Using a chemically-controlled CRISPR/Cas9 system to understand and develop new genome engineering technologies

Cindy T. Wei

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Dustin Maly, Chair

Douglas Fowler, Chair

David Shechner

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Using a chemically-controlled CRISPR/Cas9 system to understand and develop new genome engineering technologies

Cindy T. Wei

Chairs of the Supervisory Committee:

Dustin Maly

Department of Chemistry

Douglas Fowler

Department of Genome Sciences

Clustered regularly interspaced short palindromic repeat (CRISPR) systems have revolutionized our ability to investigate the genotype-phenotype relationships of specific genetic elements. The class 2 type II CRISPR system, involving a Cas9 endonuclease, has been widely adopted to aid in making specific genomic DNA changes. Precise DNA targeting by the CRISPR/Cas9 system is achieved using an RNA molecule that encodes a 20 nucleotide (nt) sequence complementary to the target site. Cas9 can be targeted to different loci in the genome by simply changing the 20 nt RNA-encoded sequence. Cas9 can then create DNA double-strand breaks (DSBs) at the target site and induce DNA repair to incorporate a specific DNA edit or uncontrolled insertions and deletions to knock out a gene. The CRISPR/Cas9 system has also been adapted to create a precise DNA targeting module that can recruit different DNA effector systems, such as transcriptional activators, DNA deaminases, and histone modifiers. While CRISPR/Cas9 has

enabled new insights into genotype-phenotype relationships, challenges remain with the formation of unwanted edits, such as off-target edits or bystander edits with base editor systems. Furthermore, there is a lack of a generalizable method to create temporally-controlled Cas9-based effector systems to allow investigation of temporally-regulated genetic elements. Here, I use a chemically-inducible Cas9 (ciCas9) to explore the *in vivo* mechanisms of Cas9 off-target editing and to develop a generalizable system to confer temporal control over a variety of Cas9-based effector systems. Using engineered chemically-controlled base editors, I dissected the kinetics of bystander editing and how base editing at one nucleotide influences subsequent base edits within the same target site. I envision the results presented here could be used to inform future efforts to study temporally-regulated genetic elements and to engineer more efficient and accurate Cas9-based genome engineering systems.

Table of Contents

List of figures

List of tables

Acknowledgements

My Ph.D. journey would not have been possible without the encouragement from my incredible support network. Every person I have had the pleasure to work with, learn from, and interact with has been instrumental in making this work possible.

First, I want to thank my Ph.D. advisors, Dustin Maly and Doug Fowler, for their support, encouragement, and patience throughout my entire graduate career. Doug and Dusty taught me how to be a curious, creative, and vigorous scientist. I appreciate their efforts in pushing me outside my scientific comfort zone and for being my personal cheerleaders at times where I needed some positivity during my projects. Without Doug and Dusty, none of this work would have been possible and I would not be the scientist I am today.

My committee was full of wonderful scientists who provided interesting perspectives about my science, helped me get out of scientific rabbit holes, and helped me get unstuck many times with my projects. Jesse Zalatan, Dave Shechner, and Suzanne Hoppins always asked insightful questions that made me think deeper about my science. Suzanne has also been a pillar of support since I joined her lab as an undergraduate student. I greatly appreciated Suzanne's regular check-in emails and our non-science conversations throughout my graduate career.

Doug and Dusty also built labs full of excellent scientists, all of whom I had the pleasure of working with every day. The members of both the Fowler and Maly labs have created a supportive and fun lab environment which made going to work every day incredibly enjoyable. Glenna Foight, Emily Dieter, and Nick Popp were always willing to chat and listen to my frustrations at happy hours (also thank you for choosing happy hours that had non-beer options). Zack Potter, who I briefly had the pleasure of sharing a lab bay with, was always fun to chat with and made my 1000's of DNA extractions less monotonous.

My outside-of-lab friends have always been there to provide an escape from science throughout my graduate career. Aimme Zhang, Ternessa Cao, Tanner Pudden, and Tony Au, my fellow Lander RAs, were always up for fun conversations and dinner parties. Taylor Wang and

John Wang were my board game and trivia buddies who made sure I took breaks away from lab and my science.

Finally, my Ph.D. journey would not have been possible without my family. My husband, Roger Tu, was my rock and he supported me in every aspect of my life throughout my entire graduate career. Roger cooked meals, cleaned our house, and took care of countless responsibilities during the most stressful stretches of my Ph.D. My mental health would not be intact without Roger's endless love, patience, and support. My parents and brother who live 11,265 km (7000 miles) away have always been my cheerleaders and who have supported my scientific career from the very beginning.

It was truly my pleasure to have had so many wonderful people in my life that supported me not only through my Ph.D. journey but also through life in general. I cannot express enough gratitude to all of you for being a part of my life. Thank you.

Dedication

To all the teachers, professors, mentors, and advisors

who inspired me to become the scientist I am today.

Chapter 1: Introduction

The genome of eukaryotic organisms consists of DNA bases that encode the instructions for every detail in building an organism. Being able to make precise targeted modifications in genomes has been of great interest to be able to understand the function and genotype-phenotype relationships of different genetic elements. Direct editing of DNA bases at precise target sites is one method of probing the direct consequences of DNA sequence on the function of a specific genetic element. Furthermore, the ability to precisely edit DNA also has therapeutic potential by correcting pathogenic mutations that result in genetic diseases. The development of precise DNA targeting systems has revolutionized our ability to make these precise DNA modifications more efficiently within the genomes of organisms.

1.1 Genome editing using targeted nucleases

Precise genome editing has been achieved through the use of targeted endonucleases to aid in the incorporation of a desired DNA edit encoded on an exogenous DNA template or to knockout a specific gene using uncontrolled DNA repair[1]. The targeted endonucleases are able to create a double-stranded break (DSB) in the DNA that signals the cell's endogenous DNA repair machinery to repair the DSB, which could then result in a DNA edit. The DSB generated by the targeted nucleases can be repaired using homology-directed repair (HDR) or through non-homologous end joining (NHEJ). HDR can result in a perfect edit by using the sister chromatid as a repair template or be exploited to incorporate a desired DNA edit into the target site using an exogenous DNA template encoding the desired edit[2]. This exogenous repair template encoding the desired edit also contains sequences homologous to the regions 5´ and 3´ of the cut site to facilitate HDR. NHEJ can result in uncontrolled insertions and deletions (indels) at the cut site and is often used to create frameshift mutations in the targeted gene and thus knocking out protein

expression[3]. The selection of DNA repair pathways depends on several cellular factors such as cell cycle stage, cell type, and the type of DSB[4].

Historically, precise targeting of DNA sequences and formation of DSBs has been achieved through the use of zinc finger proteins (ZFPs) and transcription activator-like effectors (TALEs)[5]. ZFPs and TALEs are commonly fused to a FokI endonuclease domain which then allow specific DSB generation at genome target sites defined by ZFPs and TALEs. ZFPs use modular protein domains each of which are able to recognize a 3 base pair (bp) DNA sequence[6]. Fusion of these modular ZFP domains can then allow more targeting specificity with recognition of longer DNA sequences. TALEs use a similar modular method where each protein module is able to recognize a single base pair[7]. The single base pair recognition by TALEs has expanded the sequence targeting possibilities compared to ZFPs. Both ZFPs and TALEs have allowed scientists to make precise DSBs at a predefined target site that is recognized using amino acid interactions with DNA bases.

Although ZFPs and TALEs have improved the ability to target and edit specific DNA sequences in genomes, they are limited in their sequence targeting scope and can be quite difficult to engineer. ZFPs recognize a 3 bp sequence which limits the sequences that can be targeted[6]. TALEs also pose a technical challenge to cloning due to their identical repeat sequences[5]. Furthermore, in both ZFP and TALE systems, targeting different sites in the genome requires re-engineering of a new set of proteins which can be laborious. This has hindered wide adoption of these systems by scientists for creating specific DNA edits.

The recent discovery of CRISPR/Cas systems has revolutionized our ability to target and edit specific DNA sequences in genomes. The CRISPR/Cas systems overcome many of the challenges posed by the ZFP and TALE systems. The CRISPR/Cas system consists of Cas proteins that are able to bind and cleave nucleic acids that are directed to specific target sequences using an RNA molecule[8]. The programmability of redirecting Cas endonucleases by simply changing the target sequence within the RNA molecule has resulted in wide adoption of

these systems by scientists for basic science research and for therapeutic uses to correct genetic diseases.

1.2 The CRISPR/Cas9 system for mammalian genome editing

CRISPR/Cas systems are found in nature as part of an adaptive immune system in bacteria and archaea[8]. There are two classes of CRISPR/Cas systems that are both used to cleave foreign nucleic acids upon infection and to retain a record of these nucleic acid sequences for future infections. The class I CRISPR/Cas systems consists of multiprotein component systems that enables recognition of nucleic acid sequences using a CRISPR RNA (crRNA) and subsequent cleavage of nucleic acids using a Cas3 endonuclease[9]. The class II CRISPR/Cas systems use a single Cas protein that is capable of using a crRNA to find the target nucleic acid sequence and also cleave the nucleic acid[9]. The class II CRISPR/Cas systems have been widely adapted for programmable genome engineering due to its simplicity[10]. The Cas9 and Cas12 RNA-guided endonuclease systems have been adapted for targeted DNA modifications and the Cas13 system has also been adapted for targeted RNA modifications. The CRISPR/Cas genome editing toolkit has been demonstrated in a wide variety of organisms and has been used in a plethora of biotechnology applications[11–13].

1.2.1 CRISPR/Cas9 DNA targeting and cleavage mechanisms

The CRISPR/Cas9 system is a class II CRISPR system that generates DSBs in target DNA sequences using a single Cas9 endonuclease guided by RNA molecules[14]. Target search by Cas9 begins with a protospacer-adjacent motif (PAM) which is a DNA sequence immediately 3′ of the 20 nucleotide (nt) targeting sequence[15]. Depending on the species where the CRISPR/Cas9 originated from, this PAM sequence can vary[9]. After Cas9 binding to a PAM sequence, in nature Cas9 is directed to a specified target site using two RNA molecules, the crRNA which encodes a 20 nt targeting sequence—the “protospacer” sequence—and a *trans*-

activating RNA (tracrRNA) which pairs with the crRNA and allows for complex formation with the Cas9 endonuclease[16]. To simplify the system, most DNA editing applications use a single guide RNA (sgRNA) which is a fusion of the crRNA and tracrRNA into a single RNA molecule that can simultaneously bind to Cas9 and direct it to a specific DNA target site[15]. The most commonly used CRISPR/Cas9 system for genome engineering in mammalian systems was initially discovered in *Streptococcus pyogenes* and is the CRISPR/Cas9 system described from this point forward.

