or distal to it, have less effect on overall specificity (**Figures 2.1**-**2.6** and **2.20**-**2.22**).

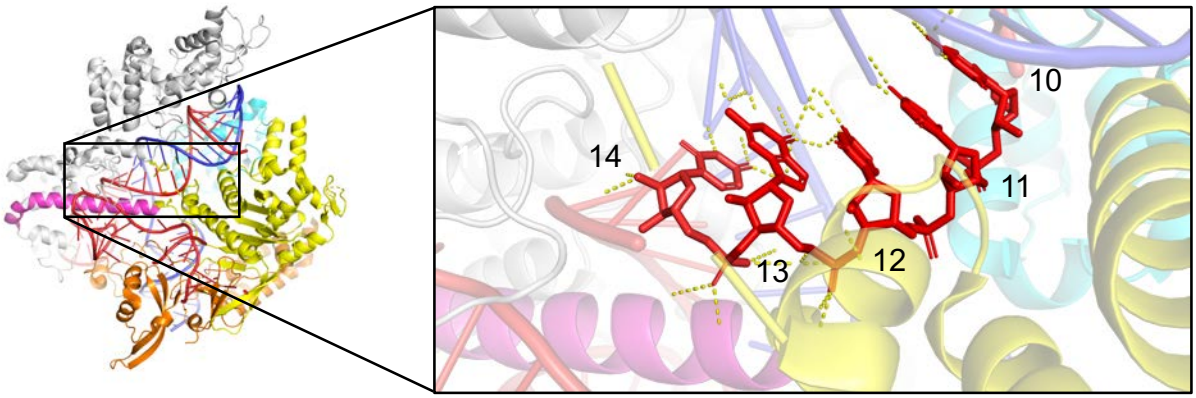


Figure 2.27. Model of Cas9 structure highlighting interactions between crRNA nucleotides 10-14 and Cas9
Pymol model of Cas9 in complex with gRNA (red) and target DNA (blue) (PDB: 4UN3) with domains colored (RuvC – yellow, Rec I and Rec II – grey, bridge helix – magenta, HNH – cyan, and PAM interacting – orange) (left). Contacts between nucleotides 10-14 of the crRNA and Cas9 and highlighted in the magnified image (right). Yellow dashed lines indicate predicted hydrogen bonds.

Our data indicates that incorporation of BNA^{NC} nucleotides into crRNAs is generally more effective than incorporation of LNA nucleotides at the same positions for improving Cas9 cleavage specificity (**Figures 2.1**, **2.5**, **2.8** and **2.20**-**2.25** and **Tables 2.5**, **2.6** and **2.9**) Since LNA bases are in fact more conformationally restricted then BNA^{NC} bases[223], it is unlikely that the specificity improvements rely solely on the degree of conformational constraint of the nucleic acid. Rather, our results suggest that other features of BNA^{NC} bases, such as their large steric bulk or chemical substituents likely contribute to improving mismatch discrimination. Previous work has reported that the nitrogen atom within BNA^{NC} bases can directly influence repulsion between negatively charged phosphate backbones on opposing DNA strands[223]. Furthermore, the increased steric bulk of BNA^{NC} substitutions vs. LNA substitutions in crRNAs could magnify the relevant conformational changes that improve specificity. Subsequent crystallographic studies will be needed to elucidate these details.

Tight target engagement is required for Cas9 DNA cleavage in cells[202]. In agreement with our *in vitro* data, we found that incorporation of BNA^{NC} and LNA nucleotides at central positions in crRNAs dramatically improves Cas9 cleavage specificity in cells. However, we also noted that the Cas9 on-target modification frequency of BNA^{NC}-modified crRNAs in cells was ~2–3-fold lower than their unmodified counterparts (**Tables 2.5**-**2.7**), in contrast to the more equal activity observed *in vitro* (**Figures 2.1** and **2.2**). This discrepancy is likely due to differences in reaction kinetics that manifest more apparently in a cellular context. For example, because Cas9 on-target cleavage is significantly delayed using BNA^{NC}-modified crRNAs (**Figure 2.17**) the probability of the Cas9 complex being ejected from DNA by cellular factors prior to inducing cleavage would be increased. This assertion is supported by our observation that LNA-modified crRNAs, whose

kinetics are less altered *in vitro* (**Figure 2.26**), induce on-target modification rates in cells closer to their unmodified crRNA counterparts (**Tables 2.7** and **2.9**). A recent study proposed that crRNA/DNA target hybridization kinetics plays a key role in Cas9 off-target discrimination[233], which could explain the enhanced specificity of BNA^{NC} substitutions compared to LNA substitutions. This assertion could be evaluated in future studies by using a protein evolution platform such as DNA-binding phage-assisted continuous evolution[13] to evolve Cas9 variants that bypass the BNA^{NC}-induced kinetic block.

Here, we show that incorporation of BNA^{NC} and LNA bases at specific positions within crRNAs broadly improves Cas9 DNA cleavage specificity *in vitro* and in cells. Furthermore, we show that these modified crRNAs enhance specificity by impairing the formation of the stable "zipped" conformation during hybridization to off-target sequences. Overall, these findings unveil a strategy for improving the specificity of the CRISPR-Cas9 system and illustrate the application of recently developed synthetic nucleic acid technologies to solving problems in enzyme specificity. We anticipate that these findings will directly contribute to the ongoing goal of improving the specificity and safety of genome-editing agents for a wide variety of experimental and clinical applications.

2.5 Methods

2.5.1 Chemical reagents and oligonucleotides

All chemicals were purchased from Sigma-Aldrich. DNA oligonucleotides and tracrRNA were purchased from Integrated DNA Technologies (IDT). Unmodified crRNAs and crRNAs containing BNAs (BNA^{NC}[N-Me]) were obtained from BioSynthesis Inc., while crRNAs

containing LNAs were purchased from Exiqon. eSpCas9 endonuclease was purchased from Sigma-Aldrich. Sequences of crRNAs and tracrRNAs used in this study are listed in **Table 2.10**. Sequences of DNA oligonucleotide used in this study are listed in **Table 2.11**.

2.5.2 Cloning and plasmid construction

Plasmid templates for *in vitro* cleavage assays were generated through ligation of inserts into *HindIII* and *XbaI* double-digested pUC19 (ThermoFisher). DNA encoding Cas9 target sites was purchased either as a gBlock from IDT (for genomic on-targets and off-targets), or as ssDNA for off-target sequences with single-nucleotide mismatches, as listed in **Table 2.11**. Forward and reverse ssDNA oligonucleotides containing single-nucleotide mismatch target sites were heated to 95 °C for 5 min, then cooled to 25 °C over the course of 1 h to assemble dsDNA prior to ligation. pET-NLS-Cas9-6xHis was purchased from Addgene (#62934) and used to express Cas9 protein for all in vitro cleavage assay experiments. pET-NLS-dCas9-6xHis (D10A/H840A) was generated using the Q5 Site Directed Mutagenesis Kit (NEB), according to the manufacturer's instructions.

2.5.3 Expression and purification of *S. pyogenes* Cas9

Recombinant Cas9 was prepared and purified as previously described[52]. Briefly, *E. coli* Rosetta 2 cells were transformed with a plasmid encoding the *S. pyogenes Cas9* gene fused to an N-terminal 6xHis-tag and NLS (Addgene #62934). Transformed bacteria were used to inoculate 5 mL of LB broth containing 50 μg mL^{-1} carbenicillin, and incubated at 37 °C overnight (~16 h). The next day, cells were diluted 1:100 into the same growth medium and grown at 37 °C until an OD_{600} of 0.6 was reached. The culture was incubated at 16 °C for 30 min after which point isopropyl-ß-D-1-thiogalactopyranoside was added to a final concentration of 0.5 mM to induce Cas9 expression.

After 16 h, cells were collected by centrifugation for 15 min at 2700 × *g* and re-suspended in lysis buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 5 mM imidazole, pH 8.0, 1 mM PMSF). The solution was incubated on ice for 30 min before proceeding. Cells were further lysed by sonication (30 s pulse-on and 60 s pulse-off for 7.5 min at 60% amplitude) with soluble lysate being obtained by centrifugation at 30,000 × *g* for 30 min. The cell lysate containing Cas9 was injected into a HisTrap FF Crude column (GE Healthcare) attached to an AKTA Start System (GE Healthcare) and washed with wash buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 10 mM imidazole, pH 8.0) until UV absorbance reached a baseline. Cas9 was eluted in elution buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 250 mM imidazole, pH 8.0) in a single step. Eluted Cas9 was exchanged into storage buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 1 mM DTT) during concentration using a 100 kDa centrifugal filter (Pall). Concentrated Cas9 was flash-frozen in liquid nitrogen and stored in aliquots at −80 °C. dCas9 was purified in a similar manner.

2.5.4 *In vitro* cleavage of on-target and off-target DNA substrates

Primers pUC19_fwd and pUC19_rev (listed in **Table 2.11**) were used to generate Cas9 substrate DNAs through PCR amplification of previously prepared plasmid templates and subsequently purified with the QIAquick PCR Purification Kit (Qiagen). Equimolar amounts of tracrRNA (IDT) and crRNA (BioSynthesis) were heated at 95 °C for 10 min, then cooled to 25 °C over the course of 1 h to prepare gRNAs. gRNAs containing BNA^{NC}-modified and LNA-modified crRNAs were prepared as described above. For each cleavage reaction, 5 nM of substrate DNAs were incubated with 150 nM, or 15 nM pre-assembled Cas9 RNP complex for 1 h at 37 °C in Cas9 cleavage buffer (5% glycerol, 0.5 mM EDTA, 1 mM DTT, 2 mM $MgCl_2$, 20 mM HEPES pH 7.5, 100 mM KCl). Reactions were halted by purifying the products using the MinElute PCR Purification Kit

(Qiagen). Cleavage products were resolved on a 1% agarose gel, and imaged on an Amersham Imager 600 (GE Healthcare). Cleavage assays using eSpCas9 (Sigma-Aldrich) were performed as described above.

2.5.5 Library for high-throughput specificity profiling

Generation of pre-selection libraries for in vitro high-throughput specificity profiling experiments were performed as previously described[52]. Briefly, 10 pmol of WAS or EMX1 lib oligonucleotides were circularized through incubation with 100 units of CircLigase II ssDNA Ligase (Epicenter) in a total reaction volume of 20 μL for 16 h at 60 °C in 1× CircLigase II Reaction Buffer. The reaction was heat inactivated by incubation at 85 °C for 10 min. 5 pmol of the crude circular ssDNA was converted into concatemeric pre-selection libraries with the illustra TempliPhi Amplification Kit (GE Healthcare) according to the manufacturer's protocol. Concatemeric pre-selection libraries were quantified with the Qubit 2.0 Fluorometer. Sequences used to generate *in vitro* pre-selection libraries are listed in **Table 2.11**.

2.5.6 *In vitro* high-throughput specificity profiling

High-throughput specificity profiling of unmodified, BNA^{NC}-modified and LNA-modified crRNAs was performed as previously described[52]. Briefly, 200 nM of concatemeric pre-selection libraries were incubated with 1000 nM Cas9 and 1000 nM gRNA or 100 nM Cas9 and 100 nM gRNA in Cas9 cleavage buffer (NEB) for 20 min at 37 °C. Pre-selection libraries were also separately incubated with 2 U of BspMI (NEB) in NEBuffer 3.1 for 1 h at 37 °C. Cas9-digested and BspMI-digested library members were purified with the QiaQuick PCR Purification Kit (Qiagen) and ligated to 10 pmol adaptor1/2(#) (post-selection) or lib adapter 1/lib adapter 2 (pre-

selection) (sequences in **Table 2.11**) with 1000 U of T4 DNA Ligase (NEB) in NEB T4 DNA Ligase Reaction Buffer for 16 h at room temperature. Adapter ligated DNA was purified using the QiaQuick PCR Purification Kit (Qiagen) and PCR amplified for 19–24 cycles with Q5 Hot Start High-Fidelity DNA Polymerase (NEB) in Q5 Reaction Buffer using primers PE2 short/sel PCR (post-selection) or primers lib seq PCR/lib fwd PCR (pre-selection) (sequences in **Table 2.11**). PCR products were gel purified and quantified using a Qubit 2.0 Fluorometer (ThermoFisher) and subject to single-read sequencing on an Illumina MiSeq. Pre-selection and post-selection sequencing data were analyzed as previously described[52]. An example sequencing read is outlined below:

*ACTGTgaangg*ACTT*AGAAGAAGAAGA*CTCGGCAGGT**ACTTGAGTCCGAGCAGAAGAA GAAGGGGTCG***AGAAGAAGAAGA*CTCGGCAGGT

Where *ACTGT* represents the 5 bp reaction barcode used for de-multiplexing individual reaction conditions within a pooled sequencing experiment, *gaangg* the half-site of the target which was cut by Cas9 during *in vitro* digestion, and *AGAAGAAGAAGA* the reverse complement of the library target barcode. The library target barcode appears repeated twice due to the concatemeric nature of the pre-selection library. CTCGGCAGGT is the constant sequence. **ACTTGAGTCCGAGCAGAAGAAGAAGGGGTCG** represents an example of the full, post-selection library member with the first and last four nucleotides being fully randomized. Following binning to isolate the post-selection library members, specificity scores may be calculated using the following formulae: Positive specificity score = (frequency of base pair at position[post-selection] – frequency of base pair at position[pre-selection]) / (1- frequency of base pair at

position[pre-selection]) and negative specificity score = (frequency of base pair at position[post-selection]) / (frequency of base pair at position[pre-selection]).

2.5.7 Electrophoretic mobility shift assay

EMSAs were performed as previously described[234], with minor modifications to the protocol. To prepare the 6-FAM-labeled dsDNA substrate, EMSA fwd and rev oligonucleotides (listed in **Table 2.11**) were mixed in a 1.5:1 molar ratio, incubated at 95 °C for 5 min, then cooled to 25 °C over the course of 1 h. DNA substrates were diluted to a working concentration of 200 nM in binding buffer (20 mM HEPES, pH 7.5, 250 mM KCl, 2 mM $MgCl_2$, 0.01% Triton X-100, 0.1 mg mL^{-1} bovine serum albumin, 10% glycerol). gRNAs were prepared as described for in vitro cleavage assays. Next, nuclease-deficient Cas9 (dCas9) was incubated with previously annealed gRNA in a 1:1 molar ratio for 10 min at 25 °C in binding buffer to form the RNP complex. 50 nM DNA substrate was incubated with 0, 10, 50, 100, 250, and 500 nM RNP for 10 min at 37 °C in binding buffer. Reactions were resolved on a 10% TBE polyacrylamide gel supplemented with 2 mM $MgCl_2$ in 1× TBE buffer supplemented with 2 mM $MgCl_2$, and imaged on a Typhoon laser gel scanner (GE Healthcare). EMSAs using BNA^{NC}-modified crRNAs were performed in a similar manner.

2.5.8 Determination of crRNA-DNA heteroduplex melting temperature

Equimolar amounts of crRNA and complementary ssDNA oligonucleotide (listed in **Table 2.11**) were mixed in Duplex Buffer (30 mM HEPES, pH 7.5, 100 mM Potassium Acetate) (IDT) to a final concentration of 2 μM. SYBR Green I was added to a final concentration of 1×. The solution was moved to a CFX96 Real-Time System (BioRad) and incubated for 5 min at 95 °C, then cooled

to 25 °C at 0.1 °C s^{-1} to anneal the RNA/DNA heteroduplex. The heteroduplex was then heated at a rate of 0.1 °C s^{-1} to 95 °C and the corresponding fluorescent signal was used to generate a melt curve.

2.5.9 Generation of Cas9 stable cells

lentiCas9-Blast viral particles were purchased from Addgene (#52962) and used to infect U2OS and HeLa cells according to the manufacturer's protocol. Briefly, on the day of infection, cells were trypsinized, counted, and diluted to a working concentration of 50,000 cells mL^{-1} in DMEM media supplemented with 10% FBS/1× pen-strep/1× glutamine (Gibco) (DMEM complete) and 10 µg mL^{-1} polybrene. Viral particles were serially diluted down to 1:500 from the original stock (2.5×10^5 TU mL^{-1}), with 500 µL of each dilution added to the corresponding wells of a 6-well plate. 1 mL of cell suspension was added to each well and incubated at 37 °C and 5% CO_2. 48 h after infection, selection was performed in DMEM-complete media supplemented with 10 µg mL^{-1} Blasticidin S HCl (Gibco). After selection, cells stably expressing Cas9 were maintained in DMEM-complete media with 5 µg mL^{-1} Blasticidin S HCl.

2.5.10 Cell culture

All cells were cultured at 37 °C in a 5% CO_2 atmosphere. U2OS-Cas9 and HeLa-Cas9 cells were cultured in high glucose DMEM media with pyruvate (Gibco) supplemented with 10% FBS/1× pen-strep/1× glutamine (Gibco) and 5 µg mL^{-1} Blasticidin S HCl (Gibco), where applicable. The U2OS and HeLa cells were previously authenticated and shown to be negative for mycoplasma at the time of purchase (ATCC).

2.5.11 Cationic lipid transfection of gRNA into stable cell lines

Cells stably expressing Cas9 were transfected with RNAiMAX and annealed gRNA according to the manufacturers' instructions to a final concentration of 30 nM. Experiments involving BNA^{NC}-modified and LNA-modified crRNAs were performed in a similar fashion.

2.5.12 Cellular cleavage assay

Genomic DNA (gDNA) from transfected cells was extracted using a DNeasy Kit (Qiagen) 48 h after transfection, according to the manufacturer's instructions, and quantified using a NanoPhotometer NP80 (Implen) spectrophotometer. Amplicon specific primer pairs (listed in **Table 2.11**) and 100 ng of gDNA was used to PCR amplify the desired target site, which was then purified with the QIAquick PCR Purification Kit (Qiagen). T7 endonuclease I (T7E1) digestion of the PCR products was performed as described in the manufacturer's protocol (NEB). Cleavage products were resolved on a 2.5% agarose gel.

2.5.13 High-throughput sequencing

100 ng gDNA isolated from cells from each transfection was amplified by PCR for 35 cycles with primers PCR1_fwd and PCR1_rev (listed in **Table 2.11**) and 2× Q5 Hot Start High Fidelity Master Mix according to the manufacturer's instructions (NEB). PCR products were purified via GeneRead Size Selection Kit (Qiagen). Purified PCR products were then amplified by PCR with Nextera XT Illumina primers for 7 cycles with 2× Q5 Hot Start High Fidelity Master Mix in Q5 Reaction Buffer (NEB), to add unique barcodes. Amplified control and treated DNA pools were subsequently purified with the GeneRead Size Selection Kit (Qiagen), quantified using the Qubit 2.0 Fluorometer (ThermoFisher), and pooled in a 1:1 ratio. The final sample underwent paired-

end sequencing using an Illumina MiSeq instrument, according to the manufacturer's protocol. Statistical analysis and determination of modification frequency was performed using CRISPR-DAV[235] and Cas-Analyzer[236].

2.5.14 Single-molecule FRET experiments

Single-molecule FRET experiments were set up and performed using a previously described protocol[229]. Briefly, to prevent non-specific binding of samples to the glass surface, all coverslips and quartz glasses were passivated using polyethylene glycol. A single-molecule flow chamber consisting of a microscope slide and coverslip sealed with epoxy and double-sided tape was assembled using rounded holes on either side of the slide as inlets and outlets for solution exchange. All imaging was performed at room temperature with the following buffer composition: 100 mM NaCl, 50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, 5% glycerol, and 0.1 mg mL^{-1} BSA. For experiments used to generate FRET histograms, the oxygen scavenger (1 mg mL^{-1} of glucose oxidase (Sigma-Aldrich), 0.04 mg mL^{-1} of catalase (Sigma-Aldrich), and 0.8% (w/v) of b-D-glucose), and the triplet quencher (~4 mM Trolox) were applied to the buffer to prevent photo-fatigue of the fluorophores. FRET histograms were obtained from the images 30 min after incubation of Cas9 RNP complex (30 nM Cas9, 10 nM gRNAs) with target containing DNA sequences (listed in **Table 2.11**). Single-molecule time trace imaging was performed as described above, with the exception of the oxygen scavenging system as well as an addition of 5% (v/v) glycerol. Time traces were acquired throughout the duration of the incubation (from 0 to 15 min) of Cas9 RNP complex with target DNA. The acquisition time for smFRET histograms and time traces were 100 and 30 ms, respectively.

2.5.15 Statistical analysis

Indel percentages shown in **Figure 2.8** were calculated as indel (%) = 100 × $(1-(1-\text{fraction}_{\text{cut}})^{0.5})$. *P*-values in **Table 2.4** were calculated through the comparison of 150,000 randomly sampled sequences from the pre-selection library (WAS or EMX1) with an equal number of sequences from the corresponding post-selection library using a Mann–Whitney test. *P*-values in **Table 2.7** were calculated using a Fisher exact test comparing crRNA transfected samples with the un-transfected control.

2.5.16 Code availability

Python scripts used for deep sequencing data processing are available upon request.

2.5.17 Data availability

Data generated for this work are included in this published article and its associated Supplementary Information files. High-throughput sequencing data files have been deposited in the NCBI SRA database and are available under accession number: SRP125574.

Table 2.10. Unmodified, BNA^{NC}- and LNA-modified crRNAs and tracrRNA sequences used

Name	Sequence (5′ -> 3′)
WAS-RNA crRNA	rUrGrGrArUrGrGrArGrGrArArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-1 crRNA	rUrGrGrArUrGrGrArGrG+ArArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-2 crRNA	rUrGrGrArUrGrGrArG+G+ArArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-3 crRNA	rUrGrGrArUrGrGrArG+G+A+ArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-4 crRNA	rUrGrGrArUrGrGrArGrGrArArUrGrArGrG+A+G+TrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-5 crRNA	rUrGrGrArUrGrGrArGrGrArArU+G+A+GrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU

Name	Sequence (5′ -> 3′)
WAS-BNA-6 crRNA	rUrGrGrArUrGrGrArG+G+A+A+TrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-7 crRNA	rUrGrGrArUrGrGrA+G+GrA+A+TrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-8 crRNA	rUrGrG+A+T+GrGrArGrGrArArU+G+A+GrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-9 crRNA	+T+GrG+A+T+GrGrArGrGrArArU+G+A+GrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-LNA-3 crRNA	rUrGrGrArUrGrGrArG*G*A*ArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-LNA-5 crRNA	rUrGrGrArUrGrGrArGrGrArArU*G*A*GrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-LNA-6 crRNA	rUrGrGrArUrGrGrArG*G*A*A*TrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-RNA crRNA	rGrArGrUrCrCrGrArGrCrArGrArArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-1 crRNA	rGrArGrUrCrCrGrArGrCrArGrArArGrArArG+A+ArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-2 crRNA	rGrArGrUrCrCrGrArGrCrArGrArA+G+A+ArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-3 crRNA	+G+A+GrUrCrCrGrArGrCrArGrArArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-4 crRNA	rGrArGrUrCrCrGrArG+C+A+GrArArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-5 crRNA	rGrArGrUrCrCrGrArGrCrA+G+A+ArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-6 crRNA	rGrArGrUrCrCrGrArG+CrA+GrA+ArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-7 crRNA	rGrArGrUrCrCrGrArGrCrArGrArArGrA+A+G+A+ArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-8 crRNA	rGrArGrU+C+C+G+ArGrCrArGrArArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-BNA-9 crRNA	rGrArGrUrCrCrGrArGrC+A+GrArArGrArA+G+A+ArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-LNA-5 crRNA	rGrArGrUrCrCrGrArGrCrA*G*A*ArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-LNA-6 crRNA	rGrArGrUrCrCrGrArG*CrA*GrA*ArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
EMX1-LNA-7 crRNA	rGrArGrUrCrCrGrArGrCrArGrArArGrA*A*G*A*ArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-RNA Cy5 crRNA	[Cy5]rUrGrGrArUrGrGrArGrGrArArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
WAS-BNA-3 Cy5 crRNA	[Cy5]rUrGrGrArUrGrGrArG+G+A+ArUrGrArGrGrArGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU

Name	Sequence (5′ -> 3′)
tracrRNA	rArGrCrArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrC rArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrUrU

A plus sign (+) indicates that the following nucleotide is a BNA^{NC}, while an asterisks (*) indicates the following nucleotide is a LNA. Nucleotides with a preceding (r) were ordered as RNA.

Table 2.11. Oligonucleotide sequences used in this study

Name	Sequence (5′ -> 3′)
	The following were sequences used to generate in vitro cleavage assay constructs
WAS_target#	GGCCGGCCGAAGCTTAGGAGGTGCGTGCTGATTCTTCCCTGTGTCTCTGGATG GATGGGTAAGAGTGGATGGAGGAATGAGGAGTTGGATGGGTGCGTAAGTGGG TGAATGGATAGGTAGATTGATAGGTATGTGGATTCTAGAGGCCGGCCG
WAS-OT1_target#	GGCCGGCCGAAGCTTCCATCCTTAGCACAGAGTTTGGCACATAGAGGAGCCTG ATCAATACCTGCTGGATGGAGGGATGAGGAGTGGGGCTGGGCCGTATCAGTC AGGATGGTGAAGTGCATAACAAACAGCCCCCACTCTAGAGGCCGGCCG
WAS-OT2_target#	GGCCGGCCGAAGCTTACATTATTTATGCTAGGAAACCCTGTTGCTGTATGATT TGTGTGTGTGAGGGGATGGAGGGATGAGGAGTGGGAAGCTGTTGACTCATGC ACATACCTGTCTCCATGGTGTCTCTGCTGTCGATCTAGAGGCCGGCCG
WAS-OT3_target#	GGCCGGCCGAAGCTTTGTAGAGTAGTCAAATTCACAGGGACAGAAAGTAGAA TAGTGGTTGCTGGGGGAGGGAGGAATGAGGAGTGGGTACAGAGTTTCATCTG GGGAAGATGAAAAAGTTCTGTAGACGGATGGTGGTCTAGAGGCCGGCCG
WAS-OT4_target#	GGCCGGCCGAAGCTTAGTAGGGTTAGGCCACGACACTCAGGCTTTCAGGACA AACAAGGAGAGGGAGGAGGGAGGAATGGGGAGTTGGTAGTTAATGGGGATG GAGTTTCAGTTTGGGATGACAAAAAAGTTCTGGAGTCTAGAGGCCGGCCG
WAS-OT5_target#	GGCCGGCCGAAGCTTGCACTGGTCACCTGTTTGAGGAAAATGGACTGTCTCCA AATTGCCCCACTCGGACGGAGGAATGGGGAGTGGGGGCTGGTTTGTGAAGGC TGTTCTCTTTCTAAAGTCTCAGTCTCTCTTTTCTCTAGAGGCCGGCCG

Name	Sequence (5′ -> 3′)
EMX1_target_fwd	GCCGAAGCTTCTGAGTCCGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1_target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCGGACTCAGAAGCTTCGGC
EMX1-OT1_target_fwd	GCCGAAGCTTCTGAGTTAGAGCAGAAGAAGAAAGGCTTCTAGAGGCC
EMX1-OT1_target_rev	GGCCTCTAGAAGCCTTTCTTCTTCTGCTCTAACTCAGAAGCTTCGGC
EMX1-OT2_target_fwd	GCCGAAGCTTCTGAGTCTAAGCAGAAGAAGAAGAGCTTCTAGAGGCC
EMX1-OT2_target_rev	GGCCTCTAGAAGCTCTTCTTCTTCTGCTTAGACTCAGAAGCTTCGGC
EMX1-OT3_target_fwd	GCCGAAGCTTCTGAGGCCGAGCAGAAGAAAGACGGCTTCTAGAGGCC
EMX1-OT3_target_rev	GGCCTCTAGAAGCCGTCTTTCTTCTGCTCGGCCTCAGAAGCTTCGGC
EMX1-OT4_target_fwd	GCCGAAGCTTCTGAGTCCTAGCAGGAGAAGAAGAGCTTCTAGAGGCC
EMX1-OT4_target_rev	GGCCTCTAGAAGCTCTTCTTCTCCTGCTAGGACTCAGAAGCTTCGGC
EMX1-OT5_target_fwd	GCCGAAGCTTCTAAGTCTGAGCACAAGAAGAATGGCTTCTAGAGGCC
EMX1-OT5_target_rev	GGCCTCTAGAAGCCATTCTTCTTGTGCTCAGACTTAGAAGCTTCGGC
WAS-m1 target_fwd	GCCGAAGCTTCTTGGATGGAGGAATGAGGAGAGGGCTTCTAGAGGCC
WAS-m1 target_rev	GGCCTCTAGAAGCCCTCTCCTCATTCCTCCATCCAAGAAGCTTCGGC
WAS-m2 target_fwd	GCCGAAGCTTCTTGGATGGAGGAATGAGGGGTGGGCTTCTAGAGGCC
WAS-m2 target_rev	GGCCTCTAGAAGCCCACCCCTCATTCCTCCATCCAAGAAGCTTCGGC
WAS-m3 target_fwd	GCCGAAGCTTCTTGGATGGAGGAATGACGAGTGGGCTTCTAGAGGCC
WAS-m3 target_rev	GGCCTCTAGAAGCCCACTCGTCATTCCTCCATCCAAGAAGCTTCGGC
WAS-m4 target_fwd	GCCGAAGCTTCTTGGATGGAGGAAAGAGGAGTGGGCTTCTAGAGGCC
WAS-m4 target_rev	GGCCTCTAGAAGCCCACTCCTCTTTCCTCCATCCAAGAAGCTTCGGC

Name	Sequence (5′ -> 3′)
WAS-m5 target_fwd	GCCGAAGCTTCTTGGATGGAGGACTGAGGAGTGGGCTTCTAGAGGCC
WAS-m5 target_rev	GGCCTCTAGAAGCCCACTCCTCAGTCCTCCATCCAAGAAGCTTCGGC
WAS-m6 target_fwd	GCCGAAGCTTCTTGGATGGAGAAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m7 target_rev	GGCCTCTAGAAGCCCACTCCTCATTTCTCCATCCAAGAAGCTTCGGC
WAS-m7 target_fwd	GCCGAAGCTTCTTGGATGGTGGAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m7 target_rev	GGCCTCTAGAAGCCCACTCCTCATTCCACCATCCAAGAAGCTTCGGC
WAS-m8 target_fwd	GCCGAAGCTTCTTGGATTGAGGAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m8 target_rev	GGCCTCTAGAAGCCCACTCCTCATTCCTCAATCCAAGAAGCTTCGGC
WAS-m9 target_fwd	GCCGAAGCTTCTTGGGTGGAGGAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m9 target_rev	GGCCTCTAGAAGCCCACTCCTCATTCCTCCACCCAAGAAGCTTCGGC
WAS-m10 target_fwd	GCCGAAGCTTCTTCGATGGAGGAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m10 target_rev	GGCCTCTAGAAGCCCACTCCTCATTCCTCCATCGAAGAAGCTTCGGC
WAS-m11 target_fwd	GCCGAAGCTTCTAGGATGGAGGAATGAGGAGTGGGCTTCTAGAGGCC
WAS-m11 target_rev	GGCCTCTAGAAGCCCACTCCTCATTCCTCCATCCTAGAAGCTTCGGC
EMX1-m1 target_fwd	GCCGAAGCTTCTGAGTCCGAGCAGAAGAAGAGGGGCTTCTAGAGGCC
EMX1-m1 target_rev	GGCCTCTAGAAGCCCCTCTTCTTCTGCTCGGACTCAGAAGCTTCGGC
EMX1-m2 target_fwd	GCCGAAGCTTCTGAGTCCGAGCAGAAGAATAAGGGCTTCTAGAGGCC
EMX1-m2 target_rev	GGCCTCTAGAAGCCCTTATTCTTCTGCTCGGACTCAGAAGCTTCGGC
EMX1-m3 target_fwd	GCCGAAGCTTCTGAGTCCGAGCAGAAGGAGAAGGGCTTCTAGAGGCC
EMX1-m3 target_rev	GGCCTCTAGAAGCCCTTCTCCTTCTGCTCGGACTCAGAAGCTTCGGC
EMX1-m4 target_fwd	GCCGAAGCTTCTGAGTCCGAGCAGAGGAAGAAGGGCTTCTAGAGGCC
EMX1-m4 target_rev	GGCCTCTAGAAGCCCTTCTTCCTCTGCTCGGACTCAGAAGCTTCGGC
EMX1-m5 target_fwd	GCCGAAGCTTCTGAGTCCGGGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m5 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCCCGGACTCAGAAGCTTCGGC
EMX1-m6 target_fwd	GCCGAAGCTTCTGAGTCCGAGTAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m6 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTACTCGGACTCAGAAGCTTCGGC

Name	Sequence (5′ -> 3′)
EMX1-m7 target_fwd	GCCGAAGCTTCTGAGTCCGGGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m7 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCCCGGACTCAGAAGCTTCGGC
EMX1-m8 target_fwd	GCCGAAGCTTCTGAGTCTGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m8 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCAGACTCAGAAGCTTCGGC
EMX1-m9 target_fwd	GCCGAAGCTTCTGAGGCCGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m9 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCGGCCTCAGAAGCTTCGGC
EMX1-m10 target_fwd	GCCGAAGCTTCTGCGTCCGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m10 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCGGACGCAGAAGCTTCGGC
EMX1-m11 target_fwd	GCCGAAGCTTCTTAGTCCGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EMX1-m11 target_rev	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCGGACTAAGAAGCTTCGGC
pUC19_fwd	GCGACACGGAAATGTTGAATACTCAT
pUC19_rev	CAGCGAGTCAGTGAGCGA
The following sequences were used for in vitro library selection experiments	
WAS_lib	/5Phos/AACACANNNNC*C*NA*C*T*C*C*T*C*A*T*T*C*C*T*C*C*A*T*C*C*A *NNNNACCTG CCGAGAACACA
EMX1_lib	/5Phos/TCTTCTNNNNC*C*NT*T*C*T*T*C*T*T*C*T*G*C*T*C*G*G*A*C*T*C* NNNNACCTGCCGAGTCTTCT
adaptor1(1)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC TGT
adaptor1(2)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCT GAA
adaptor1(3)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTG ACT
adaptor1(4)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTG CAA
adaptor1(5)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGC ATT
adaptor1(6)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCA TGA
adaptor1(7)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAT GCT
adaptor1(8)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCT AGT
adaptor1(9)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGC TAA
adaptor1(10)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCA GTT

Name	Sequence (5′ -> 3′)
adaptor1(11)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGT CAT
adaptor1(12)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC GTA
adaptor2(1)	ACA GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(2)	TTC AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(3)	AGT CAA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(4)	TTG CAA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(5)	AAT GCA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(6)	TCA TGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(7)	AGC ATA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(8)	ACT AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(9)	TTA GCA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(10)	AAC TGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(11)	ATG ACA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(12)	TAC GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
PE2_short	AAT GAT ACG GCG ACC ACC GA
WAS_sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GAA CAC A
EMX1_sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GTC TTC T
lib adaptor 1	GAC GGC ATA CGA GAT
WAS lib adaptor 2	AAC AAT CTC GTA TGC CGT CTT CTG CTT G
EMX1 lib adaptor 2	TCT TAT CTC GTA TGC CGT CTT CTG CTT G
WAS lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAA CAC A
EMX1 lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GTC TTC T
lib fwd PCR	CAA GCA GAA GAC GGC ATA CGA GAT
The following sequences were used for EMSA experiments	
WAS_EMSA fwd	CCCATCCAACTCCTCATTCCTCCATCCA
WAS_EMSA rev	TGGATGGAGGAATGAGGAGTTGGATGGGCGC[6-FAM]

Name	Sequence (5′ -> 3′)
WAS-OT3_EMSA fwd	CTGTACCCACTCCTCATTCCTCCCTCCC
WAS-OT3_EMSA rev	GGGAGGGAGGAATGAGGAGTGGGTACAGCGC[6-FAM]
EMX1_EMSA fwd	GGGAGCCCTTCTTCTTCTGCTCGGACTC
EMX1_EMSA rev	GAGTCCGAGCAGAAGAAGAAGGGCTCCCCGC[6-FAM]
EMX1-OT1_EMSA fwd	CCATGCCTTTCTTCTTCTGCTCTAACTC
EMX1-OT1_EMSA rev	GAGTTAGAGCAGAAGAAGAAAGGCATGGCGC[6-FAM]
The following sequences were used for Tm determination experiments	
WAS_Tm	TCCAACTCCTCATTCCTCCATCCAC
WAS-OT1_Tm	TCCAACTCCTCATCCCTCCATCCAC
WAS-OT2_Tm	TCCCACTCCTCATCCCTCCATCCCC
WAS-OT3_Tm	ACCCACTCCTCATTCCTCCCTCCCC
WAS-OT4_Tm	ACCAACTCCCCATTCCTCCCTCCTC
WAS-OT5_Tm	CCCCACTCCCCATTCCTCCGTCCGA
The following sequences were used for cellular cleavage experiments	
WAS_T7E1 fwd	GTGGCAGGGCTGTGATAACT
WAS_T7E1 rev	TGCTTTATCATTCACTGGCTCA
WAS-OT3_T7E1 fwd	GCTACAACATGGATGAACCTGGA
WAS-OT3_T7E1 rev	GCCTCCCAAAGTGCTGAGAT
EMX1_T7E1 fwd	CCTCAGTCTTCCCATCAGC
EMX1_T7E1 rev	GTGGGGCCATGACTCCAG
EMX1-OT1_T7E1 fwd	TGCCTTTCATCTGATGCTGT
EMX1-OT1_T7E1 rev	GGTGCCTTTTGTGGGGAGAT
WAS PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGAGAGAAGAGGAGGGCT
WAS PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCACCCATCCATCAGTCCAC

Name	Sequence (5′ -> 3′)
WAS-OT1 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTGCAGTCAAGAGCCTCG
WAS-OT1 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCTTTGCCTGGGAGAGA AA
WAS-OT2 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCAGCCTAAATCCCACCTT
WAS-OT2 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGCCCCCTCCATTGGAT A
WAS-OT3 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTACAACATGGATGAACC TGGA
WAS-OT3 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCATCACCACCATCCGT C
WAS-OT4 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCATGAAGGGAAGAGGTG C
WAS-OT4 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTATGCAACCATCACCA CC
WAS-OT5 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTCAACTGCCTGCTCTGG
WAS-OT5 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGGGCAGGAGACACAG A
EMX1 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCCTCCTGAGTTTCTCAT C
EMX1 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCCCACCCTAGTCATTGG A
EMX1-OT1 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCTTATGGCATGGCAAG A
EMX1-OT1 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAAGCTTTTCCTGACGCC
EMX1-OT2 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACCTGGGCGAGAAAGGT A
EMX1-OT2 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGCTTTCACAAGGATGCA GT
EMX1-OT3 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GTGGGAGAGAGACCCCTTCT
EMX1-OT3 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG TCATCCCGACCTTCATCCCT
EMX1-OT4 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTCTTTGGGCCTCGGCTT
EMX1-OT4 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCCCAGCTGTGCATTCTA T
EMX1-OT5 PCR1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGAAGCACCCGGATGTA G
EMX1-OT5 PCR1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCAAACAAGGTGCAGATA CAGCA
The following sequences were used for single-molecule FRET experiments	
WAS smFRET	/5Biosg/ttt ttt GAG GAA G/iAmMC6T/G CCT TGG ATG GAG GAA TGA GGA GTT GGC TCC CAT CAC ATC
WAS-OT4 smFRET	/5Biosg/ttt ttt GAG GAA G/iAmMC6T/G CCT AGG AGG GAG GAA TGG GGA GTT GGC TCC CAT CAC ATC

Sequence names ending with a number sign (#) were ordered from IDT as dsDNA gBlocks. An asterisks (*) indicates that the preceding nucleotide was incorporated as a hand mix of bases consisting of 79 mol % of the intended base, and 7 mol % of each of the other three bases. "/5Phos/" denotes a 5′ phosphate group added during synthesis.

2.6 Additional information

2.6.1 Acknowledgements

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery/Discovery Accelerator Supplement (RGPIN-2016-06381) and Alberta Glycomics Center grants to B.P.H., and a National Research Foundation of Korea (NRF) grant (No. 2018R1A2B2001422) to S.K.K. A.R.K. was supported by an NSERC USRA scholarship. We thank Sudip Subedi for technical assistance with next-generation sequencing.

2.6.2 Contributions

B.P.H. and C.R.C. conceived the study and designed experiments to assess the effect of BNA^{NC}-modified and LNA-modified crRNAs on Cas9 specificity. C.R.C. performed experiments dealing with specificity of BNA^{NC}s in vitro and in cells, and was assisted by A.R.K. K.S. and J.P. designed and performed single-molecule FRET experiments, which were supervised by S.K.K. J.J. assisted with high-throughput sequencing and performed bioinformatics analysis. All authors contributed to the writing of the manuscript.

2.6.3 Competing interests

The authors have filed a provisional patent on this work with B.P.H. and C.R.C. listed as inventors. The remaining authors declare no competing interests.

Chapter 3

High-throughput specificity profiling for the identification of Cas9 crXNAs with superior specificity

Christopher R. Cromwell[1], Juan Jovel[2], Sunil Vasu Kalmady[3], Russ Greiner[4], Basil Hubbard[1,5]

[1]Department of Pharmacology, University of Alberta, Edmonton, Canada

[2]The Metabolomics Innovation Centre, University of Alberta, Edmonton, Canada

[3]Canaidan VIGOUR Centre, University of Alberta, Edmonton, Canada

[4]Department of Computing Science, University of Alberta, Edmonton, Canada

[5]Department of Pharmacology & Toxicology, University of Toronto, Toronto, Canada

UNDER PREPARATION FOR PUBLICATION

3.1 Abstract

The minimization of off-target activity by programmable nucleases such as Cas9 is a primary goal in the field of gene editing. Incorporation of chemically-modified nucleotides within the spacer sequence of CRISPR RNAs (crRNAs) has been shown to be an effective method for modulating Cas9 specificity. However, it remains difficult to translate these findings to all sequences, as no consensus on the most effective design of crRNAs incorporating xeno nucleic acids (crXNAs) exists. Here we describe the comprehensive activity and specificity profiling of a collection of several hundred chemically modified gRNAs comprised of 7 modifications across a diverse set of 11 DNA target sequences. Our results show that the position and number of modifications within the gRNA is critical for improving specificity and maintaining on-target activity. Moreover, we show that gRNAs modified at specific positions with 2′-fluoroarabino nucleic acid (FANA) have dramatically improved Cas9 cleavage specificity profiles when compared to unmodified gRNAs. This large data set may be used to more accurately design chemical modifications for any given crRNA, so as to maximize on-target activity, and minimize off-target activity. By establishing general rules that govern their use, this work will advance the widespread use of crXNA technology.

3.2 Introduction

The clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 system was originally identified as an effector of prokaryotic adaptive immunity[209], but has since been repurposed as a powerful and versatile tool for the manipulation of genetic material[28, 35]. Since its discovery, Cas9 has been applied to the generation of knock-out and knock-in organisms and model cell lines[210, 237-239], genome-wide functional and epigenetic screens[211, 212, 240], *in vitro*

diagnostic tools[128, 129] as well as potential future therapeutics[124]. Cas9 is able to achieve sequence-specific DNA cleavage through two independent RNA oligonucleotides[32]: a CRISPR RNA (crRNA), which contains a 20 nucleotide (nt) spacer sequence complementary to the target of interest, and a trans-activating crRNA (tracrRNA), which facilitates interaction between the crRNA and Cas9 protein[35]. When hybridized, these RNAs are known as a guide RNA (gRNA)[35]. DNA target recognition is dictated by complementarity between the crRNA spacer sequence, as well as the presence of a protospacer adjacent motif (PAM) directly 3′ of the target. PAM interaction is proceeded by local DNA melting, followed by hybridization of the crRNA and target DNA strand[28, 55]. Complete hybridization of the RNA/DNA heteroduplex (which along with the displaced non-target DNA strand is known as an R-loop) induces conformational rearrangements within the HNH and RuvC nuclease domains, resulting in the introduction of a blunt-ended, double-stranded break (DSB) within the DNA target[31, 47]. While Cas9 has been demonstrated to be relatively specific, ≤5 mismatches between the crRNA and DNA target have been shown to be reliably tolerated, resulting in off-target activity at locations distinct from the target of interest[142, 216].

The widespread implementation of Cas9-based technology further underscores the importance of accurate DNA targeting, as the presence of off-target activity at undesired sequences may obfuscate future use. Numerous methods have been proposed for the enhancement of CRISPR-Cas9 specificity, including the engineering of highly specific Cas9 variants[151, 175-177, 179-182], computational prediction of gRNA specificity[162, 164] and delivery strategies aimed at reducing the resident active time of Cas9 within a cell[171-174, 241]. Additionally, several studies have also investigated the modification or otherwise alteration of the crRNA as a method for off-target

mitigation through structural changes[192, 193] or incorporation of chemically modified nucleic acids[194, 195, 203]. Several chemically modified nucleic acids have been demonstrated to significantly improve specificity of the Cas9 system when substituted into the crRNA, including bridged nucleic acids (BNAs) and locked nucleic acids (LNAs)[194], DNA[203] as well as 2′-*O*-methyl-3′-phosphonoacetate (PACE)[195]. While effective, these strategies have been shown to be highly dependent on the type, position and number of modifications made in order to elicit their effects on specificity[194, 195, 203]. As such, the use of chemically modified gRNAs has not yet been widely adopted as a method for reducing Cas9-mediated off-target activity.

Here, we describe a more comprehensive and descriptive method for modulating Cas9 DNA cleavage specificity through utilization of xeno nucleic acid (XNA)-containing crRNAs (crXNAs). Using a previously described high-throughput specificity profiling assay[13, 52, 194] and *in vitro* cleavage assay[194, 242], we were able to show global specificity and activity profiles for several hundred differentially modified crXNAs. Importantly, we were able to identify a number of crXNAs which demonstrate significant specificity enhancements and maintain their on-target activity, while also providing insight into future design considerations for effective crXNA use.

3.3 Results

3.3.1 Design of XNA-modified crRNAs

As previous studies have shown that crXNA-mediated specificity increases are dependent on how and where the spacer region of the crRNA is modified[194, 195], we initially designed a set of two independent modification schemes for either RNA- or DNA-based modification (**Figure 3.1a**, **b**). We previously demonstrated that incorporation of ≤4 modifications often resulted in the highest

levels of observed specificity improvement, while >4 resulted in abolishment of on-target activity[194]. With this in mind, we designed a modification scheme for RNA-based XNAs which divides the spacer region into three sections relative to the PAM (distal, mid and proximal), each bearing 1-4 modification positions (MOD.1-13) (**Figure 3.1a**). As DNA/RNA hybrid crRNAs have been shown to also improve Cas9 specificity[203], we designed a second modification scheme (**Figure 3.1b**) for these chimeric crRNAs specifically. To design this, we took into account previous observations that positions 1, 15, 16 and 19 of the 5′ region of the crRNA make critical hydrogen bonding interactions with the Cas9 enzyme, and therefore should not be modified (DNA.1-7 and DNA.11)[202]. However, conflicting studies have shown that this may not be absolutely necessary for activity[203], so additional DNA-based crRNAs were included to reflect this (DNA.8-10). Initially, we synthesized a set of 102 crXNAs containing one of several unique XNAs using either the RNA- or DNA-based modification scheme, against a previously described *EMX1*-based Cas9 target site[194]. XNAs were chosen based on their previous applications to either Cas9 specificity or activity, as well as general commercial availability for synthesis[188, 191, 194, 195, 202, 203, 206]. XNAs chosen included 2′-O-methyl RNA (OMe), bridged nucleic acids (BNAs), locked nucleic acids (LNAs), DNA, 2′-fluoroarabino nucleic acid (FANA), 2′-*O*-methyl phosphonoacetate (PACE) RNA, phosphorothioate RNA (RNAPS) and phosphorothioate DNA (DNAPS) (**Figure 3.1c**).

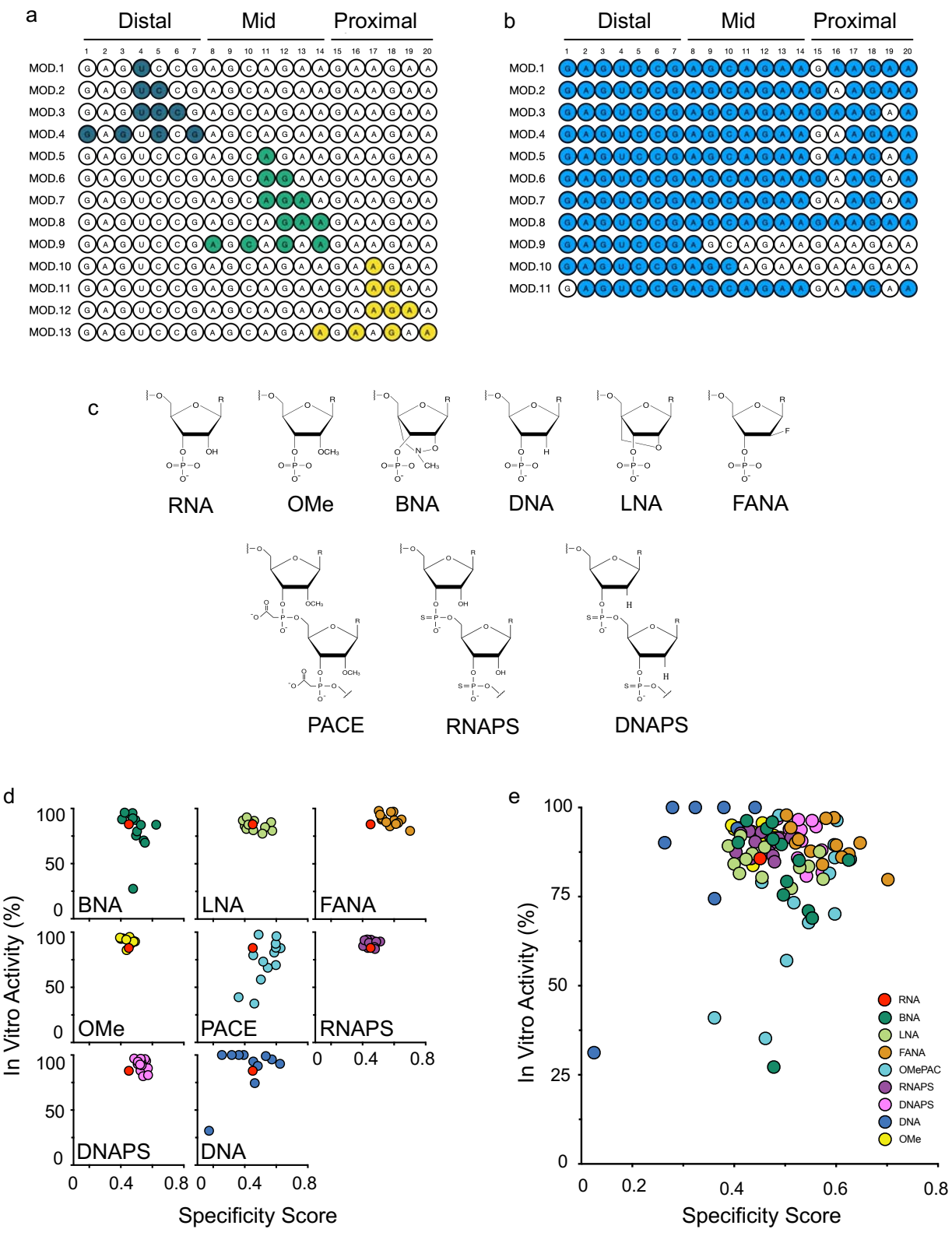

a
Distal
Mid
Proximal
MOD.1
MOD.2
MOD.3
MOD.4
MOD.5
MOD.6
MOD.7
MOD.8
MOD.9
MOD.10
MOD.11
MOD.12
MOD.13
b
Distal
Mid
Proximal
MOD.1
MOD.2
MOD.3
MOD.4
MOD.5
MOD.6
MOD.7
MOD.8
MOD.9
MOD.10
MOD.11
c
RNA
OMe
BNA
DNA
LNA
FANA
PACE
RNAPS
DNAPS
d
In Vitro Activity (%)
BNA
LNA
FANA
OMe
PACE
RNAPS
DNAPS
DNA
Specificity Score
e
In Vitro Activity (%)
Specificity Score
RNA
BNA
LNA
FANA
OMePAC
RNAPS
DNAPS
DNA
OMe

Figure 3.1. crXNAs reduce off-target cleavage while maintaining on-target activity
Diagram depicting the modification positions of (**a**) RNA-based or (**b**) DNA-based crXNAs. Colored circles are used to indicate the position of XNA substitution. The 20 nt EMX1 DNA target sequence is shown. (**c**) Chemical structures of XNAs used in this study. (**d**) Specificity and activity scores for individual crXNAs against the EMX1 target site. (**e**) Comparison of *in vitro* activity and specificity scores for all crXNAs. *In vitro* activity is shown as a mean ($n = 3$), while specificity score is shown as the average of each specificity score for each individual nucleotide along the 20 nt EMX1 spacer sequence (as determined by high-throughput specificity profiling). *In vitro* cleavage assays were performed using 25 nM of Cas9 RNP complex and 5 nM of EMX1 target DNA, while high-throughput specificity profiling assays used 1000 nM of Cas9 RNP complex and 200 nM of pre-selection library.