Cas9 has a bi-lobed structure consisting of the REC lobe that binds the sgRNA and the NUC lobe that contains the nuclease domains and PAM-interacting domain (**Fig. 1.1**)[14]. Apo-Cas9 is a highly dynamic protein that is stabilized upon binding of an sgRNA in the REC lobe[17]. The PAM for *S. pyogenes* Cas9 is a 3 bp 5´-NGG-3´ sequence that is recognized directly by two arginines at amino acid positions 1333 and 1335 (R1333 and R1335, respectively) in the PAM-interacting domain of the Cas9 endonuclease[14]. Once Cas9 has bound to the PAM sequence, it will then unwind the DNA double helix and form a DNA:RNA heteroduplex between the sgRNA protospacer sequence and the target DNA strand. Unwinding and DNA:RNA heteroduplex formation proceeds in a single direction from immediately adjacent to the PAM towards the PAM-distal 5´ end of the sgRNA protospacer sequence[18]. Based on *in vitro* biochemical studies, unwinding of DNA and DNA:RNA heteroduplex formation has been determined as the rate limiting step of Cas9 activity[19]. If Cas9 does not find a complementary DNA sequence to the sgRNA spacer, then Cas9 dissociates rapidly from DNA[20]. Once Cas9 has found the correct target site through sgRNA spacer hybridization with target DNA, Cas9 undergoes several conformational changes to position its two nuclease domains next to the scissile phosphates[21].

Cas9 contains two nuclease domains that each individually targets and cleaves a different strand of DNA. The HNH nuclease domain is highly dynamic and has to rotate nearly 180° to position the catalytic residue next to the phosphate backbone that will be cleaved[22]. The HNH nuclease domain uses a single Mg^{2+} to aid in cleavage of the target DNA strand with a key catalytic histidine at amino acid 840 (H840)[17]. The RuvC nuclease domain is less dynamic but its

nuclease activity is allosterically controlled by conformational changes in the HNH domain[22]. The RuvC nuclease cleaves the non-target DNA strand through a two-Mg^{2+} mechanism with a key catalytic aspartate at amino acid 10 (D10)[17]. Mutation of both H840 and D10 into alanine is able to create a catalytically-dead Cas9 (dCas9) that is only capable of binding to a specified DNA target site without creating a DSB.

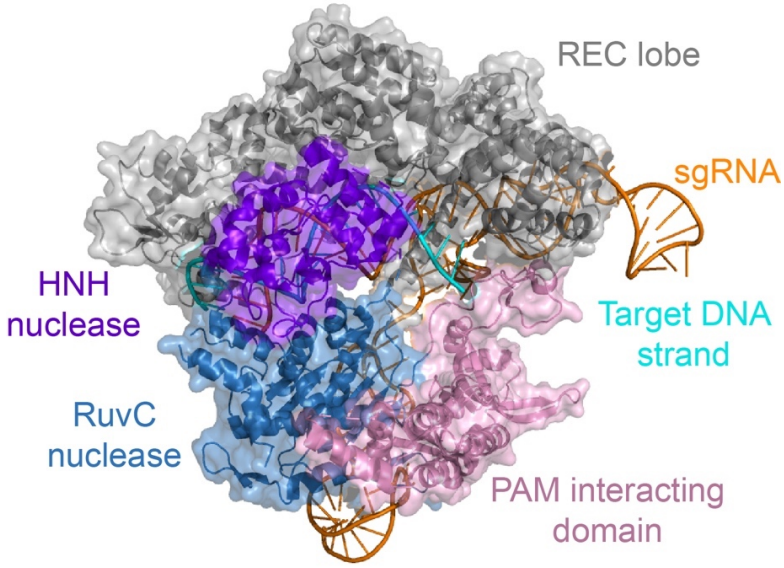

Figure 1.1. Structure of the Cas9 endonuclease bound to sgRNA and the target DNA strand. Crystal structure of the Cas9 endonuclease bound to sgRNA and the target DNA strand (PDB: 4OO8[17]). The REC lobe containing the RECI, RECII, and RECIII domains are colored in gray. The NUC lobe consisting of the HNH nuclease, RuvC nuclease, and PAM interacting domains are colored in purple, blue, and pink, respectively. The sgRNA is colored in orange and the target DNA strand is colored in aqua.

After DNA cleavage, Cas9 is thought to dissociate very slowly from DNA, thus Cas9 is , under most biologically-relevant reaction conditions, essentially a single-turnover enzyme. *In vitro* studies with purified Cas9 proteins and substrate DNA showed that Cas9 is unable to dissociate from DNA unless denatured with high concentrations of urea[18]. However, active transcription by RNA polymerase at the target site can convert Cas9 into a multi-turnover enzyme[23]. Although Cas9 dissociation is slow after DNA cleavage, Cas9 edits in a unique stoichiometric scheme where there are significantly more molecules of Cas9 protein than there are target sites in the genome for a single cell. Therefore, the single-turnover characteristic of Cas9 should not hinder

DSB formation at the intended target site. Furthermore, this means that Cas9 also does not follow Michaelis-Menten kinetics due to limited substrate availability.

The CRISPR/Cas9 system is easily adaptable to create DSBs at any location in a genome by simply changing the 20 nt protospacer sequence in the sgRNA. Furthermore, the ease of engineering catalytically-dead Cas9 or a "nickase" Cas9 that only cleaves one DNA strand has allowed for wide adaptation of the system to engineer a variety of DNA engineering tools. As a result of its simplicity and flexibility, the CRISPR/Cas9 system is currently the most used genome editing system that is highly programmable and easy to deliver to a variety of cell types and organisms.

1.2.2 Off-target genome editing by the CRISPR/Cas9 system

Although the CRISPR/Cas9 system has revolutionized the ease of targeted genome editing, challenges remain with the targeting specificity that often leads to unwanted off-target DNA editing. Off-target DNA editing poses a significant problem particularly in therapeutic applications of CRISPR/Cas9 where unwanted indel formation could be deleterious. Cas9 relies on the 20 nt sequence in the sgRNA that often lacks enough complexity to only target a single site in a genome. Mismatches within the target sequence are also often tolerated with several mismatches in the PAM-distal region still allowing DSB generation to occur[24–26]. *In vivo* genome-wide investigations into Cas9 off-target sites have revealed that these sites often contain only a handful of mismatches compared to the intended on-target site[27]. Furthermore, for applications of Cas9 that don't require DSBs, Cas9 DNA binding only relies on complementarity of 5-8 bp from the PAM which increases the number of sites that could be bound in the genome[28]. These determinants of Cas9 off-target DNA editing have been extensively investigated *in vivo* and *in vitro* to understand the mechanisms of off-target DNA editing by Cas9.

DNA binding and cleavage by Cas9 is primarily determined by the number and location of mismatches along the DNA:RNA heteroduplex. PAM mismatches are rarely tolerated as PAM

recognition by Cas9 is essential to DNA unwinding for target search[14,18]. Early studies using a few target site sequences found that mismatches were less tolerated in the 8-12 seed region as this disrupts overall Cas9 binding to a target site[24–26]. Whereas mismatches in the PAM-distal region do not disrupt DNA binding but instead disrupt DNA cleavage by Cas9. A recent *in vitro* biochemical study characterizing Cas9 binding and cleavage at thousands of mismatched target sites found similar patterns where the number and location of mismatches determined the impact on Cas9 binding versus Cas9 cleavage[29]. Target sites with one or two mismatches in the DNA:RNA heteroduplex had very little effect on the fraction of DNA bound and the fraction of DNA cleaved by Cas9 compared to a perfectly matched target sequence. Mismatches only had a modest impact on DNA binding when the mismatch was in the seed region. Other kinetic measurements of Cas9 activity have also found that multiple PAM-distal mismatches do not disrupt DNA unwinding, but instead disrupt conformational movement of Cas9 domains to form a cleavage competent state[30–32]. These studies have provided insight into the DNA sequence-based effects on Cas9 binding and cleavage mechanisms at off-target sites.

Prolonged exposure of active Cas9 to the genome as a result of using plasmid DNA to express the DNA editing components has also been hypothesized as a cause for off-target DNA editing[24,33,34]. Different approaches have been used to decrease the amount and exposure time of active Cas9 in cells. One study diluted the amount of Cas9 plasmid transfected into cells, which resulted in improved Cas9 editing specificity[24]. However, specificity was only improved with very low amounts of Cas9 plasmid which also resulted in large decreases in on-target editing magnitudes that are not useful for most DNA editing applications. Another study explored limiting exposure of active Cas9 to the genome by timing the introduction of a small anti-CRISPR protein, AcrIIA4[35]. Timed introduction of AcrIIA4 by plasmid transfection or protein nucleofection into cells undergoing active Cas9-mediated DNA editing resulted in decreased off-target editing. Both approaches can mitigate off-target editing activity by limiting the exposure of active Cas9 to the genome, albeit impacting intended on-target editing efficiency at the same time.

Introduction of Cas9 into cells as a ribonucleoprotein (RNP) using electroporation is another approach commonly used to decrease Cas9 exposure, thus decreasing off-target editing[33]. Since Cas9 is not introduced as a DNA plasmid that allows constant expression, the amount of active Cas9 in cells rapidly declines after electroporation. Thus, it is hypothesized that the Cas9 RNP can rapidly cleave on-target sites and is degraded before off-target cleavage occurs. Cas9 RNPs are now commonly used for CRISPR/Cas9 DNA editing applications due to its ease of introduction into cells, high on-target editing efficiency, and lower off-target editing outcomes.

Truncated sgRNAs (tru-sgRNAs) have also been used to mitigate off-target editing activities by Cas9[36]. sgRNAs with a protospacer as few as 17 nt have been shown to be able to create DNA edits with Cas9. At the same time, tru-sgRNAs exhibited decreased off-target editing activity. It is hypothesized that the shorter protospacer sequence allows Cas9 to be more sensitive to mismatches by reducing the binding affinity of Cas9:sgRNA for the DNA target. The tru-sgRNA would allow for more rapid rewinding of DNA and thus faster dissociation of the Cas9:tru-sgRNA complex. Tru-sgRNAs offer a simple method of mitigating off-target editing activities. However, they require empirical screening to find a length that induces sufficient on-target editing with reduced off-target editing.