3.3.2. Off-target reduction by crXNAs *in vitro* is dependent on type, position and number of modifications

To evaluate the global specificity profiles of the synthesized crXNAs, we utilized a previously described high-throughput specificity profiling assay (**Figure 3.2a**)[13, 52, 194]. In brief, this assay measures the ability of Cas9 to cleave a library of ~10^{12} potential off-target sites (as well as the on-target) containing a 10-fold coverage of all sequences with ≤8 mismatches relative to the EMX1 target. Conversely, activity against the on-target sequences was measured using an *in vitro* cleavage assay (**Figure 3.2b**). We performed both specificity profiling and activity assays against all 102 crXNAs, as well as an unmodified crRNA for reference as a baseline. Using the data set from the specificity profiling experiments, we calculated enrichment scores for each base at each position along the 20 nt spacer sequence (denoted as specificity score), in which the mean for all 20 nts was plotted against *in vitro* activity as a global profile of crXNA function (**Figure 3.1d**, **e**). On-target activity was largely maintained within the crXNAs tested, with the exception of several crXNAs containing either BNA, PACE or DNA modifications (**Figure 3.3**). Specifically, modification of positions 8, 10, 12 and 14 (MOD.9) with either BNA or PACE XNAs resulted in the most dramatic loss of on-target activity, as well as complete modification of the 20 nt spacer

with DNA (DNA.8). Effects of XNA incorporation on specificity were much more variable than activity, demonstrating both increases and decreases to the propensity for off-target cleavage based on the modification used, as well as the position within the crRNA (**Figure 3.4**). Of the tested crXNAs, those containing FANA were shown to be the most consistent for broad specificity improvements, while OMe- and RNAPS-containing crXNAs demonstrated minimal differences when compared with the unmodified control. Of the 102 crXNAs tested, 77 were determined to have specificity scores higher than the unmodified crRNA (**Table 3.1**), demonstrating the broad applicability of crXNAs as a means for improving DNA cleavage specificity of the Cas9 system.

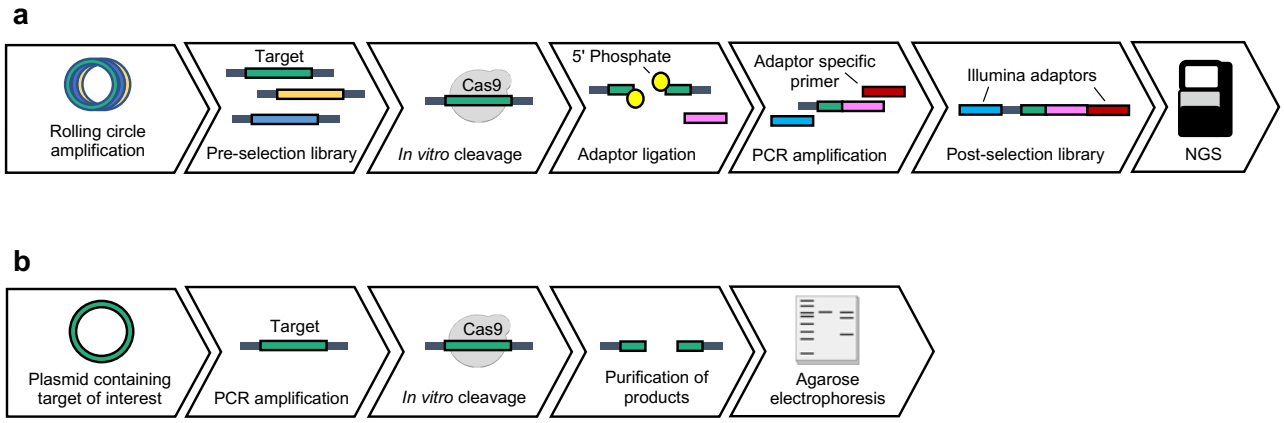


Figure 3.2. Diagram of *in vitro* specificity and activity assays
(**a**) Diagrammatic representation of the high-throughput specificity profiling assay used. Briefly, pre-selection libraries for specificity profiling are generated through rolling circle amplification, using a partially randomized, circular ssDNA as template. The resulting pre-selection library is composed of ~10^{12} potential off-target sequences based on the on-target sequence of interest. The library is digested with Cas9 in complex with a gRNA of interest. Cleaved library members contain a 5′ phosphate group, and are therefore candidates for phosphate-dependent adaptor ligation. These sequences facilitate amplification of the post-selection library, as well as adding necessary components for next-generation sequencing. (**b**) Diagrammatic representation of the *in vitro* cleavage assay. dsDNA substrates for Cas9 activity assays are generated through PCR amplification of a plasmid template containing the on-target site of interest. Cas9 complexed with a gRNA of interest is then used to digest the target DNA, followed by resolving of the cleavage products through agarose gel electrophoresis.

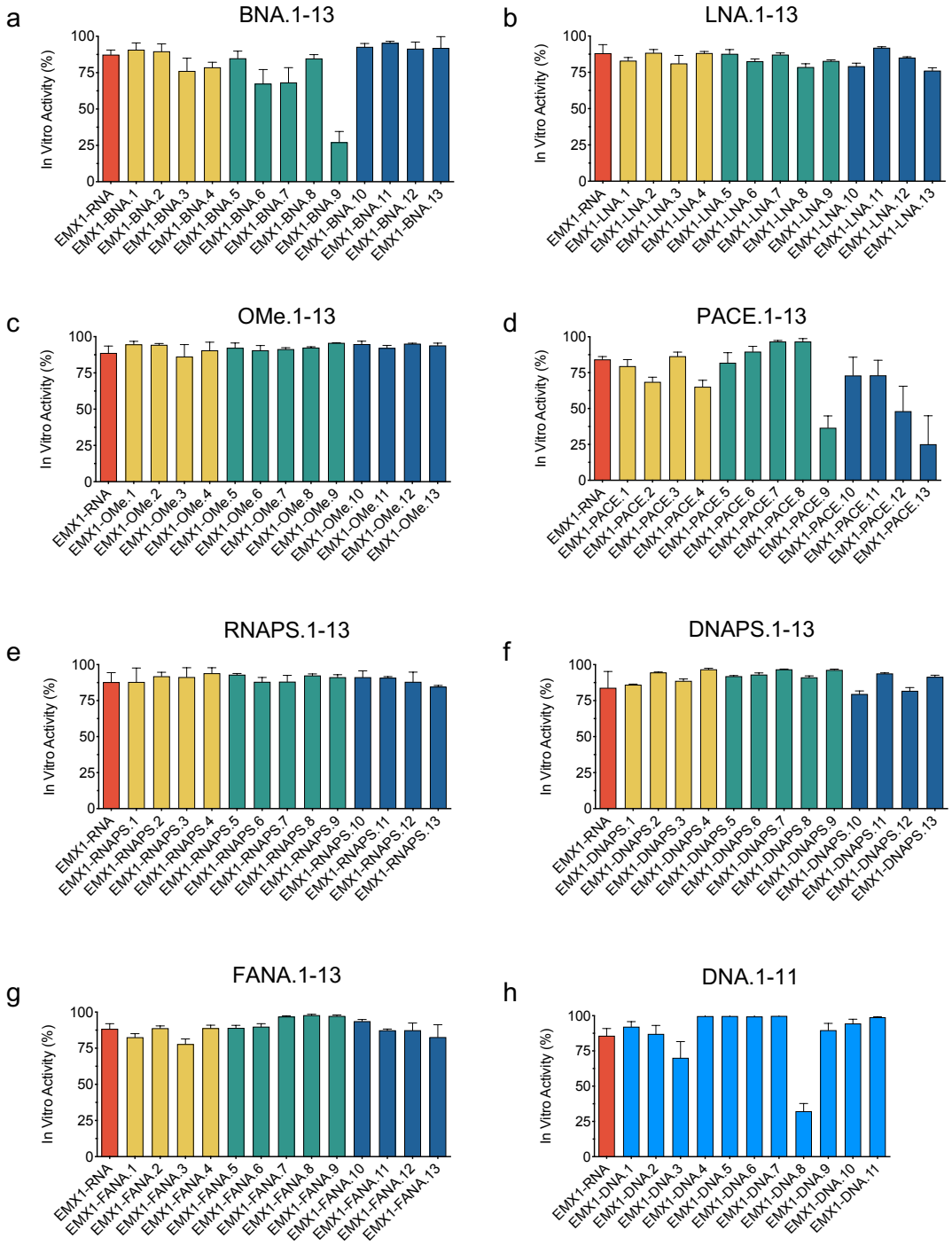


Figure 3.3. On-target activity conservation of crXNAs is dependent on modification type, number and position

Bar graphs depicting *in vitro* cleavage activity of crXNAs containing (**a**) BNA, (**b**) LNA, (**c**) OMe, (**d**) PACE, (**e**) RNAPS, (**f**) DNAPS, (**g**) FANA or (**h**) DNA modifications positioned as in **Figure 3.1a** and **b** (n = 3); Mean ± SEM shown. *In vitro* cleavage reactions were performed using 25 nM of Cas9 RNP complex and 5 nM of EMX1 target DNA.

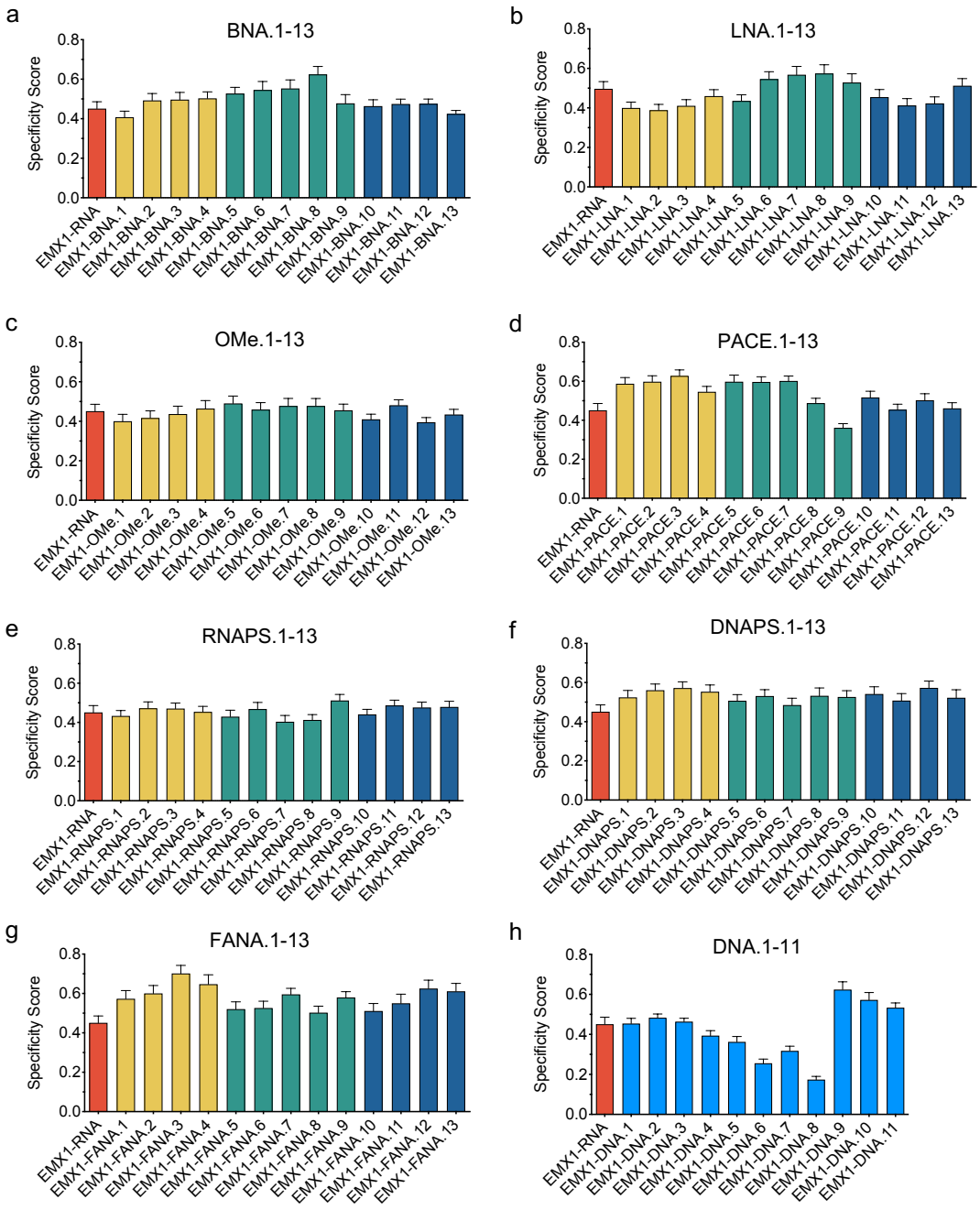


Figure 3.4. crXNA-mediated specificity increases are dependent on modification type, number and position

Bar graphs depicting *in vitro* cleavage activity of crXNAs containing (**a**) BNA, (**b**) LNA, (**c**) OMe, (**d**) PACE, (**e**) RNAPS, (**f**) DNAPS, (**g**) FANA or (**h**) DNA modifications positioned as in **Figure 3.1a** and **b**. Mean ± SEM shown; Specificity scores were calculated as an average of each enrichment score for a particular base at each position along the 20 nt spacer sequence. Specificity profiling reactions were performed using 1000 nM of Cas9 RNP complex and 200 nM of pre-selection DNA library.

Table 3.1. Specificity scores and *in vitro* cleavage assay data used to generate Figures 3.2-3.4

crXNA	Specificity Score	Specificity Score (SEM)	*In Vitro* Activity (%)	*In Vitro* Activity (SEM)
FANA.3	0.702	0.0408	79.8	2.0
FANA.4	0.647	0.0480	90.1	1.2
PACE.3	0.628	0.0305	85.5	1.7
FANA.12	0.625	0.0426	87.0	2.9
BNA.8	0.625	0.0401	85.3	1.5
DNA.9	0.624	0.0388	92.0	2.8
FANA.13	0.611	0.0400	86.1	4.9
PACE.7	0.601	0.0252	96.4	0.4
FANA.2	0.600	0.0398	89.4	0.9
PACE.2	0.597	0.0304	70.2	1.9
PACE.5	0.597	0.0341	85.9	4.0
PACE.6	0.596	0.0265	89.5	2.1
FANA.7	0.596	0.0304	97.2	0.3
PACE.1	0.587	0.0318	81.5	2.6
FANA.9	0.580	0.0292	97.0	0.4
LNA.8	0.574	0.0438	79.9	1.3
FANA.1	0.573	0.0413	84.0	1.4
DNAPS.12	0.573	0.0347	81.8	1.3
DNA.10	0.572	0.0371	96.0	1.7
DNAPS.3	0.572	0.0319	88.0	0.7
LNA.7	0.568	0.0422	87.6	0.7
DNAPS.2	0.561	0.0316	94.7	0.2
BNA.7	0.553	0.0433	69.0	5.8
DNAPS.4	0.553	0.0346	96.3	0.4
FANA.11	0.550	0.0462	87.7	0.5
PACE.4	0.546	0.0274	67.7	2.6
LNA.6	0.546	0.0375	83.5	0.9
BNA.6	0.546	0.0428	71.0	5.6
DNAPS.10	0.542	0.0365	80.8	1.2
DNA.11	0.534	0.0233	99.0	0.2
DNAPS.8	0.532	0.0400	90.6	0.6
DNAPS.6	0.531	0.0327	93.7	0.6
LNA.9	0.529	0.0441	83.2	0.4
BNA.5	0.527	0.0313	85.1	2.8
FANA.6	0.526	0.0345	91.1	1.2
DNAPS.9	0.526	0.0321	96.6	0.3
DNAPS.1	0.525	0.0351	85.8	0.2
DNAPS.13	0.522	0.0406	91.3	0.6
FANA.5	0.521	0.0370	90.1	1.0
PACE.10	0.517	0.0317	73.3	7.3
LNA.13	0.512	0.0364	77.3	1.1
RNAPS.9	0.512	0.0307	91.5	1.0

FANA.10	0.511	0.0379	94.4	0.7
DNAPS.11	0.508	0.0355	93.9	0.2
DNAPS.5	0.507	0.0313	91.8	0.2
BNA.4	0.503	0.0329	79.3	2.0
PACE.12	0.503	0.0332	57.1	10.0
FANA.8	0.503	0.0328	97.9	0.3
BNA.3	0.497	0.0364	75.5	5.1
BNA.2	0.493	0.0349	89.6	3.0
OMe.5	0.490	0.0374	92.3	1.9
PACE.8	0.488	0.0248	97.8	1.2
RNAPS.11	0.487	0.0270	91.4	0.4
DNAPS.7	0.485	0.0351	96.8	0.1
DNA.2	0.483	0.0185	90.1	3.5
OMe.11	0.481	0.0275	92.2	1.0
RNAPS.13	0.479	0.0289	84.8	0.5
BNA.9	0.478	0.0441	27.2	4.2
OMe.7	0.478	0.0377	91.4	0.6
OMe.8	0.477	0.0376	92.4	0.3
RNAPS.12	0.476	0.0275	86.5	3.9
BNA.12	0.476	0.0237	91.1	2.6
BNA.11	0.475	0.0245	95.9	0.6
RNAPS.2	0.472	0.0322	91.6	1.6
RNAPS.3	0.471	0.0282	89.6	3.7
RNAPS.6	0.468	0.0336	87.6	1.7
OMe.4	0.464	0.0403	90.5	3.2
BNA.10	0.464	0.0317	94.1	1.3
DNA.3	0.464	0.0167	74.4	6.7
PACE.13	0.461	0.0287	35.2	11.4
LNA.4	0.460	0.0326	88.9	0.6
OMe.6	0.459	0.0356	90.6	1.9
OMe.9	0.455	0.0315	95.7	0.1
LNA.10	0.455	0.0386	80.4	1.2
DNA.1	0.454	0.0259	94.1	2.1
PACE.11	0.454	0.0281	79.1	6.1
RNAPS.4	0.453	0.0283	93.1	2.2
RNA	0.451	0.0349	87.4	1.8
RNAPS.10	0.441	0.0257	90.4	2.5
OMe.3	0.437	0.0396	86.2	4.8
LNA.5	0.435	0.0314	87.1	1.7
OMe.13	0.434	0.0259	94.0	0.9
RNAPS.1	0.433	0.0274	85.7	5.6
RNAPS.5	0.429	0.0327	93.4	0.5
BNA.13	0.425	0.0171	96.3	4.5
LNA.12	0.423	0.0327	85.5	0.3
OMe.2	0.417	0.0360	94.4	0.5
LNA.11	0.413	0.0345	92.1	0.4

RNAPS.8	0.413	0.0272	92.7	0.6
LNA.3	0.410	0.0313	81.5	3.2
OMe.10	0.410	0.0261	94.9	1.2
BNA.1	0.408	0.0305	90.2	2.7
RNAPS.7	0.403	0.0322	87.5	2.5
OMe.1	0.400	0.0350	94.7	1.2
LNA.1	0.400	0.0297	84.2	1.3
OMe.12	0.395	0.0235	95.2	0.2
DNA.4	0.394	0.0247	100.0	0.3
LNA.2	0.389	0.0292	89.2	1.3
DNA.5	0.363	0.0268	100.0	0.3
PACE.9	0.361	0.0215	41.0	4.8
DNA.7	0.318	0.0234	100.0	0.2
DNA.6	0.256	0.0209	100.0	0.6
DNA.8	0.174	0.0161	31.2	3.1

Values showing average specificity scores and *in vitro* cleavage activity percentages ($n = 3$) for the 102 tested crXNAs and an unmodified crRNA control targeting EMX1; Mean ± SEM. Specificity profiling experiments were performed using 1000 nM of Cas9 RNP complex and 200 nM of pre-selection DNA library. *In vitro* activity assays were performed using 25 nM of Cas9 RNP complex and 5 nM of target DNA. crXNAs are ranked by average specificity score.

3.3.3. crXNAs display unique specificity profiles based on modified nucleotide position

Aside from the averaged specificity scores as a marker for overall off-target activity (**Figure 3.1d**, **e** and **Figure 3.4**), we were also able to determine detailed specificity profiles for each of the assayed crXNAs. To visualize the crXNA-mediated effect on off-target activity, we plotted the difference in specificity score between a given crXNA and an unmodified control for each position along the 20 nt spacer sequence (**Figures 3.5-3.12**). These results collectively revealed that type and position of XNA incorporation not only have dramatic effects on the resulting overall specificity of Cas9, but also implications for the type of off-target in which the crXNA may reduce cleavage of. For example, PACE.3 and FANA.3 (modified with a triplet of XNAs at positions 4-6 of the crRNA) were both identified as crXNAs displaying high degrees of specificity and activity (**Table 3.1**), but their respective specificity profiles differ significantly.

PACE.3 was demonstrated to be largely uniform in its specificity profile, with the largest increases in specificity situated at the 3′ end of the spacer (**Figure 3.8**). Conversely, FANA.3 showed the largest differences at the 5′ end of the spacer, diminishing relatively uniformly towards the 3′ end (**Figure 3.11**). This data indicates that while the overall specificity scores of particular crXNAs may be similar, the off-targets in which they affect may differ.

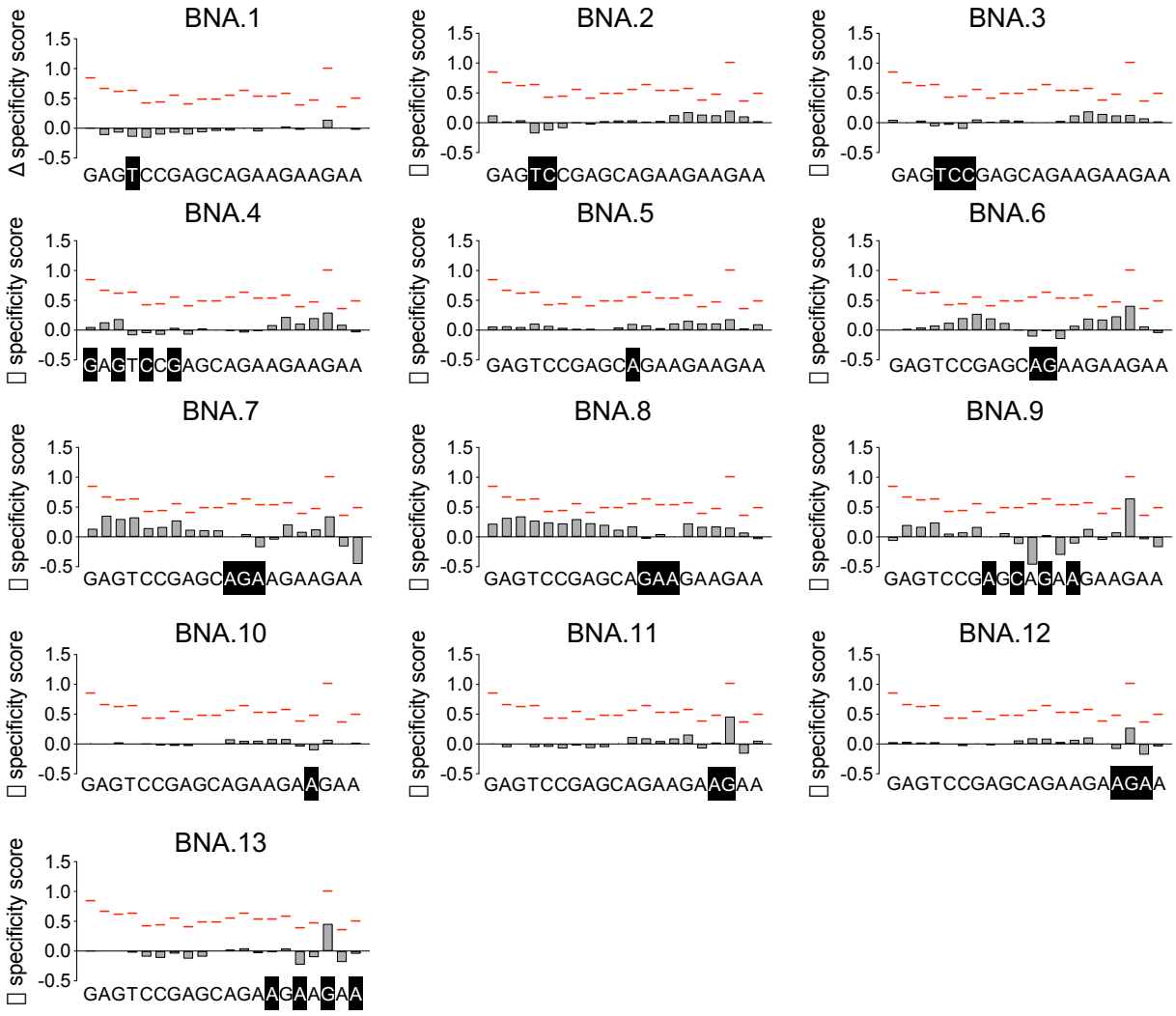


Figure 3.5. Change in specificity score of BNA-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

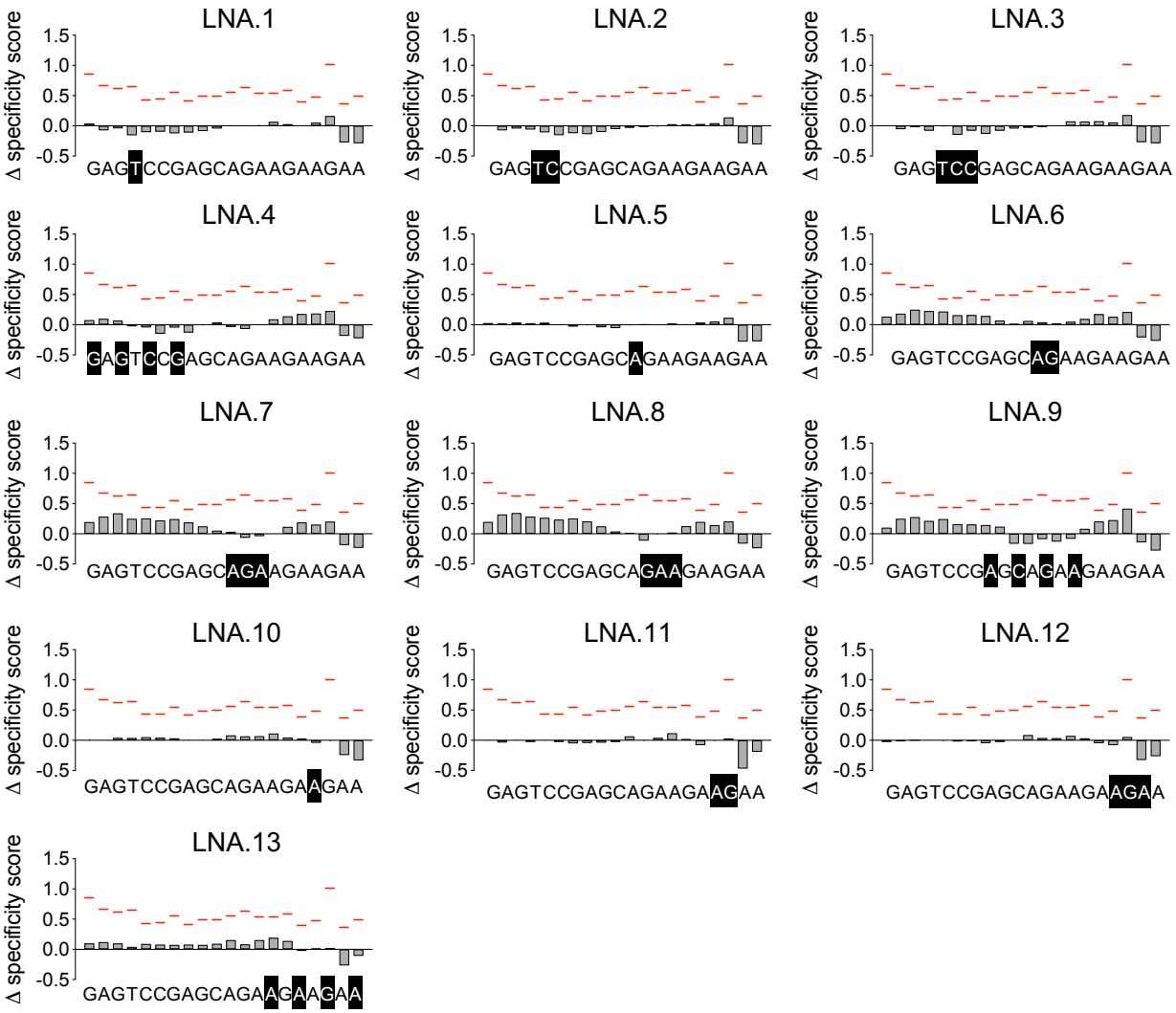


Figure 3.6. Change in specificity score of LNA-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

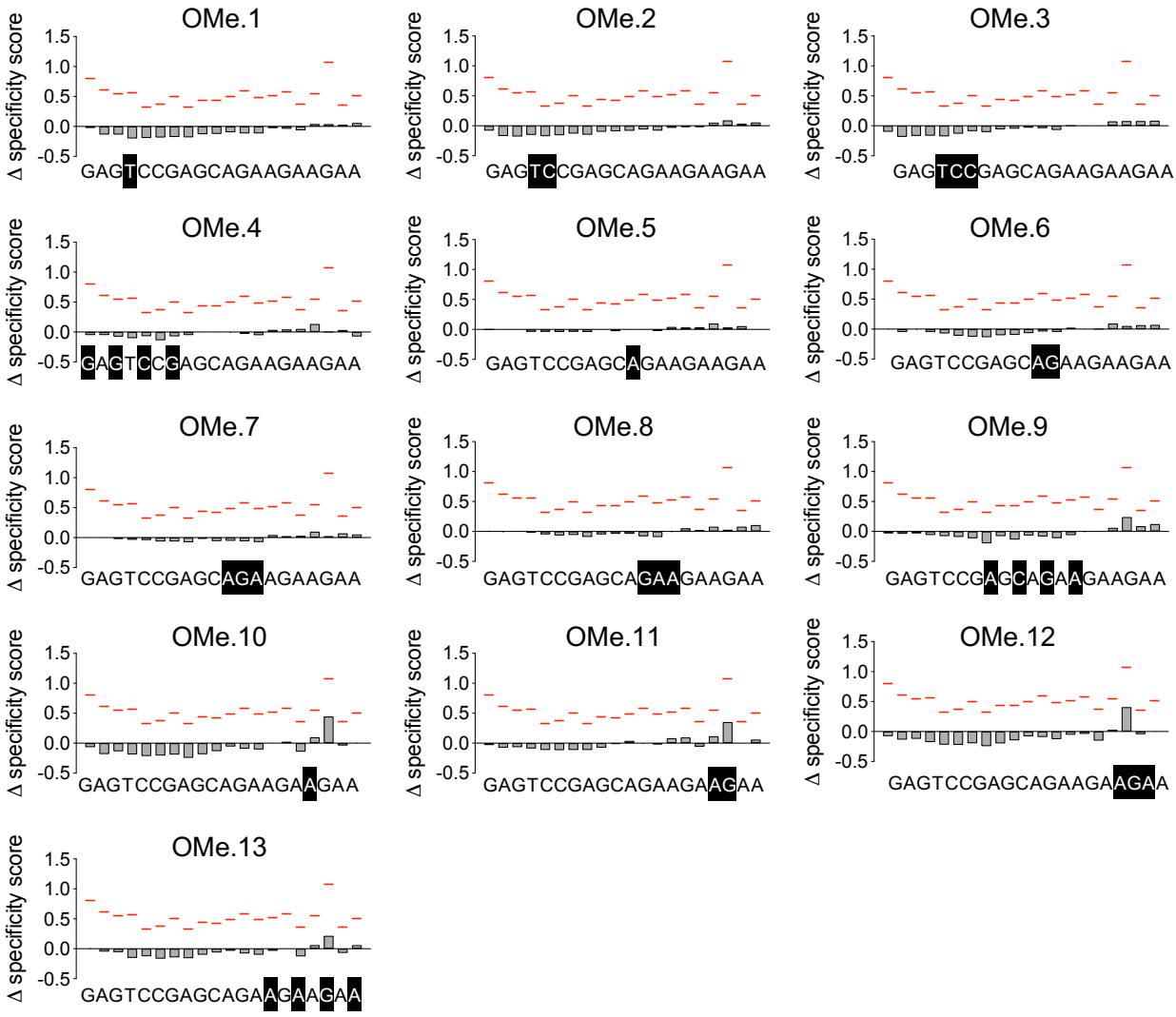


Figure 3.7. Change in specificity score of OMe-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

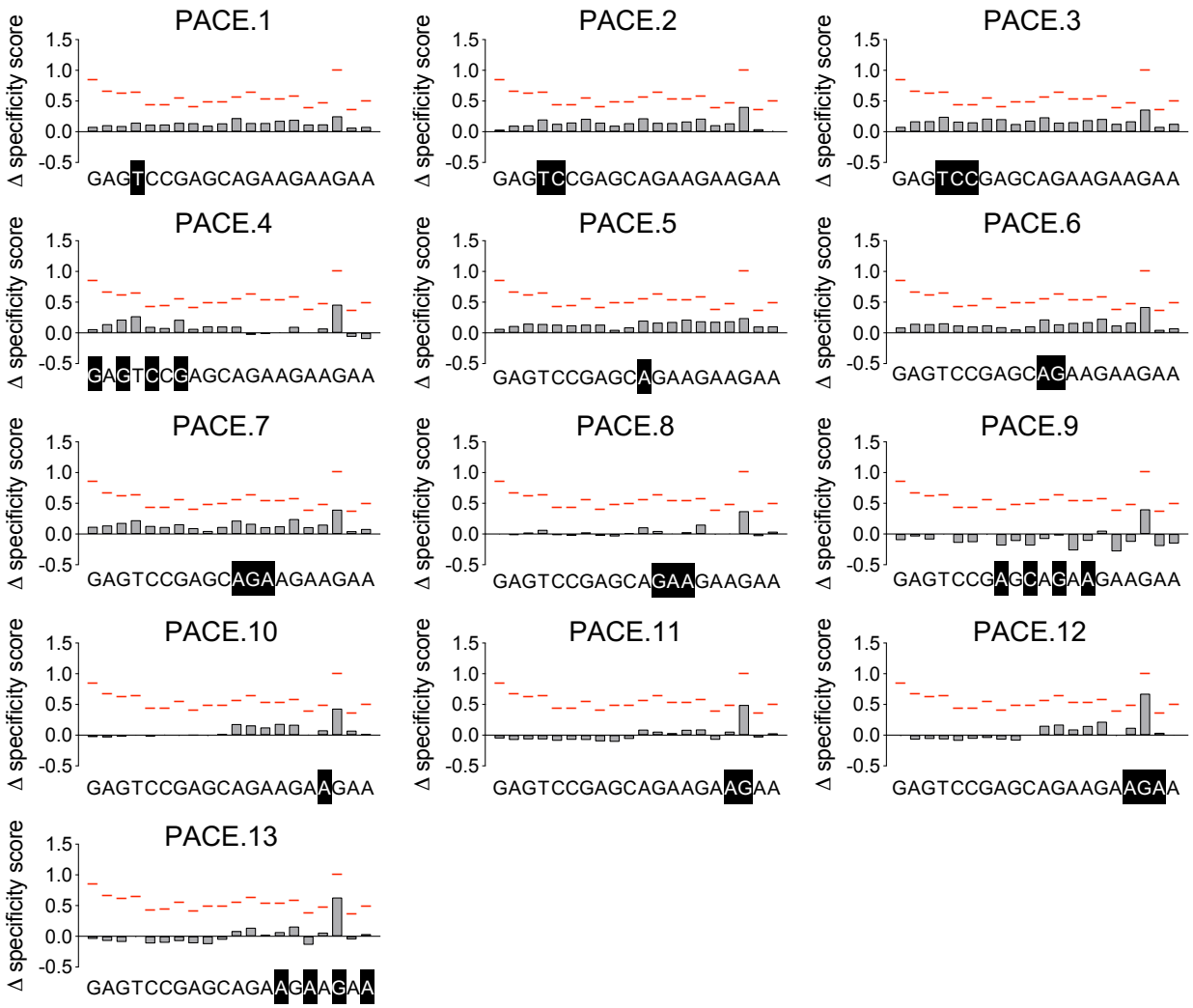


Figure 3.8. Change in specificity score of PACE-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

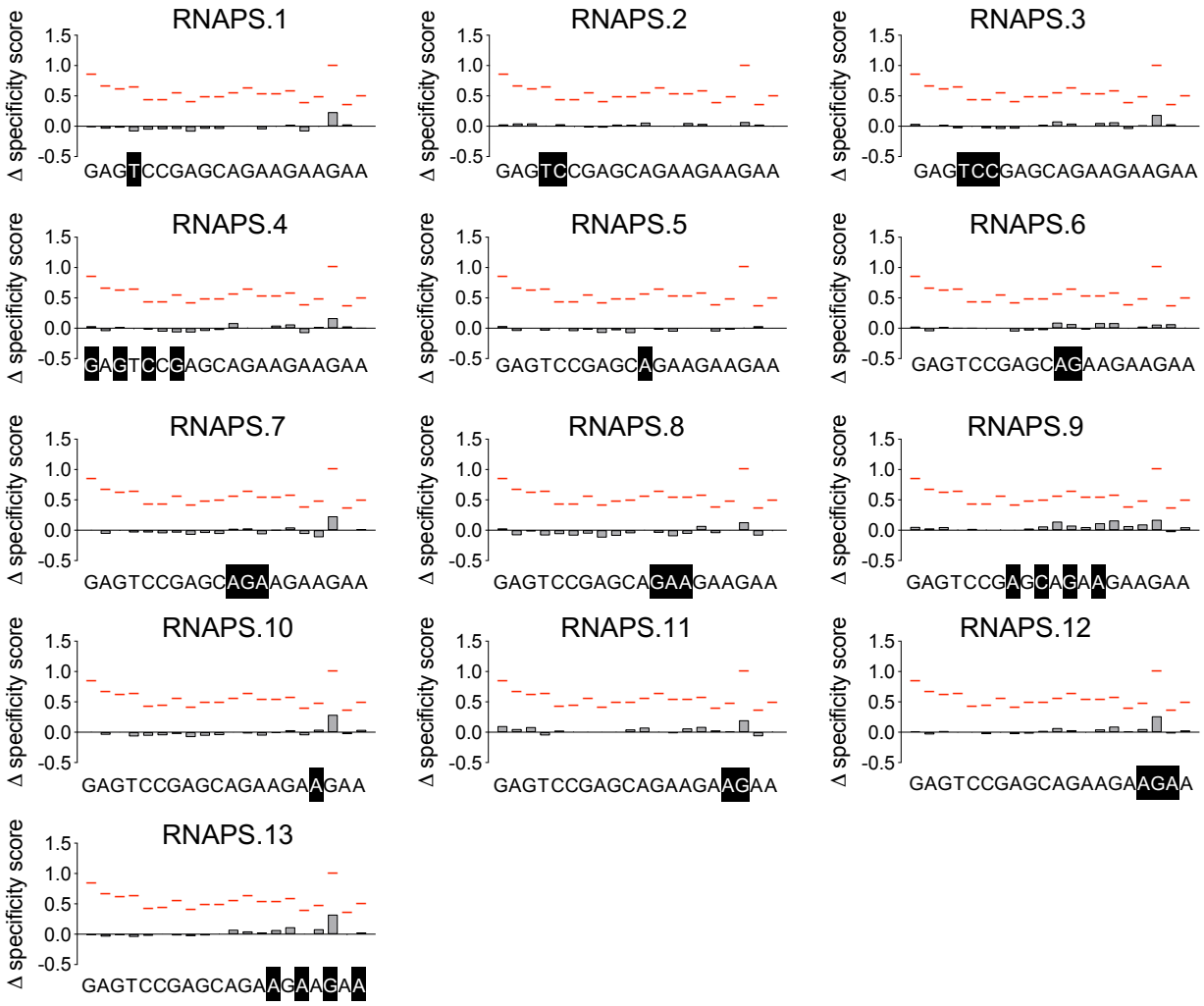


Figure 3.9. Change in specificity score of RNAPS-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

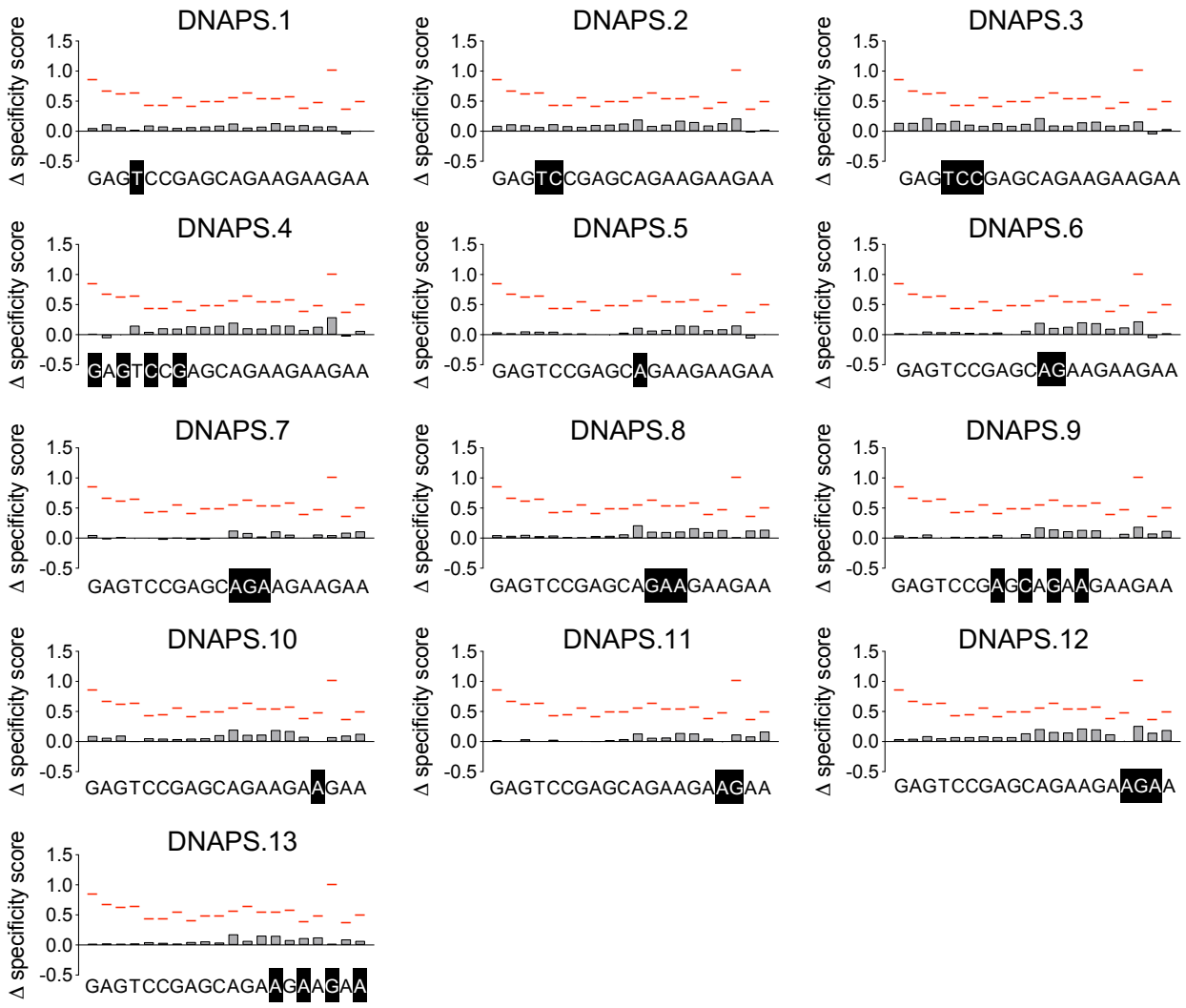


Figure 3.10. Change in specificity score of DNAPS-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

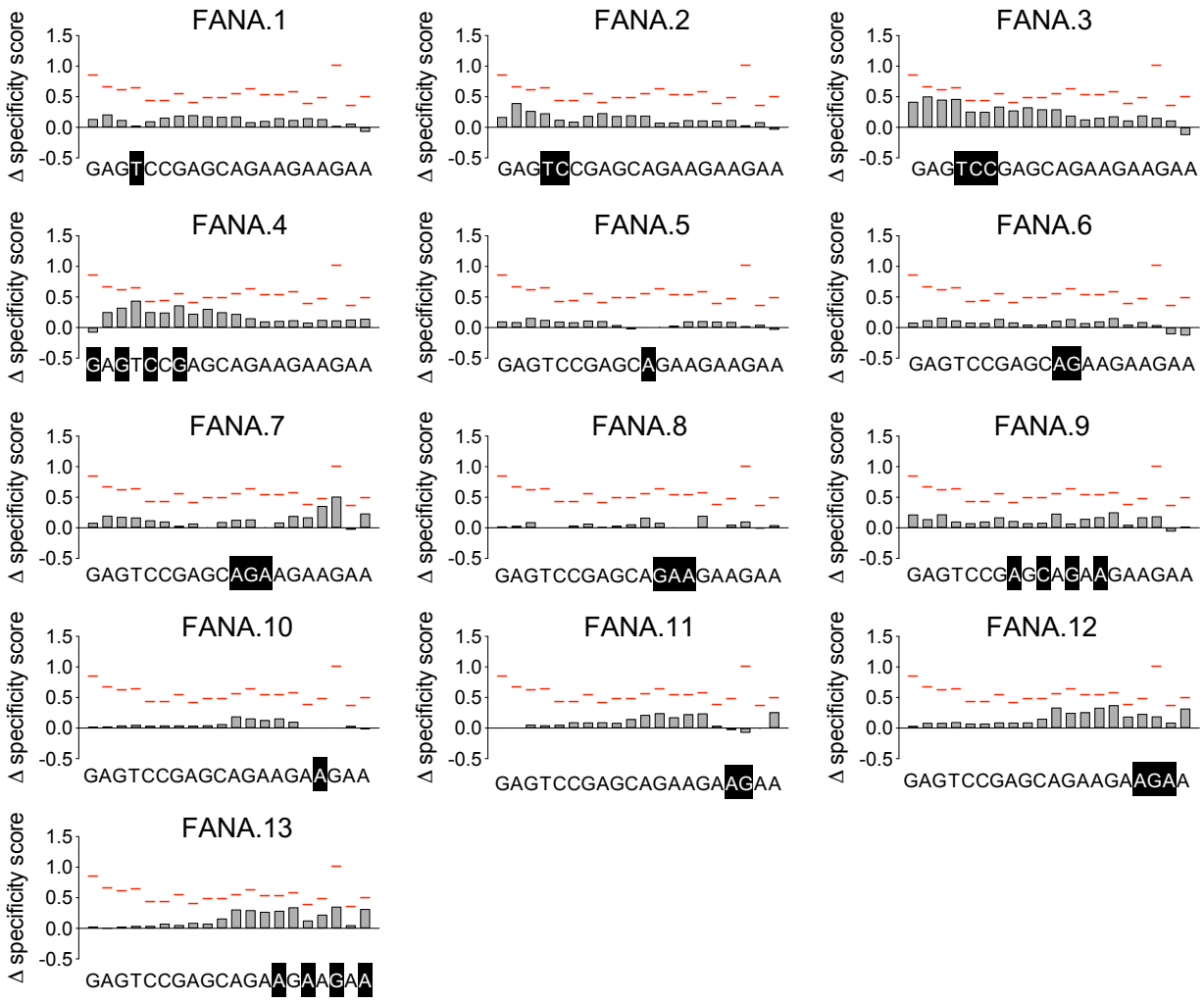


Figure 3.11. Change in specificity score of FANA-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

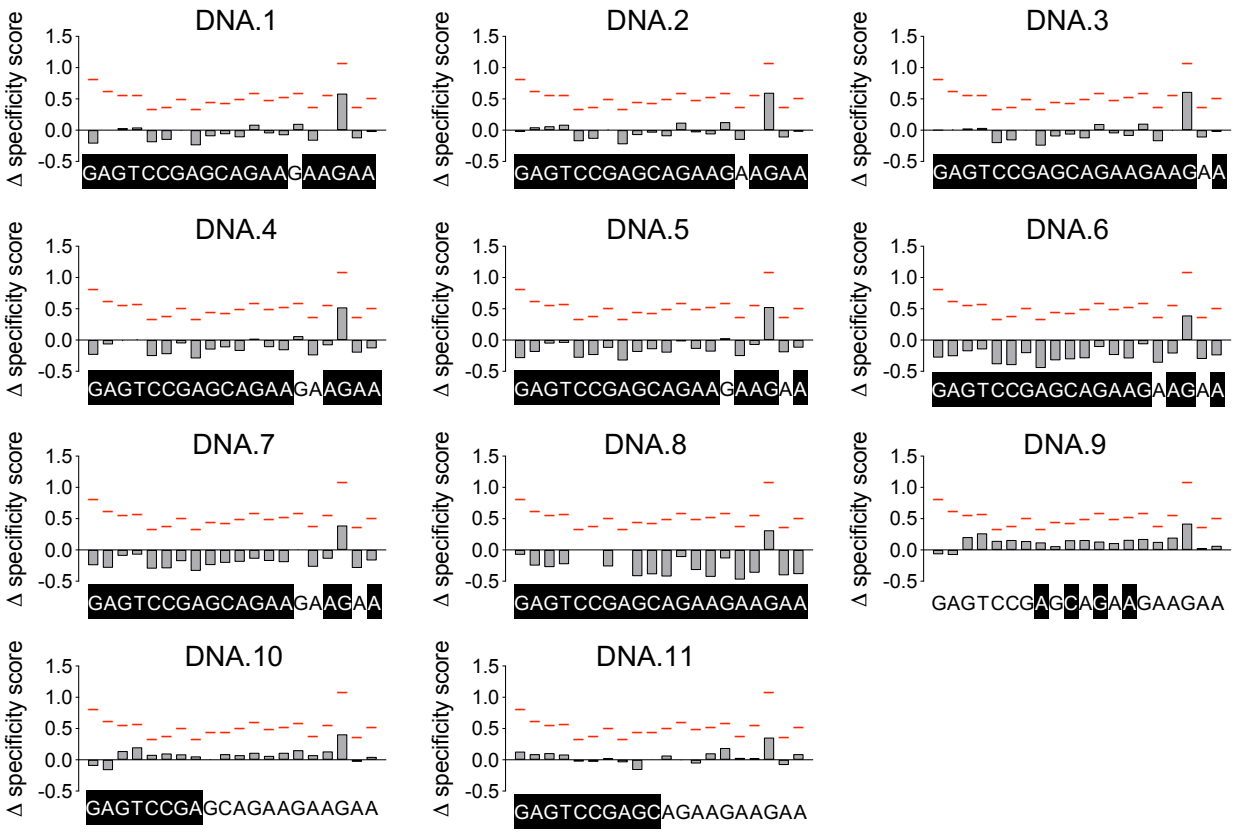


Figure 3.12. Change in specificity score of DNA-modified crXNAs
Bar graphs showing the difference in specificity score at each position along the crRNA spacer sequence relative to an unmodified crRNA control. Modification positions are indicated by a black background, while red lines indicate the maximal possible specificity increase for a given position calculated using an unmodified crRNA. A score of zero indicates no change in specificity between the samples. ΔSpecificity score was calculated as specificity score(crXNA) – specificity score(crRNA). The specificity score of each nucleotide at each position is relative to the starting pre-selection library.