Other approaches have also been employed to assess Cas9 off-target editing and inform target sequence selection. Unbiased genome-wide off-target enrichment coupled with next-generation sequencing methods have been developed to identify off-target sites corresponding to an sgRNA[27,37,38]. Data from these unbiased techniques have informed development of computational tools to predict off-target editing activities for a given sgRNA[39,40]. Structural and mechanistic information about Cas9 have been used to engineer high-fidelity variants of Cas9 that are able to better discriminate off-target DNA sequences, thus leading to less off-target DNA editing[30,41,42].

1.2.3 Improving genome editing specificity using high-fidelity Cas9 variants

Engineering to create high fidelity Cas9s has resulted in variants that are able to better discriminate on- versus off-target sites in the genome, thus resulting in fewer off-target sites accumulating unwanted DNA edits. The first two high fidelity Cas9 variants, eSpCas9 and SpCas9-HF1, both contain mutations in domains that directly interact with the DNA backbone (**Fig. 1.2**)[41,43]. These mutations were introduced into DNA-interacting domains based on the assumption that Cas9 bound to off-target sites with much lower affinity compared to on-target sites. Thus, by weakening overall DNA binding, Cas9 would preferentially bind on-target DNA sequences more stably compared to off-target sequences. eSpCas9 contains mutations in amino acid residues that interact with the backbone of the non-target DNA strand[41]. Whereas SpCas9-HF1 contains mutations in residues that interact with the backbone of the target DNA strand[43]. Both eSpCas9 and SpCas9-HF1 demonstrated increased DNA editing specificity initially hypothesized to be a result of the altered DNA binding affinity for on- versus off-target sites.

Since the engineering of eSpCas9 and SpCas9-HF1, further biochemical characterization of these high fidelity variants have revealed that their off-target discrimination mechanism is due to enhancement of a conformational-based proofreading mechanism rather than decreasing overall DNA binding affinity[30]. Single molecule fluorescence resonance energy transfer (smFRET) experiments revealed that these high fidelity Cas9 variants have altered DNA cleavage efficiencies when Cas9 is bound to off-target sites[30,31,44]. This altered cleavage efficiency is based on adjusting a conformational gating proofreading mechanism involving the REC3 domain in Cas9 which senses mismatches in the PAM-distal region of the DNA:RNA heteroduplex[30]. If no PAM-distal mismatches are detected, REC3 movement causes a simultaneous conformational change in the REC2 domain that then allows the movement of the HNH nuclease domain to position the catalytic residues at the scissile phosphate. Elucidation of this proofreading mechanism led to the engineering of Hypa-Cas9 which contains mutations in the REC3 domain that enhances its ability to detect PAM-distal mismatches (**Fig. 1.2**)[30].

Although these high-fidelity Cas9 variants have mitigated some off-target editing effects, challenges remain with their editing efficiency at on-target sites and some off-target editing is still present[30,41,43]. A variety of approaches have been employed to rationally engineer and evolve even more efficient and specific high-fidelity variants. One study hypothesized that eSpCas9, SpCas9-HF1, and Hypa-Cas9 show inefficient on-target editing when used as an RNP due to over-engineering resulting in greatly diminished DNA cleavage activity[45]. As a result, HiFi-Cas9 was created using a single mutation that is adjacent to one of the Hypa-Cas9 mutations to reinstate efficient DNA cleavage but still maintained enhanced proofreading through the REC3 domain conformational gating mechanism (**Fig. 1.2**). Directed evolution has also been used to select for Cas9 variants that are efficient in on-target DNA cleavage but have diminished off-target DNA cleavage resulting in the development of Sniper-Cas9 and evoCas9 (**Fig. 1.2**)[46,47]. These newer high-fidelity Cas9 variants have slightly improved efficiency and specificity compared to eSpCas9, SpCas9-HF1, and Hypa-Cas9. However, work still remains to achieve a balance between on-target editing efficiency and diminished off-target editing activities. This requires a deeper understanding of Cas9 on- versus off-target discrimination mechanisms *in vivo* and the genomic context and cellular factors that impact Cas9 activity.

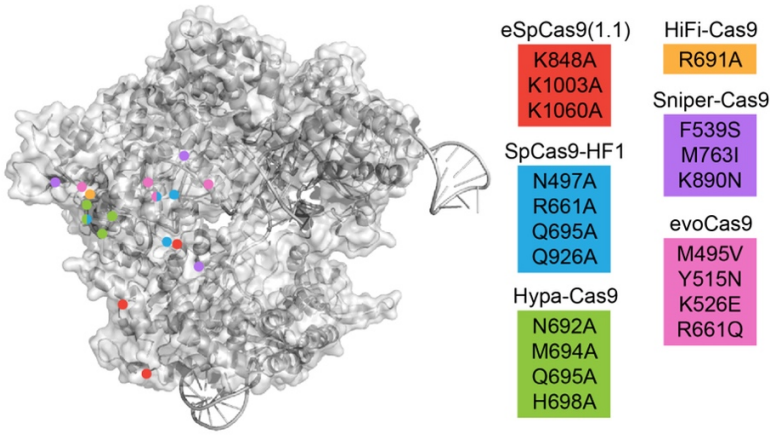

Figure 1.2. Location of high-fidelity mutations in Cas9.
Crystal structure of the Cas9 endonuclease bound to sgRNA and the target DNA strand (PDB: 4OO8[17]) colored in gray. The location of mutations corresponding to each high-fidelity Cas9 variant are represented by colored circles on the crystal structure.

1.3 Expansion of the CRISPR/Cas9-based genome editing toolbox

The wildtype Cas9 system has been widely used for DNA editing by creating a DSB at the target site encoded by the sgRNA and using endogenous HDR or NHEJ to introduce exact edits or indels at the DSB site, respectively. However, the wildtype Cas9 system has limitations that hinder its potential to make all possible DNA edits. In particular, Cas9 PAM requirements restrict the targetable sequences that can be edited. Furthermore, DSB creation at the target site often preferentially results in unwanted indels formed by NHEJ rather than the desired edit formed by HDR[48]. To overcome the need for a DSB to introduce exact DNA edits, the Cas9 endonuclease has been adapted to engineer targeted DNA editing systems that do not require DSB formation[11]. PAM-expanded Cas9 variants have also been engineered to allow Cas9 to recognize different PAM sequences and increase the number of targetable sequences in the genome[49]. These additional Cas9-based DNA editing systems have expanded the possible exact DNA edits that can be made without the potential of creating unwanted indels at the target site.

1.3.1 Increasing the targetable DNA sequences using PAM-expanded Cas9 variants

As Cas9 DNA binding target search begins with recognition of an NGG sequence through its PAM-interacting domain, accessible sites that can be edited are constrained to those immediately adjacent to a PAM sequence. Rational engineering of PAM-expanded variants have focused on introducing mutations that weaken PAM sequence binding in the PAM-interacting domain[42,50,51] or grafting PAM-interacting domains from other naturally occurring Cas9 systems to create a chimeric *S. pyogenes* Cas9[52]. Directed evolution approaches have resulted in Cas9 variants that contain mutations throughout the entire Cas9 endonuclease[53,54]. All of these engineering approaches have resulted in PAM-expanded Cas9 variants that are collectively able to target nearly any desired DNA sequence.

Targeted mutational approaches primarily focused on introducing mutations in the PAM-interacting domain to decrease specific DNA sequence interactions by Cas9. The first PAM-

expanded Cas9 variant was able to target NGA, NGAG, and NGCG PAM sequences due to mutations directly in the PAM-interacting amino acid residues[42]. A structural-based approach then created SpCas9-NG that contained similar mutations in the PAM-interacting amino acids which removed the requirement of the second G in the NGG PAM[50]. In addition to removing NGG-specific interactions, SpCas9-NG also introduced new non-base-specific interactions to compensate for the loss in Cas9:DNA binding. Thus allowing Cas9 to have less stringent PAM requirements but maintaining sufficient editing efficiency. A similar mutational approach was used to generate the near-PAMless SpRY-Cas9 variant that targets NRN/NYN PAM sequences[51]. SpRY-Cas9 also contains mutations that minimize Cas9 contact with specific PAM DNA bases and at the same time increases non-specific DNA binding interactions and improves DNA unwinding to compensate. These rationally engineered PAM-expanded variants have broadened the DNA targeting flexibility of Cas9 to allow editing at more target sites throughout the genome.

Directed evolution has also been successfully implemented to engineer PAM-expanded Cas9 variants. One directed evolution approach resulted in the Cas9-NRRH, Cas9-NRTH, and Cas9-NRCH variants which all contain mutations that are only located in the PAM-interacting domain[54]. Phage-assisted continuous evolution (PACE) was also used to create xCas9 which has greater activity at NGA and NGT PAM sequences[53]. Interestingly, xCas9 contains mutations throughout the entire Cas9 endonuclease with only a single mutation located in the PAM-interacting domain. The mechanism by which xCas9 can recognize alternative PAM sequences remains unknown. The presence of PAM-expanding mutations outside of the PAM-interacting domain suggests that initial DNA binding could rely on conformational changes throughout the entire Cas9 endonuclease.

Since engineering of PAM-expanded Cas9 variants has greatly broadened the targetable sequences within a genome, off-target editing becomes a greater concern when using these variants. To mitigate off-target editing, some PAM-expanding mutations could be combined with high fidelity mutations. For example, SpRY-Cas9 has been demonstrated to be compatible with

the high fidelity mutations in SpCas9-HF1, resulting in SpRY-HF1[51]. SpRY-HF1 showed elimination of nearly all off-target editing while maintaining non-NGG PAM target site editing. Further understanding of Cas9 targeting mechanisms *in vivo* could inform future rational engineering of a high fidelity fully-PAMless Cas9 variant.

1.3.2 Genome editing without DNA double-strand breaks using base editors

Since installing desired DNA edits at a precise location in a genome using Cas9-mediated DSB generation and endogenous HDR more often results in indel formation rather than the desired edit, base editors were developed to directly modify nucleotides to form DNA edits. Base editors are a fusion of a catalytically-impaired Cas9 and a DNA deaminase enzyme (**Fig. 1.3**). The base editor is targeted to a site in the genome using a Cas9 nickase (nCas9(D10A)) that can only cleave the target DNA strand at a specific site programmed with an sgRNA[55]. DNA deaminase enzymes fused to the nCas9(D10A) are then able to make specific DNA base changes by directly modifying the chemistry of the nucleotides.