3.3.4. Specificity and activity of crXNAs is preserved across DNA target sequences

Based on our observations, we were interested as to whether the crXNA-mediated specificity improvements were unique to the tested EMX1 target, or were applicable to a broader range of sequences. To accomplish this, we designed a series of 10 Cas9 targets computationally identified to be diverse across a number of biochemical properties, including GC content, melting temperature (T_m), nucleotide count and sequence composition (**Table 3.2**). Of the crXNAs listed in **Table 3.1**, we identified the top 20 candidates based on overall specificity score, eliminating those which showed on-target activity significantly lower than the corresponding unmodified crRNA (**Table 3.3**). We also decided to not include BNA or PACE XNAs in our subsequent experiments, largely due to difficulty in acquiring the necessary materials required for synthesis, but also because of previously demonstrated loss of activity when used in a cellular context[194]. The resulting 20 crXNAs were synthesized against the 10 Cas9 target sequences (along with 10 additional unmodified crRNAs; for a total of 200 crXNAs and 10 crRNAs) and again profiled for their activity and specificity using the methods described above (**Figure 3.13**). (It should be noted that at the time of writing this thesis, several crXNAs have yet to be synthesized and are therefore not represented in the following figures. These will be included in the final version of the proposed manuscript, and the figures will be updated accordingly.) Interestingly, our results indicated that the specificity and activity of crXNAs appears to be largely independent of sequence composition, as the diversity seen for both metrics largely mirrored what we observed for the unmodified crRNA (**Figure 3.14**). We also observed what appears to be an upper limit in terms of specificity score, with several crXNAs (FANA.2, FANA.3, FANA.4 and FANA.7) showing upper values of ~0.7 (a score of 1.0 is the theoretical maximum for perfect specificity). In almost all cases tested, we were

able to demonstrate that crXNAs led to higher observed specificities regardless of the target tested, albeit to varying degrees (**Figure 3.15**). Overall, the scores identified across the 10 divergent target sites largely reflected what we observed against EMX1, indicating that type and position of modification are the major contributing factors to specificity, not necessarily sequence composition. However, within the crXNAs tested we did identify that some crXNAs were more universally tolerable of different Cas9 DNA targets, as indicated by tighter distributions of data points shown in **Figures 3.13** and **3.14**.

Table 3.2. Cas9 target sites displaying a diverse set of biochemical properties

Gene	Sequence	T_m (°C)	GC (%)	A	T	C	G
AAVS1	GGGGCCACTAGGGACAGGAT	61.2	65	5	2	4	9
ATF1	TAGGAATCAAACACTTTTAT	44.1	25	8	7	3	2
CDH1	TGACTTGCGAGGGACGCATT	59.5	55	4	5	4	7
EIF3D	AGACGACCCTGTCATCCGCA	60.9	60	5	3	8	4
GABPA-1	GAAAAGGATAATTGAGCCCC	51.1	45	8	3	4	5
GABPA-2	TTTGGAGTCTCAGAATGTCC	51.8	45	4	7	5	4
MALAT1-1	AATGTGAAGGACTTTCGTAA	49.2	35	7	6	2	5
MALAT1-2	GGCAGGAGAGGCCAGTTGCG	63.6	70	4	2	4	10
MALAT1-3	GCTGGGGCTCAGTTGCGTAA	60.4	60	3	5	4	8
SPOUT1	CAGGCGGGCTCACCTCCGTG	65.6	75	2	3	8	7

List of Cas9 target sequences utilized with accompanying biochemical properties. T_m was determined using the IDT OligoAnalyzer Tool.

Table 3.3. Top 20 crXNAs based on high levels of observed specificity and retention of on-target activity

crXNA	Specificity Score	Specificity Score (SEM)	*In Vitro* Activity (%)	*In Vitro* Activity (SEM)
FANA.3	0.702	0.0408	79.8	2.0
FANA.4	0.647	0.0480	90.1	1.2
FANA.12	0.625	0.0426	87.0	2.9
DNA.9	0.624	0.0388	92.0	2.8
FANA.13	0.611	0.0400	86.1	4.9
FANA.2	0.600	0.0398	89.4	0.9
FANA.7	0.596	0.0304	97.2	0.3
FANA.9	0.580	0.0292	97.0	0.4
LNA.8	0.574	0.0438	79.9	1.3
FANA.1	0.573	0.0413	84.0	1.4
DNAPS.12	0.573	0.0347	81.8	1.3
DNA.10	0.572	0.0371	96.0	1.7
DNAPS.3	0.572	0.0319	88.0	0.7
LNA.7	0.568	0.0422	87.6	0.7
DNAPS.2	0.561	0.0316	94.7	0.2
DNAPS.4	0.553	0.0346	96.3	0.4
FANA.11	0.550	0.0462	87.7	0.5
LNA.6	0.546	0.0375	83.5	0.9
DNAPS.10	0.542	0.0365	80.8	1.2
DNA.11	0.534	0.0233	99.0	0.2

List of crXNAs determined to be the most effective at improving Cas9 cleavage specificity, while maintaining on-target activity *in vitro* (using a crRNA against EMX1). Mean of *in vitro* cleavage activity shown ($n = 3$). Specificity score is shown as the mean of each score for each nucleotide across the 20 nt crRNA sequence. Specificity profiling experiments were performed using 1000 nM of Cas9 RNP complex and 200 nM of pre-selection DNA library. *In vitro* activity assays were performed using 25 nM of Cas9 RNP complex and 5 nM of target DNA. crXNAs are ranked by average specificity score.

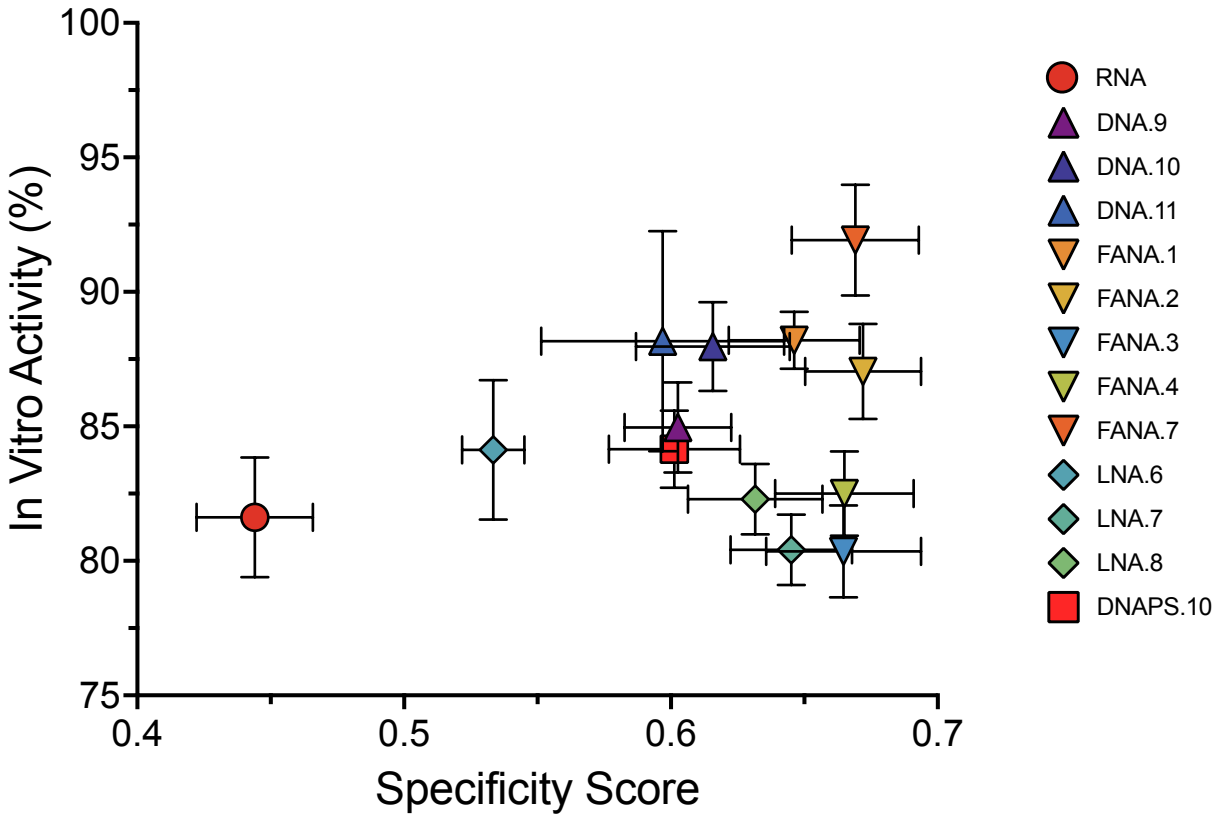


Figure 3.13. crXNAs are capable of highly specific, and highly efficient *in vitro* DNA cleavage

In vitro specificity and activity scores of individual crXNAs against 10 divergent Cas9 target sites. *In vitro* activity was determined through independent cleavage assays using a crXNA and corresponding target sequence (n = 3). Mean ± SEM for all 10 targets is shown. Specificity score is shown as a mean for all 10 targets ± SEM. *In vitro* cleavage assays were performed using 25 nM of Cas9 RNP complex and 5 nM of target DNA, while specificity profiling was performed using 1000 nM of Cas9 RNP complex and 200 nM of pre-selection library.

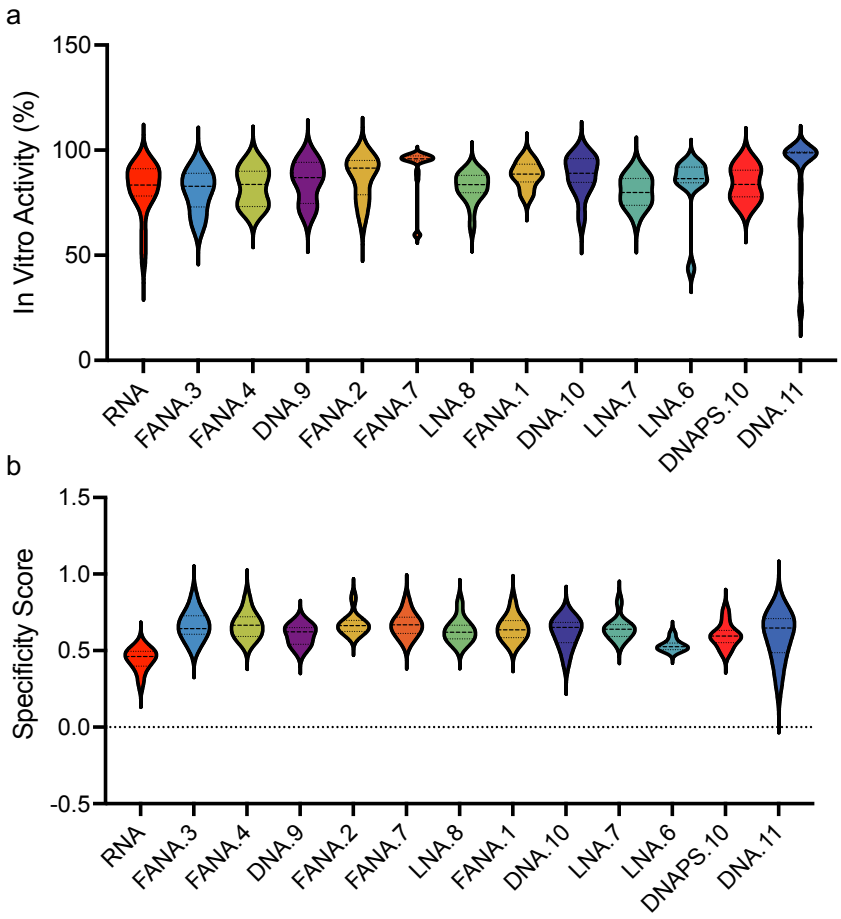


Figure 3.14. crXNA activity and specificity is preserved across diverse target sequences
Violin plots showing the diversity in (**a**) on-target activity or (**b**) specificity against 10 independent Cas9 target sequences (listed in **Table 3.2**). Dotted lines are indicative of the mean, while thinner solid lines represent the interquartile range.

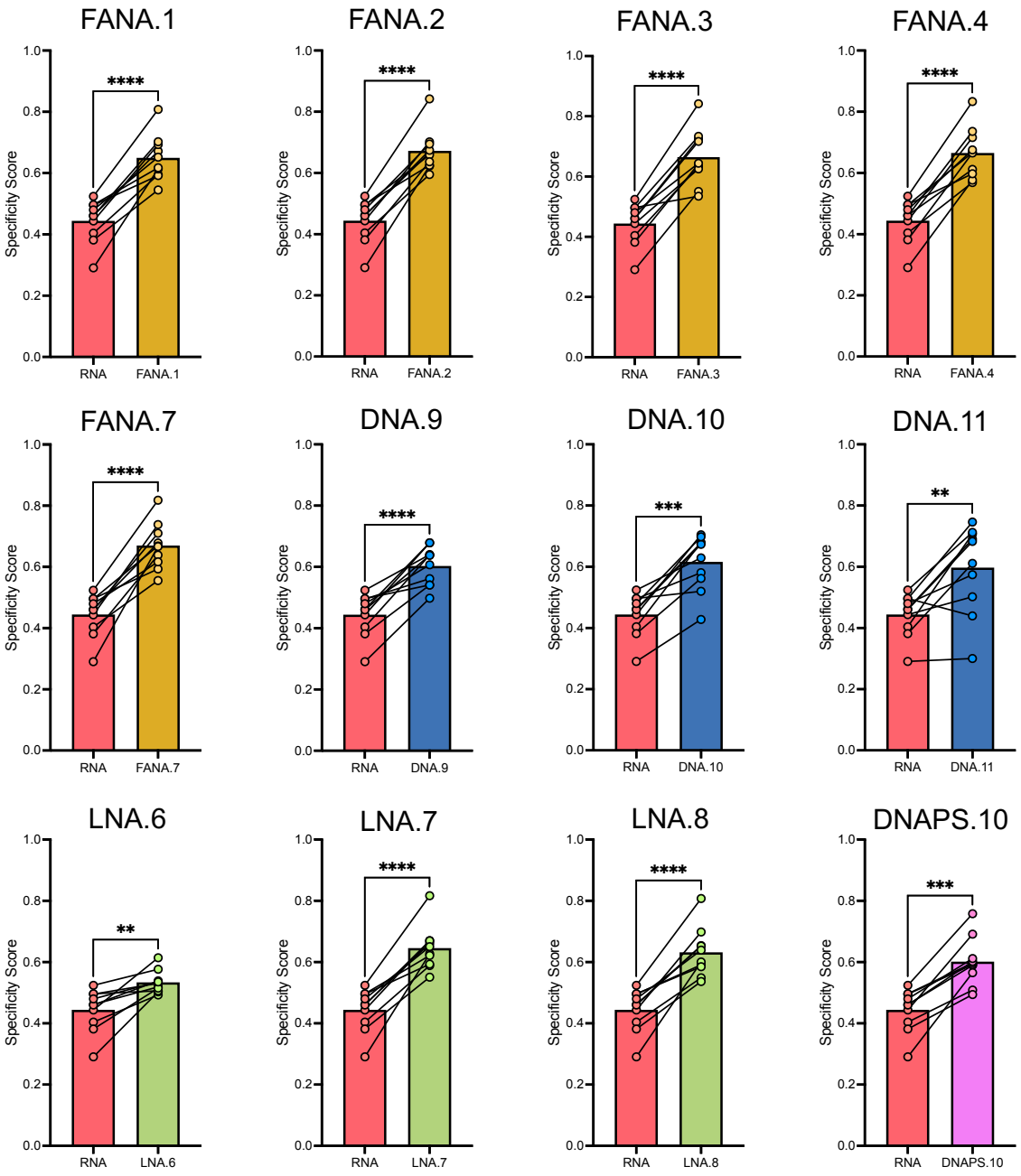


Figure 3.15. crXNAs broadly improve Cas9 specificity *in vitro*
Bar graphs showing the change in specificity between an unmodified crRNA and corresponding crXNA as determined by high-throughput specificity profiling. Lines are shown connecting the average specificity score for a particular target digested with either an unmodified or XNA-modified crRNA. Bar graphs indicate the mean for each crXNA modification pattern across all targets sites tested. Cas9 DNA target sequences are listed in **Table 3.2.** *P* values were calculated using an unpaired t test and are shown above each graph.

3.4. Discussion

Previous reports have demonstrated that Cas9 is highly tolerable of chemical modifications made to the sequence-specific spacer region of the crRNA[188, 194, 195, 243]. Consistent with these findings, we have demonstrated that incorporation of certain XNAs at defined positions within the crRNA broadly improves *in vitro* specificity, without significant loss of activity. However, this is not without restriction, as we demonstrated that careful considerations to the modification used, as well as its position of incorporation are critical to maximizing the resulting specificity improvements. We also observed several crXNAs with significant losses to on-target activity, specifically those modified with BNA or PACE (using MOD.9) (**Figure 3.1d**, **e** and **Figure 3.3a**, **d**). As both BNA and PACE XNAs display increased melting temperature of duplexes following their incorporation[194, 195], we hypothesize that this decrease in activity is due to mis-folding of the gRNA through stabilization of unwanted secondary structures (**Figure 3.16**), and therefore should be taken into account in future studies. Fully modified DNA crRNAs also showed little on-target activity, although this is consistent with previous reports demonstrating Cas9 is intolerant of a complete DNA spacer (**Figure 3.1d**, **e** and **Figure 3.3h**)[202, 203].

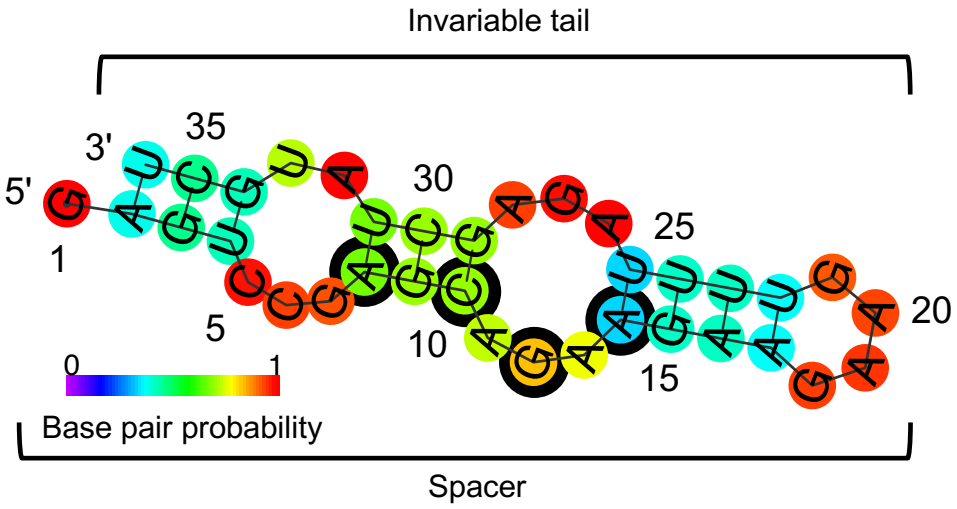


Figure 3.16. Predicted crRNA secondary structure
Graphical depiction of a predicted secondary structure for a Cas9 crRNA based on minimum free energy. Minimum free energy structure prediction was performed using *RNAfold* WebServer. Nucleotides are color coded based on their base-pairing probability. Position of XNAs for MOD.9 are indicated by a black outline around the nucleotide.

Interestingly, our results showed that a diverse set of crXNAs were capable of eliciting reduced off-target activity, including those modified with BNA, LNA, PACE or FANA XNAs (as well as chimeric DNA-RNA crRNAs) (**Figure 3.1d**, **e** and **Figure 3.4**). These XNAs have all been shown to influence the stability of duplex hybridization, either non-specifically[195] or through mismatch-induced destabilization, indicating a potential mechanism of action[220, 223, 244]. During our initial profiling of specificity, we also observed a disproportionate distribution of modification patterns within our highest performing crXNAs. Specifically, of the 20 most specific crXNAs against EMX1 (**Table 3.1**), 8 contained modifications to the distal region, 7 to the mid region and only 3 to the proximal region (the remaining 2 were DNA-based, hybrid crRNAs). This is consistent with previous reports showing that the 10-12 nt of the 3′ end of the spacer (known as the seed region) is critical for Cas9 off-target discrimination[28]. Crystal structures of Cas9 in complex with a sgRNA have demonstrated that the seed region is relatively rigid and pre-ordered as an A-form helix for specific DNA binding[28, 46]. Therefore incorporation of XNAs into this region may not have as significant an impact on specificity, as the specificity of this region is already relatively stringent.

Structural studies regarding the biochemical mechanisms of Cas9 proof-reading may also provide additional insight into the crXNA-mediated increases in specificity observed when modifying either the PAM distal, or mid regions of the crRNA spacer. When in a gRNA-bound state, the REC2 domain of Cas9 is positioned such that it occludes the HNH nuclease domain from accessing

the scissile phosphate of the target DNA strand[55]. Following DNA binding and subsequent R-loop formation, REC3 makes several contacts with the distal/mid region of the RNA/DNA heteroduplex, stabilizing it in a "kinked" conformation (which contains a 70° bend relative to the linear duplex observed prior to R-loop formation)[60]. This kinked conformation facilitates a large outward rotation of REC2, allowing the HNH domain to re-position for DNA cleavage[28, 245]. Mismatches present within the interacting regions of the heteroduplex impair REC3 docking, preventing transition to the kinked state and lock Cas9 in an inactive checkpoint conformation[60, 177]. In this mechanism, Cas9 mismatch discrimination is therefore dependent on the ability of the heteroduplex to adopt suitable geometry required for proper REC3 binding, and not necessarily the sequence composition of the target. As XNAs are known to induce changes in duplex structure following their incorporation[246], it is reasonable to assume that these may be influencing the ability of REC3 to properly recognize the heteroduplex in the presence of an off-target DNA substrate. Within the REC3 domain, 4 clusters of amino acids have been shown to interact with the RNA/DNA heteroduplex, along with an additional cluster in the HNH-RuvC-Linked 2 (L2) domain (**Figure 3.17**). Clusters 1 and 2 make significant interactions with the target DNA strand at positions 16-20 (corresponding to 1-5 of the crRNA), while clusters 3-5 interact with positions 7-12 of the target DNA (corresponding to 9-14 of the crRNA)[177]. Therefore, it is also possible that modification of these positions with specific XNAs may function to more efficiently disrupt REC3 docking (when bound to an off-target) through unfavorable duplex geometry, than other locations of modification (such as in the PAM proximal region of the crRNA). These clusters within REC3 have also been shown to be highly conserved across Cas9 enzymes derived from bacteria other than *S. pyogenes*[177], indicating the potential use of crXNAs beyond the system tested here.

Moreover, several high-fidelity Cas9 variants have taken advantage of this proof-reading mechanism, mutating residues within REC3 which interact with the heteroduplex. Specifically, Cas9-HF1 contains the mutations N497A, R661A, Q695A and Q926A[176], of which N497 and R661 make contact with the heteroduplex at positions of modification for crXNAs MOD.3, MOD.7 or MOD.8 (**Figure 3.1a** and **Figure 3.17**). Another highly specific Cas9 variant, Cas9 HSC1.1 (N588A, R765A, D835A and K1246A) contains the mutation N588A[247], which is again at a position consistent with crXNA MOD.3 (**Figure 3.1a** and **Figure 3.17**). This is mirrored by our results, which showed that FANA.3 and FANA.7 (both containing modifications to the aforementioned regions of the crRNA) are both examples of highly specific crXNAs (**Figure 3.1d**, **e**, **Figure 3.4g** and **Figure 3.14b**). High-fidelity Cas9 variants containing REC3 mutations have shown to also have reduced rates of catalysis, in addition to their observed increases in specificity[60, 248]. Kinetic experiments have shown that the conformation changes required for DNA cleavage (and mediated by REC3) are the rate-limiting step for Cas9 activation, as well as the main determinant of specificity. This is due to the observance that complete R-loop formation is largely irreversible, and the actual cleavage of DNA via the HNH and RuvC nuclease domains occurs very rapidly. Therefore, by reducing the rate at which conformational changes within Cas9 occur, high-fidelity variants shown a preference towards the release of off-target DNA, rather than its cleavage[248]. Our previous work has demonstrated that crXNAs (specifically those containing BNAs or LNAs) also display reduced reaction kinetics, lending support to the hypothesis that the mechanisms of specificity between several Cas9 variants and crXNAs may be shared. These consistencies between high-fidelity Cas9 variants and some of our highly specific crXNAs provides additional evidence for a potential mechanism of action, although additional structural studies are needed to further elucidate this.

Additionally, while our experiments have utilized synthetic libraries containing large numbers of potential off-target sequences, they have all been performed *in vitro*. Previous studies have demonstrated potential pitfalls when translating *in vitro* specificity enhancements to a cellular environment, specifically loss of on-target activity. This phenomena has been observed for several chemically modified crRNAs[194, 243], as well as for high-fidelity Cas9 variants[151]; often thought to be due to reduced reaction kinetics[60, 177, 194, 243]. Therefore, to reliably apply crXNAs to cellular genome editing applications, we must also ensure highly specific crXNAs retain not only their specificity within a cellular context, but also their on-target efficiency.

Here, we show that XNA incorporation at distinct locations within the sequence-specific region of the crRNA of Cas9 is able to broadly improve specificity, while also maintaining on-target activity *in vitro*. Additionally, we have shown that crXNAs function largely in a sequence-independent manner, able to elicit their improvements regarding off-target discrimination across a diverse set of Cas9 targets. We profiled several XNAs for their ability to modulate Cas9 specificity, identifying FANA-modified crXNAs as the most reliable for high levels of activity and off-target reduction. Overall, these findings demonstrate the broad amenability of XNA technology to the field of CRISPR-Cas9, largely in terms of their ability to address the problem of gene editing specificity. We believe these findings will continue to contribute to the pursuit of accurate and safe gene editing technologies across all aspects of their use.

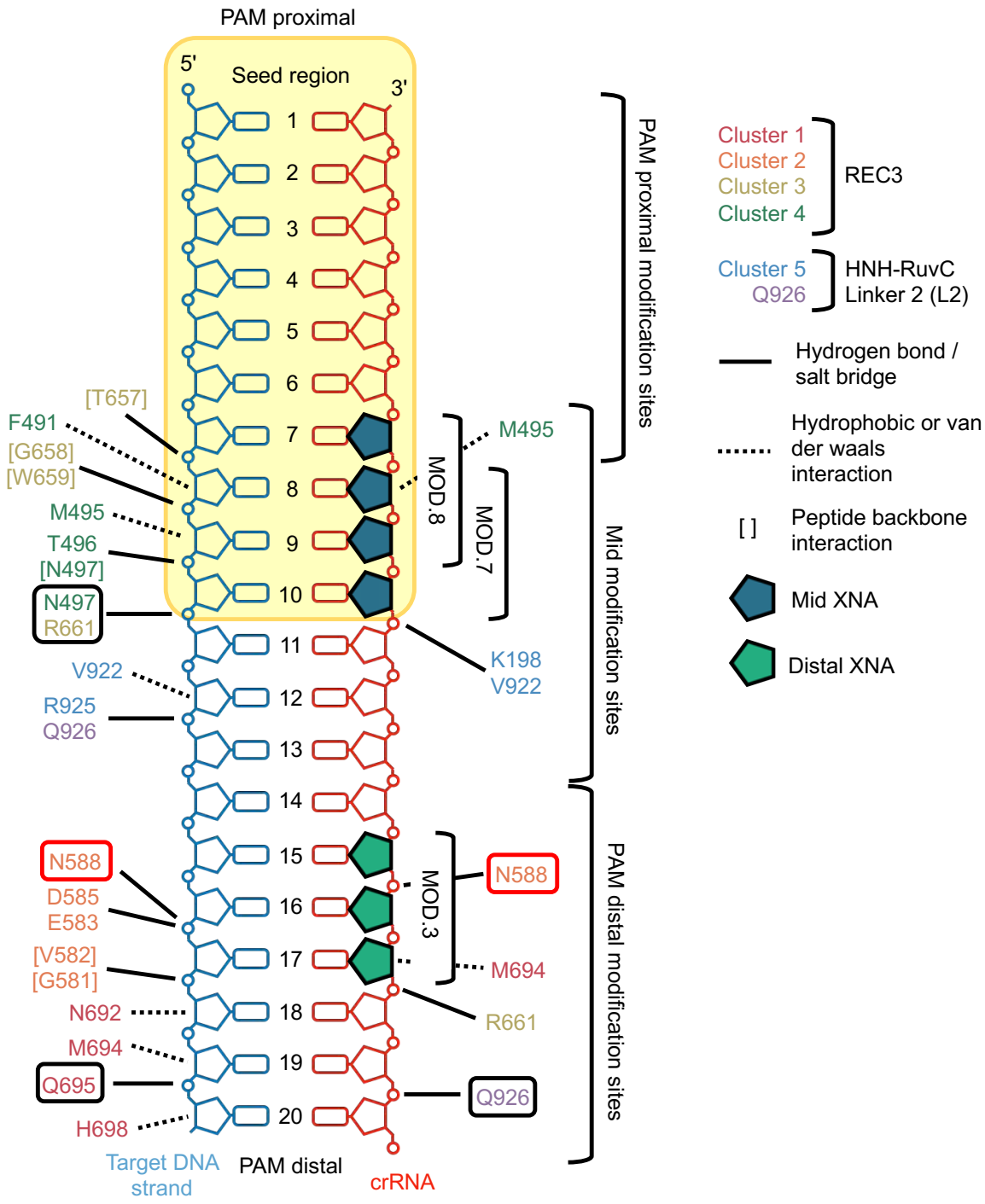


Figure 3.17. Schematic representation of the contacts made between the REC3 domain of Cas9 and the RNA/DNA heteroduplex

Amino acid residue labels surrounded by black boxes indicate positions of mutation in Cas9-HF1, while red boxes indicate mutations made in Cas9 HSC1.1. The crRNA seed region is indicated by a yellow rectangle.

3.5. Methods

3.5.1. Chemical reagents and oligonucleotides

All chemicals used were purchased from Sigma-Aldrich. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). BNA and OMe crXNAs were provided and synthesized by BioSynthesis Inc, while LNA, DNA, FANA, RNAPS and DNAPS crXNAs, as well as tracrRNAs, were provided and synthesized by Sigma-Aldrich. Sequences of oligonucleotides used in this study are listed in **Table 3.4**.

3.5.2. Plasmid construction

Plasmid templates containing Cas9 target sites for *in vitro* cleavage assays were constructed through ligation of inserts into *XbaI* and *HindIII* double-digested pUC19 (ThermoFisher) as previously described[194, 242]. DNA oligonucleotides containing the target sites of interest were purchased as ssDNA as listed in **Table 3.4**. Prior to ligation, forward and reverse ssDNA oligonucleotides were annealed by heating at 95°C for 5 min, followed by cooling to 25°C over the course of 1 hr. pMJ806 (Addgene; #39312) was purchased from Addgene and used to express Cas9 protein for all *in vitro* experiments.

3.5.3. Expression and purification of *S. pyogenes* Cas9

Recombinant Cas9 protein was prepared as previously described[249]. Briefly, Rosetta 2 DE3 *E. coli* were transformed with the plasmid pMJ806 (Addgene; #39312) encoding the *S. pyogenes cas9* gene fused to an N-terminal 6xHis-maltose binding protein (MBP) polypeptide sequence. Transformed bacteria were used to inoculate 100 mL of LB medium containing 50 μg/mL kanamycin, followed by overnight incubation (16 hrs) at 37°C in a shaking incubator. The

following day, 10 mL of starter culture was used to inoculate 1L of the same growth medium, and grown at 37°C until an OD_{600} of 0.7. The culture was then incubated at 18°C for 30 min before addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 200 μM. Induction of Cas9 was allowed to proceed for 16 hrs at 18°C. After incubation, cells were harvested by centrifugation at 2700 *x g* for 20 min at 4°C and re-suspended in lysis buffer (20 mM Tris-Cl, pH 8.0, 250 nM NaCl, 5 mM imidazole, pH 8.0, 1 mM phenylmethylsulfonylfluoride (PMSF)). Cells were lysed via sonication (30 s pulse on, and 60 s pulse off for a total of 7.5 min at 60% amplitude), with the resulting lysate clarified through centrifugation at 30,000 *x* g for 1 hr. Cas9-containing cell lysate was injected into an AKTA Start System (GE Healthcare) attached with a HisTrap FF Crude Column (GE Healthcare) and washed with wash buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 10 mM imidazole, pH 8.0) until the UV absorbance returned to baseline. Cas9 was eluted with 10 mL of elution buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 250 mM imidazole, pH 8.0) in a single step. Eluted protein concentration was estimated by measuring absorbance at 280 nm. For every 50 mg of protein, 0.5 mg of TEV protease was added to the sample. Cas9 and TEV protease were dialyzed against 4L of dialysis buffer (20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% (v/v) glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA) using dialysis tubing with a molecular weight cutoff of 15 kDa overnight at 4°C. Following dialysis, samples were centrifuged at 3200 *x* g for 5 min at 4°C to remove any precipitate. Cas9 was then loaded onto a HiTrap SP FF column (GE Healthcare) with ion exchange buffer A (20 mM HEPES-KOH, pH 7.5, 100 mM KCl) and washed with 15 mL of ion exchange buffer A. Cas9 was eluted by applying a gradient from 0-50% ion exchange buffer B (20 mM HEPES-KOH, pH 7.5, 1 M KCl) over 25 mL. Peak fractions were analyzed by SDS-PAGE, pooled and exchanged into storage buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 1 mM DTT) using a 50 kDa centrifugal filter

(Pall). Concentrated Cas9 protein was aliquoted and flash-frozen in liquid nitrogen, followed by storage at -80°C.

3.5.4. *In vitro* cleavage assays against on-target DNA substrates

Primers pUC19 Fwd and pUC19 Rev (listed in **Table 3.4**) were used to PCR amplify Cas9 substrate DNA using previously constructed plasmids as template. Following amplification, PCR amplicons were purified with the QIAquick PCR Purification Kit (Qiagen). gRNAs were prepared by combining equimolar amounts of tracrRNA and crRNA in duplex buffer (30 mM HEPES, pH 7.5, 100 mM potassium acetate), then heating at 95°C for 5 min, followed by cooling to 25°C over the course of 1 hr. All gRNAs, either unmodified or containing XNAs, were prepared as described above. Prior to addition of substrate DNA, Cas9 and gRNA were assembled in a 1:2 ratio to a final concentration of 25 nM and 50 nM respectively for 10 min at 25°C, after which DNA was added to a final concentration of 5 nM. Cleavage reactions were allowed to proceed for 1 hr at 37°C. All *in vitro* cleavage assays were performed in 1X NEBuffer 3.1. Reaction products were purified through the MinElute PCR Purification Kit (Qiagen) and resolved on a 1.2% TAE agarose gel. Images were captured on an Amersham Imager 600 (GE Healthcare). Full-length and cleavage products were quantified using densitometry (ImageJ). *In vitro* activity values were calculated using the formula: *In vitro* activity (%) = 100 x [(cleavage product 1 + cleavage product 2) / (cleavage product 1 + cleavage product 2 + full-length substrate)].

3.5.5. Preparation of *in vitro* high-throughput specificity profiling libraries

Pre-selection libraries were generated as previously described[52]. Briefly, 10 pmol of each partially randomized DNA oligonucleotide (IDT; listed in **Table 3.4**) were circularized using the

CircLigase II ssDNA Ligase Kit (Epicenter) through incubation with 100 U of CircLigase II ssDNA Ligase for 16 hrs at 60°C in 1X CircLigase II Reaction Buffer. Reactions were heat inactivated at 85°C for 15 min. 5 pmol of the crude circular ssDNA was amplified through rolling circle amplification using the illustra TempliPhi Amplification Kit (GE Healthcare) according to the manufacturers protocol. Resulting concatemeric pre-selection libraries were quantified using a Qubit 4.0 Fluorometer.

3.5.6. High-throughput specificity profiling assays

High-throughput specificity profiling of unmodified and XNA-modified crRNAs were performed as previously described[52, 194]. Briefly, 600 nM of Cas9 and 720 nM of previously prepared gRNA were pre-complexed at 25°C for 10 min in 1X NEBuffer 3.1, followed by addition of pre-selection library to a final concentration of 200 nM. Cleavage reactions were allowed to proceed for 20 min at 37°C before the reaction was halted by purification with the QIAquick PCR Purification Kit (Qiagen). In a separate reaction, pre-selection libraries were also digested with 2 U of BspMI (NEB) for 1 hr at 37°C in 1X NEBuffer 3.1. Restriction-digested pre-selection libraries were also purified with the QIAquick PCR Purification Kit (Qiagen). Cas9-digested and BspMI-digested library members were adaptor ligated to 10 pmol of adaptor1/2(#) (for post-selection) or library-specific pre-selection adaptor1/2 (for pre-selection) with 1000 U of T4 DNA Ligase (NEB) in 1X NEB T4 DNA Ligase Reaction Buffer for 16 hrs at 25°C. The next day, adaptor-ligated DNA was purified via the QIAquick PCR Purification Kit (Qiagen) and PCR amplified with Q5 Hot Start High-Fidelity DNA Polymerase (NEB) in 1X Q5 Reaction Buffer using primers post-selection fwd/library-specific post-selection rev, or primers pre-selection fwd/library-specific pre-selection rev. Libraries were amplified using a number of cycles determined to be sub-saturating. Amplified

libraries were gel purified using the MinElute Gel Extraction Kit (Qiagen), quantified with a Qubit 4.0 Fluorometer (ThermoFiser) and subject to sequencing on an Illumina HiSeq (PE150). Pre- and post-selection sequence data was analyzed as previously described[194, 242]. All sequences used are listed in **Table 3.4**.

3.5.7. Analysis of high-throughput specificity profiling data

Python scripts for high-throughput specificity profiling analysis of next-generation sequencing data are available upon request. Analysis was performed as previously described[242]. Specificity scores were calculated using the following formulae: Positive specificity score = (frequency of base pair at position[post-selection] – frequency of base pair at position[pre-selection]) / (1 – frequency of base pair at position[pre-selection]). Negative specificity score = (frequency of base pair at position[post-selection] – frequency of base pair at position[pre-selection]) / (frequency of base pair at position[post-selection]).

Table 3.4. Oligonucleotide sequences used in this study

Name	**Sequence**
Oligonucleotides used for generation of in vitro cleavage assay DNA substrate	
EMX1_F	GCCGAAGCTTCTGAGTCCGAGCAGAAGAAGAAGGGCTTCTAGAGGCC
EIF3D_F	GCCGAAGCTTCTAGACGACCCTGTCATCCGCAAGGCTTCTAGAGGCC
GABPA-1_F	GCCGAAGCTTCTGAAAAGGATAATTGAGCCCCAGGCTTCTAGAGGCC
GABPA-2_F	GCCGAAGCTTCTTTTGGAGTCTCAGAATGTCCTGGCTTCTAGAGGCC
SPOUT1_F	GCCGAAGCTTCTCAGGCGGGCTCACCTCCGTGCGGCTTCTAGAGGCC
AAVS1_F	GCCGAAGCTTCTGGGGCCACTAGGGACAGGATTGGCTTCTAGAGGCC

MALAT1-1_F	GCCGAAGCTTCTAATGTGAAGGACTTTCGTAACGGCTTCTAGAGGCC
MALAT1-2_F	GCCGAAGCTTCTGGCAGGAGAGGCCAGTTGCGGGGCTTCTAGAGGCC
MALAT1-3_F	GCCGAAGCTTCTGCTGGGGCTCAGTTGCGTAATGGCTTCTAGAGGCC
ATF1_F	GCCGAAGCTTCTTAGGAATCAAACACTTTTATTGGCTTCTAGAGGCC
CDH1_F	GCCGAAGCTTCTTGACTTGCGAGGGACGCATTGGGCTTCTAGAGGCC
EMX1_R	GGCCTCTAGAAGCCCTTCTTCTTCTGCTCGGACTCAGAAGCTTCGGC
EIF3D_R	GGCCTCTAGAAGCCTTGCGGATGACAGGGTCGTCTAGAAGCTTCGGC
GABPA-1_R	GGCCTCTAGAAGCCTGGGGCTCAATTATCCTTTTCAGAAGCTTCGGC
GABPA-2_R	GGCCTCTAGAAGCCAGGACATTCTGAGACTCCAAAAGAAGCTTCGGC
SPOUT1_R	GGCCTCTAGAAGCCGCACGGAGGTGAGCCCGCCTGAGAAGCTTCGGC
AAVS1_R	GGCCTCTAGAAGCCAATCCTGTCCCTAGTGGCCCCAGAAGCTTCGGC
MALAT1-1_R	GGCCTCTAGAAGCCGTTACGAAAGTCCTTCACATTAGAAGCTTCGGC
MALAT1-2_R	GGCCTCTAGAAGCCCCGCAACTGGCCTCTCCTGCCAGAAGCTTCGGC
MALAT1-3_R	GGCCTCTAGAAGCCATTACGCAACTGAGCCCCAGCAGAAGCTTCGGC
ATF1_R	GGCCTCTAGAAGCCAATAAAAGTGTTTGATTCCTAAGAAGCTTCGGC
CDH1_R	GGCCTCTAGAAGCCCAATGCGTCCCTCGCAAGTCAAGAAGCTTCGGC
Oligonucleotides used in high-throughput specificity profiling experiments	
EMX1	/5Phos/TCTTCTNNNNC*C*NT*T*C*T*T*C*T*T*C*T*G*C*T*C*G*G*A*C*T*C*NNNNACCTGCCGAGTCTTCT
EIF3D	/5Phos/AACGTCNNNNC*C*NT*G*C*G*G*A*T*G*A*C*A*G*G*G*T*C*G*T*C*T*NNNNACCTGCCGAGAACGTC
GABPA-1	/5Phos/ACTTAANNNNC*C*NG*G*G*G*C*T*C*A*A*T*T*A*T*C*C*T*T*T*T*C*NNNNACCTGCCGAGACTTAA
GABPA-2	/5Phos/AGCAATNNNNC*C*NG*G*A*C*A*T*T*C*T*G*A*G*A*C*T*C*C*A*A*A*NNNNACCTGCCGAGAGCAAT
SPOUT1	/5Phos/ATTAGTNNNNC*C*NC*A*C*G*G*A*G*G*T*G*A*G*C*C*C*C*G*C*C*T*G*NNNNACCTGCCGAGATTAGT
AAVS1	/5Phos/CAAGCTNNNNC*C*NA*T*C*C*T*G*T*C*C*C*T*A*G*T*G*G*C*C*C*C*NNNNACCTGCCGAGCAAGCT

MALAT1-1	/5Phos/CCTGATNNNNC*C*NT*T*A*C*G*A*A*A*G*T*C*C*T*T*C*A*C*A*T*T*NNNNACCTGCCGAGCCTGAT
MALAT1-2	/5Phos/CGCGGANNNNC*C*NC*G*C*A*A*C*T*G*G*C*C*T*C*T*C*C*T*G*C*C*NNNNACCTGCCGAGCGCGGA
MALAT1-3	/5Phos/GAACTGNNNNC*C*NT*T*A*C*G*C*A*A*C*T*G*A*G*C*C*C*C*A*G*C*NNNNACCTGCCGAGGAACTG
ATF1	/5Phos/GATTTANNNNC*C*NA*T*A*A*A*A*G*T*G*T*T*T*T*G*A*T*T*C*C*T*A*NNNNACCTGCCGAGGATTTA
CDH1	/5Phos/TGAGAANNNNC*C*NA*A*T*G*C*G*T*C*C*C*T*C*G*C*A*A*G*T*C*A*NNNNACCTGCCGAGTGAGAA
adaptor1(1)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC TGT
adaptor1(2)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCT GAA
adaptor1(3)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTG ACT
adaptor1(4)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTG CAA
adaptor1(5)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGC ATT
adaptor1(6)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCA TGA
adaptor1(7)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAT GCT
adaptor1(8)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCT AGT
adaptor1(9)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGC TAA
adaptor1(10)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCA GTT
adaptor1(11)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGT CAT
adaptor1(12)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAC GTA
adaptor2(1)	ACA GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(2)	TTC AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(3)	AGT CAA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(4)	TTG CAA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(5)	AAT GCA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(6)	TCA TGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG

adaptor2(7)	AGC ATA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(8)	ACT AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(9)	TTA GCA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(10)	AAC TGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(11)	ATG ACA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
adaptor2(12)	TAC GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GG
Post Selection Fwd	AAT GAT ACG GCG ACC ACC GA
Pre Selection Fwd	CAA GCA GAA GAC GGC ATA CGA GAT
Pre Selection Adaptor1	GAC GGC ATA CGA GAT
EMX1 Pre-Selection Adaptor2	TCTTATCTCGTATGCCGTCTTCTGCTTG
EIF3D Pre Selection Adaptor2	AAC GAT CTC GTA TGC CGT CTT CTG CTT G
GABPA-1 Pre Selection Adaptor2	ACT TAT CTC GTA TGC CGT CTT CTG CTT G
GABPA-2 Pre Selection Adaptor2	AGC AAT CTC GTA TGC CGT CTT CTG CTT G
SPOUT1 Pre Selection Adaptor2	ATT AAT CTC GTA TGC CGT CTT CTG CTT G
AAVS1 Pre Selection Adaptor2	CAA GAT CTC GTA TGC CGT CTT CTG CTT G
MALAT1-1 Pre Selection Adaptor2	CCT GAT CTC GTA TGC CGT CTT CTG CTT G
MALAT1-2 Pre Selection Adaptor2	CGC GAT CTC GTA TGC CGT CTT CTG CTT G
MALAT1-3 Pre Selection Adaptor2	GAA CAT CTC GTA TGC CGT CTT CTG CTT G
ATF1 Pre Selection Adaptor2	GAT TAT CTC GTA TGC CGT CTT CTG CTT G
CDH1 Pre Selection Adaptor2	TGA GAT CTC GTA TGC CGT CTT CTG CTT G
EMX1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GTC TTC T
EIF3D Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GAA CGT C
GABPA-1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GAC TTA A
GABPA-2 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GAG CAA T

SPOUT1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GAT TAG T
AAVS1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GCA AGC T
MALAT1-1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GCC TGA T
MALAT1-2 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GCG CGG A
MALAT1-3 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GGA ACT G
ATF1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GGA TTT A
CDH1 Post Selection Rev	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA GTG AGA A
EIF3D Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAA CGT C
GABPA-1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAC TTA A
GABPA-2 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAG CAA T
SPOUT1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAT TAG T
AAVS1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GCA AGC T
MALAT1-1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GCC TGA T
MALAT1-2 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GCG CGG A
MALAT1-3 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GGA ACT G
ATF1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GGA TTT A
CDH1 Pre Selection Rev	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GTG AGA A
EMX1 Post-Selection Rev (N701)	CAAGCAGAAGACGGCATACGAGAT TCGCCTTA ACCTGCCGAG TCTTCT

EMX1 Post-Selection Rev (N702)	CAAGCAGAAGACGGCATACGAGAT CGTACTAG ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N703)	CAAGCAGAAGACGGCATACGAGAT TTCTGCCT ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N704)	CAAGCAGAAGACGGCATACGAGAT GCTCAGGA ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N705)	CAAGCAGAAGACGGCATACGAGAT AGGAGTCC ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N706)	CAAGCAGAAGACGGCATACGAGAT CATGCCTA ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N707)	CAAGCAGAAGACGGCATACGAGAT GTAGAGAG ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N708)	CAAGCAGAAGACGGCATACGAGAT CCTCTCTG ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N709)	CAAGCAGAAGACGGCATACGAGAT AGCGTAGC ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N710)	CAAGCAGAAGACGGCATACGAGAT CAGCCTCG ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N711)	CAAGCAGAAGACGGCATACGAGAT AAGAGGCA ACCTGCCGAG TCTTCT
EMX1 Post-Selection Rev (N712)	CAAGCAGAAGACGGCATACGAGAT TCCTCTAC ACCTGCCGAG TCTTCT

An asterisks (*) denotes that the preceding nucleotide was purchased as a hand mix of bases consisting of 79% of the indicated base, and 7% of each of the other three bases. “/5Phos/” indicates a 5′ phosphate group.

Chapter 4

Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences.

Amanda R. Krysler[1*], Christopher R. Cromwell[1*], Tommy Tu[1], Juan Jovel[2] & Basil P. Hubbard[1]

[1]Department of Pharmacology, University of Alberta, Edmonton, Canada

[2]The Applied Genomics Core, Office of Research, University of Alberta, Edmonton, Canada

*Denotes equal contribution

PUBLISHED:

Krysler, A.R., Cromwell, C.R., Tu, T. *et al.* Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences. *Nat Commun* **13,** 1617 (2022). https://doi.org/10.1038/s41467-022-29202-x

4.1 Abstract

CRISPR/Cas complexes enable precise gene editing in a wide variety of organisms. While the rigid identification of DNA sequences by these systems minimizes the potential for off-target effects, it consequently poses a problem for the recognition of sequences containing naturally occurring polymorphisms. The presence of genetic variance such as single nucleotide polymorphisms (SNPs) in a gene sequence can compromise the on-target activity of CRISPR systems. Thus, when attempting to target multiple variants of a human gene, or evolved variants of a pathogen gene using a single guide RNA, more flexibility is desirable. Here, we demonstrate that Cas9 can tolerate the inclusion of universal bases in individual guide RNAs, enabling simultaneous targeting of polymorphic sequences. Crucially, we find that specificity is selectively degenerate at the site of universal base incorporation, and remains otherwise preserved. We demonstrate the applicability of this technology to targeting multiple naturally occurring human SNPs with individual guide RNAs and to the design of Cas12a/Cpf1-based DETECTR probes capable of identifying multiple evolved variants of the HIV protease gene. Our findings extend the targeting capabilities of CRISPR/Cas systems beyond their canonical spacer sequences and highlight a use of natural and synthetic universal bases.