Base editing involves several steps that require the activities of both the deaminase enzyme and nCas9(D10A) (**Fig. 1.3A**). nCas9(D10A) is important for several steps in the base editing process beyond target search and DNA binding programmed by the sgRNA. The deaminase enzyme is only able to deaminate bases on single-stranded DNA (ssDNA) substrates[55]. Thus, DNA:RNA heteroduplex formation by nCas9(D10A) is critical for base editing as it exposes the single-stranded non-target DNA strand which serves as the substrate for the deaminase enzymes. After base deamination, the ssDNA break created by nCas9(D10A) then induces the cell's endogenous base excision repair machinery to incorporate the base edit into both strands of DNA[55]. The ssDNA break activity of nCas9(D10A) has been shown to improve base editing outcomes compared to when dCas9 is fused to the deaminase enzymes[56]. As base editors are able to edit any available cytidine or adenine on the ssDNA substrate, they typically have an editing window that is defined as the positions of nucleotides that are edited to a certain

fraction or to greater magnitudes than at other positions[11]. The editing window of base editors are able to change based on the deaminase enzyme used and the Cas protein used to expose the ssDNA substrate[57–59]. Activities by both the deaminase enzyme and nCas9(D10A) are important to create desired DNA edits at specific target sites by base editing.

Currently, base editors are able to make all possible transition DNA base edits using naturally occurring cytidine deaminases and evolved adenine deaminases fused to nCas9(D10A). Cytidine base editors (CBEs) use a naturally occurring cytidine deaminase that creates cytidine-to-uridine base edits (**Fig. 1.3B**)[56]. Adenine base editors (ABEs) were then developed that used an evolved TadA deaminase (TadA*) that is able to form adenine-to-inosine base edits (**Fig. 1.3C**)[60]. DNA repair, induced by the nCas9(D10A), then converts the uridine to thymine and the inosine to guanine, thus resulting in cytidine to thymine (C-to-T) edits by CBEs and adenine to guanine (A-to-G) edits by ABEs[55]. Installation of the edit on the other strand is usually in competition with endogenous machinery that removes uridine and inosine bases from DNA[61]. Thus, CBEs often contain additional tandem uridine glycosylase inhibitor protein domains fused to nCas9(D10A) to inhibit the pathway that removes uridine bases from DNA[56,62]. Fusion of N-methylpurine DNA glycosylase to inhibit inosine removal after ABE editing did not significantly impact A-to-G editing efficiency[60]. Several different CBEs and ABEs have since been developed using alternative cytidine deaminase enzymes[63,64] or further evolved adenine deaminase enzymes that have greater editing efficiency[57,65]. Dual C-to-T and A-to-G base editors have also been engineered that are able to make both types of base edits simultaneously[66–68]. Recently, C-to-G base editors have also been reported which expand the DNA editing possibilities without the need for a DSB[69,70].

Base editing is able to create efficient DNA edits at targeted genomic sites with minimal indel formation. However, challenges remain with base editing that arise from the deaminase enzymes and also due to limitations with the nCas9(D10A) module. Since base editors are able to edit several nucleotides within an editing window, unintended editing at positions within the

target site could occur if the DNA target has multiple substrate nucleotides, termed bystander editing[11]. Bystander editing could be problematic if it leads to deleterious non-synonymous mutations in the target gene. To mitigate bystander editing, cytidine deaminases with narrower editing windows have been engineered and deaminases that are more sequence context selective can be used[58,71].

Since base editors use nCas9(D10A) to target the deaminase enzymes to specific sites in a genome, they are also limited by the NGG PAM sequence requirement. To overcome the limitation in targetable sequences, the deaminase enzymes have been coupled with engineered PAM-expanded nCas9(D10A) variants to sequences adjacent to non-NGG PAM sequences[51,53,54,72]. Base editing with Cas9 from other species with different PAM requirements has also been demonstrated[58,73]. This approach of using PAM-expanded variants coupled with an adenine deaminase has recently allowed for the correction of the Glu6Val sickle cell disease causing mutation[72].

Base editors also have Cas9-dependent and Cas9-independent off-target editing activities. Cas9-dependent off-target editing occurs at DNA off-target sites that are also edited by Cas9 endonuclease DSB generation activity[74–76]. Since nCas9(D10A) does not produce a DSB, indel formation occurs much less frequently at these off-target sites. Base editing at sgRNA-guided DNA off-target sites is also able to occur, though the off-target site must contain substrate adenine or cytidine nucleotides within the editing window for that particular deaminase for off-target editing to be detectable. These sgRNA-guided DNA off-target sites can be easily identified using genome-wide off-target detection methods commonly used with Cas9 endonuclease[27,37,38]. High-fidelity Cas9 variants have also been employed to mitigate Cas9-dependent off-target editing events[77]. Cas9-independent off-target DNA base editing has also been observed, but is rare[75,76]. These Cas9-independent base editing events often occur at DNA sites that transiently unwind and are accessible for enough time to allow deamination by the base editors. Cas9-independent off-target editing has been observed to occur more frequently with CBEs compared

to ABEs due to the faster deamination kinetics of cytidine deaminases[78,79]. Cytidine deaminases with lower Cas9-independent base editing activity have since been engineered to minimize this unwanted editing[80,81]. Finally, Cas9-independent off-target RNA editing has been observed with base editors as the deaminase enzymes were derived from RNA deaminase enzymes[82,83]. Rational engineering and directed evolution of the deaminase has led to new base editing enzymes that have high on-target DNA base editing activity and minimized off-target DNA and RNA base editing activity[57,65].

The development of base editing has vastly improved our ability to install desired edits into a specific location in the genome while minimizing unwanted editing. The combination of Cas9 variants and rationally engineered or evolved deaminase enzymes has created a variety of base editing tools with varying editing efficiencies and different base editing windows. The ability of base editors to create DNA edits without needing DSBs expands the potential of genome editing in scientific research and in therapeutic applications.

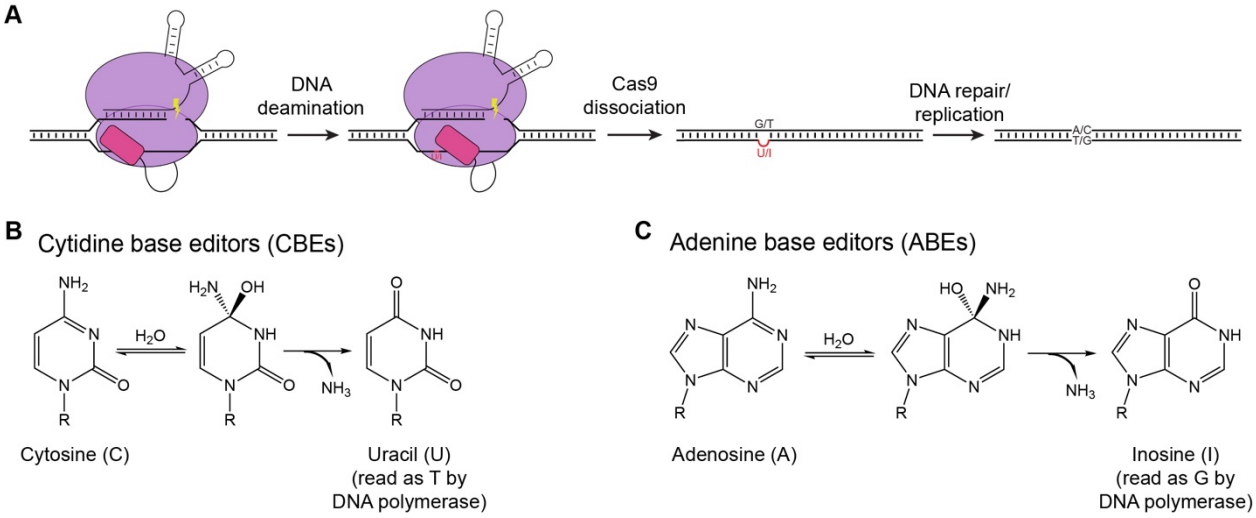

Figure 1.3. DNA base editing using Cas9 fusions to deaminase enzymes.
A) Schematic of the base editing process. Base editor binds to a target site specified by the sgRNA bound to nCas9(D10A). At the same time, nCas9(D10A) creates a ssDNA break in the target DNA strand. Then deamination by the fused deaminase enzyme creates a base edit in the exposed ssDNA. DNA repair, induced by the ssDNA break, or DNA replication causes the base edit to be installed into the other strand of DNA.
B) Cytidine deamination by cytidine base editors (CBEs).
C) Adenine deamination by adenine base editors (ABEs).

1.3.3 Prime editing with Cas9-reverse transcriptase fusion enzymes and RNA-encoded templates

Base editing allows for precise DNA editing with minimal unwanted indel formation and off-target effects. However, base editing is currently limited to DNA edits that can be performed by available deaminase enzymes which mostly only allow for transition base changes. Prime editing was recently developed to expand the variety of possible DNA edits while maintaining minimal unwanted indel formation and off-target effects (**Fig. 1.4**)[48]. Prime editing is able to install all possible transversion and transition point mutations, small insertions, and small deletions with minimal indel formation at the target site.

Prime editing relies on two components, a prime editor enzyme and a 3´ extended prime editing sgRNA (pegRNA) that encodes both the target site and the edit to be incorporated at the target site[48]. Prime editing also uses a catalyically-impaired Cas9 nickase (nCas9(H840A)) that is only able to create a ssDNA break in the non-target DNA strand. nCas9(H840A) is fused to a M-MLV reverse transcriptase which forms the PE2 enzyme. The extension of the pegRNA contains the desired DNA edit 3´ of the sgRNA scaffold sequence followed by a primer binding sequence that hybridizes to the non-target DNA strand upon nCas9(H840A) target site nicking.

Several coordinated steps are required during prime editing to successfully incorporate the desired edit into both DNA strands at the target site (**Fig. 1.4**). Upon target DNA binding, nCas9(H840A) first generates a ssDNA break which releases the non-target DNA strand[48]. The 3´ end of the cleaved non-target DNA strand then hybridizes to the primer binding sequence in the pegRNA. The 3´ end of the cleaved non-target DNA strand hybridized to the primer binding sequence can then serve as a priming sequence for the reverse transcriptase to incorporate the desired edit encoded on the pegRNA template. After reverse transcription, the newly generated DNA strand now creates a 3´ flap in addition to the original unedited DNA strand which forms a 5´ flap. 5´ flaps are preferentially excised by endogenous nucleases such as FEN1 or EXO1[84,85]. Whereas 3´ flaps are generally not the preferred substrate of these nucleases. Thus, the 5´ flap

is readily removed, resulting in one DNA strand containing the desired edit and the other DNA strand containing the unedited sequence. The edit is then incorporated into the unedited DNA strand using endogenous DNA repair machinery[48]. DNA repair to fully incorporate the edit into both DNA strands is aided by a separate nicking event of the unedited DNA strand by PE2 using an sgRNA targeting the edited sequence. This results in a nick in the unedited DNA strand and signals the DNA repair machinery to use the edited DNA strand as a repair template. Finally, this results in installation of the pegRNA-encoded edit into both strands of DNA without prolonged presence of a DSB.

The process of prime editing involves many steps and requires optimization of the pegRNA for every target site and every desired edit. Most of this optimization to enhance prime editing efficiency needs to be determined empirically. Optimization of the primer binding sequence is critical to achieving efficient editing as this determines the activity of the reverse transcriptase[48]. Current empirical evidence suggests a length of 8-15 nt is enough to achieve efficient prime editing. The reverse transcription template encoding the desired edit must also be empirically optimized to be a suitable length that allows efficient reverse transcription and incorporation into the DNA. The need to optimize both the primer binding sequence and the reverse transcription template complicates the application of prime editing.

Prime editing is a promising method that can incorporate a much wider range of desired edits into any targetable site in the genome. The use of nCas9(H840A) in prime editing also minimizes the formation of indels at the target site while achieving high levels of editing efficiency. Future systematic investigations to understand the factors impacting editing efficiency will allow for easier and wider use of the technology in research and therapeutic applications.

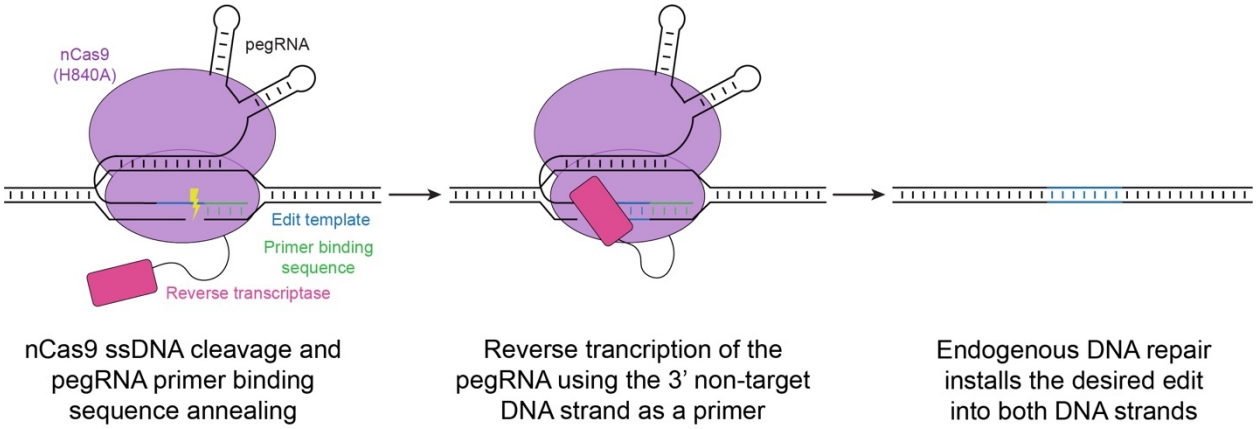

Figure 1.4. Schematic of prime editing.
Prime editing uses a nCas9(H840A) fused to a reverse transcriptase and a pegRNA. The nCas9(H840A) binds to the target site specified by the pegRNA and creates a ssDNA break in the non-target DNA strand. The primer binding sequence of the pegRNA anneals to the released 3´ end of the non-target ssDNA strand. Reverse transcription using the 3´ end of the non-target ssDNA strand then incorporates the edit encoded on the pegRNA into the DNA. Endogenous DNA repair then installs the edit into the other strand of DNA.

1.4 CRISPR/Cas9-based technologies for gene regulation

Since the discovery of the CRISPR/Cas9 system, researchers have applied the system to other aspects of genome engineering beyond applications that result in direct changes in DNA sequence. Most notable is the use of CRISPR/Cas9 to target specific genomic elements to edit the epigenome and regulate gene expression. In these applications, the CRISPR/Cas9 system is purely used for its programmable DNA targeting ability. Thus, the catalytically dead dCas9 is used in all of these systems to only allow for sgRNA-targeted DNA binding without DNA cleavage[13]. dCas9 is able to recruit gene regulation proteins to a specific loci that are then able to activate or inactivate gene expression. These dCas9-based programmable gene regulation systems have enabled investigation into factors that regulate genomic function. Furthermore, the programmable expression and repression of targeted genes has enabled elucidation of their functions without changing the DNA sequence.

Targeted gene activation can be achieved through the use of dCas9 to recruit transcriptional activation domains to the promoter region of a specific gene, termed CRISPR-

activation (CRISPRa). Early efforts to create such a system identified the transcriptional activators VP64, p65, and Rta as the ones which produced the greatest induced gene expression individually[86]. Efficient targeted transcriptional activation could be achieved with dCas9-VP64 alone, but they found that fusing the three transcriptional activators in tandem to generate the tripartite transcriptional activator Vp64-p65-Rta (VPR) produced even greater transcriptional activation. The dCas9-VP64 and dCas9-VPR systems have been widely used to study the function of specific genes and in genome-wide CRISPRa screens to identify genes that are important in certain cellular processes[87,88]. Combinatorial gene activation can also be achieved by introducing multiple sgRNAs targeting different genes[89,90]. These CRISPRa systems can also be used in cellular reprogramming applications and for the differentiation of stem cells[91].

Targeted gene inactivation can be achieved by simply using dCas9 itself to hinder RNA polymerase activity by sterically hindering transcription, termed CRISPR-interference (CRISPRi)[92]. However, this method is often inefficient in eukaryotic cells. CRISPRi using a fusion of Krüppel-associated box (KRAB) domains to dCas9, has enabled a more efficient method for gene repression[93]. KRAB domains are able to induce heterochromatin formation, thus resulting in decreased gene expression.

Targeted gene expression modifications can also be achieved using dCas9-mediated epigenome editing systems. dCas9 can be fused to enzymes that can directly acetylate or methylate histones[13,94]. Many different enzymes have been fused to dCas9 to write and erase histone modifications at specific loci in genomes[13]. For example, p300 for acetylation, HDAC3 for deacetylation, SMYD3 for methylation, and TET1 for demethylation[13,95–99].

These dCas9-based CRISPRa, CRISPRi, and epigenome editing systems have greatly expanded the toolbox for researchers. All of these tools have enabled investigation into the function of different genes and genetic elements throughout the genome without creating a DNA break or changing the DNA sequence.

1.5 Engineered CRISPR/Cas9 systems for temporally-controlled genome engineering

Precisely targeting a locus in the genome is very useful for studying the function of different genetic elements. However, many genomic elements are temporally regulated and understanding them requires targeting tools that can be temporally controlled. Thus, to be able to study these genetic elements, researchers have engineered inducible CRISPR/Cas9 systems. These inducible CRISPR/Cas9 systems have provided insight into DNA repair mechanisms, the kinetics of Cas9 editing itself, and have been used to engineer systems that record biological events in cells[100–103]. To achieve temporal control of CRISPR/Cas9 activity, a variety of methods have been employed that control sgRNA targeting ability or different steps in Cas9's activities. Furthermore, temporal control has also been applied to other Cas9-based effectors such as transcriptional activation and base editing[104–106]. Dynamic control of CRISPR/Cas9 systems has enabled a deeper investigation into temporally regulated genomic elements.

1.5.1 Engineered temporally-controlled Cas9 endonucleases

One approach to enable control of CRISPR/Cas9 activities is to control the activities of Cas9 itself. A variety of methods have been employed to limit Cas9 functionality, such as using protein destabilization domains, sequestration of active Cas9 outside the nucleus, and protein fusions to sterically hinder Cas9 activity[107,108]. These systems have varying levels of inhibition prior to stimulating Cas9 activity and different activation kinetics which could limit the applications of certain systems. Some systems have demonstrated applicability beyond gene editing such as for controlling transcriptional activation[104,109].

Control of Cas9 activity can occur at the transcriptional level with inducible promoters to control Cas9 protein expression. Doxycycline-inducible Cas9 systems were one of the first to be developed[110,111]. These systems use lentivirus- or recombination-based techniques to introduce the Cas9 gene downstream of an inducible promoter. Inducible promoter-based systems are adaptable to any Cas9 system of interest by simply changing the gene downstream of the

promoter. Thus, they are easily adaptable to provide temporal control of any Cas9-based effector system. However, since temporal control is at the level of DNA transcription, only coarse-grained control can be achieved. As Cas9 is only active after gene expression and protein translation, it could take several hours or days to achieve maximum protein levels to allow efficient Cas9 activity. Therefore, dox-inducible Cas9 systems are usually only used in applications where early time points of editing are not important and only maximum editing efficiency is desired at a defined time.

Degron-based inducible Cas9 systems have been engineered that allow temporal control of Cas9 activity at the post-translational level. These systems use a fusion of Cas9 to destabilizing domains such as dihydropholate reductase (DHFR), an FKBP-derived destabilization domain, or an estrogen receptor (ER50)[112–114]. These destabilizing domains signal for the degradation of the protein, thus inhibiting Cas9 activity. Stabilization and activation of Cas9 activity is achieved through the use of small molecules that bind to the destabilizing domain, such as TMP for DHFR, Shield-1 for an FKBP-derived destabilizer, and 4-OHT for ER50. As buildup of the Cas9 protein pool after stabilization requires DNA transcription and translation, activation of Cas9 activity is slow and could take 24 hr to reach maximum protein abundance[101]. Similar to dox-inducible systems, degron-based inducible Cas9s are more suited towards applications where precise and rapid activation of Cas9-based activities are not needed.

Cas9 can also be divided into two components to inhibit its activity, creating an inducible split Cas9 system[115]. The split-Cas9 system divides the Cas9 protein into two components in the REC2 domain. The N-terminal half of the protein contains part of the REC lobe and part of the RuvC nuclease, including the catalytic D10 residue that cleaves the non-target DNA strand, and the C-terminal half of the protein contains the remainder of the protein. Each half of the protein is then fused to a component of the FKBP-FRB rapamycin-induced dimerization system. Upon addition of rapamycin to cells expressing both halves of Cas9, dimerization of FKBP-FRB will cause reconstitution of the entire Cas9 protein, thus allowing Cas9 activity to occur. One

consideration with the split-Cas9 system is that careful stoichiometric considerations need to be made to be able to express both components equally within the same cell. Nevertheless, the split-Cas9 system is much more rapid compared to dox-inducible transcription and degron-based systems.