4.2 Introduction

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) systems play an important role in mediating adaptive immunity in prokaryotes[209] and have been effectively repurposed for gene- and RNA-editing applications[51, 78, 240, 250]. Over 400 different CRISPR-Cas homologs, comprised of diverse enzymes with unique nucleic acid binding specificities and cleavage mechanisms have been annotated[251]. These have been grouped into two distinct classes[252]. Class I

systems employ multi-subunit nuclease complexes, while Class II systems, more widely used for gene editing, rely on a single effector protein[252]. Class II systems may be further divided into subtypes (e.g., II-A, II-B, V, VI) based on architecture[252]. Cas9, a type II-A system, directs DNA cleavage using two separately expressed RNA elements: a CRISPR RNA (crRNA) that contains a 20-nucleotide (nt) sequence complementary to the target DNA sequence, and a trans-activating crRNA (tracrRNA) that bridges the Cas9-crRNA interaction[35]. Target recognition by Cas9 involves binding a protospacer adjacent motif (PAM) sequence (5′-NGG-′3 in *S. pyogenes*), followed by hybridization of the 20-nt spacer sequence to the target[32]. Formation of a fully paired duplex induces conformational changes in the RuvC and HNH nuclease domains in Cas9 that ultimately result in a double-strand DNA cleavage event (**Figure 4.1a**)[32, 55, 194]. In contrast, Cas12a (Cpf1), which is a Class II Type V system, employs a single RuvC active site to induce staggered cuts within the target and non-target strands (**Figure 4.1b**)[252]. Cas12a recognizes a T-rich PAM (5′-TTN-′3), uses a 20–24 base-pair (bp) spacer sequence, does not require a tracrRNA, and has the ability to process its own pre-crRNA[61]. Unlike Cas9, Cas12a unleashes indiscriminate single-stranded DNase activity (collateral or *trans* activity) *in vitro* upon nuclease activation[70].

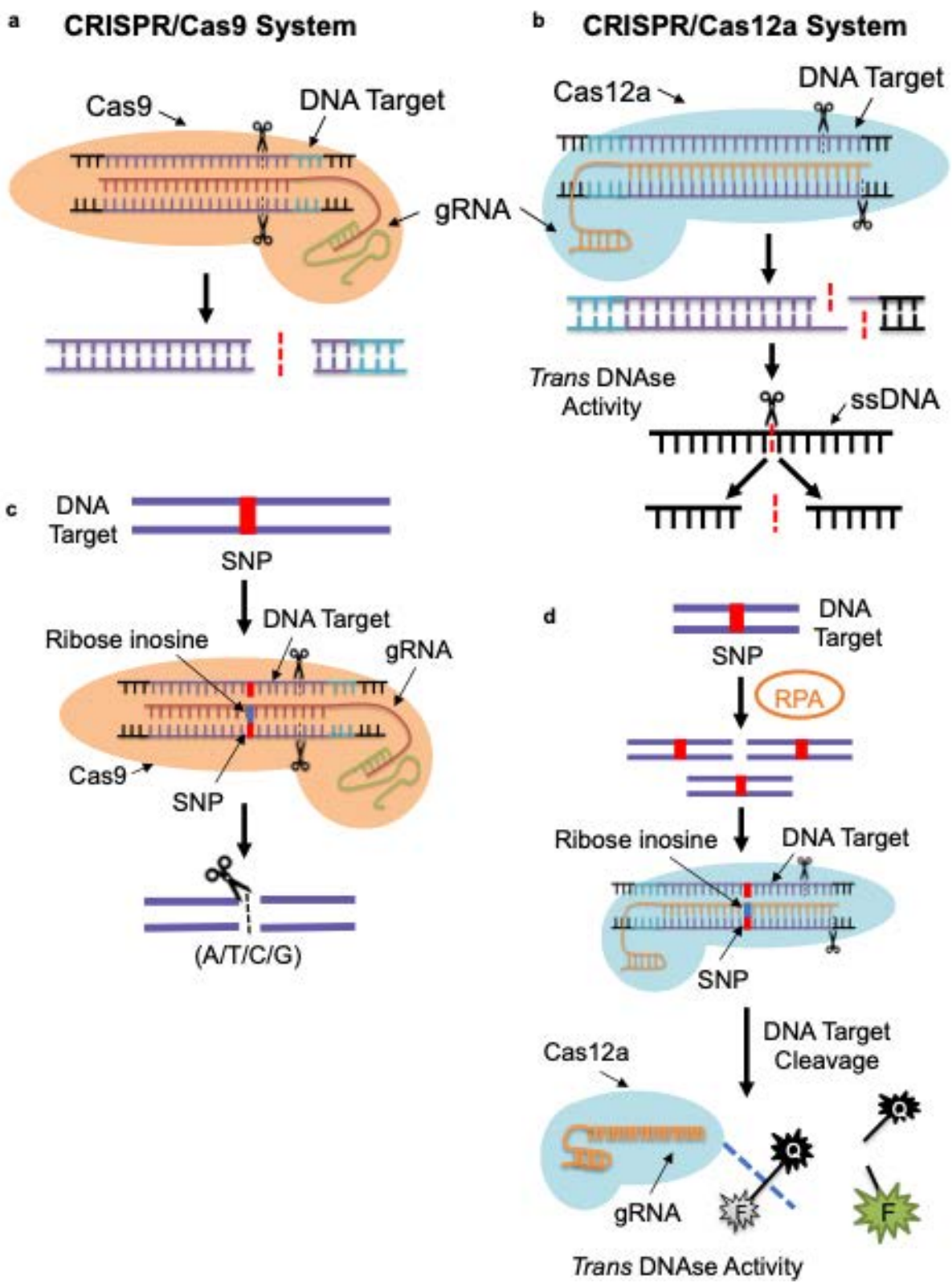

a
CRISPR/Cas9 System
Cas9
DNA Target
gRNA
b
CRISPR/Cas12a System
Cas12a
DNA Target
Trans DNAse Activity
ssDNA
c
DNA Target
SNP
DNA Target
Ribose inosine
gRNA
Cas9
SNP
(A/T/C/G)
d
DNA Target
SNP
RPA
Ribose inosine
DNA Target
SNP
DNA Target Cleavage
Cas12a
gRNA
F
F
Trans DNAse Activity

Figure 4.1. Applications of universal bases to CRISPR/Cas9 and CRISPR/Cas12a systems

(**a**) Diagram outlining the CRISPR/Cas9 DNA cleavage mechanism. After forming a ribonucleoprotein complex with a designed gRNA, Cas9 identifies its DNA target sequence based on the presence of a PAM sequence (blue section of the DNA target) and base-pair interactions between the DNA and the 20-nt spacer sequence of the gRNA. If successful pairing occurs, Cas9 induces a blunt, double-strand cut in the DNA. (**b**) Diagram outlining the CRISPR/Cas12a DNA cleavage mechanism. After forming a ribonucleoprotein complex with a designed gRNA, Cas12 identifies its DNA target sequence based on the presence of a PAM sequence (blue section of the DNA target) and base-pair interactions between the DNA and the 23-25- nt spacer sequence of the gRNA. If a highly matched sequence is found, Cas12a induces a staggered double-strand DNA cut in the target (*cis* cleavage). In addition, this process activates the collateral, or *trans* DNase activity of Cas12a which acts to subsequently degrade nearby ssDNA in a non-specific manner. Diagrams depicting applications where universal bases substituted into the gRNA could (**c**) enable the simultaneous targeting/cleavage of multiple DNA sequence variants in cells or (**d**) allow for the detection of polymorphic sequences. 'SNP' denotes single nucleotide polymorphism, 'RPA' denotes recombinase polymerase amplification, 'F' denotes fluorophore, and 'Q' denotes quencher.

Both Cas9 and Cas12a have been used to edit the genomes of numerous organisms ranging from plants to mammals[253, 254]. Moreover, both systems have demonstrated potential for use in the context of clinical therapeutics to treat human genetic disease[253, 254]. Furthermore, the collateral DNase activity of Cas12a has been exploited to generate a diagnostic platform for the detection of aberrant mutations or pathogen DNA sequences[70]. Briefly, the DNA endonuclease-targeted CRISPR trans reporter (DETECTR) system links activation of Cas12a nuclease activity to *trans* cleavage of a single-stranded DNA (ssDNA) substrate containing flanking fluorophore and quencher moieties[70]. When combined with isothermal amplification, this system achieves attomolar DNA detection sensitivity[70].

One of the primary obstacles to translating CRISPR/Cas systems to clinical applications has been concern over off-target DNA cleavage, which could have detrimental health consequences for therapeutics, and yield false-positive results for diagnostics[13]. As a result, much work has been

done to improve the specificity of these systems through protein engineering or evolution[175, 176], or engineering or chemical modification of guide RNAs[194]. For example, guide RNAs with engineered secondary structures improve Cas12a specificity[201], and incorporation of DNA[203] or bridged nucleic acids (BNA)[194] into Cas9 gRNAs improves its specificity.

While single-nt precision is desirable for many nucleic acid targeting applications[201], there are other instances where recognition of a discrete 20 bp sequence may be limiting. First, CRISPR/Cas9 can be sensitive to naturally occurring SNPs within the PAM-proximal portion of a guide sequence[229, 230, 255]. Since SNPs occur roughly every 300 bp in the human genome[256], a CRISPR/Cas9 therapeutic designed for one patient may be ineffective for another. Indeed, a recent test of 263 therapeutically-relevant guide RNAs revealed that >16% failed to cleave the on-target site in at least one of 7700 haplotypes tested[257]. Second, the high degree of natural genetic diversity present in pathogens such as HIV-1 greatly complicates antiviral treatment or diagnostic detection using CRISPR/Cas systems[258, 259]. Finally, studies have shown that even successful cleavage of HIV-1 DNA sequences using CRISPR/Cas9 can result in mutations that accelerate viral escape and render the virus resistant to the original guide RNA[260, 261]. These scenarios highlight the need for additional CRISPR/Cas capabilities that allow sequences to be targeted in a more flexible manner.

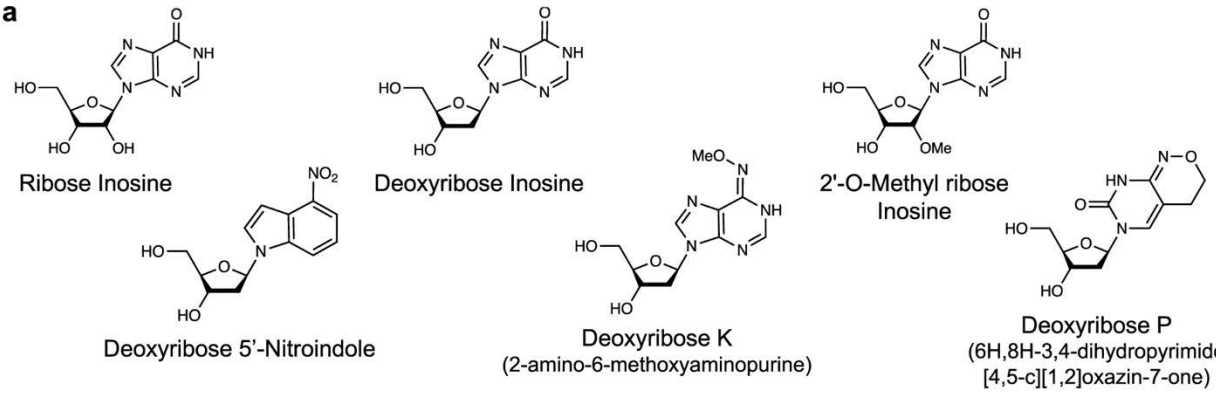
In nature, recognition of degenerate mRNA codons by the tRNA anticodon loop is achieved through the inclusion of ribose inosine (I) nts (containing the hypoxanthine base) (**Figure 4.2a**)[262, 263]. Inosine also plays a role in RNA editing[264] and acts as a DNA damage intermediate following adenosine deamination[265]. Characterized as a ‘universal base’, inosine forms 2 hydrogen bonds

with all four canonical bases with a slight I-C > I-A > I-T ≈ I-G bias in stability[262]. Inosine has been successfully applied to the design of degenerate PCR primers and diagnostic probes, as well as in DNA sequencing[262]. It can be incorporated into a nucleic acid strand either as a standard RNA or DNA nt, or as modified variant such as 2′-O-methyl (2′ OMe) RNA, which displays improved nuclease resistance and unique hybridization properties[266]. Synthetic universal bases such as deoxyribose 5′-nitroindole (**Figure 4.2a**) have also been developed[267]. This base lacks the ability to form any hydrogen bonds but adopts a standard *anti* configuration with the opposing nt and acts to stabilize hydrophobic base stacking[268]. While more destabilizing in certain contexts, 5′-nitroindole bases appear to be devoid of any base-pairing bias[268, 269]. Other synthetic bases have been developed to exhibit partial degeneracy, including deoxyribose K (2-amino-6-methoxyaminopurine) and deoxyribose P (6H,8H-3,4-dihydro-pyrimido[4,5-c] [4,5-c] [1,2]oxazin-7-one) (**Figure 4.2a**), which show a preference for C/T and A/G pairing, respectively[270, 271].

Based on their present use in nucleic acid amplification and detection technologies[262], we hypothesized that universal bases could be harnessed to impart Cas systems with the ability to target multiple polymorphic sequences using an individual guide RNA. Here, we show that multiple types of chemically distinct universal bases can be tolerated within both Cas9 and Cas12a guide RNAs, in a context-dependent manner. We show that single crRNAs containing universal bases can be used to target multiple gene variants containing naturally occurring SNPs (**Figure 4.1c**). Furthermore, we design individual probes containing universal bases that are capable of identifying multiple variants of the HIV-1 protease gene using the DETECTR system (**Figure 4.1d**). Using high-throughput specificity profiling, we show that inclusion of universal bases

imparts selective degeneracy at the site of incorporation, without otherwise altering crRNA specificity. Our results outline a new strategy for expanding the capabilities of CRISPR/Cas to the recognition of nucleic acid targets with high variability and those for which only incomplete sequence information is available.

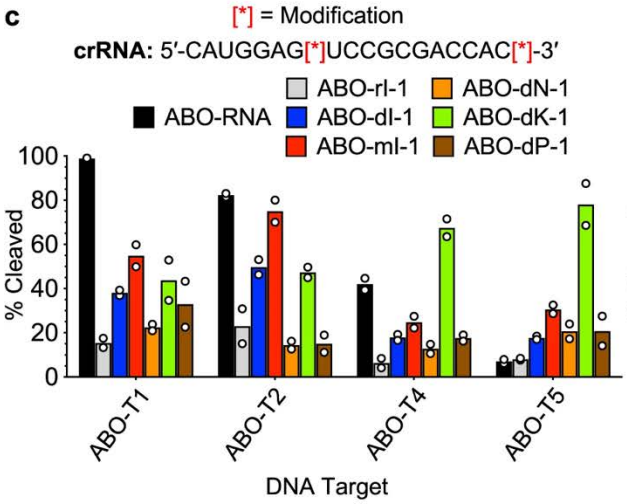
a



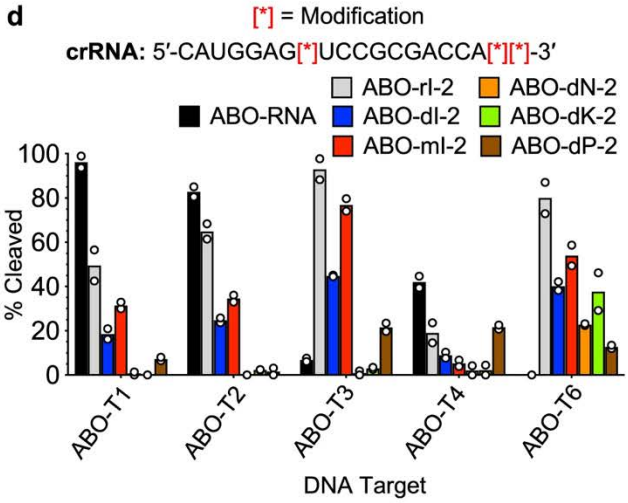
b

Target	DNA Sequence (5′→3′)	Allele Frequency (%)
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	54.21
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	25.50
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	15.98
ABO-T4	CATGGAGTTCCGCGACCAC**A**T**GG**	0.015
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**T**GG**	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**T**GG**	0.00061
ABO-T7	CATGGAG**A**TCCGCGACCA**T**G**TGG**	0.0024

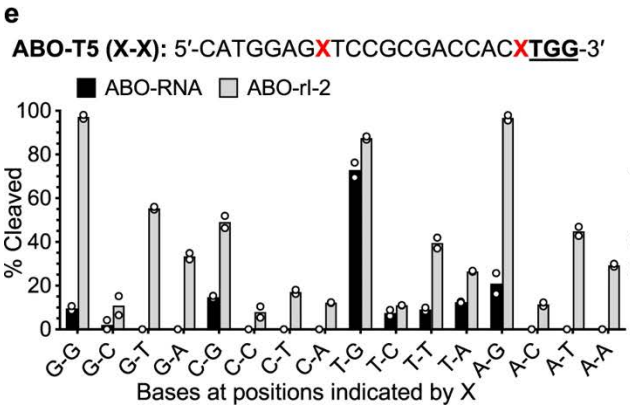
c



d



e



f

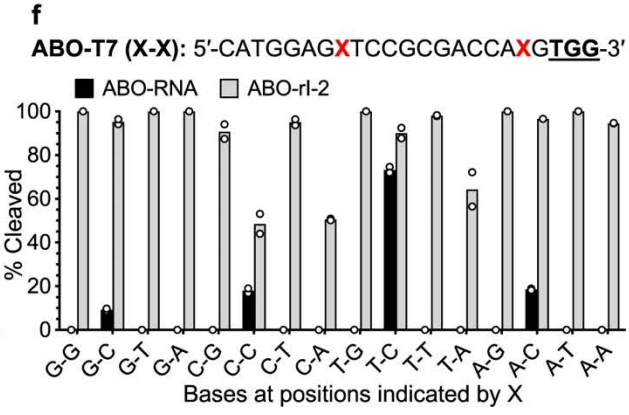


Figure 4.2. Incorporation of universal bases into Cas9 crRNAs enables the targeting of polymorphic gene variants

(**a**) Chemical structures of universal and degenerate bases used in this study. (**b**) List of DNA targets corresponding to sequences in the *ABO* gene based on clinical polymorphism data. SNPs are indicated with red lettering. Allele frequency indicates either the current tallied allele frequency or the statistically predicted frequency (for sequences containing multiple SNPs). The PAM sequence is underlined. Bar graphs showing the relative amount of DNA cleavage resulting from in vitro reactions containing Cas9 with (**c**) double or (**d**) triple modified-crRNAs and the variant DNA target sequences indicated. Locations of the universal bases in the crRNA sequence are indicated with red [*]. rI = Ribose Inosine, dI = Deoxyribose Inosine, mI = 2′-O-Methyl ribose Inosine, dN = Deoxyribose 5′-Nitroindole, dK = Deoxyribose K, dP = Deoxyribose P; Mean with individual data points shown ($n = 2$ independent experiments). Bar graphs showing the relative amount of DNA cleavage resulting from in vitro reactions containing Cas9 with ABO-RNA or ABO-rI-2 versus derivatives of the (**e**) ABO-T5 or (**f**) ABO-T7 target sequences. Base combinations listed along the *x*-axis correspond to the positions indicated by a red 'X' in the reference sequence. The PAM sequence is underlined; Mean with individual data points shown ($n = 2$ independent experiments). Reactions were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Cleavage percentages were calculated from corresponding agarose gels using densitometry software (ImageJ).

4.3 Results

4.3.1. Incorporation of universal bases into Cas9 crRNAs enables targeting of polymorphic sequences

Past studies have shown that the inclusion of sugar[194, 202] and backbone[188, 195] chemical modifications in Cas9 crRNAs can be tolerated. In addition, crRNAs containing locked/bridged nucleic acids (LNA/BNA) and DNA have been demonstrated to reduce Cas9 off-target DNA cleavage activity relative to their unmodified counterparts[194, 202]. Given these findings, we speculated that incorporation of non-canonical bases into crRNAs might also be permitted. In particular, we wondered if universal bases could be incorporated into crRNAs so as to enable Cas9 recognition of polymorphic target sequences. To test this possibility, we selected a highly polymorphic sequence from the *ABO* gene that determines the most clinically important blood

group system in mammals[272]. We generated a series of 16 DNA target sequences (ABO-T1–16), derived from prevalent alleles in the human population, containing naturally occurring single nucleotide polymorphisms (SNPs) within that region (**Figure 4.2b**, **Figure 4.3a**). Next, we tested the ability of Cas9 to cleave these sequences *in vitro* using an unmodified guide RNA (ABO-RNA) corresponding to the reference sequence (ABO-T1). Consistent with previous studies on Cas9 specificity[194, 273], we observed robust cleavage of the on-target sequence (ABO-T1) and two sequences containing single SNPs (ABO-T2, ABO-T4), but weak or absent activity on all of the other sequence variants (**Figure 4.3b**). These results reinforce the negative impact that natural genetic variation can have on Cas9 on-target activity.

a

Target	DNA Sequence (5′→3′)	# of SNPs	Target	DNA Sequence (5′→3′)	# of SNPs
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	0	**ABO-T9**	CATGGAGTTC**T**GCGACCACG**TGG**	1
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	1	**ABO-T10**	CATGGAG**A**TC**T**GCGACCACG**TGG**	2
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	1	**ABO-T11**	CATGGAGTTC**T**GCGACCA**T**G**TGG**	2
ABO-T4	CATGGAGTTCCGCGACCAC**A**T**GG**	1	**ABO-T12**	CATGGAGTTC**T**GCGACCAC**A**T**GG**	2
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**T**GG**	2	**ABO-T13**	CATGGAG**A**TC**T**GCGACCA**T**G**TGG**	3
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**T**GG**	3	**ABO-T14**	CATGGAG**A**TC**T**GCGACCAC**A**T**GG**	3
ABO-T7	CATGGAG**A**TCCGCGACCA**T**G**TGG**	2	**ABO-T15**	CATGGAGTTC**T**GCGACCA**TA**T**GG**	3
ABO-T8	CATGGAGTTCCGCGACCA**TA**T**GG**	2	**ABO-T16**	CATGGAG**A**TC**T**GCGACCA**TA**T**GG**	4

b

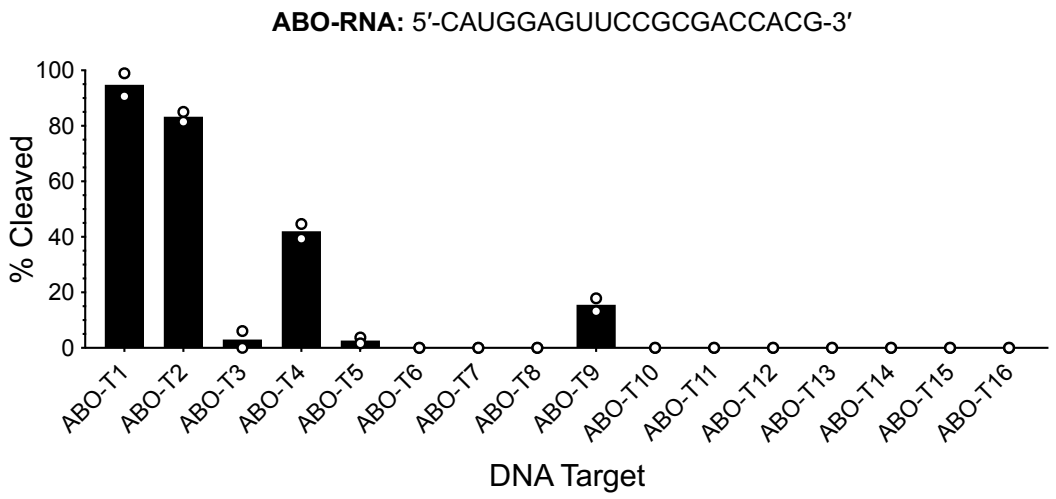


Figure 4.3. *In vitro* cleavage of *ABO* variant sequences by Cas9 using ABO-RNA
(**a**) List of DNA targets corresponding to sequences in the *ABO* gene based on clinical polymorphism data. SNPs are indicated with red lettering. The PAM sequence is underlined. (**b**) Bar graph showing the relative amount of DNA cleavage resulting from *in vitro* reactions containing Cas9 with ABO-RNA versus the indicated DNA target sequences; Mean with individual data points shown (n = 2 independent experiments). Cleavage percentages were calculated from corresponding agarose gels using densitometry software (ImageJ).

To generate guide RNAs capable of simultaneously recognizing a broader set of *ABO* sequence variants, we selected two *ABO* sequences bearing 2 (ABO-T5) or 3 (ABO-T6) polymorphisms relative to ABO-T1 (**Figure 4.2b**), and designed a panel of corresponding crRNAs in which ribose inosine (ABO-rI-1, ABO-rI-2), deoxyribose inosine (ABO-dI-1, ABO-dI-2), 2′OMe ribose inosine (ABO-mI-1, ABO-mI-2), deoxyribose 5′-nitroindole (ABO-dN-1, ABO-dN-2), deoxyribose K (ABO-dK-1, ABO-dK-2), or deoxyribose P (ABO-dP-1, ABO-dP-2) bases were substituted at positions overlapping with the SNPs (**Figure 4.2c**, **d**). Using ABO-rI-1, ABO-dI-1, ABO-mI-1, ABO-dN-1, ABO-dK-1, and ABO-dP-1, we assayed Cas9 cleavage activity on ABO-T1, the corresponding ABO-T5 double SNP variant sequence, and sequences containing each SNP in isolation (ABO-T2, ABO-T4). Using the ABO-mI-1 and ABO-dK-1 crRNAs, Cas9 cleaved ABO-T5 >5 and >10-fold more abundantly than with ABO-RNA, respectively (**Figure 4.2c**, **Figure 4.4a**, **b**). Both of these crRNAs also supported Cas9 cleavage of the single variant (ABO-T2, ABO-T4) and reference (ABO-T1) sequences (**Figure 4.2c**). Similarly, we found that ABO-rI-2, ABO-dI-2, ABO-mI-2, and ABO-dK-2 guided efficient Cas9 cleavage of ABO-T6 (> 50% compared to 0% with ABO-RNA) (**Figure 4.2d**, **Figure 4.4a**, **c**). ABO-rI-2 and ABO-mI-2 were also able to direct the cleavage of ABO-RNA and the single variant sequences ABO-T2 and ABO-T3, but not ABO-T4 (**Figure 4.2d**). These results demonstrate that universal bases with diverse chemistries can be incorporated into crRNAs to allow simultaneous targeting of complex SNP variants *in vitro*.

a

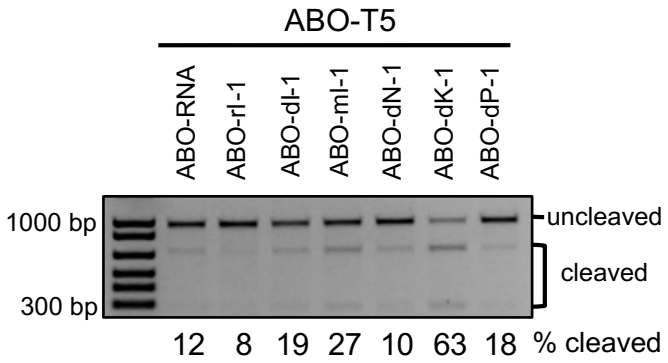
Target	DNA Sequence (5′→3′)	Allele Frequency (%)
ABO-T5	CATGGAG**A**TCCGCGACCAC**A**TGG	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TA**TGG	0.00061

b

ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCAC[*]-3′



c

ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCA[*][*]-3′

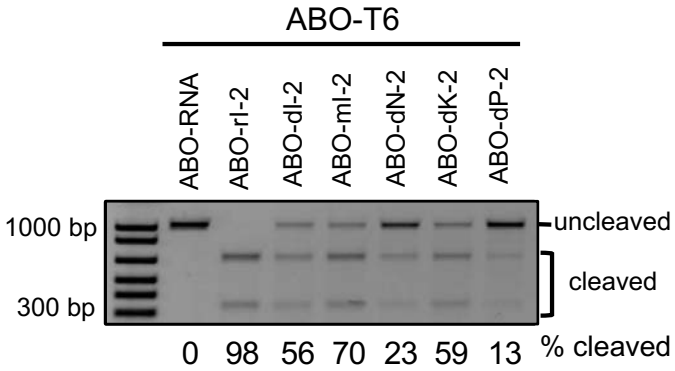


Figure 4.4. Cleavage of polymorphic sequences by Cas9 using crRNAs containing universal bases

(**a**) List of DNA target sequences. SNPs are indicated with red lettering. The PAM sequence is underlined. Representative gels showing cleavage of (**b**) ABO-T5 or (**c**) ABO-T6 DNA targets by Cas9 using the indicated crRNAs. The bottom two bands in the gel represent the cleaved DNA substrate while the top band corresponds to the undigested substrate. Reactions were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Quantification of cleavage percentages was performed using ImageJ. Cleavage experiments were performed in duplicate with similar results.

Our findings in **Figure 4.2c** and **Figure 4.2d** indicated that amongst the various universal bases we tested, inosine derivatives (ribose, deoxy, 2′OMe) appeared to be the most consistently well tolerated *in vitro*. Therefore, we chose to focus our studies on this naturally occurring non-canonical base. Unlike synthetic bases such as deoxyribose 5′-nitroindole, previous work has shown that inosine exhibits a slight base pairing preference in certain contexts[262]. We wondered if a base pairing bias might manifest in our *in vitro* Cas9 DNA cleavage reactions. To test this, we designed two sets of 16 target sequences covering all combinations of bases at the two SNP locations in ABO-T5 and ABO-T7, and evaluated cleavage of these sequences by Cas9 using either ABO-rI-2 or the unmodified crRNA. As shown in **Figure 4.2e**, Cas9 was able to cut 9 of the 16 targets with >25% efficiency using ABO-rI-2 (the remaining seven sequences were also cut at low levels), compared to only the reference sequence being cleaved to this extent using the unmodified crRNA. The results using the ABO-T7 derivative sequences were even more striking. All 16 of the derivative sequences were cleaved at >50% efficiency by Cas9 using ABO-rI-2, while only the reference sequence was cut at appreciable levels using the unmodified crRNA (three other sequences were cleaved at lower levels). These results suggest that incorporation of inosine bases into crRNAs enables targeting of all four canonical bases at the corresponding DNA target sites in a relatively unbiased and independent manner.

To characterize the patterns of inosine modifications permitted by Cas9, we synthesized an additional 13 crRNAs containing 1–4 ribose inosine modifications (**Figure 4.5a**) and tested the ability of these to direct cleavage of the ABO-T1 sequence by Cas9. We found that inclusion of a single inosine was tolerated in all instances, albeit with reduced activity, while crRNAs containing 2–4 inosine substitutions supported Cas9 cleavage of ABO-T1 in certain cases (**Figure 4.5b**).

Next, we sought to determine the cause of the reduced activity and to establish if the effect was general or target-specific. Using ABO-RNA, and two crRNAs containing two inosine modifications, ABO-rI-1 (low activity) and ABO-rI-8 (no activity on the ABO-T1 sequence using the given conditions), we performed a titration of tracrRNA:crRNA to determine if inosines within the spacer sequence might somehow impair the ability of these two RNA elements to hybridize. Altering this ratio did not result in increased activity, ruling out this possibility (**Figure 4.6a**, **b**). In addition, the low cleavage activity observed in vitro using ABO-rI-1 or ABO-rI-8 could not be augmented by increasing reaction time (**Figure 4.6c**). Based on these results, we hypothesized that the lowered activity observed using certain inosine-modified crRNAs may be due to decreased ribonucleoprotein (RNP) complex binding to the target DNA sequence. A titration of activity versus RNP concentration provided evidence to support this assertion (**Figure 4.6d**). Moreover, data from electrophoretic mobility shift assays (EMSAs) confirmed that RNP binding to ABO-T1 was substantially reduced using ABO-rI-1 and ABO-rI-8 compared to the unmodified crRNA (**Figure 4.7**). Interestingly, we found that ABO-rI-8, which did not support Cas9 cleavage of ABO-T1, did support cleavage of ABO-T7, establishing its activity on other target sequences (Figure 4.8). In all instances, we observed a strong correlation between RNP-target engagement and activity in DNA cleavage assays (**Figures 4.7** and **4.8**). Previous work has shown that I-G and I-A pairs decrease thermodynamic duplex stability by 0.84 kcal/mol and 0.52 kcal/mol compared to C-G and A-T pairs, respectively[262]. We found that in the absence of Cas9, T_m values for inosine-modified crRNA-target DNA duplexes were in fact reduced compared to the unmodified counterpart (**Figure 4.9**). Thus, it is likely that incorporation of inosines into crRNAs destabilizes Cas9-DNA target binding, although the extent to which this affects overall activity appears to be context-dependent and minimal in some cases.

a

Name	crRNA Sequence (5′→3′)	Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG	**ABO-rI-8**	CAUGGAG[I]UCCGCGACCA[I]G
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]	**ABO-rI-9**	CAUGGAGUUC[I]GCGACCA[I]G
ABO-rI-2	CAUGGAG[I]UCCGCGACCA[I][I]	**ABO-rI-10**	CAUGGAGUUC[I]GCGACCAC[I]
ABO-rI-3	CAUGGAG[I]UCCGCGACCACG	**ABO-rI-11**	CAUGGAGUUCCGCGACCA[I][I]
ABO-rI-4	CAUGGAGUUC[I]GCGACCACG	**ABO-rI-12**	CAUGGAG[I]UC[I]GCGACCA[I]G
ABO-rI-5	CAUGGAGUUCCGCGACCA[I]G	**ABO-rI-13**	CAUGGAG[I]UC[I]GCGACCAC[I]
ABO-rI-6	CAUGGAGUUCCGCGACCAC[I]	**ABO-rI-14**	CAUGGAGUUC[I]GCGACCA[I][I]
ABO-rI-7	CAUGGAG[I]UC[I]GCGACCACG	**ABO-rI-15**	CAUGGAG[I]UC[I]GCGACCA[I][I]

b

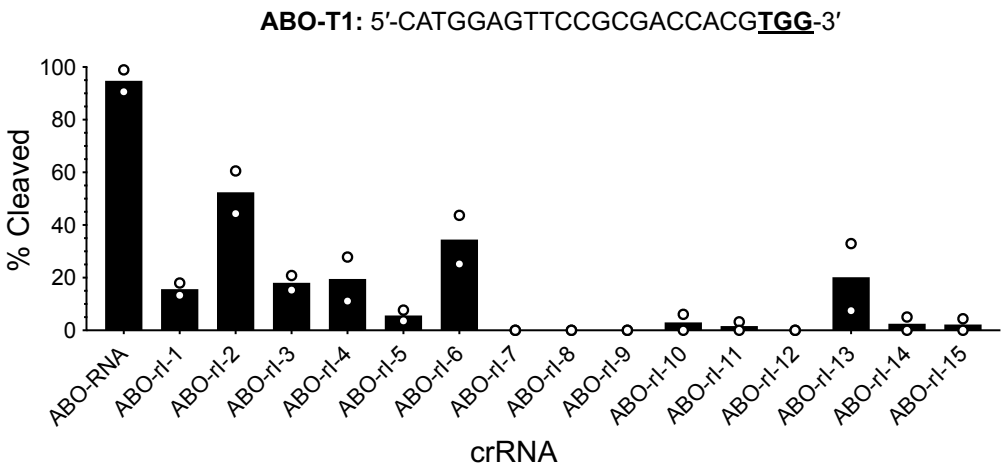


Figure 4.5. *In* vitro activity of inosine-modified crRNAs on ABO-T1 target
(**a**) List of modified crRNA sequences with inosine position(s) indicated by a red [I]. (**b**) Bar graph showing the relative amount of DNA cleavage resulting from *in vitro* reactions using the indicated inosine-modified crRNAs versus the ABO-T1 target sequence; Mean with individual data points shown (n = 2 independent experiments). Reactions were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM). Quantification of cleavage percentages was performed using ImageJ.

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]
ABO-rI-8	CAUGGAG[I]UCCGCGACCA[I]G

ABO-T1 Target: 5′- CATGGAGTTCCGCGACCACG**TGG** -3′

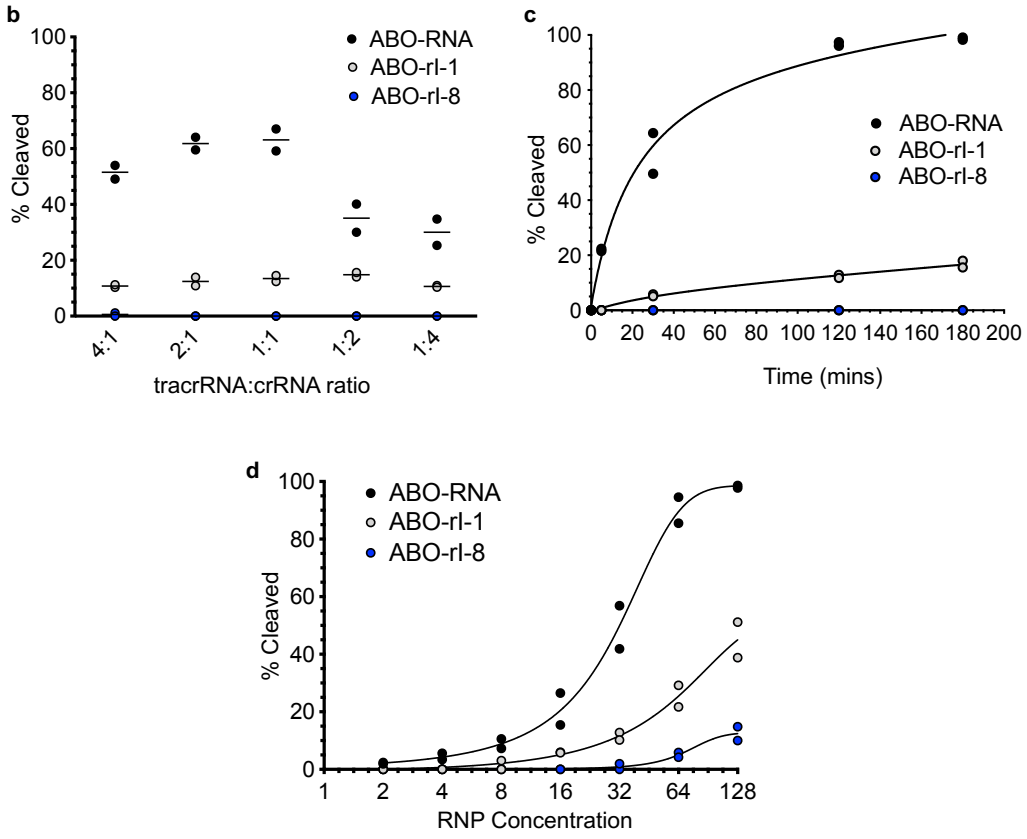


Figure 4.6. Effect of inosine modifications on the kinetics and activity of Cas9
(**a**) List of inosine-modified crRNA sequences. A red [I] indicates the position of ribose inosine modifications in the crRNA sequence. (**b**) Graph showing Cas9 cleavage activity as a function of the tracrRNA:crRNA using either ABO-RNA, ABO-rI-1, or ABO-rI-8 versus the ABO-T1 target sequence; Mean with individual data points shown (n = 2 independent experiments). (**c**) Time course of Cas9 cleavage of ABO-T1 using either ABO-RNA, ABO-rI-1, or ABO-rI-8; Individual data points shown (n = 2 independent experiments). Kinetic assays were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM) and measured a the indicated time points. d Graph showing cleavage activity resulting from a titration of Cas9 RNP complex loaded with either ABO-RNA, ABO-rI-1, or ABO-rI-8 versus the ABO-T1 target sequence; Individual data points shown (n = 2 independent experiments).

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]
ABO-rI-8	CAUGGAG[I]UCCGCGACCA[I]G

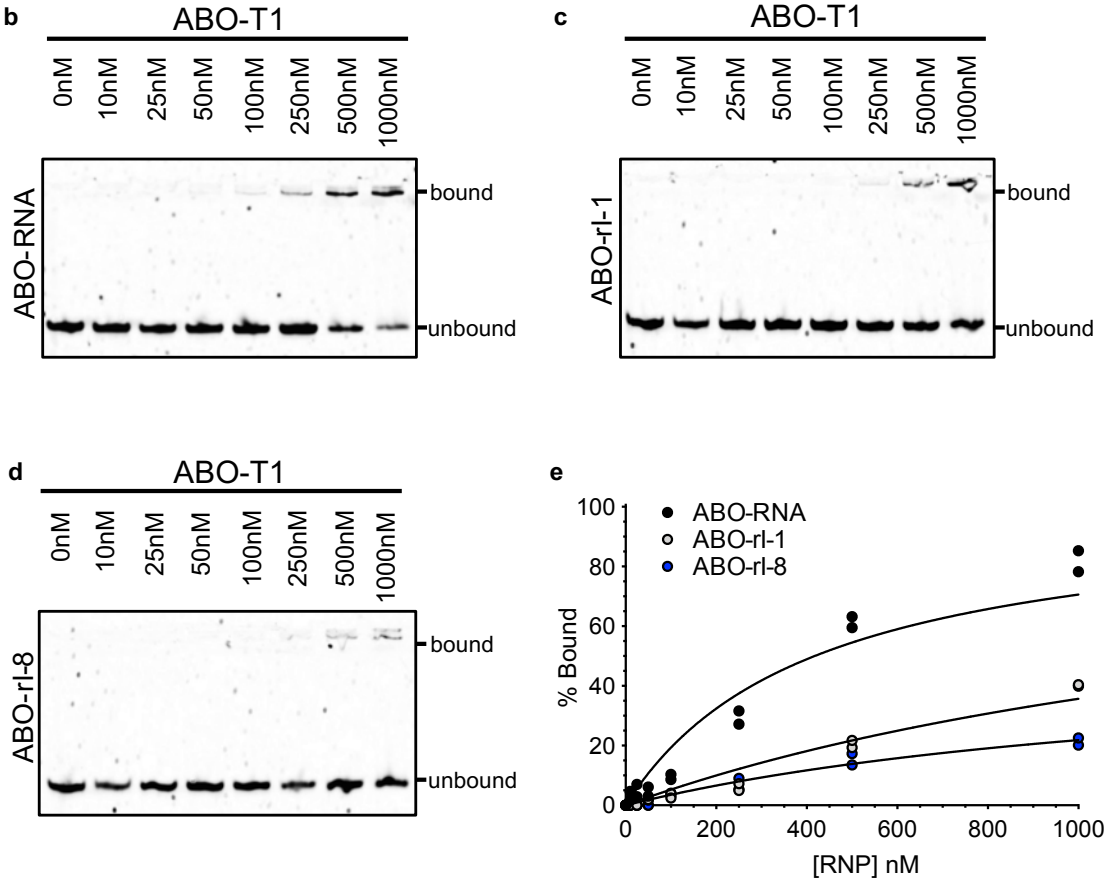


Figure 4.7. Effect of inosine modifications in crRNAs on the ability of dCas9 to bind target DNA

(**a**) List of inosine-modified crRNA sequences. Red [I] indicates the position of ribose inosine modifications in the crRNA sequence. Representative gels showing binding of nuclease-deficient Cas9 (dCas9) to ABO-T1 target DNA using (**b**) ABO-RNA, (**c**) ABO-rI-1 or (**d**) ABO-rI-8 crRNAs. The top band represents the bound DNA substrate while the bottom band corresponds to the unbound substrate. (**e**) Graph showing the DNA binding resulting from a titration of Cas9 RNP complex loaded with either ABO-RNA, ABO-rI-1, or ABO-rI-8 versus the ABO-T1 target sequence; Individual data points shown (n = 2 independent experiments). Binding experiments were performed in duplicate with similar results.

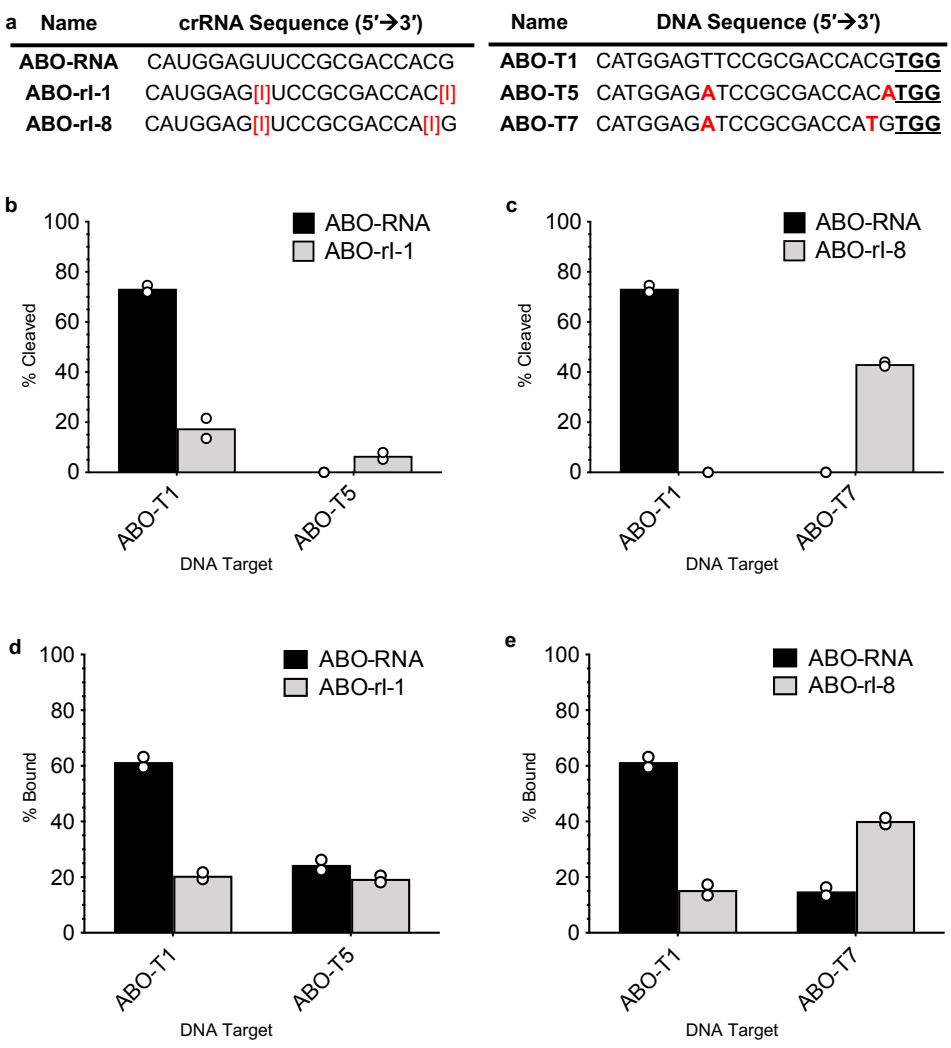


Figure 4.8. Relationship between cleavage activity and target engagement of Cas9 using inosine-modified crRNAs

(**a**) List of inosine-modified crRNA sequences and DNA target sequences. Red [I] indicates the position of ribose inosine modifications in the crRNA sequence. SNPs in DNA targets are indicated with red lettering. Bar graphs showing the relative amount of DNA cleavage resulting from *in vitro* reactions containing Cas9 with ABO-RNA or (**b**) ABO-rI-1 or (**c**) ABO-rI-8 versus the indicated target sequences; Mean with individual data points shown (n = 2 independent experiments). Bar graphs showing the relative amount of DNA binding resulting from *in vitro* reactions containing dCas9 with ABO-RNA or (**d**) ABO-rI-1 or (**e**) ABO-rI-8 versus the indicated target sequences; Mean with individual data points shown (n = 2 independent experiments). *In vitro* binding assays were performed using fixed concentrations of crRNA (750 nM) and Cas9 (500 nM). Quantification of cleavage percentages was performed using ImageJ.

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]
ABO-rI-8	CAUGGAG[I]UCCGCGACCA[I]G

b

ABO-T1 Target: 5′- CATGGAGTTCCGCGACCACG**TGG** -3′

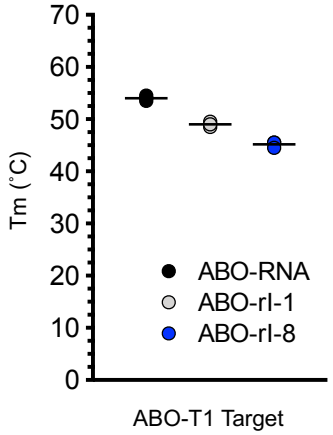


Figure 4.9. Effect of ribose inosine modifications on crRNA-DNA target heteroduplex melting temperature
(**a**) List of inosine-modified crRNA sequences. Red [I] indicates the position of ribose inosine modifications in the crRNA sequence. (**b**) Plot showing melting temperature for heteroduplexes comprised of single- stranded sequence corresponding to ABO-T1 DNA and an RNA oligonucleotide corresponding to the spacer portion of ABO-RNA, ABO-rI-1, or ABO-rI-8; Mean with individual data points shown (n = 3 independent experiments).

4.3.2 Inclusion of universal bases into crRNAs alters the specificity only at the site of incorporation

A prerequisite for the practical application of guide RNAs containing universal bases to targeting SNPs is that they must alter Cas9 specificity in a localized and predictable manner. That is to say, they should impart selective degeneracy rather than globally impacting the precision of Cas9 DNA cleavage. To evaluate this, we employed a previously described high-throughput specificity profiling assay[13, 52, 194] that measures Cas9 cleavage of a library of $>10^{12}$ off-target sequences,

containing a tenfold coverage of all sequences with ≤8 mutations relative to the ABO-T1 sequence (**Figure 4.10**). We performed the assay on the unmodified crRNA and all 15 of the ribose inosine-modified crRNAs listed in **Figure 4.5a**, as well as all of the crRNAs modified using alternative universal bases listed in **Figure 4.2c**, **d**. We used the datasets for each crRNA to calculate enrichment scores for each base at each position within the ABO-T1 sequence and generated specificity heatmaps to visualize the results. For the collection of inosine-modified crRNAs, we found that in nearly all cases the specificity profile for the crRNAs containing universal bases was similar to that of ABO-RNA at all positions except those that overlapped with the locations of the universal bases (**Figure 4.10b**, **Figures 4.11-4.13**). Moreover, substitution of the indicated bases with inosine rendered the crRNA virtually non-specific at that position (**Figure 4.10b**), and was associated with changes in specificity scores ranging from approximately −0.6 to −1.0 at those sites (**Figure 4.12**). Similar results were observed from the analysis of the crRNAs bearing deoxyribose inosine, 2′OMe ribose inosine, deoxyribose 5′-nitroindole, deoxyribose K and deoxyribose P base modifications (**Figure 4.10**, **Figures 4.14-4.17**). Overall, we found that specificity at the site of universal base incorporation was virtually abolished, while specificity at other locations appeared to be preserved, or even enhanced in certain cases (**Figure 4.15**). Substitution of the indicated PAM-distal uracil with ribose inosine (ABO-rI-1 and ABO-rI-2), deoxyribose inosine (ABO-dI-1 and ABO-dI-2), or 2′O methyl inosine (RNA-mI-1 and RNA-mI-2) rendered the crRNA non-specific at this position (**Figure 4.15**) and was associated with a difference in specificity score in excess of −0.6 (**Figure 4.16**). Similar results were observed when the indicated PAM-proximal cytosine base was replaced by a universal base, while specificity at the PAM-proximal guanine position was less affected, ostensibly due to an initial lack of specificity at this position in ABO-RNA (**Figure 4.15**, **4.16**). Finally, to generalize our findings to

other DNA target sequences, we synthesized a separate set of 8 crRNAs with inosine modifications at positions corresponding to SNPs present in a region of the major histocompatibility complex *HLA* gene. As shown in **Figures 4.18**-**4.21**, inclusion of inosine bases in this crRNA similarly abolished specificity in a site-restricted manner. Collectively, these data reveal that inclusion of universal bases in crRNAs imparts selective degeneracy at the site of incorporation without otherwise altering specificity, and that this effect extends to compositionally distinct DNA targets.

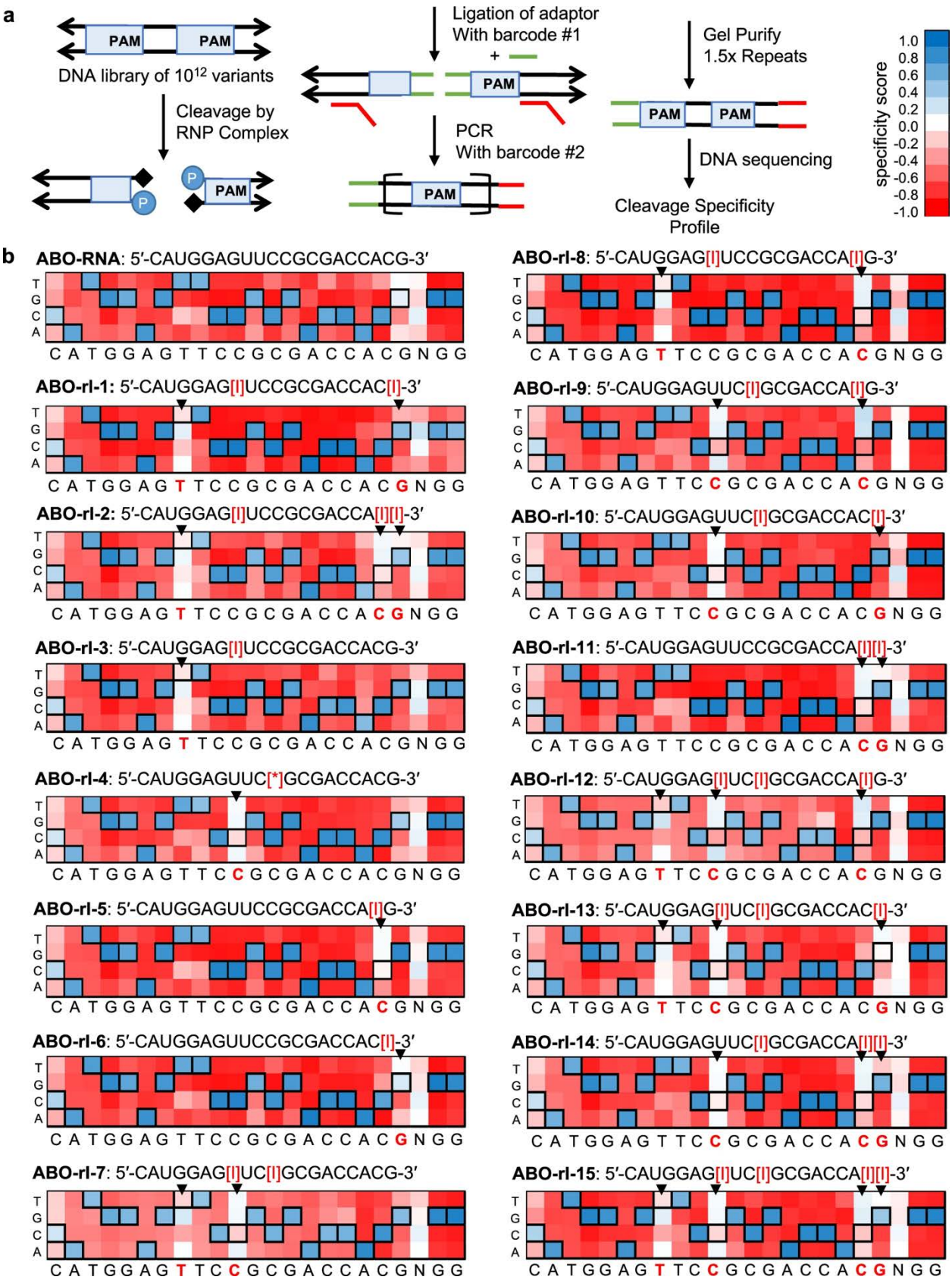

a
PAM
PAM
DNA library of 10^12 variants
Cleavage by
RNP Complex
P
P
PAM
Ligation of adaptor
With barcode #1
+
PAM
PCR
With barcode #2
PAM
Gel Purify
1.5x Repeats
PAM
PAM
DNA sequencing
Cleavage Specificity
Profile
specificity score
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
b
ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′
ABO-rI-1: 5′-CAUGGAG[I]UCCGCGACCAC[I]-3′
ABO-rI-2: 5′-CAUGGAG[I]UCCGCGACCA[I][I]-3′
ABO-rI-3: 5′-CAUGGAG[I]UCCGCGACCACG-3′
ABO-rI-4: 5′-CAUGGAGUUC[*]GCGACCACG-3′
ABO-rI-5: 5′-CAUGGAGUUCCGCGACCA[I]G-3′
ABO-rI-6: 5′-CAUGGAGUUCCGCGACCAC[I]-3′
ABO-rI-7: 5′-CAUGGAG[I]UC[I]GCGACCACG-3′
ABO-rI-8: 5′-CAUGGAG[I]UCCGCGACCA[I]G-3′
ABO-rI-9: 5′-CAUGGAGUUC[I]GCGACCA[I]G-3′
ABO-rI-10: 5′-CAUGGAGUUC[I]GCGACCAC[I]-3′
ABO-rI-11: 5′-CAUGGAGUUCCGCGACCA[I][I]-3′
ABO-rI-12: 5′-CAUGGAG[I]UC[I]GCGACCA[I]G-3′
ABO-rI-13: 5′-CAUGGAG[I]UC[I]GCGACCAC[I]-3′
ABO-rI-14: 5′-CAUGGAGUUC[I]GCGACCA[I][I]-3′
ABO-rI-15: 5′-CAUGGAG[I]UC[I]GCGACCA[I][I]-3′
T
G
C
A
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G
C A T G G A G T T C C G C G A C C A C G N G G

Figure 4.10. Inclusion of inosine bases in crRNAs affects specificity mainly at the site of incorporation

(**a**) Diagram depicting the workflow for the high-throughput specificity profiling assay. (**b**) Heatmaps corresponding to the specificity profiles of the indicated ribose inosine-modified crRNAs. The positions of inosine bases are indicated by black arrows. Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of −1.0 (dark red) correspond to 100% enrichment against a base-pair at a specific position. Black boxes denote the intended target nucleotide.

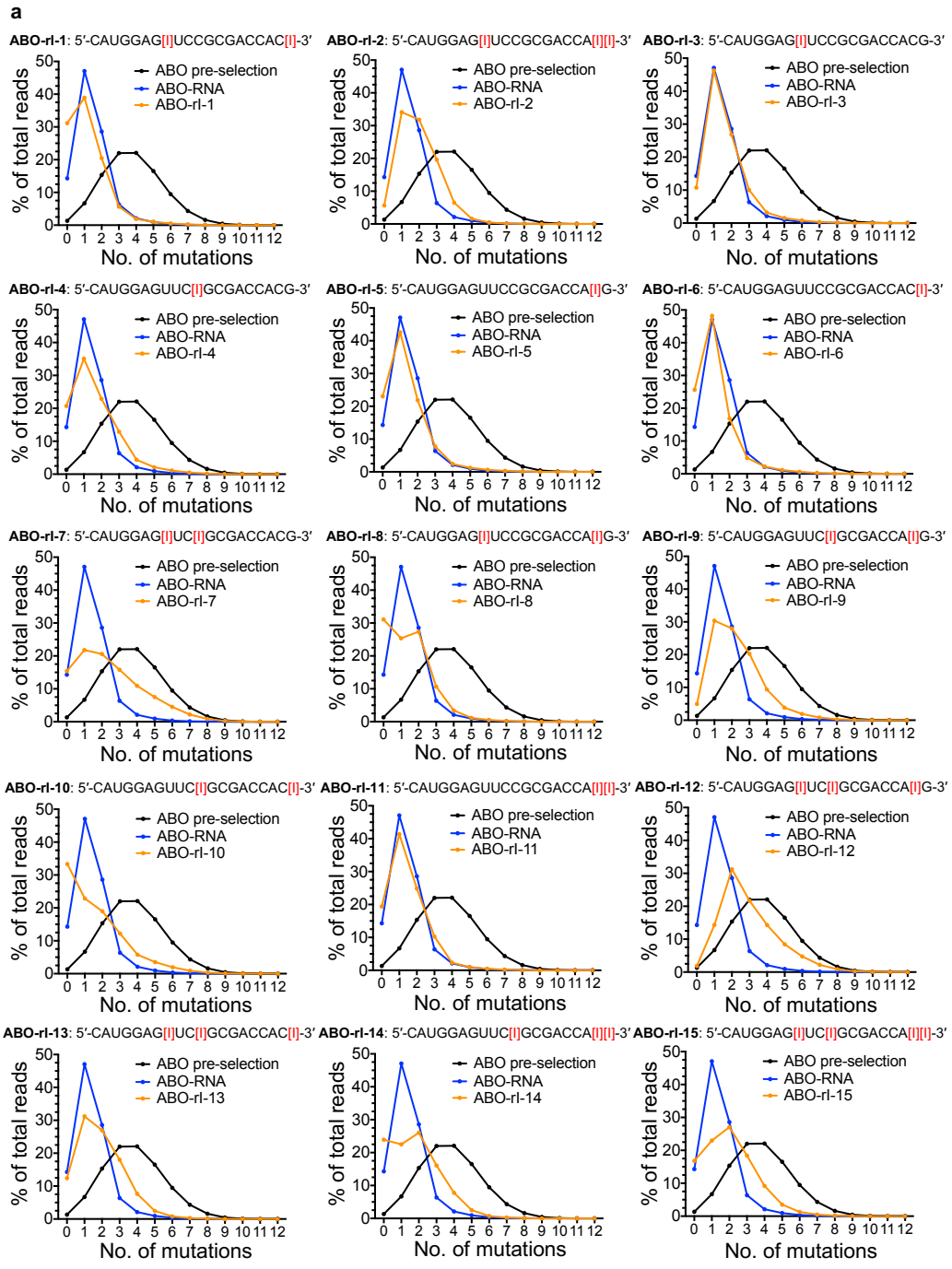


Figure 4.11. Distribution of mutations in pre- and post-selection libraries for ABO-RNA and inosine-modified crRNAs

(**a**) Graphs indicating the number of target sequence mutations in each pre- and post-selection library. ABO pre- selection (black) and ABO-RNA post-selection (blue) data are compared to the post- selection of crRNAs containing ribose inosine(s) (orange). A red [I] denotes the position of inosine modifications in the crRNA sequence. Mutations were counted for each position throughout the 20 bp target site.

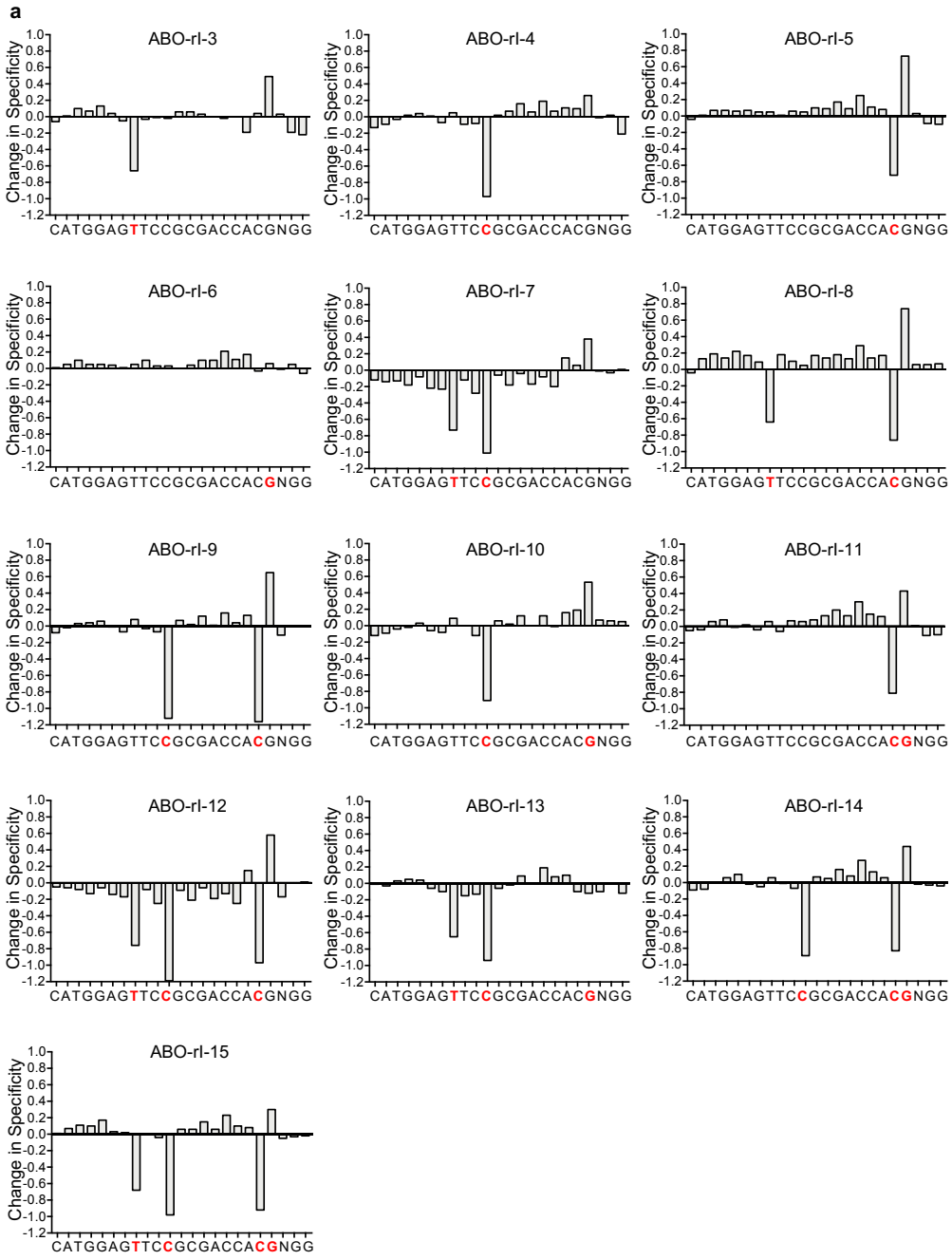


Figure 4.12. Change in specificity score of inosine-modified crRNAs compared to ABO-RNA

(**a**) Bar graphs showing the quantitative difference in specificity score at each position in the DNA target site for inosine-modified crRNAs. SNP locations in the 20 base-pair DNA target are indicated with red lettering. The PAM is shown as "NGG" on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score(ABO-RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-rI-1	CAUGGAG[I]UCCGCGACCAC[I]
ABO-rI-2	CAUGGAG[I]UCCGCGACCA[I][I]
ABO-rI-3	CAUGGAG[I]UCCGCGACCACG
ABO-rI-4	CAUGGAGUUC[I]GCGACCACG
ABO-rI-5	CAUGGAGUUCCGCGACCA[I]G
ABO-rI-6	CAUGGAGUUCCGCGACCAC[I]
ABO-rI-7	CAUGGAG[I]UC[I]GCGACCACG

Name	crRNA Sequence (5′→3′)
ABO-rI-8	CAUGGAG[I]UCCGCGACCA[I]G
ABO-rI-9	CAUGGAGUUC[I]GCGACCA[I]G
ABO-rI-10	CAUGGAGUUC[I]GCGACCAC[I]
ABO-rI-11	CAUGGAGUUCCGCGACCA[I][I]
ABO-rI-12	CAUGGAG[I]UC[I]GCGACCA[I]G
ABO-rI-13	CAUGGAG[I]UC[I]GCGACCAC[I]
ABO-rI-14	CAUGGAGUUC[I]GCGACCA[I][I]
ABO-rI-15	CAUGGAG[I]UC[I]GCGACCA[I][I]

b

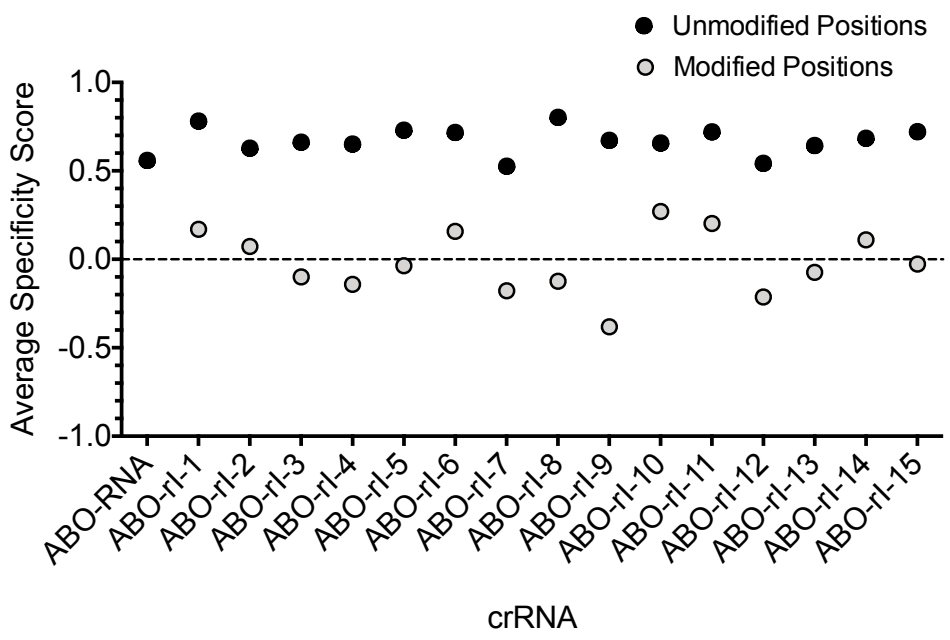


Figure 4.13. Average specificity score of ribose inosine modified positions in the *ABO* crRNAs

(**a**) List of inosine-modified crRNA sequences. Red [I] indicates the position of ribose inosine modifications in the crRNA sequence. (**b**) Graph showing the average specificity score for all non-modified positions vs. all modified positions for the indicated crRNAs. This value was calculated by averaging the specificity scores at each nucleotide position as visualized with heat maps in **Figure 4.10**. Specificity scores of 1.0 correspond to 100% enrichment for, while scores of −1.0 correspond to 100% enrichment against a base-pair at a specific position. These scores were averaged for all unmodified positions: (sum of specificity scores for each unmodified position in the crRNA) / (# of unmodified positions in the crRNA) = average unmodified specificity score (black). The average of all the modified positions was also calculated: (sum of specificity scores for each modified position in the crRNA) / (# of modified positions in the crRNA) = average modified specificity score (grey). The dotted horizontal line represents an average crRNA specificity score of 0.

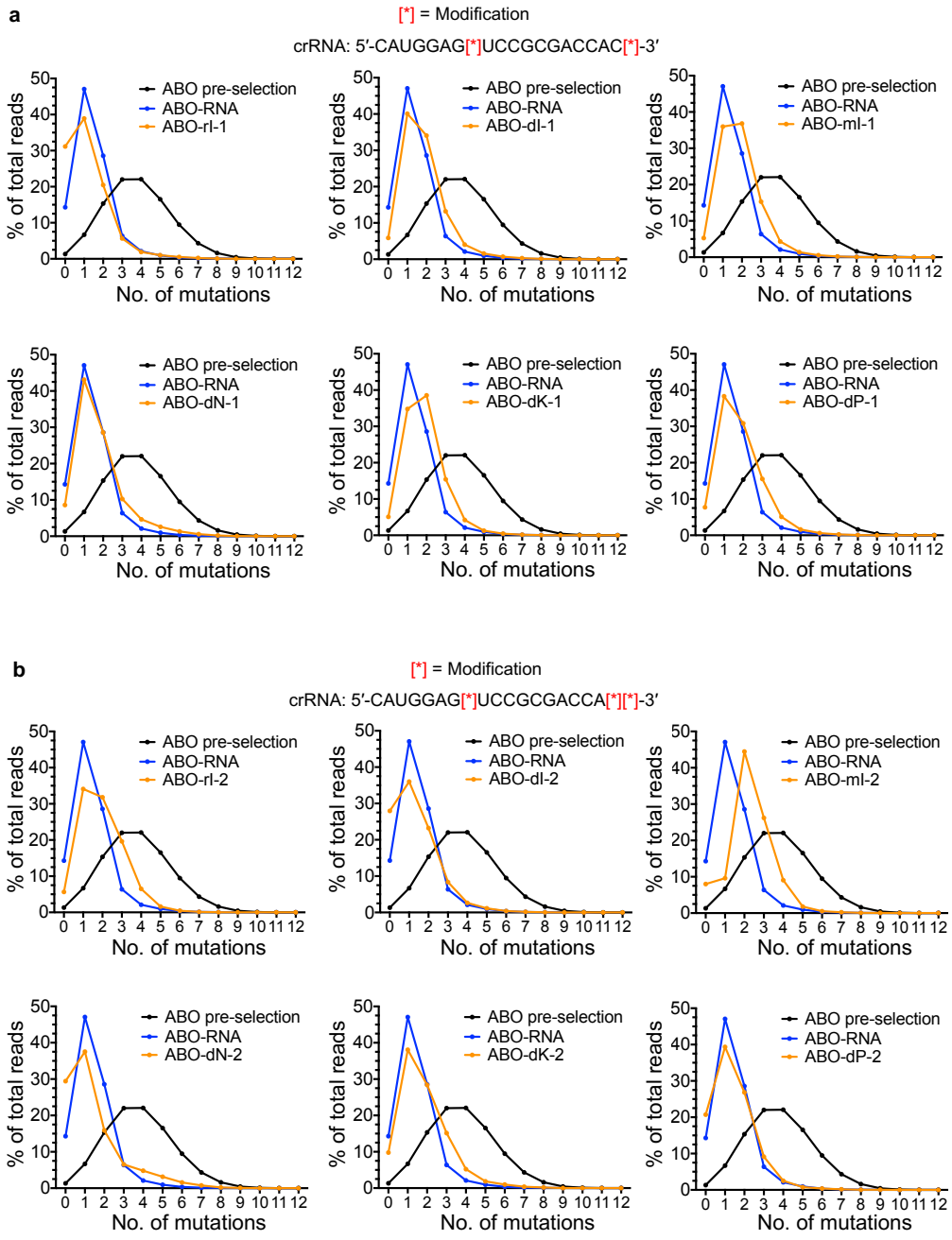


Figure 4.14. Distribution of mutations in pre- and post-selection libraries for ABO-RNA and universal base-modified crRNAs

Graphs indicating the number of target sequence mutations in each pre- and post-selection library using crRNAs modified with universal bases at (**a**) two or (**b**) three positions, as indicated. ABO pre-selection (black) and ABO-RNA post-selection (blue) data are compared to the post- selection of crRNAs containing the indicated universal base(s) (orange). A red [*] denotes the position of the universal base modification in the crRNA sequence. Mutations were counted for each position throughout the 20 bp target site.

a

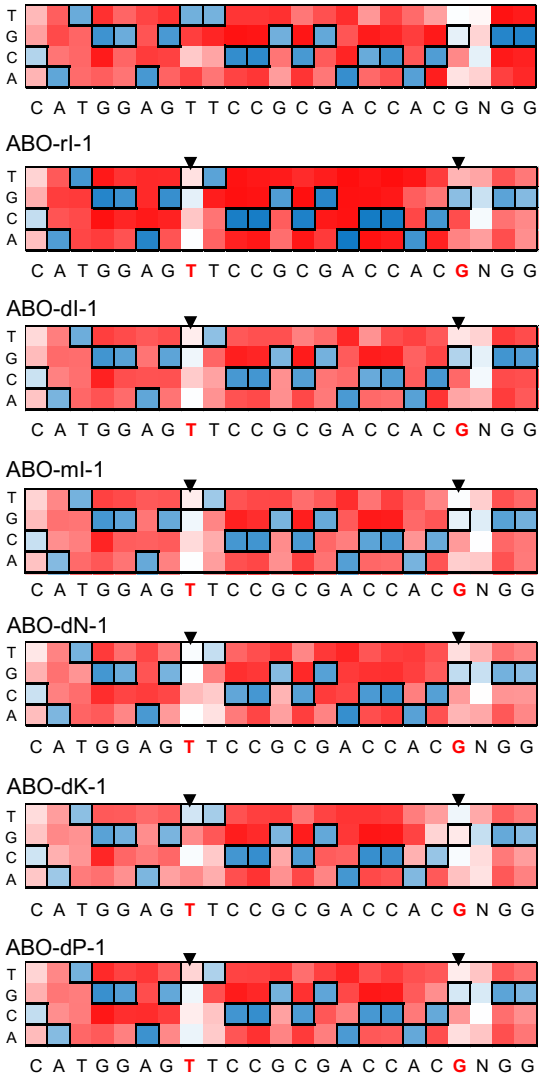
Target	DNA Sequence (5′→3′)	Allele Frequency (%)
ABO-T1	CATGGAGTTCCGCGACCACG**TGG**	54.21
ABO-T2	CATGGAG**A**TCCGCGACCACG**TGG**	25.50
ABO-T3	CATGGAGTTCCGCGACCA**T**G**TGG**	15.98
ABO-T4	CATGGAGTTCCGCGACCAC**ATGG**	0.015
ABO-T5	CATGGAG**A**TCCGCGACCAC**ATGG**	0.0038
ABO-T6	CATGGAG**A**TCCGCGACCA**TATGG**	0.00061
ABO-T7	CATGGAG**A**TCCGCGACCA**T**A**TGG**	4.08
ABO-T8	CATGGAGTTCCGCGACCA**TATGG**	0.0024

b

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCAC[*]-3′

ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′



c

[*] = Modification

crRNA: 5′-CAUGGAG[*]UCCGCGACCA[*][*]-3′

ABO-RNA: 5′-CAUGGAGUUCCGCGACCACG-3′

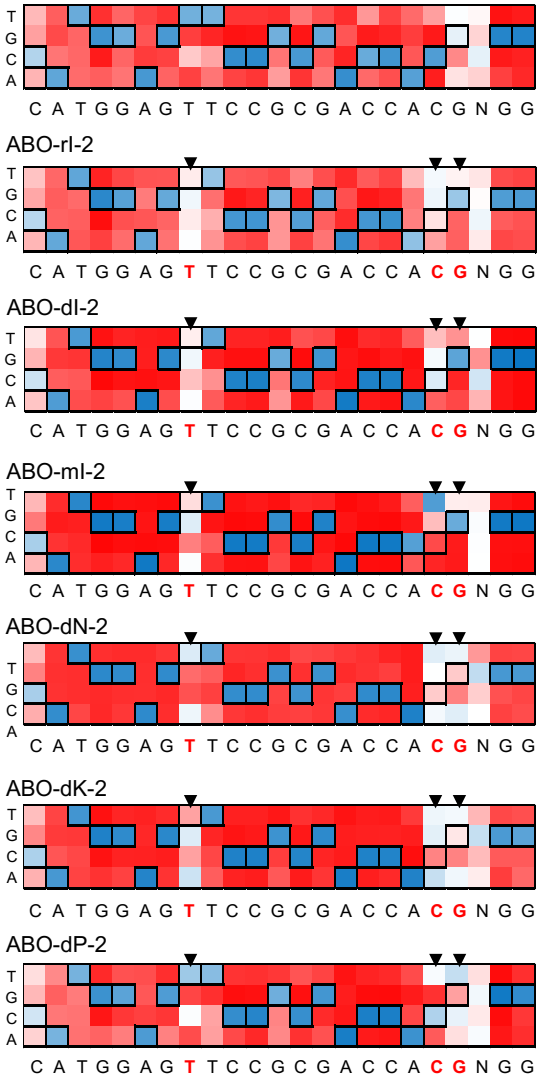


Figure 4.15. *In vitro* specificity profiles for *ABO* crRNAs containing various universal base modifications

(**a**) List of DNA targets corresponding to sequences in the *ABO* gene based on clinical polymorphism data. SNPs are indicated with red lettering. The PAM sequence is underlined. Allele frequency indicates either the current tallied allele frequency or the statistically predicted frequency (for sequences containing multiple SNPs). Heat maps corresponding to the specificity profiles of crRNAs modified with universal bases at (**b**) two or (**c**) three positions, as indicated. The positions of universal bases are indicated by black arrows. Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of −1.0 (dark red) correspond to 100% enrichment against a base-pair at a specific position. Black boxes denote the intended target nucleotide.

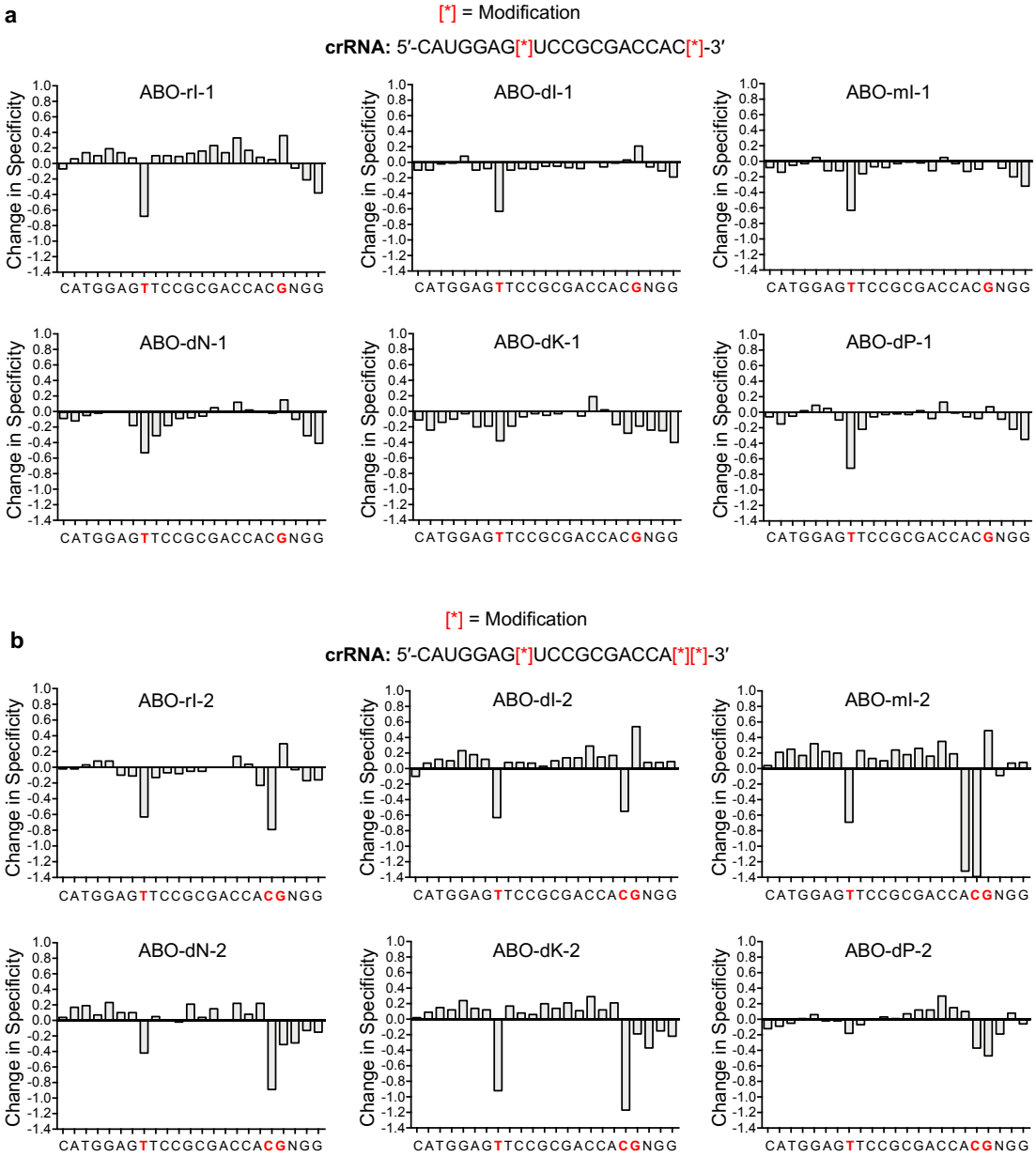


Figure 4.16. Change in specificity score of universal base-modified crRNAs compared to ABO-RNA

Bar graphs showing the quantitative difference in specificity score at each position in the DNA target site for crRNAs modified with universal bases at (**a**) two or (**b**) three positions, as indicated. SNP locations in the 20 bp DNA target are indicated with red lettering. The PAM is shown as "NGG" on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score(ABO- RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

a

Name	crRNA Sequence (5′→3′)
ABO-RNA	CAUGGAGUUCCGCGACCACG
ABO-[X]-1	CAUGGAG[*]UCCGCGACCAC[*]
ABO-[X]-2	CAUGGAG[*]UCCGCGACCA[*][*]

b

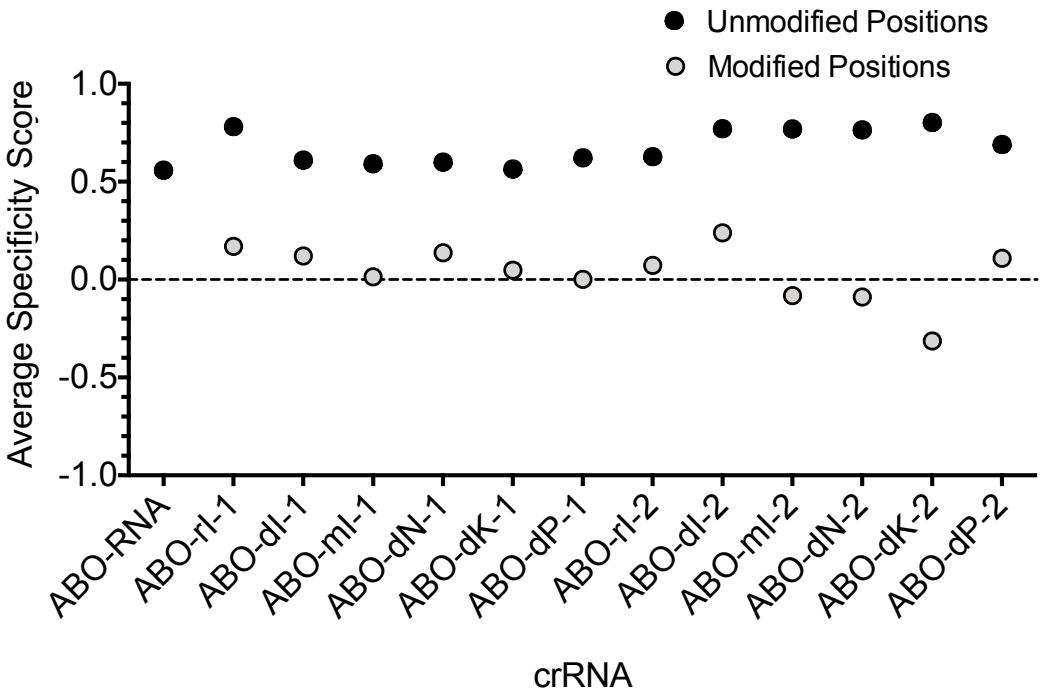


Figure 4.17. Average specificity score of universal base modified positions in the *ABO* crRNAs

(**a**) List of universal base-modified crRNA sequences. Red [I] indicates the position of the universal base modifications in the crRNA sequence. (**b**) Graph showing the average specificity score for all non-modified positions vs. all modified positions for the indicated crRNAs. This value was calculated by averaging the specificity scores at each nucleotide position as visualized with heat maps in **Figure 4.15**. Specificity scores of 1.0 correspond to 100% enrichment for, while scores of −1.0 correspond to 100% enrichment against a base-pair at a specific position. These scores were averaged for all unmodified positions: (sum of specificity scores for each unmodified position in the crRNA) / (# of unmodified positions in the crRNA) = average unmodified specificity score (black). The average of all the modified positions was also calculated: (sum of specificity scores for each modified position in the crRNA) / (# of modified positions in the crRNA) = average modified specificity score (grey). The dotted horizontal line represents an average crRNA specificity score of 0.

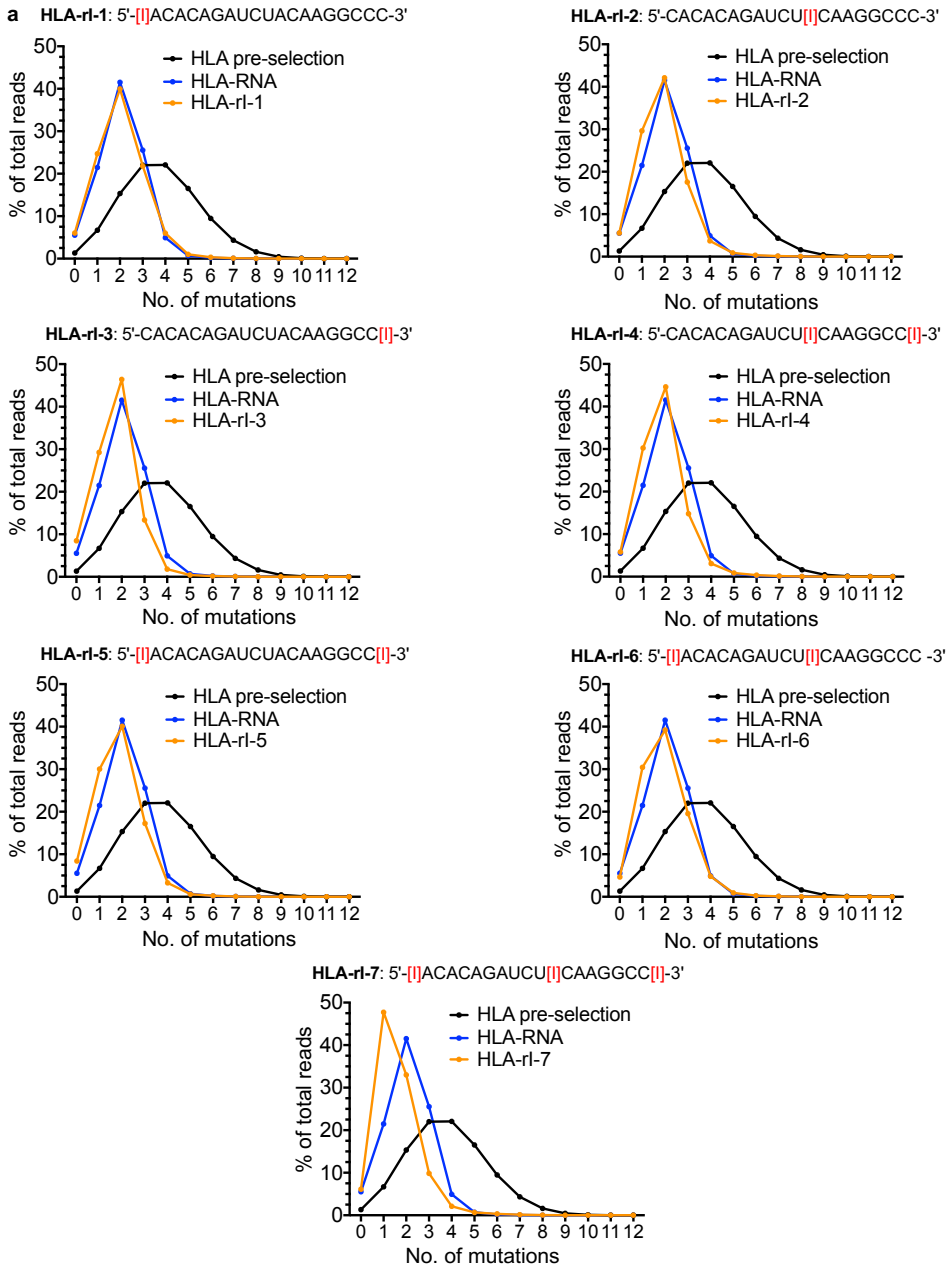


Figure 4.18. Distribution of mutations in pre- and post-selection libraries for HLA-RNA and inosine-modified crRNAs

(**a**) Graphs indicating the number of target sequence mutations in each pre- and post-selection library. HLA pre- selection (black) and HLA-RNA post-selection (blue) data are compared to the post- selection of crRNAs containing ribose inosine(s) (orange). Red [I] denotes the position of inosine modifications in the crRNA sequence. Mutations were counted for each position throughout the 20 bp target site.

a

Target	DNA Sequence (5′→3′)	Allele Frequency (%)
HLA-T1	CACACAGATCTACAAGGCCC**AGG**	0.50
HLA-T2	**G**ACACAGATCTACAAGGCCC**AGG**	45.49
HLA-T3	CACACAGATCT**C**CAAGGCCC**AGG**	36.70
HLA-T4	CACACAGATCTACAAGGCC**A**AGG	74.20
HLA-T5	CACACAGATCT**C**CAAGGCC**A**AGG	27.23
HLA-T6	**G**ACACAGATCTACAAGGCC**A**AGG	33.75
HLA-T7	**G**ACACAGATCT**C**CAAGGCCC**AGG**	16.69
HLA-T8	**G**ACACAGATCT**C**CAAGGCC**A**AGG	12.39

b

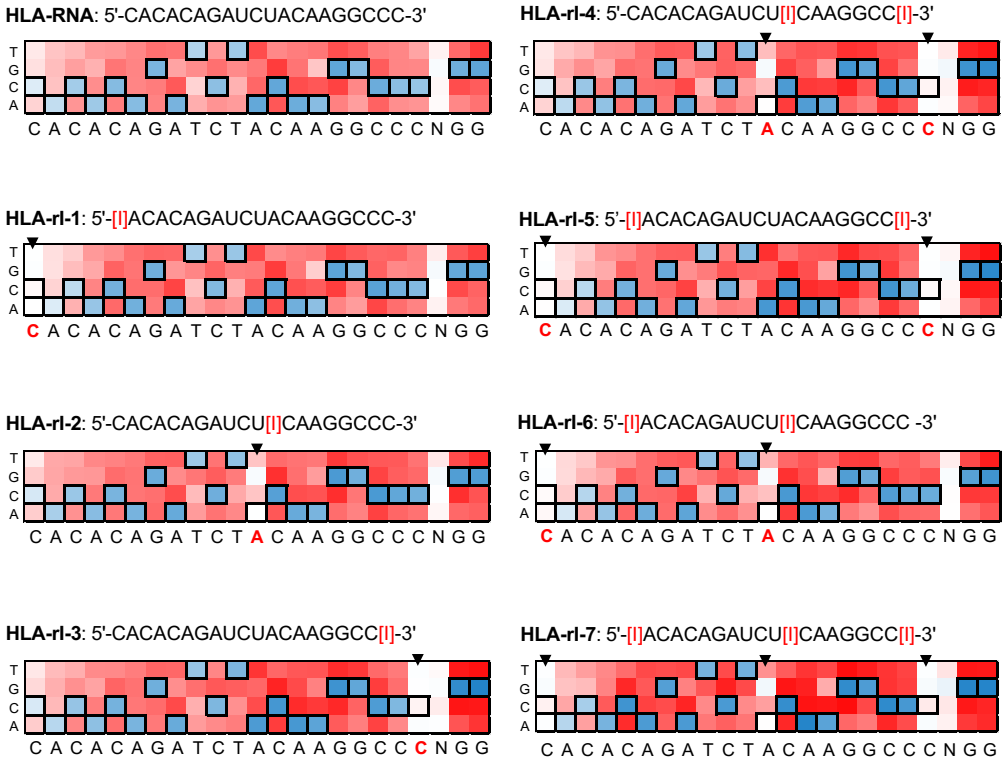


Figure 4.19. In vitro specificity profiles for HLA crRNAs containing inosine base modifications

(**a**) List of DNA targets corresponding to sequences in the *HLA* gene based on clinical polymorphism data. SNPs are indicated with red lettering. The PAM sequence is underlined. Allele frequency indicates either the current tallied allele frequency or the statistically predicted frequency (for sequences containing multiple SNPs). (**b**) Heat maps corresponding to the specificity profiles of crRNAs modified with inosine bases at the indicated positions. The positions of inosine bases are indicated by black arrows. Specificity scores of 1.0 (dark blue) correspond to 100% enrichment for, while scores of −1.0 (dark red) correspond to 100% enrichment against a base-pair at a specific position. Black boxes denote the intended target nucleotide.

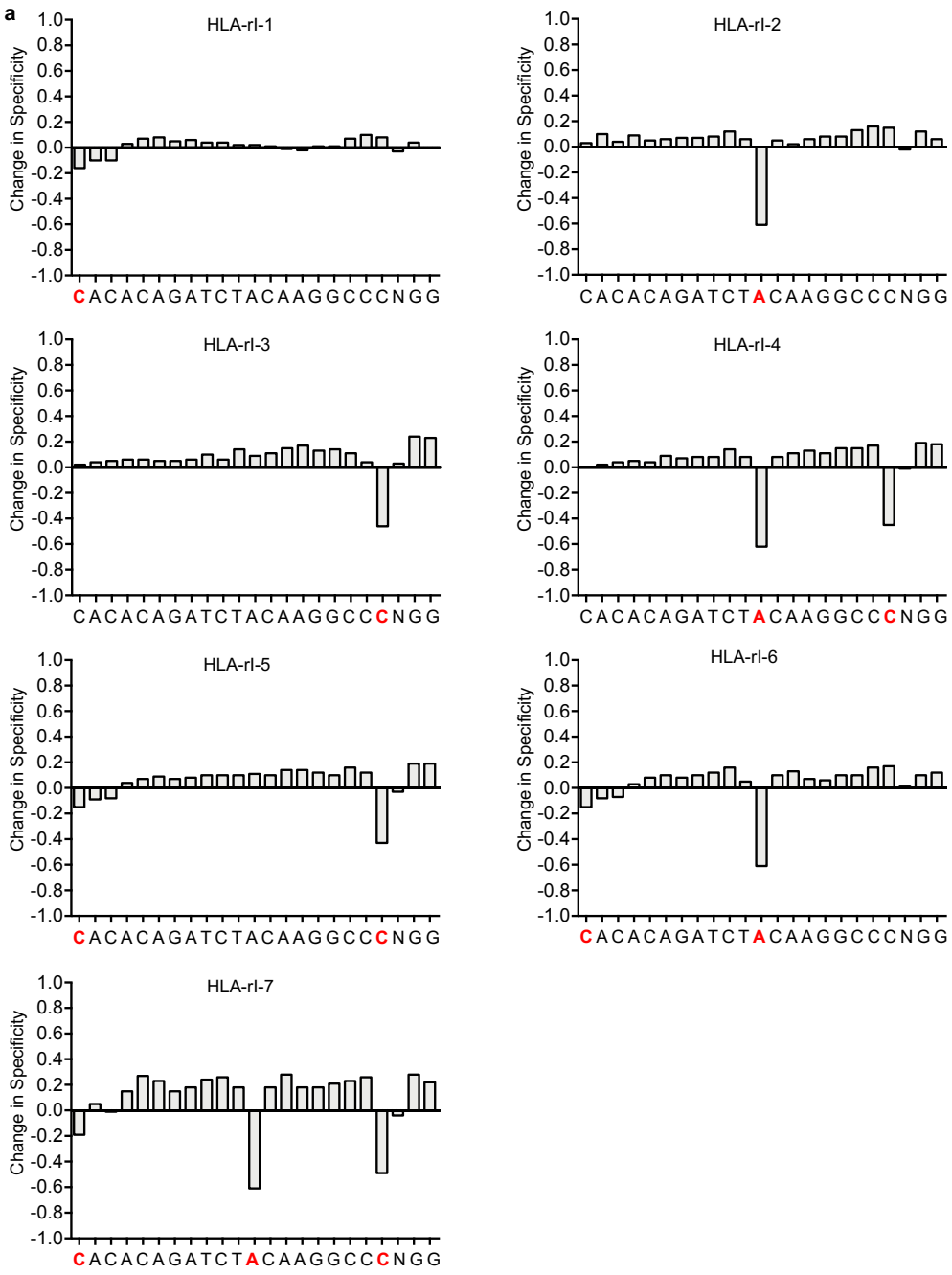


Figure 4.20. Change in specificity score of inosine-modified crRNAs compared to HLA-RNA

(**a**) Bar graphs showing the quantitative difference in specificity score at each position in the DNA target site for crRNAs modified with inosine bases as indicated. SNP locations in the 20 base-pair DNA target are indicated with red lettering. The PAM is shown as "NGG" on the 3′ end of the target. A score of zero indicates no change in specificity. The difference in specificity was calculated as the specificity score(modified)−specificity score(HLA-RNA). The specificity scoring of each nucleotide position is relative to the pre-selection control library data.

a

Name	crRNA Sequence (5′→3′)
HLA-RNA	CACACAGAUCUACAAGGCCC
HLA-rI-1	[I]ACACAGAUCUACAAGGCCC
HLA-rI-2	CACACAGAUCU[I]CAAGGCCC
HLA-rI-3	CACACAGAUCUACAAGGCC[I]
HLA-rI-4	CACACAGAUCU[I]CAAGGCC[I]
HLA-rI-5	[I]ACACAGAUCUACAAGGCC[I]
HLA-rI-6	[I]ACACAGAUCU[I]CAAGGCCC
HLA-rI-7	[I]ACACAGAUCU[I]CAAGGCC[I]

b

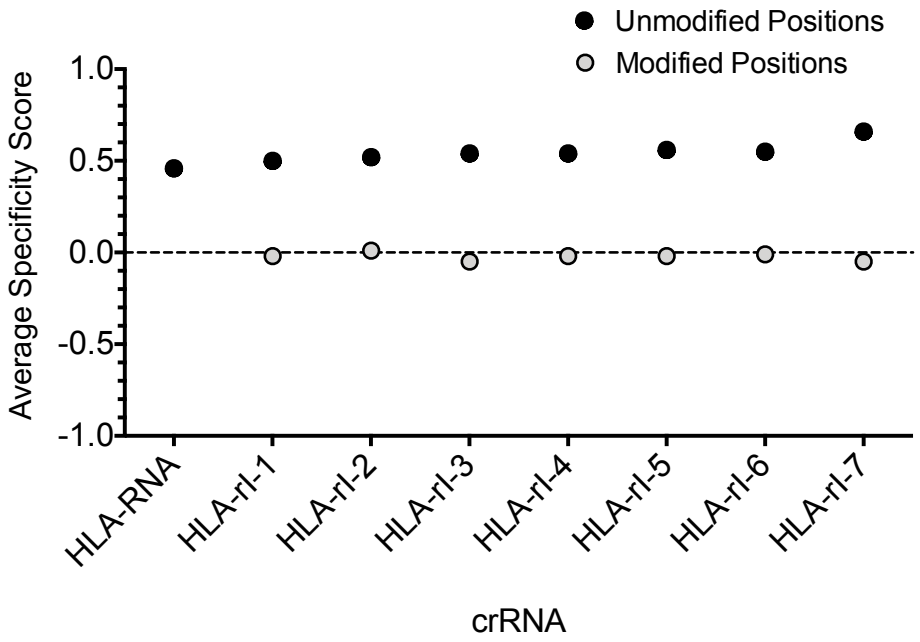
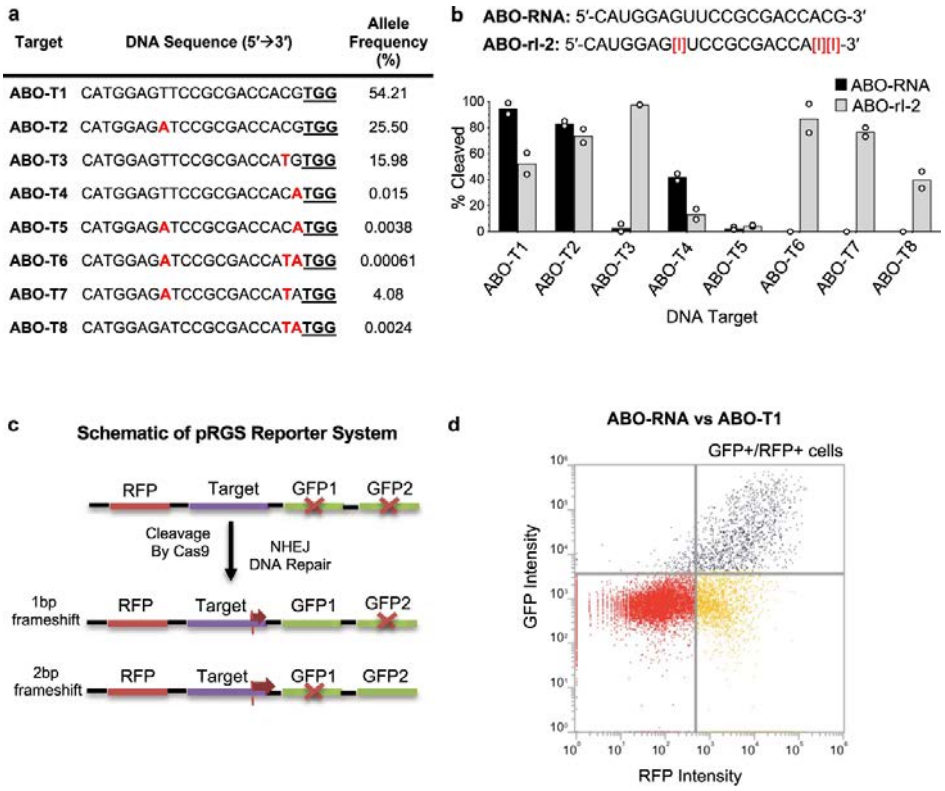


Figure 4.21. Average specificity score of ribose inosine modified positions in the *HLA* crRNAs

(**a**) List of inosine-modified crRNA sequences. Red [I] indicates the position of ribose inosine modifications in the crRNA sequence. (**b**) Graph showing the average specificity score for all non-modified positions vs. all modified positions for the indicated crRNAs. This value was calculated by averaging the specificity scores at each nucleotide position as visualized with heat maps in **Figure 4.19**. Specificity scores of 1.0 correspond to 100% enrichment for, while scores of −1.0 correspond to 100% enrichment against a base-pair at a specific position. These scores were averaged for all unmodified positions: (sum of specificity scores for each unmodified position in the crRNA) / (# of unmodified positions in the crRNA) = average unmodified specificity score (black). The average of all the modified positions was also calculated: (sum of specificity scores for each modified position in the crRNA) / (# of modified positions in the crRNA) = average modified specificity score (grey). The dotted horizontal line represents an average crRNA specificity score of 0.