Translocation of the Cas9 system to the nucleus is another method utilized to control Cas9 activity. In these systems, the entire Cas9 system or parts of a split-Cas9 system is sequestered outside the nucleus using an estrogen receptor (ERT2)[116–118]. Binding of a small molecule, 4-HT, to ERT2 allows the Cas9 to be translocated into the nucleus where it is then able to act on the genome. Sequestration of Cas9 editing components outside of the nucleus also decreases the background editing activity. For example, the split-Cas9 system also utilizes nuclear translocation for both N- and C-terminal Cas9 components to decrease the background editing prior to dimerization[115]. Finally, translocation of Cas9 into the nucleus is also more rapid than dox-inducible expression.

Using protein domains or small peptides to control Cas9 activity have also been engineered and some systems appear to have much faster kinetics than the other strategies. Insertion of these protein domains either sterically hinders Cas9 activity or it constrains the structure of Cas9 to be inactive. Intein-Cas9 uses the ligand-dependent self-cleaving protein intein to block Cas9 activity[119]. Sites for intein insertion were screened and it was found that insertion into the REC lobe of Cas9 produced the greatest editing magnitude while still minimizing background editing. Cleavage of intein from Cas9 occurs upon the addition of a small molecule, 4-HT. After the intein is cleaved, Cas9 is released and fully active. An allosterically-activated Cas9 (arC9) has been developed that also relies on protein domain insertion into the REC lobe, in this case an estrogen receptor (ER) responsive to the small molecule 4-HT[120]. ER fusion to Cas9 causes it to be constrained in an inactive state due to the structure of the ER. Addition of 4-HT induces a conformational change in the ER which then transduces an allosteric signal for Cas9 to form an active conformation. Intein-Cas9 and arC9 both involve insertion of a protein domain

into a site in Cas9 that doesn't fully disrupt activity when activated. An investigation to reveal more of these sites using circular permutation of Cas9 was performed and resulted in development of protease-activated inducible Cas9 systems[59]. In this study, they fused the original N- and C-termini of Cas9 together using a small peptide linker. At the same time, they used a library-based approach to find new N- and C-termini of Cas9 that were tolerated in other locations throughout the protein. They found that the REC2 domain of Cas9 was able to tolerate formation of new N- and C-termini. Next, they discovered that if the peptide linker at the original termini was short enough, Cas9 could be constrained into an inactive conformation. Using protease-cleavable peptides to fuse the original Cas9 termini, they created inducible Cas9 variants that are constrained in an inactive conformation when translated. Upon exposure to a protease specific to the peptide-cleavable linker, Cas9 activity was able to be restored by relieving the conformational constraints. These inducible Cas9 systems that rely on induced conformational changes could have much faster activation kinetics compared to the systems that rely on Cas9 translocation and dimerization. However, direct comparisons have not been made.

We developed a chemically-controlled Cas9, ciCas9, which can be rapidly activated and rheostatically tuned to enable rapid, precise, and dose-dependent DSB generation[100,121]. ciCas9 uses an autoinhibitory interaction between a Bcl-xL protein and BH3 peptide to control Cas9 activity. To inhibit Cas9 activity, the REC2 domain of Cas9 was replaced with Bcl-xL and the BH3 peptide was appended at the C-terminus of Cas9. Upon translation, Bcl-xL and BH3 form a high affinity interaction, sterically hindering Cas9 activity. The interaction between Bcl-xL and BH3 can be rapidly disrupted using potent small molecules that bind to Bcl-xL and disrupt the interaction with BH3. Early studies with ciCas9 revealed that DSB formation could be detected as early as 10 min after small molecule introduction. The first generation of ciCas9 utilized a G22 residue in the BH3 peptide (ciCas9(G22)) that would allow rapid activation of Cas9 activity using the small molecule A385358 (A3)[100]. Subsequent engineering of ciCas9 revealed that an L22 residue in the BH3 peptide (ciCas9(L22), hereafter ciCas9) decreased unwanted background editing in the

absence of small molecule activator[121]. Furthermore, a more potent small molecule activator, A-1155463 (A115), was available to induce ciCas9 activity. Our ciCas9 system has been used to study the kinetics of DSB formation by Cas9 and the subsequent timing of DNA repair at the DSB site[100].

Photoactivatable approaches to temporally control Cas9 activity have also been employed. Photoactivation and split protein dimerization methods have been combined to create photoactivatable Cas9 (paCas9) using photoactivatable dimerization domains[122]. Cas9 activity is induced upon blue light activation and subsequent dimerization. A different photoactivatable approach to control Cas9 activity was engineered by fusing photoactivatable LOV2 domains onto the small protein Cas9 inhibitor, AcrIIA4[123]. Blue-light activates a conformational change in LOV2 and subsequently AcrIIA4 which then causes AcrIIA4 to release from Cas9, resulting in Cas9 activation. These photoactivatable systems allow rapid activation for Cas9 activity. However, these switches are binary and do not allow fine tuning of the level of Cas9 activity compared to the small molecule activated inducible Cas9 systems.

The myriad of inducible Cas9 systems have provided scientists with a plethora of approaches to confer temporal control over Cas9 activity. Furthermore, some of the inducible systems allow for reversible control of Cas9 activity. For example, the arC9 system has demonstrated reversibility after removal of the small molecule stabilizer from cell culture[120]. However, the deactivation process is often slow and doesn't completely turn off Cas9 activities. Finally, in most of the engineering of these systems, temporal control was only demonstrated for Cas9 DNA editing. Being able to temporally control all Cas9-based effector activities would immensely add to the Cas9 genome engineering toolbox.

1.5.2 Engineered temporally-controlled Cas9-based effector systems

Direct temporal control has also been engineered for Cas9-based transcriptional activation and base editing systems. Temporally-controlled transcriptional activation systems have used

chemically-inducible dimerization techniques to control recruitment of transcriptional activation domains to dCas9[104,105]. In these systems, one component of the dimerizer system is fused to dCas9, and the other component to VPR. Upon addition of the small molecule dimerizer, VPR can then be recruited to a specific site in the genome specified by dCas9 targeting. It has also been demonstrated that transcriptional activation can be rheostatically tuned by titrating the amount of small molecule dimerizer added to cells.

Temporally-controlled base editing has been demonstrated using split-engineered base editors (seBEs)[106]. In these systems, the base editing deaminase enzyme was split into two protein components. Each half of the base editor was then fused to one of the components of the FKBP-FRB chemically-inducible dimerization systems. One of the halves was also fused to nCas9(D10A) to allow for precise genome targeting. Rapamycin addition causes reconstitution of the active deaminase enzyme and allows localization of the base editor to the target site specified by the dCas9 for targeted base editing to occur. The method used to engineer seBEs is difficult to apply to other Cas9-based effectors or even other deaminase enzymes. A new split Cas9-based effector will need to be screened for functionality every time a new system is engineered.

Although some temporally-controlled Cas9-based effector systems have been developed, the methods used to engineer the systems are not widely applicable to all types of effectors. Some engineered temporally-controlled Cas9 proteins could be adapted for these other systems, but few have been demonstrated. A generalizable method to confer temporal control over all Cas9-based effector systems would simplify and vastly expand the temporally-controlled CRISPR/Cas9 toolbox.

1.5.3 Engineered sgRNAs for temporally-controlled Cas9 activity

A generalizable way to temporally-control Cas9 activity is to control the functionality of the sgRNA. Aptazyme-embedded guide RNAs (agRNAs) have been developed that block the 20 nt protospacer sequence[124]. In this system, the sgRNA is fused to a ligand-activated self-cleaving

RNA that contains a complementary sequence to the 20 nt protospacer. Thus upon transcription of the agRNA, the protospacer is blocked and Cas9 target search and binding cannot occur. Upon addition of the small molecules theopylline or guanine, the aptazyme will cleave itself and release the sgRNA to be able to direct Cas9 to its target site. However, agRNAs do have some background editing activity prior to ligand activation. More engineering of the aptazyme RNA is needed to fully eliminate background activity. One advantage of the agRNA system is that a variety of ligand-activated aptazymes with different small molecule specificities exist in nature and have been developed by scientists. This allows for multiplexed approaches to conditionally induce Cas9 activity at different target sites.

A different approach using ligand-activated RNA aptamers has also been employed to control Cas9 activity by stabilizing or destabilizing the sgRNA[125]. These aptamer-inserted sgRNAs (ligRNAs) have an RNA aptamer inserted into the stem loops in the sgRNA. They found that depending on the location of aptamer insertion, the aptamer could confer ligand-activating or ligand-deactivating control over Cas9 activity. Activation and deactivation of Cas9 activity was also dose-dependent and occurred within a few minutes of small molecule addition. However, the study also found that these ligRNAs only worked well in prokaryotic systems and more engineering is needed to function within eukaryotic cells.

Chemical modification of the sgRNA to block the protospacer has also been employed to create the light-inducible very fast CRISPR (vfCRISPR) system[102]. In this system, two or three uridine nucleotides in the PAM-distal region of the protospacer sequence are replaced with 6-nitropiperonyloxymethyl-modified deoxynucleotide thymine caged nucleotides, forming a caged sgRNA (cgRNA). The cgRNA can then be “uncaged” using light and Cas9 targeting can occur. The vfCRISPR system is currently the fastest inducible CRISPR/Cas9 system that is able to induce DSBs at the target site within seconds of light activation. However, the biggest limitation to the vfCRISPR system is that the cgRNA cannot be genetically encoded. Thus, Cas9:cgRNA must be delivered into cells as an RNP. Nevertheless, vfCRISPR has been used in conjunction

with time-resolved chromatin immunoprecipitation to reveal DSB DNA repair mechanisms. vfCRISPR provides a highly temporally-precise method to controlling Cas9 activity.

Engineered sgRNAs provide an alternative method for precise temporally-controlled Cas9 activity. In addition, as an sgRNA is needed in all Cas9-based effector systems, these methods could be widely applicable to confer temporal control.

1.6 Using a chemically-controlled CRISPR/Cas9 system to understand and develop new genome engineering technologies

The discovery and development of CRISPR/Cas9 as a precise genome targeting technology has empowered scientists to easily investigate any locus in the genomes of many organisms. CRISPR/Cas9 has also enabled new therapeutic possibilities such as directly correcting disease causing mutations[72]. However, limitations remain in understanding and fully eliminating off-target editing in human cells that hinder the full potential of CRISPR/Cas9 applications. In **Chapter 2**, I describe the use of our ciCas9 system to explore the *in vivo* mechanisms of Cas9 off-target editing by investigating the editing kinetics, Cas9 binding affinity, and Cas9 cleavage ability at on- versus off-target sites in the human genome.