4.3.3 crRNAs containing universal bases can direct Cas9 cleavage of polymorphic sequences in cells, but with limitations

Knowing that inclusion of universal bases in crRNAs could impart selective degeneracy while broadly maintaining cleavage specificity *in vitro*, we sought to determine if our results could be translated to cells. As an initial test, we adapted a plasmid-based fluorescence reporter system[274] and used it to measure the cleavage of eight heterologous *ABO* sequences in cells. First, we selected ABO-rI-2, which bears three inosine modifications, and tested its ability to direct Cas9 cleavage of ABO-T1, the corresponding triple SNP variant (ABO-T6), three double SNP sequences (ABO-T5, 7, 8) and three single SNP sequences (ABO-T2, T3, T4) *in vitro* (**Figure 4.22a**). We found ABO-rI-2 directed >50% cleavage of 6/8 sequences tested, the exceptions being ABO-T4 (~20%) and ABO-T5 (<10%) (**Figure 4.22b**). In contrast, ABO-RNA only supported robust Cas9 cleavage of >50% of its matched sequence (ABO-T1) and ABO-T2 (**Figure 4.22b**). Next, we cloned all of the target DNAs into a plasmid in which sequences were flanked by an in-frame mRFP gene at the 5′ end and two out-of-frame eGFP genes at the 3′ end (**Figure 4.22c**). Past work has shown that double-strand breaks formed in the intervening target sequence can be repaired by non-homologous end-joining (NHEJ), resulting in frameshift mutations that generate a multifluorescent mRFP-eGFP fusion protein (**Figure 4.22c**)[274]. We co-transfected all eight constructs with either ABO-RNA or ABO-rI-2 into HeLa cells stably expressing Cas9 and used fluorescence-activated cell sorting (FACS) to quantify the resulting cell populations (**Figure 4.22d**). Using the ABO-RNA, Cas9 cleaved 3/8 sequences with >20% efficiency (ABO-T1, ABO-T2, ABO-T4). However, 7/8 sequences were cleaved with >20% efficiency, and ABO-T4 RNA was cleaved at 16% efficiency when the ABO-rI-2 guide RNA was used (**Figure 4.22e**).



e

Target	ABO-RNA (%GFP+/%RFP+)	ABO-rI-2 (%GFP+/%RFP+)
ABO-T1	36.82 ± 0.75	26.59 ± 0.73
ABO-T2	27.06 ± 0.30	29.40 ± 0.94
ABO-T3	19.27 ± 1.31	39.24 ± 0.33
ABO-T4	25.86 ± 0.43	15.75 ± 0.67
ABO-T5	13.94 ± 1.82	21.43 ± 1.02
ABO-T6	5.21 ± 0.84	35.76 ± 1.09
ABO-T7	6.01 ± 0.56	46.35 ± 0.52
ABO-T8	5.45 ± 0.46	26.65 ± 1.48

Figure 4.22. crRNAs containing inosine bases direct simultaneous cleavage of polymorphic gene variants in cells

(**a**) List of *ABO* variant DNA target sequences (ABO-T1-T8) assayed in cells. Positions of SNPs are indicated with red lettering. The PAM sequence is underlined. (**b**) Bar graphs showing the relative amount of DNA cleavage resulting from in vitro reactions containing Cas9 with ABO-RNA or ABO-rI-2 versus the indicated DNA target sequences. Assays were performed using fixed concentrations of gRNA (80 nM) and Cas9 (40 nM); Mean with individual data points shown ($n = 2$ independent experiments). (**c**) Schematic outlining the framework for the fluorescence-based assay used to evaluate cleavage of the *ABO* variant target sequences in cells. (**d**) Representative FACS plot showing the distribution of RFP and GFP positive cells. Dual positive cells appear in the top right quadrant. (**e**) Table showing normalized %GFP + /all %RFP + events corresponding to cleavage of the indicated target sequences in cells by Cas9 using either ABO-RNA or ABO-rI-2; Mean ± S.D. shown ($n = 3$ independent samples).

To examine the utility of this approach for targeting endogenous sites in cells, we sequenced several loci containing PAM sequences in 293 T and HeLa cells that were predicted to contain SNPs based on the Ensembl[275] and HEK293T[276] reference genomes. We identified a homozygous sequence within the *HLA-C* gene differing at 2 base positions between 293 T and HeLa cells (**Figure 4.23a**). We generated a crRNA corresponding to the *HLA-C* sequence in 293 T cells (HLA-C-RNA), and verified its ability to direct Cas9 cleavage of the *HLA-C* gene in 293 T cells (HLA-C-T1) but not HeLa cells (HLA-C-T2) (**Figure 4.23b**). We then synthesized crRNAs containing ribose inosine, deoxyribose inosine, or deoxyribose P bases at positions overlapping with the locations of the mismatches in the HLA-C-T1 and HLA-C-T2 sequences (**Figure 4.23c**). We tested the ability of these universal base-modified crRNAs and the unmodified crRNA to direct Cas9 cleavage of the two target sequences *in vitro*. We found that Cas9 was able to robustly cut both the HeLa and 293 T *HLA-C* sequences when the rI, dI, and dP base-modified crRNAs were used (**Figure 4.23d**). The unmodified crRNA induced ~60% cleavage of its corresponding target (HLA-C-T1), but only ~30% when HLA-C-T2 was used as a substrate (**Figure 4.23d**). The fact that Cas9 cleavage of this off-target sequence was absent in cells using the unmodified crRNA is consistent with previous reports in literature showing higher stringency against off-target cutting in cells[194, 202, 206]. Finally, we tested the ability of the unmodified and the rI-, dI-, and dP-modified crRNAs to direct Cas9 cleavage of the *HLA-C* locus in 293 T and HeLa cells. As previously observed, we found that the HLA-C-RNA supported Cas9 cleavage of the *HLA-C* locus in 293 T cells (~40%) but not HeLa cells (0%) (**Figure 4.23e**). In contrast to our *in vitro* findings, HLA-C-rI and HLA-C-dI showed either weak or undetectable Cas9 cleavage activity in both 293 T and HeLa cells. However, HLA-C-dP was able to direct Cas9 cleavage of the *HLA-C* locus in both 293T (~40%) and HeLa cells (~12%). Collectively, these results demonstrate the potential for

universal base-modified crRNAs to drive Cas9 cleavage of polymorphic sequences in cells, but also reveal some limitations to their general use.

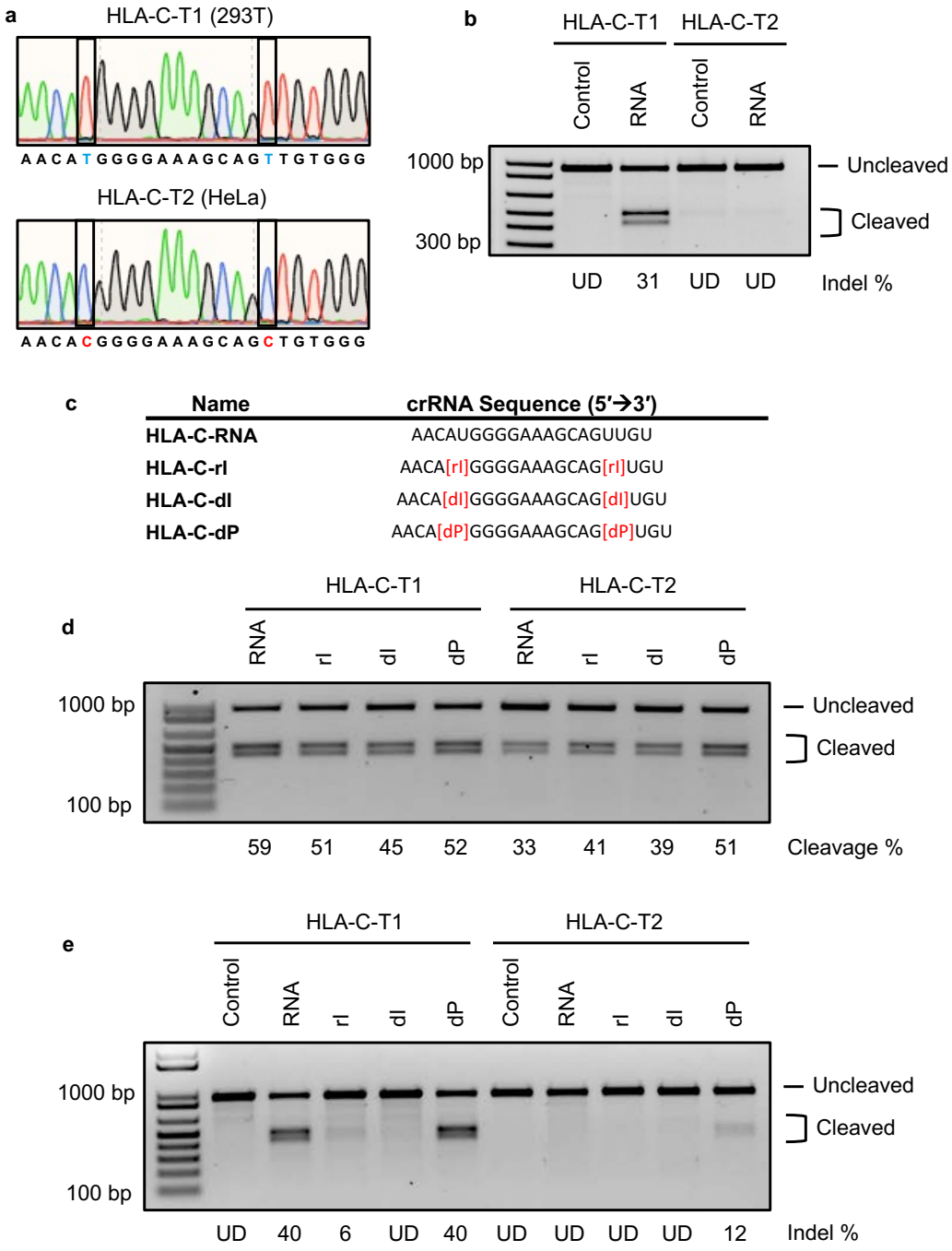


Figure 4.23. Cas9 cleavage activity using HLA-C-RNA and universal base-modified crRNAs *in vitro* and on the endogenous *HLA-C* locus in 293T and HeLa cells
(**a**) Sanger sequencing trace of endogenous SNPs present within the *HLA-C* locus of 293T and HeLa cells. (**b**) Gel showing the cellular Cas9 cleavage efficiencies of HLA-C-T1 or HLA-C-T2 sequences using an umodified HLA-C-RNA crRNA as determined by T7 endonuclease I digestion. Controls were harvested from cells stably expressing Cas9 that were not transfected with a guide RNA. (**c**) List of universal base- modified crRNA sequences. Red [X] indicates the position of the modified base within the crRNA sequence. (**d**) Gel showing the relative *in vitro* cleavage efficiencies of Cas9 RNPs complexed with HLA-C-RNA or universal base-modified crRNAs against the HLA- C-T1 or HLA-C-T2 DNA targets. (**e**) Gel showing the relative cellular cleavage efficiences of Cas9 against HLA-C-T1 or HLA-C-T2 when complexed with HLA-C-RNA or universal base-modified crRNAs as determined by T7 endonuclease I digestion. Controls were harvested from cells stably expressing Cas9 that were not transfected with a guide RNA. UD denotes undetectable. Both *in vitro* and cellular cleavage assays of HLA-C- T1/T2 were performed in duplicate with similar results.

We wondered if the discrepancy between the activity of the HLA-C-rI and HLA-C-dI crRNAs *in vitro* and in cells could be the result of delayed Cas9 cleavage kinetics. Previous work has shown that modification of the ribose sugar in crRNAs can lead to slower enzyme kinetics that manifests as reduced activity in cells[194]. To test this hypothesis, we performed a Cas9 cleavage time course on DNA substrates corresponding to either HLA-C-T1 or HLA-C-T2 using HLA-C-RNA or HLA-C-rI, -dI, or -dP crRNAs. As shown in **Figure 4.24**, we found that Cas9 cleavage of HLA-C-T1 using HLA-C-rI or HLA-C-dI crRNAs was slower than with HLA-C-RNA or HLA-C-dP by a factor of ~4 fold. Furthermore, we found that Cas9 cleavage of the HLA-C-T2 substrate using HLA-C-dP was substantially quicker than cleavage using the HLA-C-RNA, -rI, and -dI crRNAs (**Figure 4.24**). This strong correlation between cellular modification rates and *in vitro* kinetics suggests that delayed enzyme kinetics could underlie the low activity of the HLA-C-rI and HLA-C-dI crRNAs in cells.

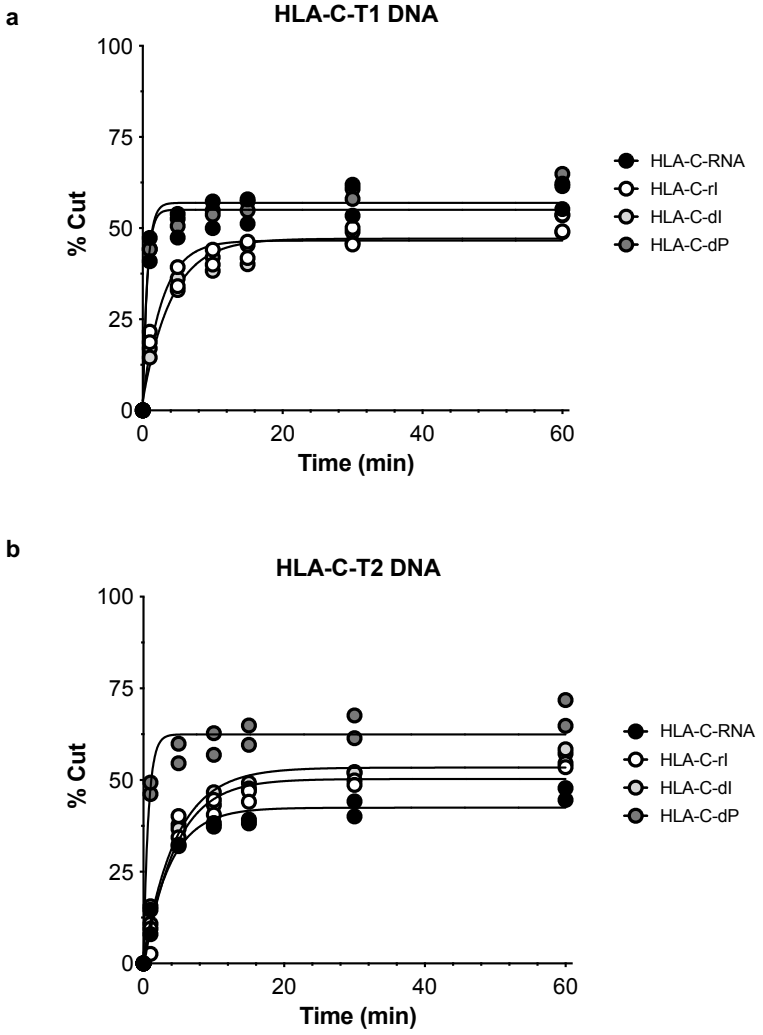


Figure 4.24. *In vitro* Cas9 cleavage kinetics using HLA-C-RNA or universal base-modified crRNAs
Time course showing *in vitro* cleavage activity of Cas9 on (**a**) HLA-C-T1 or (**b**) HLA-C-T2 DNA targets using HLA-C-RNA or the indicated universal base-modified crRNAs. Experiments were performed with a concentration of 10 nM Cas9, 20 nM gRNA and 5 nM DNA target; Individual data points shown (n = 2 independent experiments).

4.3.4 DETECTR probes containing universal bases identify evolved variants of a pathogen

In addition to its use as a gene-editing agent, Cas12a/Cpf1 has also successfully been harnessed for diagnostic purposes as part of the DETECTR system[70]. Point-of-need technologies using this

platform to diagnose swine flu[277] as well as COVID-19[278] have now been deployed. However, the prospect of viral evolution presents a unique challenge for the identification of these pathogens, as mutations could subvert detection by Cas12a guide probes designed to target only reference sequences, leading to false negative results. We hypothesized that inosine bases could be incorporated into Cas12a guide RNAs to impart them with selectively degenerate targeting capabilities in order to circumvent this limitation. To test this possibility, we selected a DNA sequence from the HIV-1 protease gene and identified seven clinically-relevant sequence variants bearing 1, 2, or 3 SNPs encoding mutations that confer resistance of the virus toward HIV protease inhibitor drugs[279, 280] (**Figure 4.25a**). Next, we synthesized two crRNAs, HIV-RNA to target the canonical sequence, and HIV-rI-1, which contains three inosine substitutions designed to enable flexible targeting of both the canonical and evolved variant sequences. An *in vitro* cleavage assay of all target sequences using HIV-RNA with Cas12a revealed that HIV-T1, HIV-T3, HIV-T4, HIV-T7 were cleaved at efficiencies of 55%, 30%, 55%, and ~30%, respectively (**Figure 4.25b**, **c**). In stark contrast, all eight sequences were fully cleaved when HIV-rI-1 was used as the guide RNA (**Figure 4.25b**, **c**), supporting our assertion and revealing a high degree of tolerance for the presence of inosine substitutions in Cas12a guide RNAs. To ensure that the lack of cleavage activity observed with HIV-RNA on sequences such as HIV-T8 was not simply due to insufficient RNP, we performed titrations of RNP concentration. Consistent with our model, we found that overall Cas12a cleavage activity (combined *cis* and *trans*) was comparable between HIV-RNA and HIV-rI-1 using the HIV-T1 substrate (**Figure 4.26a**, **b**). However, HIV-RNA was unable to direct cleavage of HIV-T8, in contrast to HIV-rI-1, which induced complete cleavage of this substrate at an RNP concentration of ~25 nM (**Figure 4.26c**, **d**). Subsequently, we ported these probes into the DETECTR system, outlined in **Figure 4.25d**. To simulate pathogen DNA, we

cloned each of our eight target sequences into pUC19 plasmids and performed recombinant polymerase amplification (RPA) as described in the protocol[70]. Next, we set up individual reactions containing each DNA sample paired with either HIV-RNA or HIV-rI-1 probes in the presence of a fluorescent detection substrate. As shown in **Figure 4.25e**, the HIV-rI-1 probe positively identified all eight of the HIV-1 variant sequences, while the HIV-RNA probe only identified three sequences and provided false negatives for the other five variants. These findings provide justification for the use of universal base-modified crRNAs in CRISPR-based diagnostic platforms.

Viral escape due to mutation of the target site to a variant is a major roadblock to using CRISPR therapeutics as antivirals[281]. Based on our results demonstrating effective targeting of polymorphic sequences using HIV-rI-1 *in vitro*, we wondered if this crRNA could direct Cas12a cleavage of variant viral sequences in cells. To test this possibility, we used the Flp/FRT system[282] to stably integrate single copies of the HIV-T1 and HIV-T8 sequences into 293 cells (**Figure 4.27a**, **b**). We found that the unmodified HIV-RNA crRNA directed robust Cas12a cleavage of the HIV-T1 site (~28%) but virtually no cleavage at the HIV-T8 site (<3%) (**Figure 4.27c**). In contrast, the HIV-rI-1 RNA induced cleavage of both sites with relatively equal efficacy (HIV-T1: 6%, HIV-T8: 8%) (**Figure 4.27c**). This corresponds to a change in HIV-T1:HIV-T8 cleavage preference of >12-fold. Importantly, we did not detect any DNA cleavage using either crRNA at two predicted genomic off-target sites (**Figure 4.27d**, **e**). These data demonstrate that crRNAs containing inosine modifications can be used in combination with Cas12a to cleave polymorphic sequences in cells, albeit with reduced activity.

a

Target	DNA Sequence (5′→3′)	Protease Inhibitor Drug Resistant Mutation(s)
HIV-T1	TTTGATAAAACCTCCAATTCCCCCTAT	N/A
HIV-T2	TTTGATAAGACCTCCAATTCCCCCTAT	F53L
HIV-T3	TTTGATAAAACCTCCAATTCCCACTAT	G48V
HIV-T4	TTTGATAAAACCTCCAATTCCCCCTAC	I47V
HIV-T5	TTTGATAAGACCTCCAATTCCCACTAT	F53L, G48V
HIV-T6	TTTGATAAGACCTCCAATTCCCCCTAC	F53L, I47V
HIV-T7	TTTGATAAAACCTCCAATTCCCACTAC	G48V, I47V,
HIV-T8	TTTGATAAGACCTCCAATTCCCACTAC	F53L, G48V, I47V

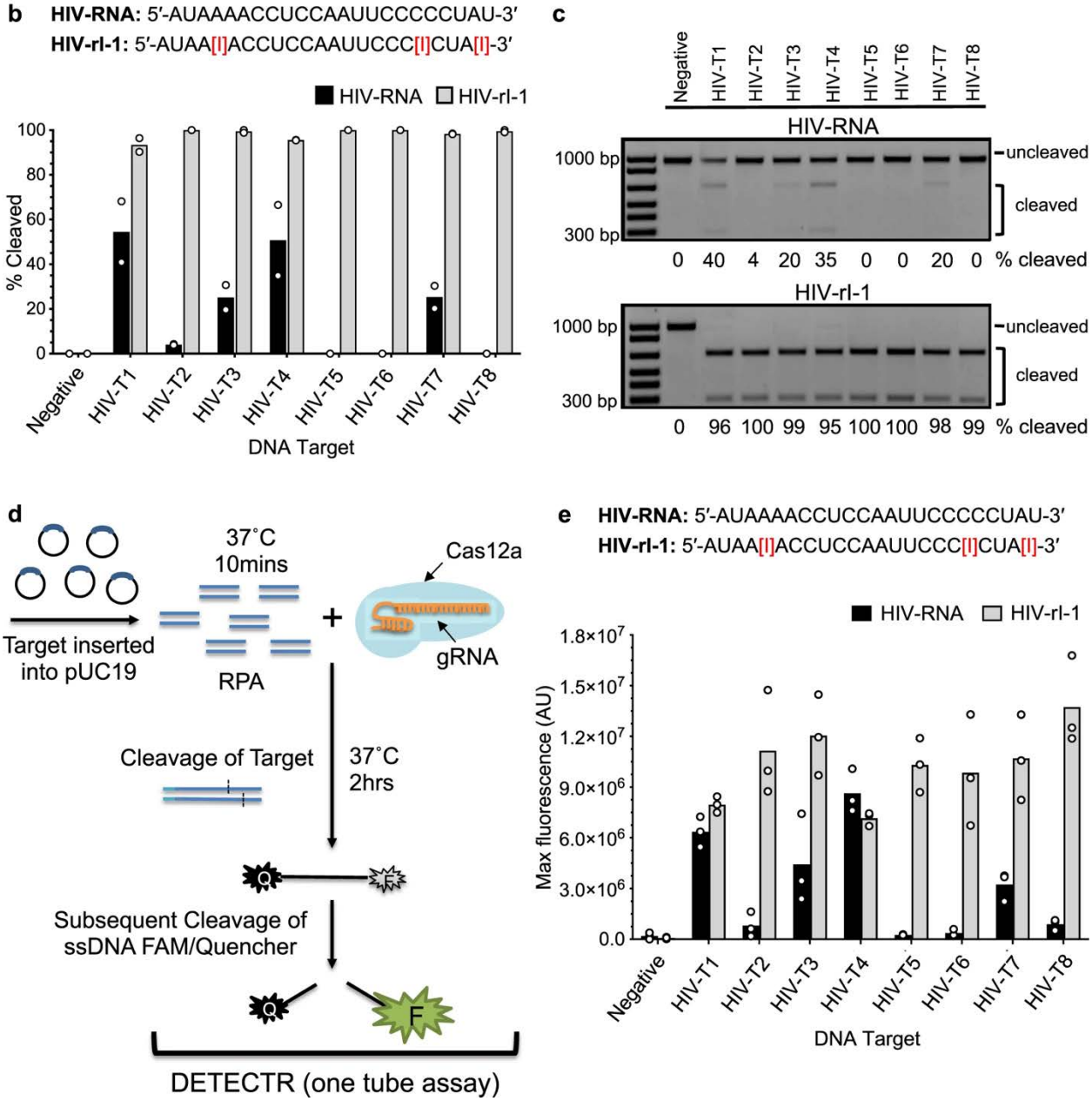


Figure 4.25. Incorporation of inosine bases into crRNA probes for the Cas12a-based DETECTR system enables the detection of evolved HIV-1 DNA target sequences
(**a**) List of DNA target sequences derived from the HIV-1 protease gene containing evolved SNPs detected in patient samples. SNPs position(s) are indicated with red lettering. The PAM sequence is underlined. (**b**) Bar graphs showing the relative amount of DNA cleavage resulting from in vitro reactions containing Cas9 with HIV-RNA or HIV-rI-1 versus the indicated DNA target sequences. A scrambled crRNA with the sequence 5′-AUUCUUGCUCUGCUCUCUUCGUC-′3 was used as a negative control. Assays were performed using fixed concentrations of crRNA (125 nM) and Cas12a (100 nM); Mean with individual data points shown ($n = 2$ independent experiments). (**c**) Representative gels of the in vitro cleavage assay results for Cas12a with HIV-RNA or HIV-rI-1 versus the indicated DNA target sequences. The bottom two bands in the gel represent the cleaved DNA substrate while the top band corresponds to the undigested substrate. Cleavage experiments were performed in duplicate with similar results. (**d**) Diagram outlining the DETECTR assay. e Bar graph indicating the fluorescence signal obtained in the DETECTR assay using Cas12a in combination with either HIV-RNA or HIV-rI-1 and samples containing the indicated target sequences. Max fluorescence values were normalized to background; Mean with individual data points shown ($n = 3$ independent experiments).

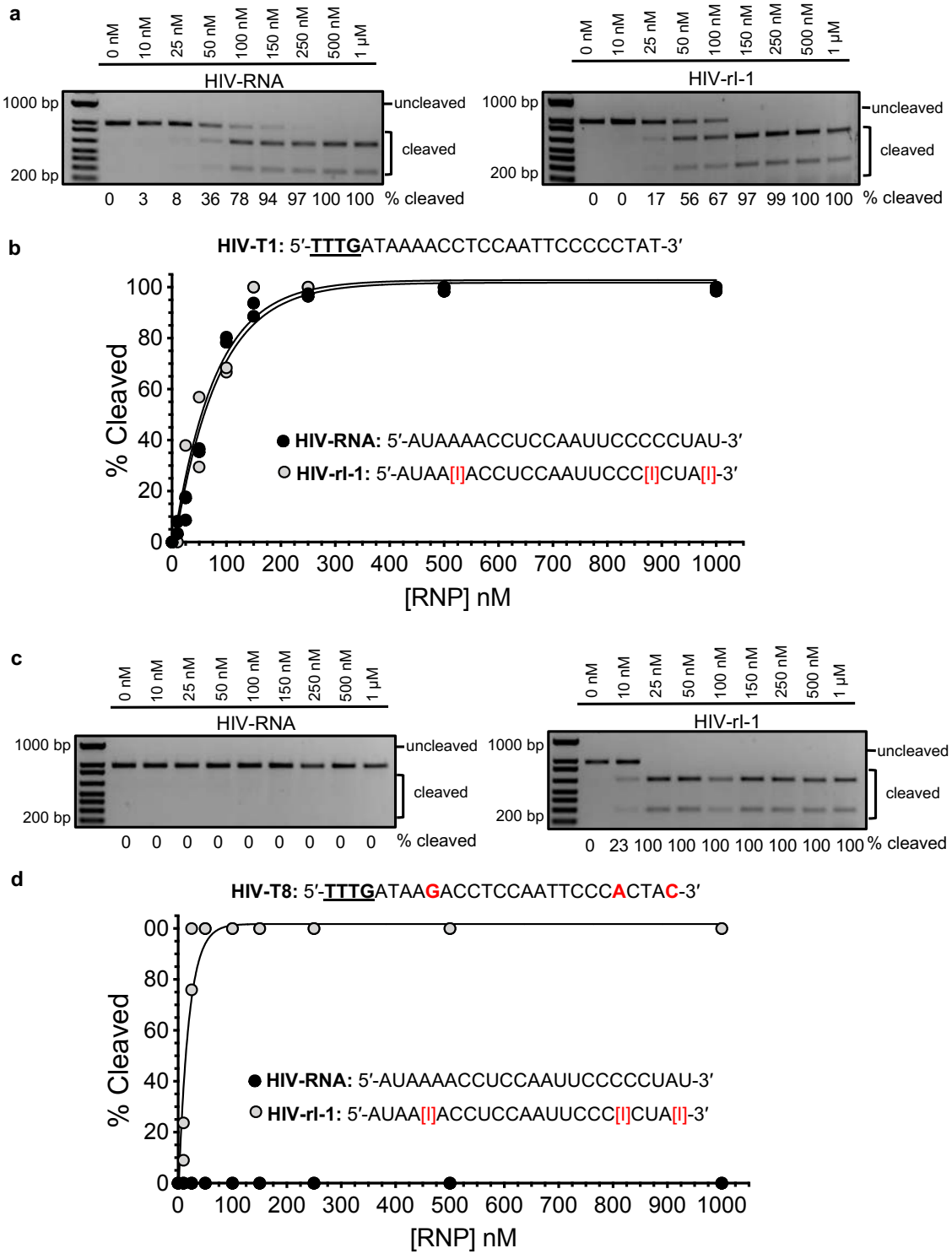

a
0 nM
10 nM
25 nM
50 nM
100 nM
150 nM
250 nM
500 nM
1 µM
HIV-RNA
1000 bp
200 bp
uncleaved
cleaved
0 3 8 36 78 94 97 100 100 % cleaved
HIV-rI-1
0 0 17 56 67 97 99 100 100 % cleaved
b
HIV-T1: 5′-TTTGATAAAACCTCCAATTCCCCCTAT-3′
% Cleaved
HIV-RNA: 5′-AUAAAACCUCCAAUUCCCCCUAU-3′
HIV-rI-1: 5′-AUAA[I]ACCUCCAAUUCCC[I]CUA[I]-3′
[RNP] nM
c
0 0 0 0 0 0 0 0 0 % cleaved
0 23 100 100 100 100 100 100 100 % cleaved
d
HIV-T8: 5′-TTTGATAAGACCTCCAATTCCCACTACC-3′

Figure 4.26. Titration of Cas12a RNP containing HIV-RNA or HIV-rI-1 against select target sequences in vitro

(**a**) Representative gels showing a titration of Cas12a RNP containing HIV-RNA or HIV-rI-1 crRNAs against the HIV-T1 target sequence. The bottom two bands in the gel represent the cleaved DNA substrate while the top band corresponds to the undigested substrate. (**b**) Plot quantifying the results of experiments in (**a**). DNA target sequences with SNPs are indicated with red lettering. [I] in the crRNA sequence indicates a ribose inosine position. Quantification of cleavage percentages was performed using ImageJ; Individual data points shown (n = 2 independent experiments). (**c**) Representative gels showing a titration of Cas12a RNP containing HIV-RNA or HIV-rI-1 crRNAs against the HIV-T8 target sequence. The bottom two bands in the gel represent the cleaved DNA substrate while the top band corresponds to the undigested substrate. (**d**) Plot quantifying the results of experiments in (**c**). DNA target sequences with SNPs are indicated with red lettering. Red [I] in the crRNA sequence indicates a ribose inosine position. Quantification of cleavage percentages was performed using ImageJ; Individual data points shown (n = 2 independent experiments). Cleavage experiments were performed in duplicate with similar results.

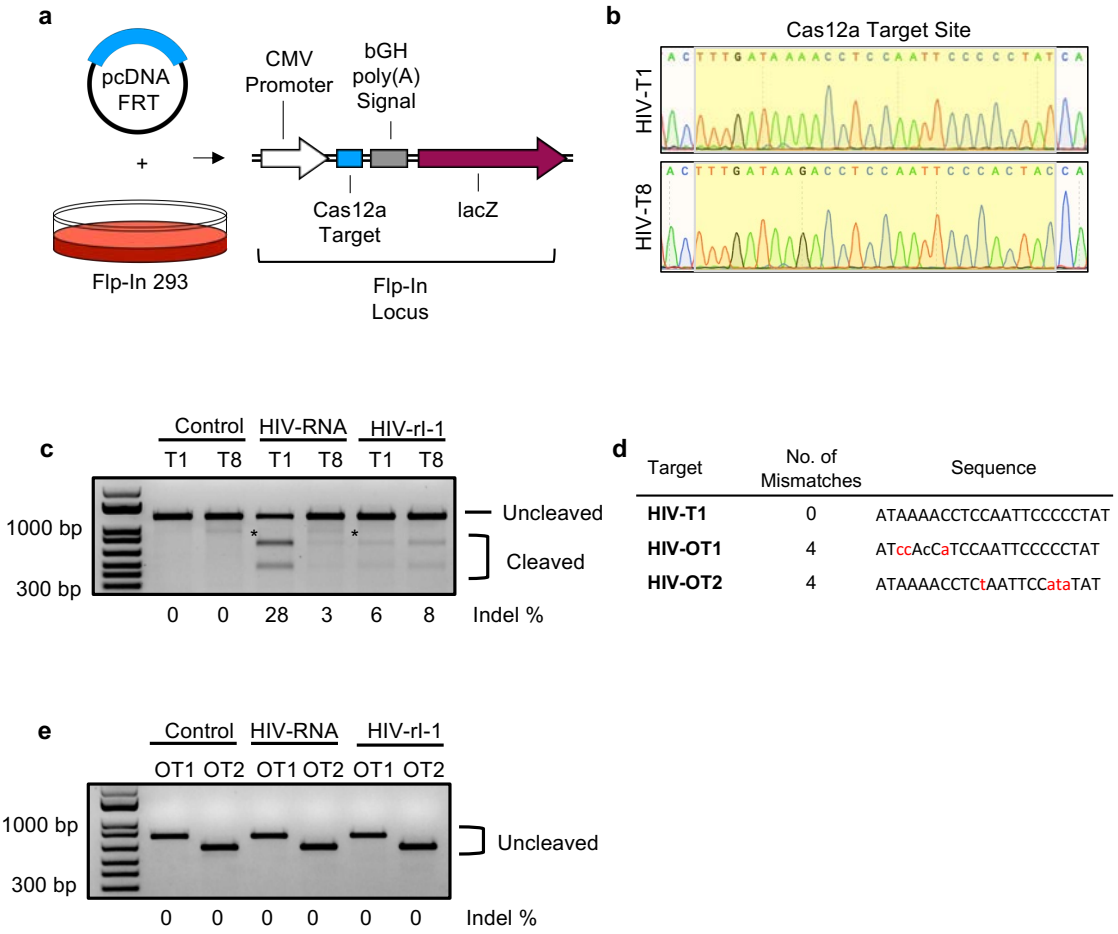


Figure 4.27. Cleavage of HIV target sequences by Cas12a with unmodified and inosine modified crRNAs in cells

(**a**) Diagram illustrating the generation of Flp-In 293 cells containing single-copy genomic HIV DNA target sequences. (**b**) Sanger sequencing trace confirming DNA target integration. (**c**) Gel representing cellular Cas12a cleavage efficiencies of the HIV-T1 or HIV-T8 sequences using either the unmodified or ribose inosine-modified HIV crRNAs, as determined by T7 endonuclease I digestion. Control transfections were performed without crRNAs. Indel percentages were determined using densitometry (ImageJ) and are shown below each lane. (**d**) Potential genomic off-target sequences (OT1: Chr17:40142160, OT2: Chr1:213961893) corresponding to the HIV target sequence (HIV-T1) predicted using Cas-OFFinder ((http://www.rgenome.net/cas-offinder/)1. (**e**) Gel representing cellular Cas12a cleavage efficiencies of the OT1 or OT2 sequences using either the unmodified or ribose inosine-modified *HIV* crRNAs, as determined by T7 endonuclease I digestion. Control transfections were performed without crRNAs. Indel percentages were determined using densitometry (ImageJ) and are shown below each lane. Cellular cleavage experiments were performed in duplicate with similar results.

4.4 Discussion

Cas9 tolerates a number of chemical alterations within the guide segment of its crRNA, including sugar modifications such as 2′O methylation[188], 2′O-4′C linkages[194], and 2′ deoxyribose modifications[202], as well as phosphate backbone modifications such as phosphorothioate[188] and phosphonoacetate[195]. In fact, 2′ deoxyribose modification of the crRNA was reported to be tolerated in all locations of the spacer segment except position 16[202]. This flexibility is quite remarkable given that the enzyme forms four direct amino acid contacts with sugar moieties within the spacer sequence, and over ten interactions with the phosphate backbone in this region[46]. Crystal structures have elucidated only one major Cas9 amino acid (Tyr1013)-base interaction, which occurs at position 1 in the crRNA targeting region[46]. Our data demonstrate that several classes of chemically unrelated non-canonical bases may be tolerated within Cas9 targeting sequences, even in positions where direct contact between Cas9 and the base or sugar backbone of the crRNA are being made[46]. For example, crRNAs containing modified bases at position 1 (a Cas9-base interaction)[46] (**Figures 4.18**-**4.21**), and position 19 (a Cas9-sugar interaction)[46] (**Figures 4.2** and **4.10**), were still able to direct Cas9 cleavage of target DNA, albeit with reduced activity compared the unmodified RNA in some cases.

We show that incorporation of multiple types of universal bases into Cas9 crRNAs abolishes specificity at the site of incorporation but otherwise preserves specificity (**Figures 4.10**, **4.15** and **4.19**). These findings are in agreement with results from PCR-based studies in which incorporation of inosine or synthetic universal bases into detection or amplification primers confers partial degeneracy[267, 283]. Similarly, our data indicating that the presence of universal bases in crRNAs may lower *in vitro* Cas9 activity via reduced DNA binding (**Figure 4.2** and **Figures 4.5**-**4.8**), are

reminiscent of the decreases in PCR efficiency observed using inosine-containing primers[284]. This decreased efficiency results from a decrease in duplex thermodynamic stability resulting from A-U to I-U and G-U to I-C transitions[284]. Also in agreement with our findings, the magnitude of this effect appears to be context dependent and is influenced by the nearest-neighbor 5′ and 3′ bases, following a decreasing stability trend of G-C > C-G > A-T > T-A[262]. Interestingly, we observed that Cas9 activity was lower with 2′ deoxyribose inosine substitutions compared to 2′ OMe ribose inosine substitutions (**Figure 4.2c**, **d**). Since previous reports have suggested that DNA/DNA hybrids are in fact more thermodynamically stable than 2′-O-methyl RNA/DNA duplexes[266], this could be due to a conformational/steric effect[194]. Other universal bases examined in this work such as deoxyribose 5′-nitroindole have been reported to be more destabilizing than inosine, due to the inability to form hydrogen bonds[269]. However, this effect is also context dependent, as short contiguous stretches of 5′-nitroindole are more tolerated than contiguous stretches of inosine in PCR amplification reactions[267]. Collectively, our data suggest that universal bases behave similarly in the context of crRNAs as they do in other types of nucleic acid probes. Thus, it is likely that many of the established rules governing their effective placement in an oligonucleotide are transferrable to this new application. However, future studies will be needed to test this assertion, and to determine how broadly these base modification schemes can be applied across different crRNAs.

The 1000 genomes project identified over 85 million SNPs, 3.6 million short indels, and 60,000 structural variants, underscoring the vast myriad of human genetic diversity[285]. While current CRISPR/Cas technology is not adequately equipped to deal with the challenge, this study puts forth a possible solution. We show that the incorporation of universal bases into individual crRNAs

can enable simultaneous targeting of a clinically relevant polymorphic gene in vitro and in cells (**Figures 4.2**, **4.10**, **4.22** and **4.23**). Given the extensive costs associated with personalized-medicine based clinical trials, CRISPR guide RNAs with partial degeneracy could be designed to circumvent natural genetic variation and enable all individuals in a patient population to be treated using a single, heavily validated therapeutic. Similarly, this technology could be applied in the lab to the development of guide RNAs capable of directing cleavage of a gene sequence across multiple different species for which evolutionary divergence may have occurred. However, our data suggest that certain limitations will first need to be overcome to realize this full potential. We noted in several instances that universal base-modified crRNAs yielded reduced or absent Cas9 activity in cells, despite showing strong activity in vitro (**Figure 4.23**). Interestingly, we found that these same crRNAs induced a delay in Cas9 cleavage kinetics *in vitro*, and that there was a direct correlation between slower cleavage kinetics and lower cellular modification rates (**Figure 4.24**). These findings are not unprecedented[194]. Previous work has shown that LNA and BNA substitutions in crRNAs also induce a delay in Cas9 enzyme kinetics that is associated with lower cleavage activity in cells[194]. It has been proposed that delayed kinetics could increase the probability that Cas9 is ejected from DNA by cellular factors prior to cutting, thereby reducing modification rates[194]. Prospective studies could address this issue by identifying ways to increase the residence time of Cas9 on DNA in cells, or by generating enzymes with kinetic properties more tailored to this application through protein evolution or engineering.

Several viruses have mutation rates that are up to 1 million times higher than their hosts[286]. This statistic highlights the obstacle that genetic variation presents for detecting and targeting pathogens using CRISPR systems. One study aiming to treat HIV-1 infection using CRISPR/Cas9

documented effective viral escape through evolved mutations[281]. Here we show that universal bases can be incorporated into Cas12a guide RNAs to enable detection of evolved viral gene sequences using the DETECTR platform (**Figure 4.25**). Unlike Cas9, where rI-modified crRNAs resulted in lower *in vitro* activity in certain cases (**Figure 4.2**), we found that Cas12a activity remained robust when coupled with universal base-modified crRNAs (**Figure 4.25**), making this an ideal application of the technology. We speculate that this could be due to the fact that our Cas12a experiments measured combined *cis* and *trans* enzyme activity. It is conceivable that small to moderate decreases in *cis* cleavage activity could be fully masked by Cas12a collateral activity. Overall, we envision that this technology could be used to help reduce the false negative detection rate of the DETECTR system by imparting the platform with the flexibility to take into account pathogen evolution, either documented or predicted.

We demonstrate the potential to use universal base-modified crRNAs with Cas12a to target polymorphic viral sequences in cells (**Figure 4.27**). While the rI-modified crRNA we used resulted in the expected degenerate sequence cutting in cells, it did show reduced activity (**Figure 4.27**). This could be due to the apparent lack of Cas12a *trans* activity in cells[287], or also the result of slower enzyme kinetics, as described above for Cas9.

In addition to targeting known polymorphic sequences, this technology could be used to target pathogens for which only incomplete sequence information is known. For example, future studies could assess if contiguous stretches of universal bases can be incorporated into crRNAs to reduce the requisite length of the spacer sequence. Theoretically, this would enable targeting of shorter sequences in emergent pathogens for which all 20 bp of the sequence may be unavailable.

This work details the first demonstration that incorporation of non-canonical universal bases into Cas9/Cas12a guide RNAs can be tolerated and impart selectively degenerate specificity. We demonstrate the applicability of this technology to targeting a series of polymorphic gene variants in vitro and in cells using a single guide RNA. Furthermore, we delineate how this technology can be applied to diagnostics to circumvent false-negative results caused by pathogen evolution. By relaxing the current restrictions of guide RNA targeting, we anticipate that this study will expand the operative capabilities of Cas9, Cas12a/Cpf1, and potentially other CRISPR systems.

4.5 Methods

4.5.1 Chemical reagents and design and synthesis of crRNAs

Unless otherwise noted, all chemical reagents were purchased from Sigma-Aldrich. DNA oligonucleotides and tracrRNA were purchased from Integrated DNA Technologies (IDT). crRNAs were rationally designed based on clinical polymorphism data for the *HLA-B*[288], *ABO*[288], and HIV[280] gene sets. *HLA-C* crRNAs were designed as described in the manuscript. Two of the four SNPs chosen for the *ABO* target site are found in the most common *ABO* alleles and are linked to changes in blood type[272]. The polymorphisms seen in the HIV gene set are linked to the formation of drug-resistant mutations in a domain of the viral protease[279]. Cas9 crRNAs were designed based on the presence of a 3′-NGG PAM directly adjacent to a 20 bp target site for the *HLA-B*, *HLA-C* and *ABO* genes. Cas12a crRNAs were designed based on a 5′-TTTN sequence directly adjacent to the 23 bp target site for the HIV protease gene. Sequences for these crRNAs can be found in **Table 4.1**. Chemical synthesis of the crRNAs was performed by Bio-Synthesis Inc. and GeneLink Inc.

4.5.2 Preparation of DNA targets

Forward and reverse ssDNA target inserts were designed for Cas9 target sites. The oligos used to make the DNA targets are listed in **Table 4.2**. The forward and reverse ssDNA sequences were annealed by heating to 95 °C for 5 mins, then cooling to 25 °C over 1 h. Next, pUC19 plasmid (Invitrogen) and the annealed dsDNA target inserts were double-digested with *HindIII* and *XbaI*(NEB). These were ligated and then transformed into DH5α *E. coli*. Proper insertion was confirmed by performing Sanger sequencing. DNA targets for in vitro experiments concerning *HLA-C* were prepared through PCR amplification of genomic DNA (gDNA) from 293 T or HeLa cells using primers listed in **Table 4.2**.

4.5.3 Expression and purification of *S. pyogenes* Cas9

Recombinant Cas9 and dCas9 were purified as previously described[52]. Briefly, *E. coli* Rosetta (DE3) cells were transformed with a plasmid encoding either *S. pyogenes Cas9* or catalytically-dead *S. pyogenes Cas9* (dCas9) fused to an N-terminal 6xHis-tag, MBP, and TEV site (Addgene #39312 and #39318, respectively). 25 mL of LB broth containing 25 μg mL^{-1} of kanamycin was inoculated and grown overnight (~16 hrs) at 37 °C. These cells were diluted 1:100 in the same growth media and grown at 37 °C until an OD_{600} of 0.8 before moving to 18 °C for 30 mins. Protein production was induced by the addition of isopropyl-ß-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After induction for 16 h, the cells were harvested by centrifugation for 15 min at 2700 × *g* and resuspended in 15 ml/L culture lysis buffer (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 5 mM imidazole, pH 8.0) supplemented with lysozyme and 0.1 M PMSF. This was incubated on ice for 30 mins before being further lysed by sonication (30 sec pulse-on and 60 secs

pulse-off for 7.5 min at 60% amplitude) and centrifuged at 30,000 × *g* for 1 h to obtain cleared lysate. The lysate was applied to a 1 mL HisTrap FF Crude column (GE Healthcare) attached to an AKTA Start System (GE Healthcare), washed (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 10 mM imidazole, pH 8.0), and eluted with a single concentration of imidazole (20 mM Tris-Cl, pH 8.0, 250 mM NaCl, 250 mM imidazole, pH 8.0). Fractions containing Cas9 were pooled, TEV protease was added, and this was dialyzed into ion-exchange buffer overnight (20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% (v/v) glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA). After dialysis, the sample was centrifuged to remove cleaved MBP. The supernatant was loaded onto a 1 mL HiTrap SP FF column (GE Healthcare), washed (20 mM HEPES-KOH, pH 7.5, 100 mM NaCl), and eluted with a 0–50% gradient of NaCl (20 mM HEPES-KOH, pH 7.5, 1 M NaCl). Fractions containing purified Cas9 were concentrated using a 50 kDa centrifugal filter (Pall). During concentration, the buffer was exchanged into storage buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 1 mM DTT). Concentrated protein was aliquoted and stored at −80 °C.

4.5.4 *In vitro* cleavage assays (Cas9)

In vitro DNA cleavage assays were performed as previously described[194]. Briefly, plasmid templates containing DNA targets were amplified with pUC19F/R primers listed in **Table 4.2**. gRNAs were created by mixing equimolar amounts of tracrRNA (IDT) and crRNA (GeneLink) in Nuclease Free Duplex Buffer (IDT), and then heating to 95 °C for 5 min before cooling to 25 °C over 1 h. Sequences for crRNAs and tracrRNA are listed in **Table 4.1**. Each reaction consisted of the amplified 5 nM DNA target with 40 nM Cas9 protein and 80 nM gRNA, unless otherwise stated. Initially, Cas9 and gRNA were incubated in 1× NEB 3.1 buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9) at 25 °C for 10 mins. Subsequently, the DNA

template was added and the reaction was incubated at 37 °C for 3 h (or in the time indicated in the figure legend for kinetic experiments). Reactions were stopped by purifying the DNA with a MinElute PCR Purification Kit (Qiagen). Cleavage products were run on a 1% agarose gel and imaged with an Amersham Imager 600 (GE Healthcare). Densitometry was performed using Image J.

4.5.5 Library construction for high-throughput specificity profiling

Pre-selection libraries were generated as previously described[194]. Briefly, 10 pmol of each partially randomized oligo (IDT, sequences are listed in **Table 4.1**) was circularized with CircLigase II ssDNA Ligase Kit (Epicenter). 5 pmol of the circularized ssDNA was used as a template for the Illustra TempliPhi Amplification Kit (GE Healthcare) according to the manufacturer's protocol. The resulting amplified libraries were quantified with a Qubit 2.0 Fluorometer (Invitrogen).

4.5.6 *In vitro* high-throughput specificity profiling

A specificity profile of the modified crRNAs was created as previously described[194]. Briefly, 200 nM of the pre-selection library was incubated with 1000 nM gRNA and 1000 nM Cas9 in NEB Buffer 3.1 for 1 h at 37 °C to create the post-selection library. In addition, 200 nM of the library was incubated with 2U of BspMI using the same reaction conditions as above, to create the final pre-selection library. Both digestion reactions were purified using a QiaQuick PCR Purification Kit (Qiagen) and ligated to 10 pmol of barcoded adaptor S50X-F/R (post-selection) or lib_adapter1 with ABO/HLA_lib_adapter2 (pre-selection) using 1000U of T4 DNA Ligase (NEB) for 16 hrs at room temperature. Ligation reactions were purified using the MinElute PCR Purification Kit (Qiagen) then amplified using primer PE2_short with barcoded primer HLA/ABO-N70X (post-

selection) or primer lib_PCR_F with barcoded ABO/HLA_PCR_R (pre-selection) using Q5 Hot Start High-Fidelity Master Mix (NEB). Products were gel extracted and purified using MinElute Gel Extraction Kit (Qiagen) and quantified with a Qubit 2.0 Fluorometer. Finished libraries were run on a HiSeq 2000 (Novogene), demultiplexed, and analyzed as previously described[52]. The sequences used for this protocol are listed in **Table 4.2**.

4.5.7 Electrophoretic mobility shift assay (EMSA)

EMSA assays were performed as previously described[194], with minor modifications. Briefly, the dsDNA target was created by annealing the ssDNA target and Cy-5 labelled non-target strands in a 1.5:1 ratio. dsDNA substrate, gRNA, and dCas9 were diluted in the binding buffer to working concentrations (20 mM HEPES, pH 7.5, 50 mM KCl, 2 mM $MgCl_2$, 0.01% Triton X-100, 0.1 mg mL^{-1} bovine serum albumin, 10% glycerol) and 50 µg/mL of heparin was added. 500 nM dCas9 and 750 nM gRNA were incubated at room temperature for 10 mins. For titration reactions, 0, 10, 25, 50, 100, 250, 500 nM & 1 µM of dCas9 and 1.5× gRNA was used. 10 nM Cy-5 labelled annealed DNA substrate was then added and incubated for 1 h at 37 °C. The reaction was run on a 10% TBE-2mM $MgCl_2$ polyacrylamide gel at 4 °C and imaged with an Amersham Imager 600 (GE Healthcare). Densitometry was used to measure percent binding (Image J).

4.5.8 Determination of crRNA-DNA heteroduplex melting temperature

Equimolar amounts of crRNA and complementary ssDNA were combined in Duplex Buffer (30 mM HEPES, pH 7.5, 100 mM Potassium Acetate) (IDT) to a final concentration of 2 µM. 100× SYBR Green I was then added to yield a final concentration of 10×. The solution was added to a CFX96 Real-Time System (BioRad). The following program was run to anneal the RNA/DNA

heteroduplex: 5 min at 95 °C followed by cooling to 25 °C at 0.1 °C s^{-1}. To measure the melting temperature, the heteroduplex was heated to 45 °C and then subsequently heated at a rate of 0.1 °C s^{-1} to 95 °C. The SYBR Green I fluorescent signal was used to generate a melt curve from which a T_m value was determined.

4.5.9 Cell-based RFP/GFP reporter assay

Target sites were cloned into a pRGS backbone (PNA Bio Inc.) containing an RFP reporter and two out-of-frame GFP reporters, as previously described[289]. gRNA was annealed as described above. HeLa-Cas9 cells (previously authenticated and shown to be free of mycoplasma)[194] were cultured in high-glucose DMEM media with pyruvate (Gibco) supplemented with 10% FBS/1× pen-strep/1× glutamine (Gibco) and 5 µg mL^{-1} Blasticidin S HCl (Gibco) at 37 °C in 5% CO_2. Transfection of the HeLa-Cas9 cells was performed using DharmaFECT Duo (Dharmacon), according to manufacturer instructions for the CRISPR system. The degree of target sequence cleavage was calculated based on the %GFP + /%RFP + cells using an Attune NxT Flow Cytometer (Invitrogen).

4.5.10 Expression and purification of humanized Lachnospiraceae bacterium Cas12a/Cpf1

Humanized Cpf1 was purified as previously described[290]. Briefly, *E. coli* Rosetta (DE3) pLyseS (EMD Millipore) cells were transformed with a plasmid encoding humanized *Lachnospiraceae bacterium* Cpf1 fused to an N-terminal 6xHis-tag, MBP, TEV site, and C-terminal NLS and HA tag (Addgene # 90096). 25 mL of Terrific broth containing 100 µg mL^{-1} of carbenicillin was inoculated and grown overnight (~16 hrs) at 37 °C. These cells were diluted 1:100 in the same growth media and grown at 37 °C until OD_{600} of 0.2. This was moved to 21 °C and grown until an

OD_{600} of 0.6 before induction with IPTG to a final concentration of 0.5 mM for 14–18 h. After induction, the cells were harvested by centrifugation for 15 min at 2700 × *g* and resuspended in 50 mL/L culture of lysis buffer (50 mM HEPES pH 7, 2 M NaCl, 5 mM MgCl2, 20 mM imidazole, pH 8.0), supplemented with lysozyme and 0.1 M PMSF. Cell lysis and protein purification were performed as described above. lbCpf1 was stored in Cpf1 storage buffer (50 mM Tris-HCl pH7.5, 2 mM DTT, 5% glycerol, 500 mM NaCl).

4.5.11 *In vitro* cleavage assays (Cas12a)

In vitro DNA cleavage reactions for Cas12a were performed as described above with slight modifications. Each reaction consisted of amplified 10 nM DNA target with 100 nM Cas12a protein and 125 nM gRNA. Reactions were incubated at 37 °C for 30 mins (or the time indicated in the figure legends for kinetic experiments). For experiments involving titrations of Cas12a RNP, each reaction consisted of amplified 10 nM DNA target with 0 nM, 10 nM, 25 nM, 50 nM, 100 nM, 150 nM, 250 nM, 500 nM, or 1 μM of Cas12a protein and 1.25× gRNA. Reactions were incubated at 37 °C for 30 mins. Sequences for the crRNAs used in these experiments are listed in **Table 4.1**.

4.5.12 DETECTR assay

DETECTR assays were performed as previously described[70], with minor modifications. Briefly, target constructs were created with a pUC19 backbone as described above. Recombinase Polymerase Amplification (RPA) reactions were performed using the target plasmid constructs as the template and pUC19 RPA F/R primers. This reaction was incubated at 37 °C for 10 mins. 250 nM LbCas12a, 312.5 nM crRNA, and 250 nM ssDNA-FQ reporter were incubated at 25 °C for 10 mins and added directly to the reaction. Subsequently, reactions were incubated at 37 °C in

a fluorescent plate reader (Spectramax i3, Molecular Devices) for 2 hrs with measurements taken every 2 mins (λ_{ex}: 535 nm; λ_{em}: 595 nm). The sequences used for this experiment are listed in **Table 4.2**.

4.5.13 Generation of Flp-In 293-Cas12a-HIV-T1/T8 cells

Parental Flp-In 293 cells were obtained from ThermoFisher (catalog number R75007) and were cultured in high-glucose DMEM media containing pyruvate (Gibco), supplemented with 10% FBS, 1X Penicillin-Streptomycin (Gibco) and 100 µg/mL Zeocin (Invitrogen). For Flp-In 293-Cas12a-HIV-T1/T8 cells, Zeocin was substituted with 100 µg/mL Hygromycin B (Invitrogen) and 300 µg/mL Geneticin (Gibco). All cells were cultured in a 5% CO_2 atmosphere. Stable integration of HIV-T1 and -T8 Cas12a target sites into parental Flp-In 293 cells was performed as described by the manufacturer. Briefly, oligonucleotides containing the Cas12a target site of interest (listed in **Table 4.2**) were annealed and ligated into BamHI / XhoI double-digested pcDNA5 plasmid containing an FRT site (Addgene #127108). Sequence-verified pcDNA5-HIV-T1/T8 plasmids were then co-transfected with pCSFLPe (Addgene #31130) at a ratio of 1:9 (w/w) into Flp-In 293 cells using Effectene Transfection Reagent (Qiagen) according to the manufacturer's instructions. 24 hrs after transfection, media was replaced with fresh DMEM lacking Zeocin. The next day cells were split at a confluence of 30% into media containing 100 µg/mL Hygromycin B. Following 2 weeks of selection, cells were singly sorted on a BD FACS Aria III instrument by the Flow Cytometry Core at the University of Alberta into the wells of a 96-well plate for monoclonal expansion. Viral particles for Cas12a expression were generated in 293 T cells transfected with plenti-Lb-Cas12a-2xNLS (Addgene #155046), psPAX2 (Addgene #12260) and pMD2.G (Addgene #12259) using Effectene Transfection Reagent (Qiagen) according to the

manufacturer's instructions. Supernatant containing the viral particles was harvested 48 hrs after transfection and filter sterilized before being used to transduce Flp-In 293-HIV-T1 and Flp-In 293-HIV-T8 cells. 48 h after infection, media was replaced with DMEM containing 300 µg/mL Geneticin.

4.5.14 Lipid transfection of crRNAs into stable cell lines

Cells stably expressing Cas9 or Cas12a were transfected with crRNAs to a final concentration of 60 nM using Lipofectamine RNAiMAX (Invitrogen) according to manufacturer's instructions.

4.5.15 Cellular cleavage assays

72 hrs after transfection, gDNA from transfected cells was isolated using the DNeasy Kit (Qiagen) and quantified using a NanoPhotometer NP80 (Implen) spectrophotometer. Target-specific primers (listed in Table 4.2) were used to PCR amplify the desired site, with 100 ng of gDNA used as template. PCR products were purified using QIAquick PCR Purification Kit (Qiagen). 200 ng of product was subject to T7 endonuclease I (T7E1) digestion as described by the manufacturer (NEB). Cleavage assays were resolved on a 2% TAE agarose gel.

4.5.16 Calculations, statistics & reproducibility

Indel percentages were calculated using the formula $\text{indel}\ (\%) = 100 \times (1 - (1 - \text{fraction}_{\text{cut}})^{0.5})$. Replicate numbers and measures of variance are included in the Figure legends. Experiments were not randomized, but the high-throughput specificity profiling assays employed large libraries of partially randomized sequences. No statistical method was used to predetermine sample size, and

no data were excluded from the analyses. The investigators were not blinded to allocation during experiments and outcome assessment.

4.5.17 Data availability

All high-throughput sequencing data files associated with this paper have been deposited in the NCBI SRA database and are available under accession number: PRJNA669024. Databases that were used in the selection of cellular target sequences are publicly accessible: Ensembl (https://uswest.ensembl.org/index.html) and HEK293 Genome (http://www.hek293genome.org/v1/index.php).

4.5.18 Code availability

Python scripts used for data analysis are available upon request.

Table 4.1. Sequences of crRNAs and tracrRNA

Name	Sequence (5′→3′)
HLA-RNA	rCrArCrArCrArGrArUrCrUrArCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-1	[rI]rArCrArCrArGrArUrCrUrArCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-2	rCrArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-3	rCrArCrArCrArGrArUrCrUrArCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-4	rCrArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-5	[rI]rArCrArCrArGrArUrCrUrArCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-6	[rI]rArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrCrCrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-rI-7	[rI]rArCrArCrArGrArUrCrU[rI]rCrArArGrGrCrC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU

ABO-RNA	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-1	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dI-1	rCrArUrGrGrArG[dI]rUrCrCrGrCrGrArCrCrArC[dI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-mI-1	rCrArUrGrGrArG[mI]rUrCrCrGrCrGrArCrCrArC[mI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dN-1	rCrArUrGrGrArG[5NitInd]rUrCrCrGrCrGrArCrCrArC[5NitInd]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dK-1	rCrArUrGrGrArG[dK]rUrCrCrGrCrGrArCrCrArC[dK]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dP-1	rCrArUrGrGrArG[dP]rUrCrCrGrCrGrArCrCrArC[dP]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-2	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dI-2	rCrArUrGrGrArG[dI]rUrCrCrGrCrGrArCrCrA[dI][dI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-mI-2	rCrArUrGrGrArG[mI]rUrCrCrGrCrGrArCrCrA[mI][mI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dN-2	rCrArUrGrGrArG[5NitInd]rUrCrCrGrCrGrArCrCrA[5NitInd][5NitInd]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dK-2	rCrArUrGrGrArG[dK]rUrCrCrGrCrGrArCrCrA[dK][dK]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-dP-2	rCrArUrGrGrArG[dP]rUrCrCrGrCrGrArCrCrA[dP][dP]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-3	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-4	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-5	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-6	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-7	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrArCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-8	rCrArUrGrGrArG[rI]rUrCrCrGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-9	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-10	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-11	rCrArUrGrGrArGrUrUrCrCrGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU

ABO-rI-12	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrA[rI]rGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-13	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrArC[rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-14	rCrArUrGrGrArGrUrUrC[rI]rGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-rI-15	rCrArUrGrGrArG[rI]rUrC[rI]rGrCrGrArCrCrA[rI][rI]rGrUrUrUrUrArGrArGrCrUrArUrGrCrU
ABO-Negative	rGrArGrUrCrCrGrArGrCrArGrArArGrArArGrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-C-RNA	rArArCrArUrGrGrGrGrArArArGrCrArGrUrUrGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-C-rI	rArArCrA[rI]rGrGrGrGrArArArGrCrArG[rI]rUrGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-C-dI	rArArCrA[dI]rGrGrGrGrArArArGrCrArG[dI]rUrGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HLA-C-dP	rArArCrA[dP]rGrGrGrGrArArArGrCrArG[dP]rUrGrUrGrUrUrUrUrArGrArGrCrUrArUrGrCrU
HIV-RNA	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArArArArCrCrUrCrCrArArUrUrCrCrCrCrUrArU
HIV-rI-1	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArUrArA[rI]rArCrCrUrCrCrArArUrUrCrCrC[rI]rCrUrA[rI]
tracrRNA	rArGrCrArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrUrU

[rI] = Ribose inosine, [dI] = Deoxyribose inosine, [mI] = 2′O methyl ribose inosine, [dN] = Deoxyribose 5′- nitroindole, [dK] = Deoxyribose K, [dP] = Deoxyribose P, and /Cy5/ = Cyanine 5.

Table 4.2. Sequences of oligonucleotides

Name	**Sequences (5′→3′)**
Oligos used to create in vitro cleavage assay DNA target constructs.	
ABO-T1-F	GCCGAAGCTTCTCCACGTGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T1-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCACGTGGAGAAGCTTCGGC

ABO-T2-F	GCCGAAGCTTCTCCACGTGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T2-R	GGCCTCTAGAAGCATGGAGATCCGCGACCACGTGGAGAAGCTTCGGC
ABO-T3-F	GCCGAAGCTTCTCCACATGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T3-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCATGTGGAGAAGCTTCGGC
ABO-T4-F	GCCGAAGCTTCTCCATGTGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T4-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCACATGGAGAAGCTTCGGC
ABO-T5-F	GCCGAAGCTTCTCCATGTGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T5-R	GGCCTCTAGAAGCATGGAGATCCGCGACCACATGGAGAAGCTTCGGC
ABO-T6-F	GCCGAAGCTTCTCCATATGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T6-R	GGCCTCTAGAAGCATGGAGATCCGCGACCATATGGAGAAGCTTCGGC
ABO-T7-F	GCCGAAGCTTCTCCACATGGTCGCGGATCTCCATGCTTCTAGAGGCC
ABO-T7-R	GGCCTCTAGAAGCATGGAGATCCGCGACCATGTGGAGAAGCTTCGGC
ABO-T8-F	GCCGAAGCTTCTCCATATGGTCGCGGAACTCCATGCTTCTAGAGGCC
ABO-T8-R	GGCCTCTAGAAGCATGGAGTTCCGCGACCATATGGAGAAGCTTCGGC
ABO-T9-F	GCCGAAGCTTCTCCACGTGGTCGCAGAACTCCATGCTTCTAGAGGCC
ABO-T9-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCACGTGGAGAAGCTTCGGC
ABO-T10-F	GCCGAAGCTTCTCCACGTGGTCGCAGATCTCCATGCTTCTAGAGGCC
ABO-T10-R	GGCCTCTAGAAGCATGGAGATCTGCGACCACGTGGAGAAGCTTCGGC
ABO-T11-F	GCCGAAGCTTCTCCACATGGTCGCAGAACTCCATGCTTCTAGAGGCC
ABO-T11-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCATGTGGAGAAGCTTCGGC
ABO-T12-F	GCCGAAGCTTCTCCATGTGGTCGCAGAACTCCATGCTTCTAGAGGCC
ABO-T12-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCACATGGAGAAGCTTCGGC
ABO-T13-F	GCCGAAGCTTCTCCACATGGTCGCAGATCTCCATGCTTCTAGAGGCC

ABO-T13-R	GGCCTCTAGAAGCATGGAGATCTGCGACCATGTGGAGAAGCT TCGGC
ABO-T14-F	GCCGAAGCTTCTCCATGTGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T14-R	GGCCTCTAGAAGCATGGAGATCTGCGACCACATGGAGAAGCT TCGGC
ABO-T15-F	GCCGAAGCTTCTCCATATGGTCGCAGAACTCCATGCTTCTAGA GGCC
ABO-T15-R	GGCCTCTAGAAGCATGGAGTTCTGCGACCATATGGAGAAGCT TCGGC
ABO-T16-F	GCCGAAGCTTCTCCATATGGTCGCAGATCTCCATGCTTCTAGA GGCC
ABO-T16-R	GGCCTCTAGAAGCATGGAGATCTGCGACCATATGGAGAAGCT TCGGC
ABO-T5 (G-G)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T5 (G-G)-R	GGC CTC TAG AAG CAT GGA GGT CCG CGA CCA CGT GGA GAA GCT TCG GC
ABO-T5 (G-C)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T5 (G-C)-R	GGC CTC TAG AAG CAT GGA GCT CCG CGA CCA CGT GGA GAA GCT TCG GC
ABO-T5 (G-T)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T5 (G-T)-R	GGC CTC TAG AAG CAT GGA GTT CCG CGA CCA CGT GGA GAA GCT TCG GC
ABO-T5 (G-A)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T5 (G-A)-R	GGC CTC TAG AAG CAT GGA GAT CCG CGA CCA CGT GGA GAA GCT TCG GC
ABO-T5 (C-G)-F	GCC GAA GCT TCT CCA GGT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T5 (C-G)-R	GGC CTC TAG AAG CAT GGA GGT CCG CGA CCA CCT GGA GAA GCT TCG GC
ABO-T5 (C-C)-F	GCC GAA GCT TCT CCA GGT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T5 (C-C)-R	GGC CTC TAG AAG CAT GGA GCT CCG CGA CCA CCT GGA GAA GCT TCG GC
ABO-T5 (C-T)-F	GCC GAA GCT TCT CCA GGT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T5 (C-T)-R	GGC CTC TAG AAG CAT GGA GTT CCG CGA CCA CCT GGA GAA GCT TCG GC
ABO-T5 (C-A)-F	GCC GAA GCT TCT CCA GGT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC

ABO-T5 (C-A)-R	GGC CTC TAG AAG CAT GGA GAT CCG CGA CCA CCT GGA GAA GCT TCG GC
ABO-T5 (T-G)-F	GCC GAA GCT TCT CCA AGT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T5 (T-G)-R	GGC CTC TAG AAG CAT GGA GGT CCG CGA CCA CTT GGA GAA GCT TCG GC
ABO-T5 (T-C)-F	GCC GAA GCT TCT CCA AGT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T5 (T-C)-R	GGC CTC TAG AAG CAT GGA GCT CCG CGA CCA CTT GGA GAA GCT TCG GC
ABO-T5 (T-T)-F	GCC GAA GCT TCT CCA AGT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T5 (T-T)-R	GGC CTC TAG AAG CAT GGA GTT CCG CGA CCA CTT GGA GAA GCT TCG GC
ABO-T5 (T-A)-F	GCC GAA GCT TCT CCA AGT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T5 (T-A)-R	GGC CTC TAG AAG CAT GGA GAT CCG CGA CCA CTT GGA GAA GCT TCG GC
ABO-T5 (A-G)-F	GCC GAA GCT TCT CCA TGT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T5 (A-G)-R	GGC CTC TAG AAG CAT GGA GGT CCG CGA CCA CAT GGA GAA GCT TCG GC
ABO-T5 (A-C)-F	GCC GAA GCT TCT CCA TGT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T5 (A-C)-R	GGC CTC TAG AAG CAT GGA GCT CCG CGA CCA CAT GGA GAA GCT TCG GC
ABO-T5 (A-T)-F	GCC GAA GCT TCT CCA TGT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T5 (A-T)-R	GGC CTC TAG AAG CATGGAGTTCCGCGACCACATGGA GAA GCT TCG GC
ABO-T5 (A-A)-F	GCC GAA GCT TCT CCATGTGGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T5 (A-A)-R	GGC CTC TAG AAG CATGGAGATCCGCGACCACATGGA GAA GCT TCG GC
ABO-T7 (G-G)-F	GCC GAA GCT TCT CCA CCT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T7 (G-G)-R	GGC CTC TAG AAG CATGGAGGTCCGCGACCAGGTGGA GAA GCT TCG GC
ABO-T7 (G-C)-F	GCC GAA GCT TCT CCA CCT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T7 (G-C)-R	GGC CTC TAG AAG CATGGAGCTCCGCGACCAGGTGGA GAA GCT TCG GC
ABO-T7 (G-T)-F	GCC GAA GCT TCT CCA CCT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC

ABO-T7 (G-T)-R	GGC CTC TAG AAG CATGGAGTTCCGCGACCAGGTGGA GAA GCT TCG GC
ABO-T7 (G-A)-F	GCC GAA GCT TCT CCA CCT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T7 (G-A)-R	GGC CTC TAG AAG CATGGAGATCCGCGACCAGGTGGA GAA GCT TCG GC
ABO-T7 (C-G)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T7 (C-G)-R	GGC CTC TAG AAG CATGGAGGTCCGCGACCACGTGGA GAA GCT TCG GC
ABO-T7 (C-C)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T7 (C-C)-R	GGC CTC TAG AAG CATGGAGCTCCGCGACCACGTGGA GAA GCT TCG GC
ABO-T7 (C-T)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T7 (C-T)-R	GGC CTC TAG AAG CAT GGA GTT CCG CGA CCA CGT GGA GAA GCT TCG GC
ABO-T7 (C-A)-F	GCC GAA GCT TCT CCA CGT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T7 (C-A)-R	GGC CTC TAG AAG CATGGAGATCCGCGACCACGTGGA GAA GCT TCG GC
ABO-T7 (T-G)-F	GCC GAA GCT TCT CCA CAT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T7 (T-G)-R	GGC CTC TAG AAG CATGGAGGTCCGCGACCATGTGGA GAA GCT TCG GC
ABO-T7 (T-C)-F	GCC GAA GCT TCT CCA CAT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T7 (T-C)-R	GGC CTC TAG AAG CATGGAGCTCCGCGACCATGTGGA GAA GCT TCG GC
ABO-T7 (T-T)-F	GCC GAA GCT TCT CCA CAT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T7 (T-T)-R	GGC CTC TAG AAG CATGGAGTTCCGCGACCATGTGGA GAA GCT TCG GC
ABO-T7 (T-A)-F	GCC GAA GCT TCT CCA CAT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T7 (T-A)-R	GGC CTC TAG AAG CATGGAGATCCGCGACCATGTGGA GAA GCT TCG GC
ABO-T7 (A-G)-F	GCC GAA GCT TCT CCA CTT GGT CGC GGA CCT CCA TGC TTC TAG AGG CC
ABO-T7 (A-G)-R	GGC CTC TAG AAG CATGGAGGTCCGCGACCAAGTGGA GAA GCT TCG GC
ABO-T7 (A-C)-F	GCC GAA GCT TCT CCA CTT GGT CGC GGA GCT CCA TGC TTC TAG AGG CC
ABO-T7 (A-C)-R	GGC CTC TAG AAG CATGGAGCTCCGCGACCAAGTGGA GAA GCT TCG GC

ABO-T7 (A-T)-F	GCC GAA GCT TCT CCA CTT GGT CGC GGA ACT CCA TGC TTC TAG AGG CC
ABO-T7 (A-T)-R	GGC CTC TAG AAG CATGGAGTTCCGCGACCAAGTGGA GAA GCT TCG GC
ABO-T7 (A-A)-F	GCC GAA GCT TCT CCA CTT GGT CGC GGA TCT CCA TGC TTC TAG AGG CC
ABO-T7 (A-A)-R	GGC CTC TAG AAG CATGGAGATCCGCGACCAAGTGGA GAA GCT TCG GC
HIV-T1-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCCCTATC TTC TAG AGG CC
HIV-T1-R	GGC CTC TAG AAG ATAGGGGGAATTGGAGGTTTTATCAAAA GAA GCT TCG GC
HIV-T2-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCCCTATC TTC TAG AGG CC
HIV-T2-R	GGC CTC TAG AAG ATAGGGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
HIV-T3-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCACTATC TTC TAG AGG CC
HIV-T3-R	GGC CTC TAG AAG ATAGTGGGAATTGGAGGTTTTATCAAAA GAA GCT TCG GC
HIV-T4-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCCCTACC TTC TAG AGG CC
HIV-T4-R	GGC CTC TAG AAG GTAGGGGGAATTGGAGGTTTTATCAAA A GAA GCT TCG GC
HIV-T5-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCACTATC TTC TAG AGG CC
HIV-T5-R	GGC CTC TAG AAG ATAGTGGGAATTGGAGGTCTTATCAAAA GAA GCT TCG GC
HIV-T6-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCCCTACC TTC TAG AGG CC
HIV-T6-R	GGC CTC TAG AAG GTAGGGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
HIV-T7-F	GCC GAA GCT TCT TTTGATAAAACCTCCAATTCCCACTACC TTC TAG AGG CC
HIV-T7-R	GGC CTC TAG AAG GTAGTGGGAATTGGAGGTTTTATCAAAA GAA GCT TCG GC
HIV-T8-F	GCC GAA GCT TCT TTTGATAAGACCTCCAATTCCCACTACC TTC TAG AGG CC
HIV-T8-R	GGC CTC TAG AAG GTAGTGGGAATTGGAGGTCTTATCAAA A GAA GCT TCG GC
Negative-F	GCC GAA GCT TCT TTTGATTCTTGCTCTGCTCTCTTCGTCC TTC TAG AGG CC
Negative-R	GGC CTC TAG AAG GACGAAGAGAGCAGAGCAAGAATCAAA A GAA GCT TCG GC

pUC19_F	CAGCGAGTCAGTGAGCGA
pUC19_R	GCGACACGGAAATGTTGAATACTCAT
HLA-C-F	ACACACTCGAAACGTCCCAA
HLA-C-R	AAGTCCTTCTGGAGCCCTTC

Name	**Sequences (5′→3′)**
Oligos used in *in vitro* high throughput assay experiments.	
HLA-library	/5′Phos/*TTGTGT*NNNNC*C*NG*G*G*C*C*T*T*G*T*A*G*A*T*C*T*G*T*G*T*G*T*G*NNNNACCTGCCGAG*TTGTGT*
ABO-library	/5′Phos/*AGAGAA*NNNNC*C*NC*G*T*G*G*T*C*G*C*G*G*A*A*C*T*C*C*A*T*G*NNNNACCTGCCGAG*AGAGAA*
S501-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***TAGATCGC***
S501-R	GCGATCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S502-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*C****TCTCTAT***
S502-R	ATAGAGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S503-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***TATCCTCT***
S503-R	AGAGGATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S504-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***AGAGTAGA***
S504-R	TCTACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S505-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***GTAAGGAG***
S505-R	CTCCTTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S506-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***ACTGCATA***
S506-R	TATGCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S507-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***AAGGAGTA***
S507-R	TACTCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
S508-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT***CTAAGCCT***

S508-R	AGGCTTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT GTAGATCTCGGTGG
HLA-N701	CAAGCAGAAGACGGCATACGAGAT***TCGCCTTA***ACCTGCC GAG*TTGTGT*
HLA-N702	CAAGCAGAAGACGGCATACGAGAT***CGTACTAG***ACCTGCC GAG*TTGTGT*
HLA-N703	CAAGCAGAAGACGGCATACGAGAT***TTCTGCCT***ACCTGCCG AG*TTGTGT*
ABO-N705	CAAGCAGAAGACGGCATACGAGAT***AGGAGTCC***ACCTGCC GAG*AGAGAA*
ABO-N706	CAAGCAGAAGACGGCATACGAGAT***CATGCCTA***ACCTGCC GAG*AGAGAA*
ABO-N707	CAAGCAGAAGACGGCATACGAGAT***GTAGAGAG***ACCTGCC GAG*AGAGAA*
ABO-N708	CAAGCAGAAGACGGCATACGAGAT***CCTCTCTG***ACCTGCC GAG*AGAGAA*
PE2_short	AAT GAT ACG GCG ACC ACC GA
HLA_sel_PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA G*TT GTG T*
ABO_sel_PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACC TGC CGA G*AG AGA A*
Lib_adaptor1	GAC GGC ATA CGA GAT
HLA_lib_adaptor2	*TTG T*AT CTC GTA TGC CGT CTT CTG CTT G
ABO_lib_adaptor2	*AGA G*AT CTC GTA TGC CGT CTT CTG CTT G
lib_PCR_F	CAA GCA GAA GAC GGC ATA CGA GAT
HLA_PCR_R	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA G*TT GTG T*
ABO_PCR_R	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA G*AG AGA A*

Name	**Sequences (5′→3′)**
Oligos used to create DNA target constructs in a cell reporter assay	
ABO-T1-REP-F	AATTCC CATGGAGTTCCGCGACCACGTGG AGGAG
ABO-T1-REP-R	GATCCTCCT CCACGTGGTCGCGGAACTCCATG GG
ABO-T2-REP-F	AATTCC CATGGAGATCCGCGACCACGTGG AGGAG
ABO-T2-REP-R	GATCCTCCT CCACGTGGTCGCGGATCTCCATG GG
ABO-T3-REP-F	AATTCC CATGGAGTTCCGCGACCATGTGG AGGAG
ABO-T3-REP-R	GATCCTCCT CCACATGGTCGCGGAACTCCATG GG
ABO-T4-REP-F	AATTCC CATGGAGTTCCGCGACCACATGG AGGAG
ABO-T4-REP-R	GATCCTCCT CCATGTGGTCGCGGAACTCCATG GG
ABO-T5-REP-F	AATTCC CATGGAGATCCGCGACCACATGG AGGAG

ABO-T5-REP-R	GATCCTCCT CCATGTGGTCGCGGATCTCCATG GG
ABO-T6-REP-F	AATTCC CATGGAGATCCGCGACCATATGG AGGAG
ABO-T6-REP-R	GATCCTCCT CCATATGGTCGCGGATCTCCATG GG
ABO-T7-REP-F	AATTCC CATGGAGATCCGCGACCATGTGG AGGAG
ABO-T7-REP-R	GATCCTCCT CCACATGGTCGCGGATCTCCATG GG
ABO-T8-REP-F	AATTCC CATGGAGTTCCGCGACCATATGG AGGAG
ABO-T8-REP-R	GATCCTCCT CCATATGGTCGCGGAACTCCATG GG

Name	**Sequences (5′→3′)**
Oligo used as ssDNA target in melting temperature assay experiments.	
ABO-T1-Tm	CCCACGTGGTCGCGGAACTCCATGT

Name	**Sequences (5′→3′)**
Oligos used for dsDNA target in EMSA assay experiments.	
ABO-T1-EMSA-F	CAATA CCACGTGGTCGCGGAACTCCATG
ABO-T1-EMSA-R	CATGGAGTTCCGCGACCACGTGG TATTGCGC/Cy5/
ABO-T5-EMSA-F	CAATA CCATGTGGTCGCGGATCTCCATG
ABO-T5-EMSA-R	CATGGAGATCCGCGACCACATGG TATTGCGC/Cy5/
ABO-T7-EMSA-F	CAATA CCACATGGTCGCGGATCTCCATG
ABO-T7-EMSA-R	CATGGAGATCCGCGACCATGTGG TATTGCGC/Cy5/
ABO-Neg-EMSA-F	CAATA CCTGTGCCTTGTAGATCTGTGTG
ABO-Neg-EMSA-R	CACACAGATCTACAAGGCACAGG TATTGCGC/Cy5/

Name	**Sequences (5′→3′)**
Oligos used in cellular cleavage assay experiments.	
HLA-C-F	ACACACTCGAAACGTCCCAA
HLA-C-R	AAGTCCTTCTGGAGCCCTTC
HIV-T1-Flp-F	GATCCCGAGTGAAGATGGAAACCAAAAATGATAGGGGGA ATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATCG GCCGC
HIV-T1-Flp-R	TCGAGCGGCCGATCTGATCATACTGTCTTACTTTGATAAA ACCTCCAATTCCCCCTATCATTTTTGGTTTCCATCTTCACT CGG
HIV-T8-Flp-F	GATCCCGAGTGAAGATGGAAACCAAAAATGGTAGTGGGA ATTGGAGGTCTTATCAAAGTAAGACAGTATGATCAGATCG GCCGC

HIV-T8-Flp-R	TCGAGCGGCCGATCTGATCATACTGTCTTACTTTGATAAG ACCTCCAATTCCCACTACCATTTTTGGTTTCCATCTTCACT CGG
FlpLocus-F	CGATGTACGGGCCAGATATAC
FlpLocus-R	AGGGAAGAAAGCGAAAGGAG
HIV-OT1-F	CGTGTACACACCTTCGTTGC
HIV-OT1-R	TCCCGACTGCCTAAGATGGA
HIV-OT2-F	GCTCTCTTGCCCATGGAGTT
HIV-OT2-R	AATTGGGGCCTTGAGACCAG

NNNNNN = Library barcode, ***NNNNNNNN*** = target barcode

An asterisk (*) indicates that the preceding nucleotide was incorporated as a hand mix of bases consisting of 79 mol % of the indicated base, and 7 mol % of each of the other three natural bases. "/5Phos/" denotes a 5′ phosphate group added to the sequence.

/Cy5/ refers to Cyanine 5.

4.6 Additional information

4.6.1 Acknowledgements

This work was supported by a Project Grant from the Canadian Institutes for Health Research (CIHR) (CIHR-PS-408552) and a Discovery/Discovery Accelerator Supplement grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2016-06381) to B.P.H. A.R.K. was supported by an NSERC CGS-M award. The authors thank Jinho Park and Prof. Seong Keun Kim (Seoul National University) for experimental advice.

4.6.2 Contributions

B.P.H., A.R.K., and C.R.C. conceived the study and designed experiments to examine how the inclusion of universal bases into crRNAs affects Cas9/Cas12a activity and specificity. A.R.K. performed all experiments studying the activity, kinetics, and thermodynamics of universal-base modified crRNAs *in vitro* and was assisted by C.R.C. and T.T. C.R.C. performed experiments

testing the activity and specificity of Cas9 and Cas12a crRNAs in cells. J.J. assisted with high-throughput sequencing and subsequent data analysis. All authors contributed to the writing of the paper.

4.6.3 Competing interests

The authors declare the following competing interests: B.P.H. and A.R.K. are inventors on a provisional patent (#63279550) covering the use of universal bases in CRISPR guide RNAs. The authors declare no other competing interests.

Chapter 5

Summary and Conclusions

5.1 Current challenges facing CRISPR-Cas9 based gene editing

Engineered nucleases such as ZFNs, TALENs and CRISPR-Cas systems have always been intimately tied to the notion of a therapeutic gene-editing process capable of correcting disease-causing mutations. However, the modularity and simplicity of implementation associated with CRISPR-Cas systems have allowed these technologies to expand across several disciplines, including the generation of model cell lines and engineered organisms[86, 210, 291], *in vitro* nucleic acid detection platforms for pathogen identification[70, 131, 292], large-scale genetic[293] and epigenetic screens[212], as well as population-scale modifications through gene drives[294-296] as well as agricultural applications[297-299] among others. While the rise of CRISPR has facilitated this surge in the ability to modulate genetic information, there are several challenges which still remain before the technology may be fully realized. These include such factors as: reliable delivery of the necessary components *in vivo*[300], auto-immune responses against bacterial-derived Cas9 proteins[301, 302], induction of DNA-repair pathways following Cas9-mediated DSBs[303, 304], off-target activity at highly homologous sequences[216], and the presence of polymorphisms between individuals, or high levels of population diversity which may negatively impact efficient Cas-based targeting[257].

The work presented herein (in addition to previous literature) addresses the latter, specifically issues pertaining to Cas9 off-target activity (**Chapters 2** and **3**), and the ability to target multiple polymorphic sequences simultaneously (**Chapter 4**). Despite the significant advancements made

towards higher-fidelity Cas9-based gene editing, the presence of off-target activity at undesired loci still remains a concern. Several studies aimed at describing the landscape of Cas9 specificity have demonstrated that the number of off-targets for a particular crRNA is largely dependent on the uniqueness of the DNA target in question (i.e., how many similar sites are present within the genome)[305], as well as epigenetic factors situated locally around the site[306]. They have also shown that the number of off-targets ranges anywhere from a single detectable site, to >50 in some instances[147]. Nuclease-deficient dCas9 has been shown to be markedly more prone to off-target binding (when compared to wild-type Cas9 nuclease activity), demonstrating over >1000 dCas9-bound off-target sites for some sgRNAs tested (as measured by ChIP-seq)[307].

The introduction of these Cas9-mediated DSBs (at on- and off-target sites) is capable of introducing indels, structural variants (indels larger than 50 bp), large deletions up to 9.5 kbp in length, as well as complex genomic rearrangements[308-311]. These genomic perturbations are capable of being passed down generationally, with one study demonstrating that 26% of a Cas9-edited zebrafish's offspring contained off-target mutations, while 9% contained larger structural variations[309]. Cas9-induced DSBs are also capable of activating the p53-mediated DNA damage response pathways, resulting in cell cycle arrest and eventual cell death[312]. Because of this, cells which are deficient in p53 outperform wild-type cells in terms of Cas9-mediated efficiencies[312], selecting for cells which therefore may be more vulnerable to tumourigenic mutations. Therefore, controlling DNA damage responses following Cas9 treatment, as well as minimizing the number of introduced DSBs capable of activating p53, is integral for future studies regarding Cas9's therapeutic application. Although a number of efforts have been made towards improving specificity of the system, the presence of off-target activity still remains a significant barrier,

especially regarding clinical applications of CRISPR. While other challenges are also present, the multifaceted and detrimental effects of off-target activity mediated by Cas9 (as well as other CRISPR-Cas systems) further underscores the importance of developing genome engineering tools with reliable and accurate targeting capabilities.

5.2 Improving Cas9 endonuclease specificity through XNA-modified crRNAs

While several approaches have been described for improving specificity of CRISPR-Cas9 systems, the work herein focuses on the RNA/DNA heteroduplex, specifically through the use of chemically modified nucleotides within the sequence-specific region of the Cas9 crRNA. Two studies (**Chapter 2** and **Chapter 3**) were performed to assess the design, implementation and functionality of XNA-modified crRNAs for use with Cas9.

The results from **Chapter 2** were among the first to demonstrate the applicability of XNAs as a means of increasing CRISPR-based cleavage specificity, and the first reported use of BNAs or LNAs within the crRNA of Cas9[194]. Importantly, we showed that XNAs previously used to increase mismatch discrimination and/or stability in technologies distinct from CRISPR (SNP detection, siRNA, ribozymes, aptamers, etc…)[221, 222, 313, 314] are capable of transferring these qualities to the RNA/DNA heteroduplex formed during Cas9 target binding. Interestingly, these results also showed that previously validated methods for designing highly specific LNA-based probes, were ineffective when applied to the crRNA of Cas9. When incorporating LNAs into a probe for SNP detection, it has previously been demonstrated that a triplet of modifications surrounding the polymorphic nucleotide leads to the highest level of mismatch discrimination[220]. However, our results indicated that this had the inverse effect, stabilizing the mismatch-containing

off-target (WAS-OT1, shown in **Figure 2.1b**), showing no discernible difference in specificity at these target sites (relative to an unmodified control) (**Figures 2.1b**, **d** and **Figures 2.20a, c**). This was further confirmed by high-throughput specificity profiling data, which showed that the position of modification with either BNA or LNA results in either no observable increases, or decreases in specificity locally at the site of incorporation (**Figures 2.5**, **2.6**, and **2.22**). Collectively, these results showed that while characteristics of BNAs and LNAs are capable of being transferred to CRISPR-based systems such as Cas9, the design of said modifications is independent of previous applications, and therefore must be carefully considered in order to elicit the desired effects.

In addition to the aforementioned design considerations, we also identified limitations associated with the use of BNA- or LNA-modified crRNAs. While specificity was maintained across all contexts tested, we did observe a reduction in on-target activity within cells that was not seen *in vitro* (**Figure 2.1**, **2.2** and **2.8**, and **Tables 2.5** and **2.6**). This loss of on-target activity was determined to be likely due to delayed reaction kinetics (**Figures 2.17-2.19 and 2.26c**, **d**), and is consistent with other studies aimed at increasing the specificity of Cas9[151, 176, 177, 248], or utilizing chemically modified gRNAs[243]. Through smFRET-based experiments, we were able to show that BNA-modified crXNAs prevent Cas9 from stably maintaining a high-energy, zipped conformation (distinguished by complete formation of the R-loop) required for DNA cleavage when bound to an off-target DNA. However, we also identified that of the Cas9 RNP complexes within this high-energy state, those bound to BNA crXNAs existed in a more transient state. Approximately 30% of Cas9 molecules complexed with BNA crXNAs were found to be transiently within this high-energy state, as opposed to ~20% for wild-type crRNA bound Cas9 (when incubated with an on-

target DNA) (**Figure 2.16**), potentially manifesting as the observed delayed cleavage kinetics in our results. Specifically, R-loop formation is required for REC3 domain docking, which induces the necessary nuclease domain rearrangements for DNA cleavage[60, 177]. Our results show that BNA crXNAs impede Cas9 from maintaining a stable, zipped conformation when bound to both on- and off-targets (although to different degrees). By preventing stable R-loop formation, this inhibits REC3 docking and prevents subsequential domain rearrangements[177]. Several high-fidelity Cas9 variants also achieve increased specificity through disruption of REC3 binding to the RNA/DNA heteroduplex, including HypaCas9[177] and HiFi Cas9[176]. Interestingly, these variants also display reduced reaction kinetics, and a loss of on-target activity[60, 151, 176, 177], potentially linking the manner in which specificity is improved between crXNAs and engineered Cas9 variants.

Our results in **Chapter 2** demonstrated that the number and position of modifications within the sequence-specific region of the Cas9 crRNA is critical to the resulting fidelity and activity of the crXNA (when using BNAs or LNAs). In **Chapter 3**, we expanded our initial scope (which focused exclusively on the use of BNAs and LNAs) to several additional XNAs previously used to either stabilize RNA oligonucleotides, or improve mismatch discrimination (shown in **Figure 3.1c**). Of the XNAs tested, we were able to identify several specific modifications capable of broad-spectrum specificity improvements based on *in vitro* high-throughput specificity profiling data (**Figure 3.1d**, **e**). These results more comprehensively showed the specificity profiles of previously utilized XNAs such as PACE[195], DNA[202, 203], LNA and BNA modifications[194], while also being the first reported use of DNAPS and FANA XNAs for increased Cas9 specificity. We were also able to demonstrate that crXNA specificity and activity is largely sequence-independent, functioning across a diverse set of Cas9 target sites without significant variation in most examples

(**Figure 3.14b**). However, we did show that while a single crXNA modification pattern may function across a variety of sequences, differentially modified crXNAs do display distinct specificity profiles regarding the off-targets which they are able to prevent cleavage of (**Figures 3.5-3.12**). This is important, as it has significance towards how crXNA technology may be applied to CRISPR-Cas systems in future studies and applications. More specifically, based on the composition and position of mismatches present within natural off-targets for a given Cas9 target site, different crXNAs may be more, or less appropriate for their ability to remove said off-targets. While we were able to show examples of crXNAs with modifications across the 20 nt spacer sequence, we did observe a bias towards modification of the PAM distal, and mid regions. These regions of the RNA/DNA heteroduplex make several contacts with the REC3 domain of Cas9 (**Figure 3.17**)[177]. Therefore, it is possible that XNA-mediated specificity is due to disruption of REC3 correctly docking to the heteroduplex, although additional structural studies are required to confirm this.

5.3 Multi-sequence targeting through universal base-modified Cas9/Cas12a crRNAs

In **Chapter 4**, we explored alternative uses of chemically modified crRNAs, specifically the ability to utilize universal bases as a means for the simultaneous targeting of multiple polymorphic sequences without loss of overall specificity. While single-site genome engineering is desirable, the presence of natural polymorphic sequences obfuscates the use of CRISPR-Cas systems therapeutically without prior sequencing of the target region, or delivery of a cocktail of gRNAs containing possible variant sequences. Both of these have associated pitfalls, as preliminary sequencing adds financial cost and time to potential therapies, while delivery of multiple gRNAs increases the risk of off-target activity. Additionally, CRISPR systems have been

proposed as potential anti-viral therapies[5], and used effectively as point-of-contact tools for detection of pathogenic material[292]. Due to the high level of sequence diversity within both bacterial and viral pathogens, the ability to target multiple sites simultaneously not only increases the effectiveness of potential anti-bacterial and anti-virals (by providing a more broad targeting spectrum), but also reduces the number of false negatives associated with CRISPR-based diagnostic platforms.

Our results showed that through the modification of Cas9 crRNAs with one of several universal bases (shown in **Figure 4.2a**), we were able to achieve selective degeneracy limited to the site of their incorporation. Importantly, we were also able to demonstrate that these modifications did not otherwise alter the specificity of Cas9, or introduce new off-targets (**Figures 4.10b** and **4.15**). While all of the universal bases tested were able to impart non-specific interactions at the site of their incorporation, we determined that ribose inosine and dK/dP modifications were the most reliable for multi-sequence targeting. However, as seen with our previous results in **Chapter 2**, we did observe a loss of on-target activity within cells when using crRNAs modified with these nucleotides. Interestingly, like the BNA-modified crRNAs in **Chapter 2**, we did also observe a reduction in reaction kinetics *in vitro* (**Figure 4.24**), which may be correlated with the loss of cellular activity.

We were also able to show that the universal base ribose inosine (rI) is not only tolerated in Cas9, but also in Cas12a; and is able to elicit efficient multi-sequence targeting (**Figure 4.25b**, **c**). These results were significant, as it demonstrated that the use of universal base-modified crRNAs were able to be applied to Cas12a-based *in vitro* nucleic acid detection platforms such as DETECTR.

We were able to show that our universal base-modified crRNAs out performed wild-type crRNAs when used within the DETECTR assay to identify polymorphisms within the HIV-1 protease gene. More specifically, when using unmodified crRNAs, we failed to detect 4 of the 8 HIV-1 variant sequences tested. However, using an rI-modified Cas12a crRNA, we efficiently identified all 8 variants (as well as displaying stronger signals for each sample) (**Figure 4.25e**). These results demonstrated that while chemically modified nucleotides may be used to broadly increase specificity (**Chapters 2** and **3**), they may also be applied to additional applications, as well as Cas systems distinct from Cas9.

5.4 Remaining questions and future directions

While we were able to identify several XNAs capable of increasing the specificity of the CRISPR-Cas9 system in **Chapter 2** and **Chapter 3**, we did identify several limitations to the technology. The foremost being the apparent loss of on-target activity observed for BNA- and LNA-modified crXNAs in **Chapter 2**, as well as universal base-modified crRNAs in **Chapter 4**. In **Chapter 3**, we expanded our study of crXNAs to include a more diverse array of chemically modified nucleotides, but have yet to confirm whether or not these also share any observable loss of on-target activity in cells. As functioning in a cellular context is important for any gene editing tool, the testing of these crXNAs in cells for their ability to maintain the activity and specificity observed *in vitro*, is critical for future implementation.

Secondly, our results also determined that for crXNAs to improve specificity, they must be carefully designed with regards to type, and position of XNA modification. In conjunction with this, we also showed that different crXNAs display unique specificity profiles, and therefore are

effective at reducing off-target sequences differentially. In order for crXNAs to be more widely adopted, there must be tools capable of accurately predicting how and where to modify a particular crRNA of interest. This is especially important, as there are currently no available methods for reliably describing the design of high-fidelity, modified crRNAs. To-date, several studies have utilized previously published, large-scale Cas9 specificity experiments to train machine learning-based tools for the accurate prediction of gRNAs capable of high levels of specificity and/or activity[305, 315-317]. This methodology is amenable to the significant amount of high-throughput specificity profiling data outlined in **Chapter 3**. More specifically, we plan to utilize these data sets to train a machine learning algorithm such that it is capable of determining the specific patterns of off-targets which are unique to a given crXNA. These models may then be linked to previously described programs capable of identifying sequences of high homology to a target of interest (i.e., potential off-targets) within a specific genomes for a given Cas9 target[318]. In this manner, we aim to develop a program which takes a Cas9 target site as input, determine potential off-target sites based on sequence similarity to a reference genome (as previously described)[159], followed by comparison with our machine learning models to identify a specific crXNA most likely to eliminate the identified off-targets, while maintaining on-target activity.

For chemically modified crRNAs to be more widely adopted as either a means of increasing the fidelity of CRISPR-Cas systems, or for simultaneous multi-sequence targeting, on-target efficiency is a critical component and must be addressed. While several studies have identified possible biochemical mechanisms which explain the observed reduction in on-target activity for engineered Cas9 variants[60, 177, 248], there has yet to be any work done regarding this phenomena in chemically modified crRNAs. Many of the crXNAs tested in **Chapter 3** are modified in positions

which make direct contact with the REC3 domain of Cas9[177]. Therefore, it is possible that some crXNAs are functioning through biochemical mechanisms similar to those of HypaCas9[177] or Cas9-HF1[176]; specifically through increasing the energetic barrier required for proper REC3 docking to the RNA/DNA heteroduplex. However, manipulation of REC3 is not the only means by which specificity may be increased. In **Chapter 2**, we demonstrated that BNA crXNAs mediate increased specificity through preventing formation of a stable R-loop (**Figure 5.1**)[194]. Interestingly, in conjunction with previous literature, our results are consistent with the observation that higher levels of specificity are correlated with reduced reaction kinetics (and subsequent loss of on-target activity). In Cas9 variants, these delayed cleavage kinetics are hypothesized to be due to slower conformational rearrangements of the HNH nuclease domain, an unfavourable equilibrium required for conformational changes, and/or mis-alignment of catalytic residues responsible for DNA cleavage[248]. While our results are consistent with improved specificity and reduced reaction kinetics, additional studies are needed to confirm this. By determining the underlying biochemical mechanisms which result in the delayed on-target activity of Cas9 bound to crXNAs, we may be able to improve on-target efficiency through rational engineering of a novel variant, or through previously described directed evolution techniques[13]. Additionally, a new hyperactive Cas9 variant (known as TurboCas9) was recently generated through a combination of rational design and directed evolution[319]. Containing mutations within the HNH domain, TurboCas9 has been demonstrated to achieve significantly higher on-targeting efficiencies within cells, while maintaining comparable specificity (when compared to wild-type Cas9). This hyperactive variant is one example of how new studies may function synergistically with the modified crRNAs described in **Chapters 2**-**4**, resulting in increased fidelity and a rescue of on-target activity; although additional experimentation is required to confirm this.