Being able to temporally control CRISPR/Cas9-based effector activities have further expanded the CRISPR/Cas9 toolbox. Temporally-controlled CRISPR/Cas9 systems have allowed scientists to interrogate temporally-regulated biological processes and have provided a new[101,102]. In addition, temporally-controlled CRISPR/Cas9-based effector systems have enabled new methods of investigating genetic elements without formation of a DSB[104–106]. However, current approaches to engineer these systems use a patchwork of methods and there has been no demonstration of a temporally-controlled Cas9 system that can be widely adapted to form temporally-controlled versions of all Cas9-based effector systems. In **Chapter 3**, I describe the versatility of the ciCas9 system to act as a DNA binding switch that is able to confer temporal

control to a variety of Cas9-based effector systems. I demonstrate the use of ciCas9 to engineer temporally-controlled transcriptional activation, base editing, and prime editing technologies.

Chapter 2: Dissecting the mechanism of *in vivo* CRISPR/Cas9 off-target DNA editing using a chemically-controlled Cas9

A version of this chapter is in preparation for publication.

Wei, C.T., Stephany, J.J., Popp, N.A., Fowler, D.M., Maly, D.J.

2.1 Introduction

CRISPR/Cas9 is a powerful tool for genome editing, enabling precise installation of desired DNA changes. Genome editing with the CRISPR/Cas9 system relies on two main components: the Cas9 endonuclease and the single guide RNA (sgRNA)[16,126]. The Cas9 endonuclease is able to make a DNA double strand break (DSB) at a precise target site within the genome specified by a 20 nt sequence encoded by the sgRNA. Uncontrolled insertions and deletions (indels) can form at the DSB site as a result of non-homologous end joining (NHEJ) or a precise DNA edit can be incorporated through homology directed repair (HDR) of an exogenous DNA template[2,127].

Even though the location of the DSB is specified by the sgRNA, CRISPR/Cas9 methods suffer from off-target editing. Off-target editing is harmful in many applications of genome editing where unintended indel formation could lead to unwanted outcomes. For example, in therapeutic applications, off-target editing could lead to unintended editing of oncogenes and tumor suppressors which could result in deleterious outcomes. Off-target editing also poses a problem for researchers conducting genome-wide CRISPR screens where an observed phenotype may be a result of an off-target edit rather than the intended target.

Approaches focusing on both the DNA target and the Cas9 nuclease itself have been used to understand and minimize off-target editing. Techniques have been developed to detect genome-wide off-target editing to identify Cas9 off-target sites for a given sgRNA[27,37,38], typically

revealing hundreds of off-target sites. Early systematic studies into the target sequence determinants using DNA mismatch target libraries revealed that mismatches were poorly tolerated in the region 8-12 nucleotides from the PAM because they disrupt Cas9 binding to a target site[24–26]. Mismatches in the PAM-distal region are better tolerated. *In vitro* studies of Cas9 binding and cleavage using a library of mismatched DNA target substrates have since been reported and reflect the results of the earlier studies[29,128]. Other *in vitro* biochemical investigations using purified Cas9 components have elucidated the protein conformational changes and DNA unwinding and cleavage kinetics that occur during editing at on- versus off-target DNA substrates[19,30–32]. These *in vitro* mechanistic studies found that mismatches in the PAM-distal region disrupt the cleavage kinetics of Cas9, thus allowing Cas9 to dissociate from an off-target site prior to creating a DSB. Together, these studies informed development of computational tools to predict sgRNA off-target editing and have enabled rational engineering of high-fidelity Cas9 variants[30,39–41,43].

Although techniques have been developed to mitigate off-target editing, off-target editing remains a challenge. For example, high-fidelity Cas9 variants effectively eliminate most but not all off-target editing. In addition, the efficiency of on-target editing by high-fidelity variants is variable even for sgRNAs targeting the same gene[43]. Computational tools to predict off-target editing perform poorly, often predicting sites that show no editing and missing sites that have high off-target editing[129]. Thus, an in-depth understanding of how Cas9 discriminates between on- and off-target sites, as well as between different off-target sites, to achieve the editing magnitudes we observe is needed to facilitate development of improved Cas9 high-fidelity variants and off-target editing predictions. Furthermore, investigating target site discrimination *in vivo* is necessary to account for cellular and genomic factors that influence editing, such as chromatin state[130].

Thus, we used ciCas9, a chemically-inducible Cas9 variant, to dissect the factors influencing target discrimination *in vivo*[100]. We measured editing kinetics and conducted experiments to measure relative DNA binding by Cas9 at three genomic on-target sites and ten corresponding off-target sites. Surprisingly, off-target sites and on-target sites generally had

similar overall editing kinetics and relative binding affinities. We hypothesize that Cas9 likely forms a cleavage-competent conformation less frequently at these genomic off-target sites, thus resulting in lowered cleavage efficiency and lower editing magnitude compared to the on-target site. To complement this in-depth, mechanistic look at a limited number of endogenous genomic loci, we also measured Cas9 binding and editing for a total of 180 single nucleotide variants of three synthetic target sites. Single mismatches in the seed region of these synthetic targets disrupted the Cas9 whereas single mismatches in the PAM-distal region maintained similar binding affinities as the perfectly matched target sequence. As at the endogenous genomic loci, PAM-distal mismatches lowered editing magnitude due to altered Cas9 cleavage efficiency. This study provides a better understanding of Cas9 *in vivo* off-target editing mechanisms and could inform future efforts in mitigating off-target editing effects.

2.2 Results

2.2.1 ciCas9 can be used to study on- and off-target editing mechanisms of Cas9 *in vivo*

To dissect the kinetic mechanisms of Cas9 target discrimination *in vivo*, we used ciCas9, a chemically inducible variant of *S. pyogenes* Cas9 whose activity can be rapidly induced and rheostatically tuned to control DSB formation[100,121,131]. ciCas9 employs an autoinhibitory interaction between Bcl-xL and a BH3 peptide that have been appended to WT-Cas9 (**Fig. 2.1A**). This autoinhibitory interaction can be rapidly disrupted by a Bcl-xL inhibitor, A-1155463 (A115).

We first evaluated the similarity of on- and off-target editing between WT-Cas9 and WT-ciCas9. We compared their specificities by measuring the appearance of indels at seven off-target sites corresponding to three different sgRNA targets using targeted next-generation sequencing (**Appendix B Fig. B.1**). We defined specificity as the ratio of on-target indels to off-target indels. Similar to previous studies, we observe robust indel formation after activation of ciCas9 with A115 (**Appendix B Fig. B.1A**)[100]. The specificities of WT-Cas9 and WT-ciCas9 did not differ

significantly at the off-target sites investigated, suggesting that WT-ciCas9 edits off-target sites similarly to WT-Cas9 (**Appendix B Fig. B.1B**).

To evaluate the mechanistic similarities in off-target editing between WT-ciCas9 and WT-Cas9, we asked if previously identified Cas9 specificity enhancing mutations function in ciCas9. Indeed, the specificity enhancing mutations from eSpCas9(1.1) reduced the indel frequency at off-target sites when introduced into ciCas9[41,100]. Since the initial report of ciCas9, a new high-fidelity Cas9 variant, Hypa-Cas9, has been reported[30]. Hypa-Cas9 was engineered to have diminished cleavage activity at off-target sites via the introduction of four mutations that enhance the Cas9 proofreading mechanism governing catalytic competency. We introduced these four mutations into ciCas9 to create Hypa-ciCas9 and found that Hypa-ciCas9 showed robust editing at on-target sites (**Appendix B Fig. B.1A**). At off-target sites where indel frequencies were measurable, Hypa-ciCas9 increased editing specificity similar to Hypa-Cas9 (**Appendix B Fig. B.1B**). The similarity in specificity of WT-Cas9 and WT-ciCas9 suggests that ciCas9 targets on- and off-target sites comparably *in vivo*. Furthermore, the introduction of high fidelity mutations into ciCas9 with resulting specificity improvements similar to Cas9 suggests that ciCas9 maintains similar editing and proofreading mechanisms. Thus, the DNA targeting mechanism of ciCas9 appears comparable to Cas9 enabling us to use ciCas9 to investigate on- and off-target editing mechanisms.

2.2.2 Changes in editing rates do not explain off-target discrimination by Cas9

One possible mechanism of on- versus off-target discrimination by Cas9 is differential DSB generation kinetics. Assuming that DNA repair kinetics is comparable at on- versus off-target sites, measuring indel formation over time can reveal DSB generation kinetics. Previously, a degron-fused Cas9 or the anti-CRISPR AcrIIA4 were used to show that early inhibition of Cas9 activity can result in diminished editing at off-target sites relative to on-target sites[34,35]. This improved specificity with shorter exposure to active Cas9 editing suggests that the specificity of

editing changes over time. Specifically, the on-target site accumulates indels rapidly while off-target sites accumulate indels later, with an increasing rate over time. The difference in indel accumulation rate would result in a decrease in Cas9 specificity over time (**Fig. 2.1B**). However, these studies did not measure indel outcomes until several days after inhibiting Cas9. At these late time points, editing is often saturated and has reached a plateau. Furthermore, kinetic studies of Cas9 editing have focused on on-target sites but not off-target sites during this active editing period[100–102,121]. Thus, we sought to dissect the kinetic differences between Cas9 on- and off-target editing during the early, active period of indel accumulation.

First, we examined Cas9 indel outcomes and specificity at 24 and 48 hours after transfection of Cas9 and sgRNA into HEK293T cells. We chose three sgRNAs with corresponding off-target sites that are edited at magnitudes detectable by targeted next-generation sequencing[30,37,132]. Between 24 and 48 hrs, we observed a two-fold increase in indel frequency suggesting that Cas9 was still actively editing these target sites (**Figs. 2.1A-C**). For all sgRNAs studied, we observed no significant decrease in specificity at 48 hr Cas9 editing compared to 24 hr (Mann-Whitney U-test $p > 0.05$, **Fig. 2.1D**). For example, FANCF on-target indel frequency increased from 21.4% to 44.4% and FANCF off-target (OT) indel frequency increased from 17.1% to 37.2% between 24 and 48 hrs. Meanwhile, the specificity of FANCF off-target editing remained similar at 1.25 and 1.19 between 24 and 48 hrs. The similar specificity at 24 and 48 hrs despite accumulating indels suggests that there is no change in specificity of Cas9 editing over time. Thus, cellular processes such as DNA repair, protein accumulation, and cell growth may account for the conclusions drawn from late time point experiments using degron-Cas9 or AcrIIA4.