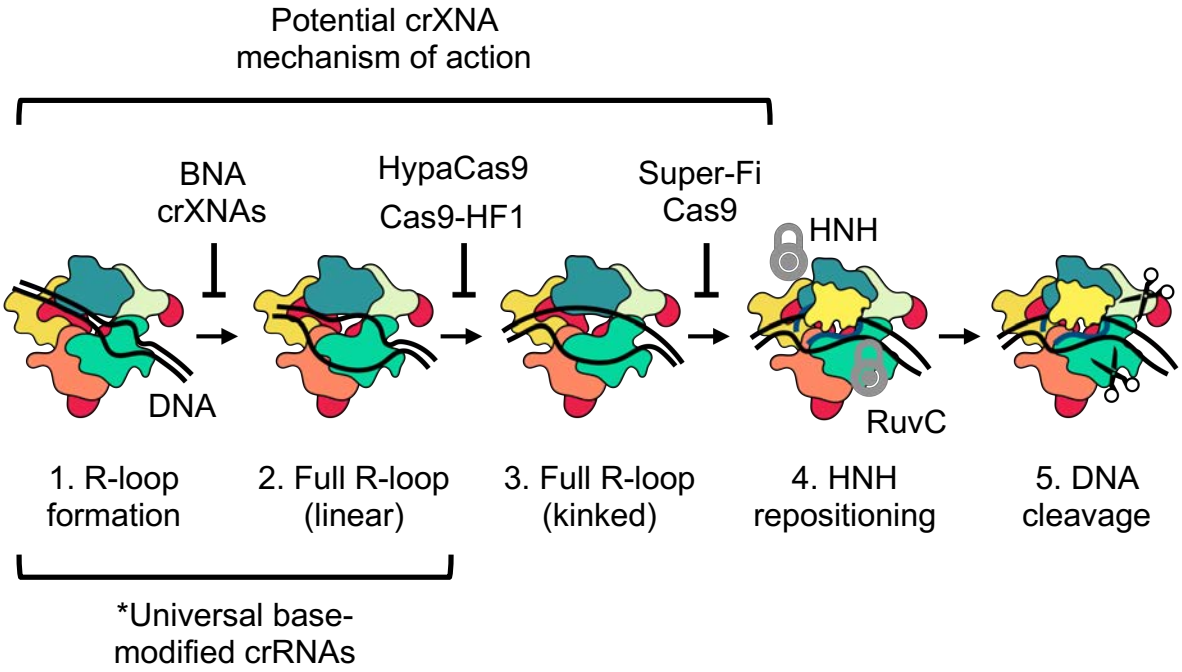


Figure 5.1. Potential mechanisms for Cas9 endonuclease specificity enhancements
Cas9 specificity may be achieved through a variety of mechanisms. crXNAs containing BNAs have been previously shown to impede the ability of Cas9 to maintain a stable R-loop, therefore preventing transition to a cleavage competent conformation. Engineered variants such as HypaCas9 and Cas9-HF1 have been demonstrated to reduce the number of interactions between REC3 and the RNA/DNA heteroduplex, therefore raising the energetic barrier for HNH repositioning. Super-Fi Cas9 contains mutations within the linker regions (L1 and L2) which stabilize the distal end of the heteroduplex, functioning to reduce the tolerance for mismatches within this region. While there is evidence for several of these mechanisms within the crXNAs studied, additional experimentation is required to elucidate the exact mechanism by which specificity is increased. *Universal base-modified crRNAs display reduced melting temperatures, and therefore may impede R-loop formation resulting in the observed decrease in activity within cells.

While the work detailed in **Chapters 2** and **3** focused on enhancing specificity of the CRISPR-Cas9 system, the use of universal base-modified crRNAs in **Chapter 4** outlined a novel application for chemically modified nucleotides which is independent of off-target mitigation. Although there is additional work to be done regarding improving the cellular efficiencies of such modified crRNAs, we were able to demonstrate their applicability to other CRISPR-based technologies such

as *in vitro* diagnostic platforms like DETECTR[70]. As the world continues to combat the SARS-CoV-2 pandemic, the need for technologies which are capable of reliably and confidently detecting highly variable pathogenic material continues to increase. While we were able to successfully demonstrate the use of chemically modified nucleotides for variant detection through the Cas12-based DETECTR assay, additional work is required to fully realize the potential for this technology. This includes optimization of the design of crRNAs containing universal bases, as well as how these may be applied to other CRISPR-based diagnostic tools, such as those which utilize Cas13 for RNA detection[320]. Additionally, while we have shown that universal base-modified crRNAs are capable of recognizing viral DNA containing polymorphisms, it is interesting to speculate as to whether or not these modified crRNAs may be used to improve anti-viral therapies. Current strategies have shown that the targeting of viral genetic material with CRISPR-Cas systems acts as a form of selection pressure for viruses bearing mutations that prevent efficient endonuclease targeting. Specifically, for Cas9-based HIV therapies it has been demonstrated that a significant proportion of mutations which lead to viral escape are single nucleotide substitutions present at the site of DNA cleavage following Cas9 digestion[260]. Therefore, it is possible that by placing universal bases at key positions within the crRNA of Cas9, we may be able to reduce the number of mutation-containing viruses able to escape Cas9 targeting, and therefore increase therapeutic efficiency. This may be true for other CRISPR-based anti-virals as well as antibiotics in cases where high levels of variation within the pathogenic population leads to therapy failure, although additional experimentation is required to confirm this.

5.5 Final comments

In the work presented herein, we reported how chemically modified nucleotides may be used to modulate specificity in two distinct applications. We believe that these results represent a strong foundation for the continued study of chemically modified crRNAs, and how they may continue to be used to augment the specificity of CRISPR-Cas systems. While beyond the scope of this thesis, chemically modified nucleotides have been used to impart a number of characteristics to Cas9-based gene editing which are independent of specificity; including photo-activatable Cas9 systems[321], novel strategies for the synthesis of large-scale crRNA libraries[322], multi-colour cell sorting through fluorescently labelled sgRNAs[323], and reduced levels of cellular toxicity[324] (among other uses). This diverse array of applications reflects the importance of chemically modified nucleic acid technology for the future of CRISPR-Cas based systems. While focused primarily on specificity, the results presented also function to comprehensively profile the positional dependence of chemically modified nucleotide incorporation on the ability of Cas9 to maintain on-target activity. As such, we believe that these results are not only important for future crXNAs with high fidelity, but also for design considerations wherever chemical modification of the crRNA is used. In addition to the specificity enhancing properties of XNAs, we have also demonstrated that crXNAs are capable of imparting novel properties to the CRISPR-Cas system; exemplified through the ability to target multiple polymorphic sequences using universal base-modified crRNAs. This not only has significant implications for future gene-editing based therapeutics, but also towards the growing field of CRISPR-based diagnostic platforms and anti-viral therapies.

Along with previously described strategies such as protein engineering, crXNAs have been demonstrated to be an effective strategy for both the enhancement of Cas9-mediated DNA cleavage specificity, as well as the modulation of sequences which are able to be recognized by the system. However, the solution to off-target activity will not likely be a unilateral one. More specifically, it is likely that a combination of protein engineering, computational predictions, improved delivery, as well as modification to the gRNA that will allow for the highest degrees of both on-target efficiency, and desired specificity.

References

1. Siu, K.H. and W. Chen, *Riboregulated toehold-gated gRNA for programmable CRISPR-Cas9 function.* Nat Chem Biol, 2019. **15**(3): p. 217-220.
2. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
3. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
4. Uddin, F., C.M. Rudin, and T. Sen, *CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future.* Front Oncol, 2020. **10**: p. 1387.
5. Baddeley, H.J.E. and M. Isalan, *The Application of CRISPR/Cas Systems for Antiviral Therapy.* Front Genome Ed, 2021. **3**: p. 745559.
6. Aslam, B., et al., *CRISPR-Cas system: a potential alternative tool to cope antibiotic resistance.* Antimicrob Resist Infect Control, 2020. **9**(1): p. 131.
7. Danna, K. and D. Nathans, *Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae.* Proc Natl Acad Sci U S A, 1971. **68**(12): p. 2913-7.
8. Rouet, P., F. Smih, and M. Jasin, *Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.* Mol Cell Biol, 1994. **14**(12): p. 8096-106.
9. Klug, A. and D. Rhodes, *Zinc fingers: a novel protein fold for nucleic acid recognition.* Cold Spring Harb Symp Quant Biol, 1987. **52**: p. 473-82.
10. Boch, J., et al., *Breaking the code of DNA binding specificity of TAL-type III effectors.* Science, 2009. **326**(5959): p. 1509-12.

11. Chandrasegaran, S. and D. Carroll, *Origins of Programmable Nucleases for Genome Engineering.* J Mol Biol, 2016. **428**(5 Pt B): p. 963-89.

12. Li, T. and B. Yang, *TAL effector nuclease (TALEN) engineering.* Methods Mol Biol, 2013. **978**: p. 63-72.

13. Hubbard, B.P., et al., *Continuous directed evolution of DNA-binding proteins to improve TALEN specificity.* Nat Methods, 2015. **12**(10): p. 939-42.

14. Wolfe, S.A., L. Nekludova, and C.O. Pabo, *DNA recognition by Cys2His2 zinc finger proteins.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 183-212.

15. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.

16. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.

17. Tautz, D., *Notes on the definition and nomenclature of tandemly repetitive DNA sequences.* EXS, 1993. **67**: p. 21-8.

18. Olsen, U.D.J.G.I.H.T.B.E.P.P.R.P.W.S.S.T.D.N.C.J.-F., et al., *Initial sequencing and analysis of the human genome.* nature, 2001. **409**(6822): p. 860-921.

19. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology (Reading), 2005. **151**(Pt 8): p. 2551-2561.

20. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-82.

21. Pourcel, C., G. Salvignol, and G. Vergnaud, *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.* Microbiology (Reading), 2005. **151**(Pt 3): p. 653-663.

22. Mojica, F.J., et al., *Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.* Mol Microbiol, 2000. **36**(1): p. 244-6.

23. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.* Biol Direct, 2006. **1**: p. 7.

24. Harnett, J., et al., *Lactic Acid Bacteria | Streptococcus thermophilus*, in *Encyclopedia of Dairy Sciences (Second Edition)*, J.W. Fuquay, Editor. 2011, Academic Press: San Diego. p. 143-148.

25. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.

26. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.

27. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.

28. Jiang, F. and J.A. Doudna, *CRISPR-Cas9 Structures and Mechanisms.* Annu Rev Biophys, 2017. **46**: p. 505-529.

29. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.

30. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.

31. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.

32. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

33. Sapranauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.* Nucleic Acids Res, 2011. **39**(21): p. 9275-82.

34. Jinek, M., et al., *RNA-programmed genome editing in human cells.* Elife, 2013. **2**: p. e00471.

35. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

36. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

37. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Curr Opin Microbiol, 2017. **37**: p. 67-78.

38. Wright, A.V., J.K. Nunez, and J.A. Doudna, *Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering.* Cell, 2016. **164**(1-2): p. 29-44.

39. Makarova, K.S., et al., *Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants.* Nat Rev Microbiol, 2020. **18**(2): p. 67-83.

40. van Houte, S., A. Buckling, and E.R. Westra, *Evolutionary Ecology of Prokaryotic Immune Mechanisms.* Microbiol Mol Biol Rev, 2016. **80**(3): p. 745-63.

41. Murovec, J., Z. Pirc, and B. Yang, *New variants of CRISPR RNA-guided genome editing enzymes.* Plant Biotechnol J, 2017. **15**(8): p. 917-926.

42. Kim, E., et al., *In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni.* Nat Commun, 2017. **8**: p. 14500.

43. Chen, H., J. Choi, and S. Bailey, *Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease.* J Biol Chem, 2014. **289**(19): p. 13284-94.

44. Heler, R., et al., *Cas9 specifies functional viral targets during CRISPR-Cas adaptation.* Nature, 2015. **519**(7542): p. 199-202.

45. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science, 2014. **343**(6176): p. 1247997.

46. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-49.

47. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.

48. Yang, W., *An equivalent metal ion in one- and two-metal-ion catalysis.* Nat Struct Mol Biol, 2008. **15**(11): p. 1228-31.

49. Semenova, E., et al., *Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10098-103.

50. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10092-7.

51. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213): p. 1258096.

52. Pattanayak, V., et al., *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.* Nat Biotechnol, 2013. **31**(9): p. 839-43.

53. Kunne, T., D.C. Swarts, and S.J. Brouns, *Planting the seed: target recognition of short guide RNAs.* Trends Microbiol, 2014. **22**(2): p. 74-83.

54. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-73.

55. Jiang, F., et al., *STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition.* Science, 2015. **348**(6242): p. 1477-81.

56. Jiang, F., et al., *Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.* Science, 2016. **351**(6275): p. 867-71.

57. Ma, H., et al., *CRISPR-Cas9 nuclear dynamics and target recognition in living cells.* J Cell Biol, 2016. **214**(5): p. 529-37.

58. Knight, S.C., et al., *Dynamics of CRISPR-Cas9 genome interrogation in living cells.* Science, 2015. **350**(6262): p. 823-6.

59. Sternberg, S.H., et al., *Conformational control of DNA target cleavage by CRISPR-Cas9.* Nature, 2015. **527**(7576): p. 110-3.

60. Bravo, J.P.K., et al., *Structural basis for mismatch surveillance by CRISPR-Cas9.* Nature, 2022. **603**(7900): p. 343-347.

61. Zetsche, B., et al., *Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system.* Cell, 2015. **163**(3): p. 759-71.

62. Stella, S., P. Alcon, and G. Montoya, *Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage.* Nature, 2017. **546**(7659): p. 559-563.

63. Charpentier, E., et al., *Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity.* FEMS Microbiol Rev, 2015. **39**(3): p. 428-41.

64. Fonfara, I., et al., *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.* Nature, 2016. **532**(7600): p. 517-21.

65. Paul, B. and G. Montoya, *CRISPR-Cas12a: Functional overview and applications.* Biomed J, 2020. **43**(1): p. 8-17.

66. Swarts, D.C., J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a.* Mol Cell, 2017. **66**(2): p. 221-233 e4.

67. Jones, S.K., Jr., et al., *Massively parallel kinetic profiling of natural and engineered CRISPR nucleases.* Nat Biotechnol, 2021. **39**(1): p. 84-93.

68. Kim, D., et al., *Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells.* Nat Biotechnol, 2016. **34**(8): p. 863-8.

69. Gao, L., et al., *Engineered Cpf1 variants with altered PAM specificities.* Nat Biotechnol, 2017. **35**(8): p. 789-792.

70. Chen, J.S., et al., *CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity.* Science, 2018. **360**(6387): p. 436-439.

71. Stella, S., et al., *Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity.* Cell, 2018. **175**(7): p. 1856-1871 e21.

72. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.* Science, 2016. **353**(6299): p. aaf5573.

73. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Mol Cell, 2015. **60**(3): p. 385-97.

74. Konermann, S., et al., *Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors.* Cell, 2018. **173**(3): p. 665-676 e14.

75. Yan, W.X., et al., *Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein.* Mol Cell, 2018. **70**(2): p. 327-339 e5.

76. O'Connell, M.R., *Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems.* J Mol Biol, 2019. **431**(1): p. 66-87.

77. Xu, C., et al., *Programmable RNA editing with compact CRISPR-Cas13 systems from uncultivated microbes.* Nat Methods, 2021. **18**(5): p. 499-506.

78. Abudayyeh, O.O., et al., *RNA targeting with CRISPR-Cas13.* Nature, 2017. **550**(7675): p. 280-284.

79. Zhang, K., et al., *CRISPR/Cas13d-Mediated Microbial RNA Knockdown.* Front Bioeng Biotechnol, 2020. **8**: p. 856.

80. Kushawah, G., et al., *CRISPR-Cas13d Induces Efficient mRNA Knockdown in Animal Embryos.* Dev Cell, 2020. **54**(6): p. 805-817 e7.

81. Aman, R., et al., *RNA virus interference via CRISPR/Cas13a system in plants.* Genome Biol, 2018. **19**(1): p. 1.

82. Ozcan, A., et al., *Programmable RNA targeting with the single-protein CRISPR effector Cas7-11.* Nature, 2021. **597**(7878): p. 720-725.

83. Wyman, C. and R. Kanaar, *DNA double-strand break repair: all's well that ends well.* Annu Rev Genet, 2006. **40**: p. 363-83.

84. Heyer, W.D., K.T. Ehmsen, and J. Liu, *Regulation of homologous recombination in eukaryotes.* Annu Rev Genet, 2010. **44**: p. 113-39.

85. Symington, L.S. and J. Gautier, *Double-strand break end resection and repair pathway choice.* Annu Rev Genet, 2011. **45**: p. 247-71.

86. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.

87. Essletzbichler, P., et al., *Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line.* Genome Res, 2014. **24**(12): p. 2059-65.

88. Miyaoka, Y., et al., *Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing.* Sci Rep, 2016. **6**: p. 23549.

89. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* Elife, 2014. **3**: p. e04766.

90. Pinder, J., J. Salsman, and G. Dellaire, *Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing.* Nucleic Acids Res, 2015. **43**(19): p. 9379-92.

91. Sfeir, A. and L.S. Symington, *Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?* Trends Biochem Sci, 2015. **40**(11): p. 701-714.

92. Bhargava, R., D.O. Onyango, and J.M. Stark, *Regulation of Single-Strand Annealing and its Role in Genome Maintenance.* Trends Genet, 2016. **32**(9): p. 566-575.

93. *Control of hereditary diseases. Report of a WHO Scientific Group.* World Health Organ Tech Rep Ser, 1996. **865**: p. 1-84.

94. Paquet, D., et al., *Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.* Nature, 2016. **533**(7601): p. 125-9.

95. Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.* Nature, 2016. **533**(7603): p. 420-4.

96. Nishida, K., et al., *Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.* Science, 2016. **353**(6305).

97. Ma, Y., et al., *Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells.* Nat Methods, 2016. **13**(12): p. 1029-1035.

98. Rees, H.A., et al., *Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery.* Nat Commun, 2017. **8**: p. 15790.

99. Kim, K., et al., *Highly efficient RNA-guided base editing in mouse embryos.* Nat Biotechnol, 2017. **35**(5): p. 435-437.

100. Rossidis, A.C., et al., *In utero CRISPR-mediated therapeutic editing of metabolic genes.* Nat Med, 2018. **24**(10): p. 1513-1518.

101. Li, G., et al., *Highly efficient and precise base editing in discarded human tripronuclear embryos.* Protein Cell, 2017. **8**(10): p. 776-779.

102. Komor, A.C., et al., *Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity.* Sci Adv, 2017. **3**(8): p. eaao4774.

103. Wang, X., et al., *Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion.* Nat Biotechnol, 2018. **36**(10): p. 946-949.

104. Gehrke, J.M., et al., *An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities.* Nat Biotechnol, 2018. **36**(10): p. 977-982.

105. Zong, Y., et al., *Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A.* Nat Biotechnol, 2018.

106. Thakore, P.I., et al., *Editing the epigenome: technologies for programmable transcription and epigenetic modulation.* Nat Methods, 2016. **13**(2): p. 127-37.

107. Larson, M.H., et al., *CRISPR interference (CRISPRi) for sequence-specific control of gene expression.* Nat Protoc, 2013. **8**(11): p. 2180-96.

108. Gebelein, B. and R. Urrutia, *Sequence-specific transcriptional repression by KS1, a multiple-zinc-finger-Kruppel-associated box protein.* Mol Cell Biol, 2001. **21**(3): p. 928-39.

109. Kearns, N.A., et al., *Functional annotation of native enhancers with a Cas9-histone demethylase fusion.* Nat Methods, 2015. **12**(5): p. 401-403.

110. Thakore, P.I., et al., *Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements.* Nat Methods, 2015. **12**(12): p. 1143-9.

111. Amabile, A., et al., *Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing.* Cell, 2016. **167**(1): p. 219-232 e14.

112. Perez-Pinera, P., et al., *RNA-guided gene activation by CRISPR-Cas9-based transcription factors.* Nat Methods, 2013. **10**(10): p. 973-6.

113. Maeder, M.L., et al., *CRISPR RNA-guided activation of endogenous human genes.* Nat Methods, 2013. **10**(10): p. 977-9.

114. Farzadfard, F., S.D. Perli, and T.K. Lu, *Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas.* ACS Synth Biol, 2013. **2**(10): p. 604-13.

115. Hilton, I.B., et al., *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.* Nat Biotechnol, 2015. **33**(5): p. 510-7.

116. Choudhury, S.R., et al., *CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter.* Oncotarget, 2016. **7**(29): p. 46545-46556.

117. Miller, J.F. and M. Sadelain, *The journey from discoveries in fundamental immunology to cancer immunotherapy.* Cancer Cell, 2015. **27**(4): p. 439-49.

118. Baylis, F. and M. McLeod, *First-in-human Phase 1 CRISPR Gene Editing Cancer Trials: Are We Ready?* Curr Gene Ther, 2017. **17**(4): p. 309-319.

119. Qasim, W., et al., *Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells.* Sci Transl Med, 2017. **9**(374).

120. Salmaninejad, A., et al., *PD-1 and cancer: molecular mechanisms and polymorphisms.* Immunogenetics, 2018. **70**(2): p. 73-86.

121. Nicholaou, T., et al., *Directions in the immune targeting of cancer: lessons learned from the cancer-testis Ag NY-ESO-1.* Immunol Cell Biol, 2006. **84**(3): p. 303-17.

122. Basak, A., et al., *BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations.* J Clin Invest, 2015. **125**(6): p. 2363-8.

123. Basak, A. and V.G. Sankaran, *Regulation of the fetal hemoglobin silencing factor BCL11A.* Ann N Y Acad Sci, 2016. **1368**(1): p. 25-30.

124. Maeder, M.L., et al., *Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10.* Nat Med, 2019. **25**(2): p. 229-233.

125. Freije, C.A., et al., *Programmable Inhibition and Detection of RNA Viruses Using Cas13.* Mol Cell, 2019. **76**(5): p. 826-837 e11.

126. Abbott, T.R., et al., *Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza.* Cell, 2020. **181**(4): p. 865-876 e12.

127. Fareh, M., et al., *Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance.* Nat Commun, 2021. **12**(1): p. 4270.

128. Pardee, K., et al., *Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components.* Cell, 2016. **165**(5): p. 1255-1266.

129. Jiao, C., et al., *Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9.* Science, 2021. **372**(6545): p. 941-948.

130. Green, A.A., et al., *Toehold switches: de-novo-designed regulators of gene expression.* Cell, 2014. **159**(4): p. 925-39.

131. Kellner, M.J., et al., *SHERLOCK: nucleic acid detection with CRISPR nucleases.* Nat Protoc, 2019. **14**(10): p. 2986-3012.

132. Chandu, D., et al., *Development of a Rapid Point-of-Use DNA Test for the Screening of Genuity(R) Roundup Ready 2 Yield(R) Soybean in Seed Samples.* Biomed Res Int, 2016. **2016**: p. 3145921.

133. Gootenberg, J.S., et al., *Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6.* Science, 2018. **360**(6387): p. 439-444.

134. Joung, J., et al., *Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing.* N Engl J Med, 2020. **383**(15): p. 1492-1494.

135. Li, S.Y., et al., *CRISPR-Cas12a-assisted nucleic acid detection.* Cell Discov, 2018. **4**: p. 20.

136. Li, L., et al., *HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation.* ACS Synth Biol, 2019. **8**(10): p. 2228-2237.

137. Notomi, T., et al., *Loop-mediated isothermal amplification of DNA.* Nucleic Acids Res, 2000. **28**(12): p. E63.

138. Harrington, L.B., et al., *Programmed DNA destruction by miniature CRISPR-Cas14 enzymes.* Science, 2018. **362**(6416): p. 839-842.

139. Aquino-Jarquin, G., *CRISPR-Cas14 is now part of the artillery for gene editing and molecular diagnostic.* Nanomedicine, 2019. **18**: p. 428-431.

140. He, Q., et al., *High-throughput and all-solution phase African Swine Fever Virus (ASFV) detection using CRISPR-Cas12a and fluorescence based point-of-care system.* Biosens Bioelectron, 2020. **154**: p. 112068.

141. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nat Biotechnol, 2013. **31**(9): p. 822-6.

142. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

143. Cromwell, C.R., J. Jovel, and B.P. Hubbard, *Methods for Measuring CRISPR/Cas9 DNA Cleavage in Cells.* Methods Mol Biol, 2021. **2162**: p. 197-213.

144. Heigwer, F., G. Kerr, and M. Boutros, *E-CRISP: fast CRISPR target site identification.* Nat Methods, 2014. **11**(2): p. 122-3.

145. Wang, X., et al., *Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors.* Nat Biotechnol, 2015. **33**(2): p. 175-8.

146. Gabriel, R., et al., *An unbiased genome-wide analysis of zinc-finger nuclease specificity.* Nat Biotechnol, 2011. **29**(9): p. 816-23.

147. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-197.

148. Liang, S.Q., et al., *Genome-wide detection of CRISPR editing in vivo using GUIDE-tag.* Nat Commun, 2022. **13**(1): p. 437.

149. Gu, B., E. Posfai, and J. Rossant, *Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos.* Nat Biotechnol, 2018. **36**(7): p. 632-637.

150. Nobles, C.L., et al., *iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity.* Genome Biol, 2019. **20**(1): p. 14.

151. Schmid-Burgk, J.L., et al., *Highly Parallel Profiling of Cas9 Variant Specificity.* Mol Cell, 2020. **78**(4): p. 794-800 e8.

152. Picelli, S., et al., *Tn5 transposase and tagmentation procedures for massively scaled sequencing projects.* Genome Res, 2014. **24**(12): p. 2033-40.

153. Wienert, B., et al., *Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq.* Science, 2019. **364**(6437): p. 286-289.

154. Connelly, J.C. and D.R. Leach, *Tethering on the brink: the evolutionarily conserved Mre11-Rad50 complex.* Trends Biochem Sci, 2002. **27**(8): p. 410-8.

155. Moreno-Herrero, F., et al., *Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA.* Nature, 2005. **437**(7057): p. 440-3.

156. Borde, V., *The multiple roles of the Mre11 complex for meiotic recombination.* Chromosome Res, 2007. **15**(5): p. 551-63.

157. Pattanayak, V., et al., *Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection.* Nat Methods, 2011. **8**(9): p. 765-70.

158. Guilinger, J.P., et al., *Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity.* Nat Methods, 2014. **11**(4): p. 429-35.

159. Bae, S., J. Park, and J.S. Kim, *Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.* Bioinformatics, 2014. **30**(10): p. 1473-5.

160. Xiao, A., et al., *CasOT: a genome-wide Cas9/gRNA off-target searching tool.* Bioinformatics, 2014. **30**(8): p. 1180-1182.

161. Jacquin, A.L.S., D.T. Odom, and M. Lukk, *Crisflash: open-source software to generate CRISPR guide RNAs against genomes annotated with individual variation.* Bioinformatics, 2019. **35**(17): p. 3146-3147.

162. Stemmer, M., et al., *CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.* PLoS One, 2015. **10**(4): p. e0124633.

163. Fine, E.J., et al., *An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage.* Nucleic Acids Res, 2014. **42**(6): p. e42.

164. Cradick, T.J., et al., *COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites.* Mol Ther Nucleic Acids, 2014. **3**: p. e214.

165. Singh, R., et al., *Cas9-chromatin binding information enables more accurate CRISPR off-target prediction.* Nucleic Acids Res, 2015. **43**(18): p. e118.

166. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.

167. Peng, H., et al., *Recognition of CRISPR/Cas9 off-target sites through ensemble learning of uneven mismatch distributions.* Bioinformatics, 2018. **34**(17): p. i757-i765.

168. Listgarten, J., et al., *Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs.* Nat Biomed Eng, 2018. **2**(1): p. 38-47.

169. Abadi, S., et al., *A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action.* PLoS Comput Biol, 2017. **13**(10): p. e1005807.

170. Kim, S., et al., *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* Genome Res, 2014. **24**(6): p. 1012-9.

171. Zuris, J.A., et al., *Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo.* Nat Biotechnol, 2015. **33**(1): p. 73-80.

172. Zetsche, B., S.E. Volz, and F. Zhang, *A split-Cas9 architecture for inducible genome editing and transcription modulation.* Nat Biotechnol, 2015. **33**(2): p. 139-42.

173. Davis, K.M., et al., *Small molecule-triggered Cas9 protein with improved genome-editing specificity.* Nat Chem Biol, 2015. **11**(5): p. 316-8.

174. Nihongaki, Y., et al., *Photoactivatable CRISPR-Cas9 for optogenetic genome editing.* Nat Biotechnol, 2015. **33**(7): p. 755-60.

175. Slaymaker, I.M., et al., *Rationally engineered Cas9 nucleases with improved specificity.* Science, 2016. **351**(6268): p. 84-8.

176. Kleinstiver, B.P., et al., *High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.* Nature, 2016. **529**(7587): p. 490-5.

177. Chen, J.S., et al., *Enhanced proofreading governs CRISPR-Cas9 targeting accuracy.* Nature, 2017. **550**(7676): p. 407-410.

178. Bravo, J.P.K., et al., *Structural basis for broad anti-phage immunity by DISARM.* Nat Commun, 2022. **13**(1): p. 2987.

179. Casini, A., et al., *A highly specific SpCas9 variant is identified by in vivo screening in yeast.* Nat Biotechnol, 2018. **36**(3): p. 265-271.

180. Vakulskas, C.A., et al., *A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells.* Nat Med, 2018. **24**(8): p. 1216-1224.

181. Hu, J.H., et al., *Evolved Cas9 variants with broad PAM compatibility and high DNA specificity.* Nature, 2018. **556**(7699): p. 57-63.

182. Lee, J.K., et al., *Directed evolution of CRISPR-Cas9 to increase its specificity.* Nat Commun, 2018. **9**(1): p. 3048.

183. Esvelt, K.M., J.C. Carlson, and D.R. Liu, *A system for the continuous directed evolution of biomolecules.* Nature, 2011. **472**(7344): p. 499-503.

184. Brodel, A.K., M. Isalan, and A. Jaramillo, *Engineering of biomolecules by bacteriophage directed evolution.* Curr Opin Biotechnol, 2018. **51**: p. 32-38.

185. Dove, S.L., J.K. Joung, and A. Hochschild, *Activation of prokaryotic transcription through arbitrary protein-protein contacts.* Nature, 1997. **386**(6625): p. 627-30.

186. Chen, Z. and H. Zhao, *A highly sensitive selection method for directed evolution of homing endonucleases.* Nucleic Acids Res, 2005. **33**(18): p. e154.

187. Basila, M., M.L. Kelley, and A.V.B. Smith, *Minimal 2'-O-methyl phosphorothioate linkage modification pattern of synthetic guide RNAs for increased stability and efficient CRISPR-Cas9 gene editing avoiding cellular toxicity.* PLoS One, 2017. **12**(11): p. e0188593.

188. Hendel, A., et al., *Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells.* Nat Biotechnol, 2015. **33**(9): p. 985-989.

189. Rahdar, M., et al., *Synthetic CRISPR RNA-Cas9-guided genome editing in human cells.* Proc Natl Acad Sci U S A, 2015. **112**(51): p. E7110-7.

190. Mir, A., et al., *Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing.* Nat Commun, 2018. **9**(1): p. 2641.

191. Yin, H., et al., *Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing.* Nat Biotechnol, 2017. **35**(12): p. 1179-1187.

192. Cho, S.W., et al., *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.* Genome Res, 2014. **24**(1): p. 132-41.

193. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nat Biotechnol, 2014. **32**(3): p. 279-284.

194. Cromwell, C.R., et al., *Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity.* Nat Commun, 2018. **9**(1): p. 1448.

195. Ryan, D.E., et al., *Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs.* Nucleic Acids Res, 2018. **46**(2): p. 792-803.

196. Nahar, S., et al., *A G-quadruplex motif at the 3′ end of sgRNAs improves CRISPR-Cas9 based genome editing efficiency.* Chem Commun (Camb), 2018. **54**(19): p. 2377-2380.

197. Finn, J.D., et al., *A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing.* Cell Rep, 2018. **22**(9): p. 2227-2235.

198. Ma, H., et al., *Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow.* Nat Biotechnol, 2016. **34**(5): p. 528-30.

199. Shao, S., et al., *Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system.* Nucleic Acids Res, 2016. **44**(9): p. e86.

200. Ryu, S.M., et al., *Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy.* Nat Biotechnol, 2018. **36**(6): p. 536-539.

201. Kocak, D.D., et al., *Increasing the specificity of CRISPR systems with engineered RNA secondary structures.* Nat Biotechnol, 2019. **37**(6): p. 657-666.

202. Rueda, F.O., et al., *Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease.* Nat Commun, 2017. **8**(1): p. 1610.

203. Yin, H., et al., *Partial DNA-guided Cas9 enables genome editing with reduced off-target activity.* Nat Chem Biol, 2018. **14**(3): p. 311-316.

204. Kartje, Z.J., et al., *Chimeric Guides Probe and Enhance Cas9 Biochemical Activity.* Biochemistry, 2018. **57**(21): p. 3027-3031.

205. Simone, B.W., et al., *Chimeric RNA: DNA TracrRNA Improves Homology-Directed Repair In Vitro and In Vivo.* CRISPR J, 2022. **5**(1): p. 40-52.

206. O'Reilly, D., et al., *Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity.* Nucleic Acids Res, 2019. **47**(2): p. 546-558.

207. Ageely, E.A., et al., *Gene editing with CRISPR-Cas12a guides possessing ribose-modified pseudoknot handles.* Nat Commun, 2021. **12**(1): p. 6591.

208. Mendez-Mancilla, A., et al., *Chemically modified guide RNAs enhance CRISPR-Cas13 knockdown in human cells.* Cell Chem Biol, 2022. **29**(2): p. 321-327 e4.

209. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-70.

210. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nat Biotechnol, 2014. **32**(4): p. 347-55.

211. Joung, J., et al., *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nat Protoc, 2017. **12**(4): p. 828-863.

212. Klann, T.S., et al., *CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome.* Nat Biotechnol, 2017. **35**(6): p. 561-568.

213. Cheong, T.C., M. Compagno, and R. Chiarle, *Editing of mouse and human immunoglobulin genes by CRISPR-Cas9 system.* Nat Commun, 2016. **7**: p. 10934.

214. Hoban, M.D., et al., *CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells.* Mol Ther, 2016. **24**(9): p. 1561-9.

215. O'Geen, H., et al., *A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture.* Nucleic Acids Res, 2015. **43**(6): p. 3389-404.

216. O'Geen, H., A.S. Yu, and D.J. Segal, *How specific is CRISPR/Cas9 really?* Current Opinion in Chemical Biology, 2015. **29**: p. 72-78.

217. Xu, X., D. Duan, and S.J. Chen, *CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment.* Sci Rep, 2017. **7**(1): p. 143.

218. Tycko, J., V.E. Myer, and P.D. Hsu, *Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity.* Mol Cell, 2016. **63**(3): p. 355-70.

219. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nat Biotechnol, 2014. **32**(6): p. 577-582.

220. You, Y., et al., *Design of LNA probes that improve mismatch discrimination.* Nucleic Acids Res, 2006. **34**(8): p. e60.

221. Vester, B. and J. Wengel, *LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA.* Biochemistry, 2004. **43**(42): p. 13233-41.

222. Elmen, J., et al., *Locked nucleic acid (LNA) mediated improvements in siRNA stability and functionality.* Nucleic Acids Res, 2005. **33**(1): p. 439-47.

223. Rahman, S.M., et al., *Design, synthesis, and properties of 2',4'-BNA(NC): a bridged nucleic acid analogue.* J Am Chem Soc, 2008. **130**(14): p. 4886-96.

224. Obika, S., et al., *2'-O,4'-C-Methylene bridged nucleic acid (2',4'-BNA): synthesis and triplex-forming properties.* Bioorg Med Chem, 2001. **9**(4): p. 1001-11.

225. Gotfredsen, C.H., P. Schultze, and J. Fegon, *Solution Structure of an Intramolecular Pyrimidine−Purine−Pyrimidine Triplex Containing an RNA Third Strand.* J. Am. Chem. Soc., 1998. **120**(18): p. 4281-4289.

226. Swayze, E.E., et al., *Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals.* Nucleic Acids Res, 2007. **35**(2): p. 687-700.

227. Manning, K.S., et al., *BNA(NC) Gapmers Revert Splicing and Reduce RNA Foci with Low Toxicity in Myotonic Dystrophy Cells.* ACS Chem Biol, 2017. **12**(10): p. 2503-2509.

228. Kim, D., et al., *Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq.* Genome Res, 2016. **26**(3): p. 406-15.

229. Christie, K.A., et al., *Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders.* Sci Rep, 2017. **7**(1): p. 16174.

230. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems.* Nat Biotechnol, 2013. **31**(3): p. 233-9.

231. Guschin, D.Y., et al., *A rapid and general assay for monitoring endogenous gene modification.* Methods Mol Biol, 2010. **649**: p. 247-56.

232. Lim, Y., et al., *Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease.* Nat Commun, 2016. **7**: p. 13350.

233. Klein, M., et al., *Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules.* Cell Rep, 2018. **22**(6): p. 1413-1423.

234. Anders, C., O. Niewoehner, and M. Jinek, *In Vitro Reconstitution and Crystallization of Cas9 Endonuclease Bound to a Guide RNA and a DNA Target.* Methods Enzymol, 2015. **558**: p. 515-537.

235. Wang, X., et al., *CRISPR-DAV: CRISPR NGS data analysis and visualization pipeline.* Bioinformatics, 2017. **33**(23): p. 3811-3812.

236. Park, J., et al., *Cas-analyzer: an online tool for assessing genome editing results using NGS data.* Bioinformatics, 2017. **33**(2): p. 286-288.

237. Ali, M.A.M., et al., *Matrix metalloproteinase-2 mediates ribosomal RNA transcription by cleaving nucleolar histones.* FEBS J, 2021. **288**(23): p. 6736-6751.

238. Boukouris, A.E., et al., *A reversible metabolic stress-sensitive regulation of CRMP2A orchestrates EMT/stemness and increases metastatic potential in cancer.* Cell Rep, 2022. **38**(11): p. 110511.

239. Bhullar, K.S., et al., *Tripeptide IRW Upregulates NAMPT Protein Levels in Cells and Obese C57BL/6J Mice.* J Agric Food Chem, 2021. **69**(5): p. 1555-1566.

240. Kerek, E.M., C.R. Cromwell, and B.P. Hubbard, *Identification of Drug Resistance Genes Using a Pooled Lentiviral CRISPR/Cas9 Screening Approach.* Methods Mol Biol, 2021. **2381**: p. 227-242.

241. Tu, Z., et al., *Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos.* Sci Rep, 2017. **7**: p. 42081.

242. Cromwell, C.R. and B.P. Hubbard, *In Vitro Assays for Comparing the Specificity of First- and Next-Generation CRISPR/Cas9 Systems.* Crispr Guide Rna Design, 2021. **2162**: p. 215-232.

243. Krysler, A.R., et al., *Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences.* Nat Commun, 2022. **13**(1): p. 1617.

244. Watts, J.K. and M.J. Damha, *2′ F-Arabinonucleic acids (2′ F-ANA)—History, properties, and new frontiers.* Canadian Journal of Chemistry, 2008. **86**(7): p. 641-656.

245. Palermo, G., et al., *Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations.* ACS Cent Sci, 2016. **2**(10): p. 756-763.

246. Anosova, I., et al., *The structural diversity of artificial genetic polymers.* Nucleic Acids Res, 2016. **44**(3): p. 1007-21.

247. Zuo, Z., et al., *Rational Engineering of CRISPR-Cas9 Nuclease to Attenuate Position-Dependent Off-Target Effects.* CRISPR J, 2022. **5**(2): p. 329-340.

248. Liu, M.S., et al., *Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA.* Nat Commun, 2020. **11**(1): p. 3576.

249. Anders, C. and M. Jinek, *In vitro enzymology of Cas9.* Methods Enzymol, 2014. **546**: p. 1-20.

250. Kerek, E.M., C.R. Cromwell, and B.P. Hubbard, *CRISPR Lights up In Situ Protein Evolution.* Cell Chemical Biology, 2020. **27**(5): p. 475-478.

251. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-77.

252. Moon, S.B., et al., *Recent advances in the CRISPR genome editing tool set.* Exp Mol Med, 2019. **51**(11): p. 1-11.

253. Alok, A., et al., *The Rise of the CRISPR/Cpf1 System for Efficient Genome Editing in Plants.* Front Plant Sci, 2020. **11**: p. 264.

254. Nakade, S., T. Yamamoto, and T. Sakuma, *Cas9, Cpf1 and C2c1/2/3-What's next?* Bioengineered, 2017. **8**(3): p. 265-273.

255. Canver, M.C., et al., *Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci.* Nat Genet, 2017. **49**(4): p. 625-634.

256. Lindpaintner, K., *Genetics in drug discovery and development: challenge and promise of individualizing treatment in common complex diseases.* Br Med Bull, 1999. **55**(2): p. 471-91.

257. Lessard, S., et al., *Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci.* Proc Natl Acad Sci U S A, 2017. **114**(52): p. E11257-E11266.

258. Cuevas, J.M., et al., *Extremely High Mutation Rate of HIV-1 In Vivo.* PLoS Biol, 2015. **13**(9): p. e1002251.

259. Darcis, G., et al., *The Impact of HIV-1 Genetic Diversity on CRISPR-Cas9 Antiviral Activity and Viral Escape.* Viruses, 2019. **11**(3).

260. Wang, Z., et al., *CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape.* Cell Rep, 2016. **15**(3): p. 481-489.

261. De Silva Feelixge, H.S., et al., *CRISPR/Cas9 and Genome Editing for Viral Disease-Is Resistance Futile?* ACS Infect Dis, 2018. **4**(6): p. 871-880.

262. Watkins, N.E., Jr. and J. SantaLucia, Jr., *Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes.* Nucleic Acids Res, 2005. **33**(19): p. 6258-67.

263. Crick, F., *Codon--anticodon pairing: the wobble hypothesis.* 1966.

264. Clutterbuck, D.R., et al., *A bioinformatic screen for novel A-I RNA editing sites reveals recoding editing in BC10.* Bioinformatics, 2005. **21**(11): p. 2590-5.

265. Karran, P. and T. Lindahl, *Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus.* Biochemistry, 1980. **19**(26): p. 6005-11.

266. Tsourkas, A., M.A. Behlke, and G. Bao, *Hybridization of 2'-O-methyl and 2'-deoxy molecular beacons to RNA and DNA targets.* Nucleic Acids Res, 2003. **31**(6): p. 5168-74.

267. Loakes, D., et al., *3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR.* Nucleic Acids Res, 1995. **23**(13): p. 2361-6.

268. Gallego, J. and D. Loakes, *Solution structure and dynamics of DNA duplexes containing the universal base analogues 5-nitroindole and 5-nitroindole 3-carboxamide.* Nucleic Acids Res, 2007. **35**(9): p. 2904-12.

269. Vallone, P.M. and A.S. Benight, *Melting studies of short DNA hairpins containing the universal base 5-nitroindole.* Nucleic Acids Res, 1999. **27**(17): p. 3589-96.

270. Lin, P.K. and D.M. Brown, *Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues.* Nucleic Acids Res, 1989. **17**(24): p. 10373-83.

271. Lin, P.K. and D.M. Brown, *Synthesis of oligodeoxyribonucleotides containing degenerate bases and their use as primers in the polymerase chain reaction.* Nucleic Acids Res, 1992. **20**(19): p. 5149-52.

272. Yip, S.P., *Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles.* Blood, 2000. **95**(4): p. 1487-92.

273. Ivanov, I.E., et al., *Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling.* Proc Natl Acad Sci U S A, 2020. **117**(11): p. 5853-5860.

274. Ramakrishna, S., et al., *Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations.* Nat Commun, 2014. **5**: p. 3378.

275. Howe, K.L., et al., *Ensembl 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D884-D891.

276. Lin, Y.C., et al., *Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations.* Nat Commun, 2014. **5**: p. 4767.

277. Bai, J., et al., *Cas12a-Based On-Site and Rapid Nucleic Acid Detection of African Swine Fever.* Front Microbiol, 2019. **10**: p. 2830.

278. Broughton, J.P., et al., *CRISPR-Cas12-based detection of SARS-CoV-2.* Nat Biotechnol, 2020. **38**(7): p. 870-874.

279. Shafer, R.W., et al., *Human immunodeficiency virus reverse transcriptase and protease sequence database.* Nucleic Acids Res, 2000. **28**(1): p. 346-8.

280. Shafer, R.W., *Rationale and uses of a public HIV drug-resistance database.* J Infect Dis, 2006. **194 Suppl 1**: p. S51-8.

281. Wang, G., et al., *CRISPR-Cas9 can inhibit HIV-1 replication but NHEJ repair facilitates virus escape.* Molecular Therapy, 2016. **24**(3): p. 522-526.

282. Phan, Q.V., et al., *Site-specific chromosomal gene insertion: Flp recombinase versus Cas9 nuclease.* Sci Rep, 2017. **7**(1): p. 17771.

283. Christopherson, C., J. Sninsky, and S. Kwok, *The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies.* Nucleic Acids Res, 1997. **25**(3): p. 654-8.

284. Wright, D.J., C.R. Force, and B.M. Znosko, *Stability of RNA duplexes containing inosine.cytosine pairs.* Nucleic Acids Res, 2018. **46**(22): p. 12099-12108.

285. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

286. Duffy, S., *Why are RNA virus mutation rates so damn high?* PLoS Biol, 2018. **16**(8): p. e3000003.

287. Teng, F., et al., *Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds.* Genome Biol, 2019. **20**(1): p. 15.

288. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.

289. Kim, H., et al., *Surrogate reporters for enrichment of cells with nuclease-induced mutations.* Nat Methods, 2011. **8**(11): p. 941-3.

290. Zetsche, B., et al., *Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array.* Nat Biotechnol, 2017. **35**(1): p. 31-34.

291. Zsogon, A., et al., *De novo domestication of wild tomato using genome editing.* Nat Biotechnol, 2018.

292. Kaminski, M.M., et al., *CRISPR-based diagnostics.* Nat Biomed Eng, 2021. **5**(7): p. 643-656.

293. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-4.

294. Gantz, V.M., et al., *Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi.* Proc Natl Acad Sci U S A, 2015. **112**(49): p. E6736-43.

295. DiCarlo, J.E., et al., *Safeguarding CRISPR-Cas9 gene drives in yeast.* Nat Biotechnol, 2015. **33**(12): p. 1250-1255.

296. Hammond, A., et al., *A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae.* Nat Biotechnol, 2016. **34**(1): p. 78-83.

297. Komatsu, A., et al., *Production of Herbicide-Sensitive Strain to Prevent Volunteer Rice Infestation Using a CRISPR-Cas9 Cytidine Deaminase Fusion.* Front Plant Sci, 2020. **11**: p. 925.

298. Gao, H., et al., *Complex Trait Loci in Maize Enabled by CRISPR-Cas9 Mediated Gene Insertion.* Front Plant Sci, 2020. **11**: p. 535.

299. Gonzalez, M.N., et al., *Reduced Enzymatic Browning in Potato Tubers by Specific Editing of a Polyphenol Oxidase Gene via Ribonucleoprotein Complexes Delivery of the CRISPR/Cas9 System.* Front Plant Sci, 2019. **10**: p. 1649.

300. Rasul, M.F., et al., *Strategies to overcome the main challenges of the use of CRISPR/Cas9 as a replacement for cancer therapy.* Mol Cancer, 2022. **21**(1): p. 64.

301. Charlesworth, C.T., et al., *Identification of preexisting adaptive immunity to Cas9 proteins in humans.* Nat Med, 2019. **25**(2): p. 249-254.

302. Crudele, J.M. and J.S. Chamberlain, *Cas9 immunity creates challenges for CRISPR gene editing therapies.* Nat Commun, 2018. **9**(1): p. 3497.

303. Ihry, R.J., et al., *p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.* Nat Med, 2018. **24**(7): p. 939-946.

304. Enache, O.M., et al., *Cas9 activates the p53 pathway and selects for p53-inactivating mutations.* Nat Genet, 2020. **52**(7): p. 662-668.

305. Lin, J. and K.C. Wong, *Off-target predictions in CRISPR-Cas9 gene editing using deep learning.* Bioinformatics, 2018. **34**(17): p. i656-i663.

306. Daer, R.M., et al., *The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells.* ACS Synth Biol, 2017. **6**(3): p. 428-438.

307. Kuscu, C., et al., *Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.* Nat Biotechnol, 2014. **32**(7): p. 677-83.

308. Liu, M., et al., *Global detection of DNA repair outcomes induced by CRISPR-Cas9.* Nucleic Acids Res, 2021. **49**(15): p. 8732-8742.

309. Hoijer, I., et al., *CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations.* Nat Commun, 2022. **13**(1): p. 627.

310. Zuccaro, M.V., et al., *Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos.* Cell, 2020. **183**(6): p. 1650-1664 e15.

311. Papathanasiou, S., et al., *Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing.* Nat Commun, 2021. **12**(1): p. 5855.

312. Haapaniemi, E., et al., *CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response.* Nat Med, 2018. **24**(7): p. 927-930.

313. Christiansen, J.K., et al., *LNA nucleotides improve cleavage efficiency of singular and binary hammerhead ribozymes.* Bioorg Med Chem, 2007. **15**(18): p. 6135-43.

314. Karlsen, K.K. and J. Wengel, *Locked nucleic acid and aptamers.* Nucleic Acid Ther, 2012. **22**(6): p. 366-70.

315. Liu, Q., et al., *Deep learning improves the ability of sgRNA off-target propensity prediction.* BMC Bioinformatics, 2020. **21**(1): p. 51.

316. Chuai, G., et al., *DeepCRISPR: optimized CRISPR guide RNA design by deep learning.* Genome Biol, 2018. **19**(1): p. 80.

317. Rahman, M.K. and M.S. Rahman, *CRISPRpred: a flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems.* PloS one, 2017. **12**(8): p. e0181943.

318. Hwang, G.H., J.S. Kim, and S. Bae, *Web-Based CRISPR Toolkits: Cas-OFFinder, Cas-Designer, and Cas-Analyzer.* Methods Mol Biol, 2021. **2162**: p. 23-33.

319. Vos, P.D., et al., *Computationally designed hyperactive Cas9 enzymes.* Nat Commun, 2022. **13**(1): p. 3023.

320. Palaz, F., et al., *CRISPR-Cas13 System as a Promising and Versatile Tool for Cancer Diagnosis, Therapy, and Research.* ACS Synth Biol, 2021. **10**(6): p. 1245-1267.

321. Semikolenova, O., et al., *Photoactivatable nanoCRISPR/Cas9 System Based on crRNA Reversibly Immobilized on Carbon Nanoparticles.* Int J Mol Sci, 2021. **22**(20).

322. Taemaitree, L., et al., *An artificial triazole backbone linkage provides a split-and-click strategy to bioactive chemically modified CRISPR sgRNA.* Nat Commun, 2019. **10**(1): p. 1610.

323. Park, H., et al., *CRISPR-Click Enables Dual-Gene Editing with Modular Synthetic sgRNAs.* Bioconjug Chem, 2022. **33**(5): p. 858-868.

324. Schubert, M.S., et al., *Chemical Modification of CRISPR gRNAs Eliminate type I Interferon Responses in Human Peripheral Blood Mononuclear Cells.* J Cytokine Biol, 2018. **3**(1).