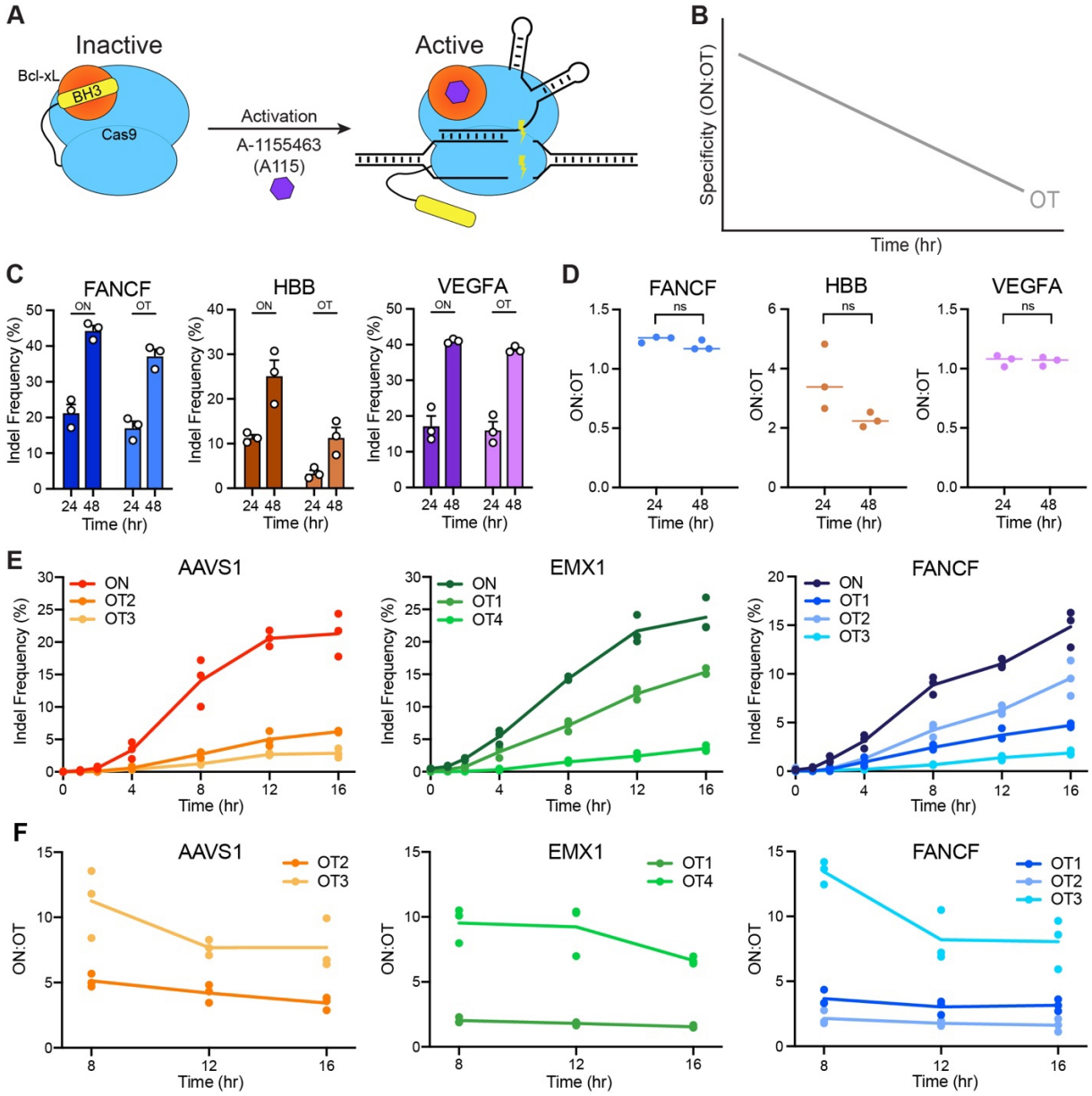

Figure 2.1. Overall Cas9 editing kinetics do not explain on- vs. off-target editing.

A) Schematic of the chemically-inducible (ciCas9) system.

B) Schematic of time courses for specificity measured as a ON:OT of indel frequencies.

C) Time courses of Cas9 editing measured by indel frequency at on- and off-target sites with sgRNAs targeting FANCF, HBB, and VEGFA. Data represented as mean ON:OT ± SEM of 3 cell culture replicates.

D) Specificity (ON:OT of indel frequency) of Cas9 editing at off-target sites using sgRNAs targeting FANCF, HBB, and VEGFA. $p > 0.05$ for all three off-target sites, Mann-Whitney U-test comparing ON:OT at 24 and 48 hr.

E) Time courses of ciCas9 editing measured by indel frequency at on- and off-target sites with sgRNAs targeting AAVS1, EMX1, and FANCF. ciCas9 was activated with 1 μM of A115.

F) Time course of specificity (ON:OT of indel frequency) from ciCas9 editing at off-target sites with sgRNAs targeting AAVS1, EMX1, and FANCF.

We next investigated even earlier time points using ciCas9. We monitored editing at three previously characterized on-target sites and seven corresponding off-target sites for 16 hrs after A115 activation (**Fig. 2.1E; Appendix B Fig. B.2**). Similar to our previous studies, on-target indels accumulated linearly soon after A115 addition. Indels at corresponding off-target sites also accumulated linearly soon after A115 addition, similar to the on-target site. The linear increase in indel frequency at both on- and off-target sites strongly suggests that there is no change in the rate of indel accumulation over time. Although the indels at on- versus off-target sites accumulate at different rates, the ratio of on- to off-target indel frequency remains largely constant over time, starting as early as 8 hr after A115 activation of ciCas9 (**Fig. 2.1F**). The indel frequency at off-target sites prior to 8 hrs is too small for quantitative calculations of specificity. This observed constant specificity over time also reflects the constant rate of indel accumulation at off-target sites relative to on-target sites. The unchanging rate of indel accumulation suggests that specificity does not appear to be a time-dependent phenomenon and that the overall editing kinetics of Cas9 at on- and off-target sites appear to be similar. Thus, there is not a time window of editing prior to saturation where specificity is enhanced.

2.2.3 Cas9 binding strength is similar at on- and off-target sites, and is not the major determinant of off-target discrimination

Given that on- and off-target editing kinetics appear to be similar and do not account for on-target specificity, we explored whether other aspects of Cas9 editing could account for off-target discrimination. The off-target sites most commonly detected in Cas9 editing are sites that only differ by a few mismatches compared to the on-target site. We investigated how DNA binding affinity and DNA cleavage efficiency of Cas9 contributes to off-target discrimination at these sites with few mismatches.

Previous studies to dissect the activity of Cas9 at on- vs. off-target sequences have all been performed *in vitro* with purified Cas9 components where binding and cleavage can be

investigated separately. Under these conditions, DNA cleavage by Cas9 is very rapid and the rate limiting step for Cas9 activity is DNA unwinding during R-loop formation when Cas9 is at the on-target site[19]. For some high-fidelity Cas9 variants, DNA binding appears to be similar but the DNA cleavage rate decreases when these variants are bound to off-target or mismatched DNA substrates[31,32]. The decrease in cleavage rate appears to be due to the variants forming the cleavage-competent conformation less often when bound to mismatched DNA substrates[30]. These *in vitro* studies suggest that the cleavage activitiy of Cas9 is the key to off-target discrimination, and, by tuning cleavage competency, off-target discrimination can be enhanced. However, *in vitro* studies lack cellular and genomic context. Previously, binding and cleavage could not be separately measured *in vivo* because indel frequency integrates both DNA binding and cleavage.

To overcome this limitation we used ciCas9 to measure the relative binding affinity of ciCas9:sgRNA to different targets by titrating the amount of small molecule activator added to the cells[121]. Here, we hypothesized that titration of the small molecule activator would titrate the amount of ciCas9 that is able to bind to the target site. To test this hypothesis, we used CRISPR-activation (CRISPRa) with a catalytically-dead ciCas9 fused to the transcriptional activators VP64-p65-Rta (VPR), dciCas9-VPR, to study ciCas9's ability to bind DNA in its inactive and active states. We first targeted dciCas9-VPR to the promoter region of the CXCR4 gene and monitored expression of CXCR4 using fluorescent antibody staining and flow cytometry (**Fig. 2.2A**). When A115 is absent, the expression level of CXCR4 is at baseline (**Fig. 2.2B**). CXCR4 expression is only increased when A115 is present to open dciCas9-VPR and allow it to bind to DNA. We then constructed a synthetic transcriptional reporter locus with a defined target sequence upstream of an EGFP reporter to evaluate the dose-dependence of dciCas9-VPR DNA binding (**Appendix B Fig. B.3**). Titration of A115 dciCas9-VPR yielded a monotonic response in EGFP fluorescence (**Fig. 2.2C**). This suggests that the titration of indels we previously observed is a result of modulation of the amount of ciCas9 able to bind the target site. Thus, we can study the relative

binding of ciCas9 at on- and off-target sites by measuring indel frequency and computing a dose-response of editing. An EC_{50} can also be measured from the dose-response to approximate relative binding affinities of Cas9 to on- and off-target sites from the dose-response of ciCas9 editing (**Fig. 2.2D**).

First, we measured indel frequencies at the on- and off-target sites with 1 μM A115 where we expect fully opened and active ciCas9 to be present. As expected, the on- and off-target sites with fully open ciCas9 have very different indel magnitudes, yet the off-target sequences only differ from the corresponding on-target sequence by a few mismatches (**Fig. 2.2E, Appendix B Fig. B.4**). To probe how binding differences might relate to these indel outcomes, we used our ciCas9 titration assay to study ciCas9 binding at these on- and off-target sites. To ensure enough indels had accumulated with fully open ciCas9 to allow for a wide dynamic range for measuring a dose response, we investigated ciCas9 binding at 12 hr after A115 addition. The dose response curves for all three sgRNA targets show that ciCas9 binds to on- and off-target sites very similarly (**Figs. 2.2F**). In addition, the computed EC_{50}s show very similar apparent binding affinities of ciCas9 to on- and off-target sites (**Table 2.1**). Minimal changes in relative binding affinity are also observed over time for all on- and off-target sites, suggesting that we are measuring relative binding at equilibrium (**Figs. 2.2G, Appendix B Fig. B.5, Tables 2.2-3**). The similarities in apparent binding affinities of ciCas9 at on- and off-target sites for all three sgRNAs do not explain the diminished indel frequencies at off-target sites for fully active ciCas9. Therefore, at least for these sgRNAs, binding is not the determining factor in Cas9 off-target discrimination. Thus, off-target discrimination must be dictated by factors other than binding affinity differences, such as Cas9's cleavage efficiency after DNA binding.

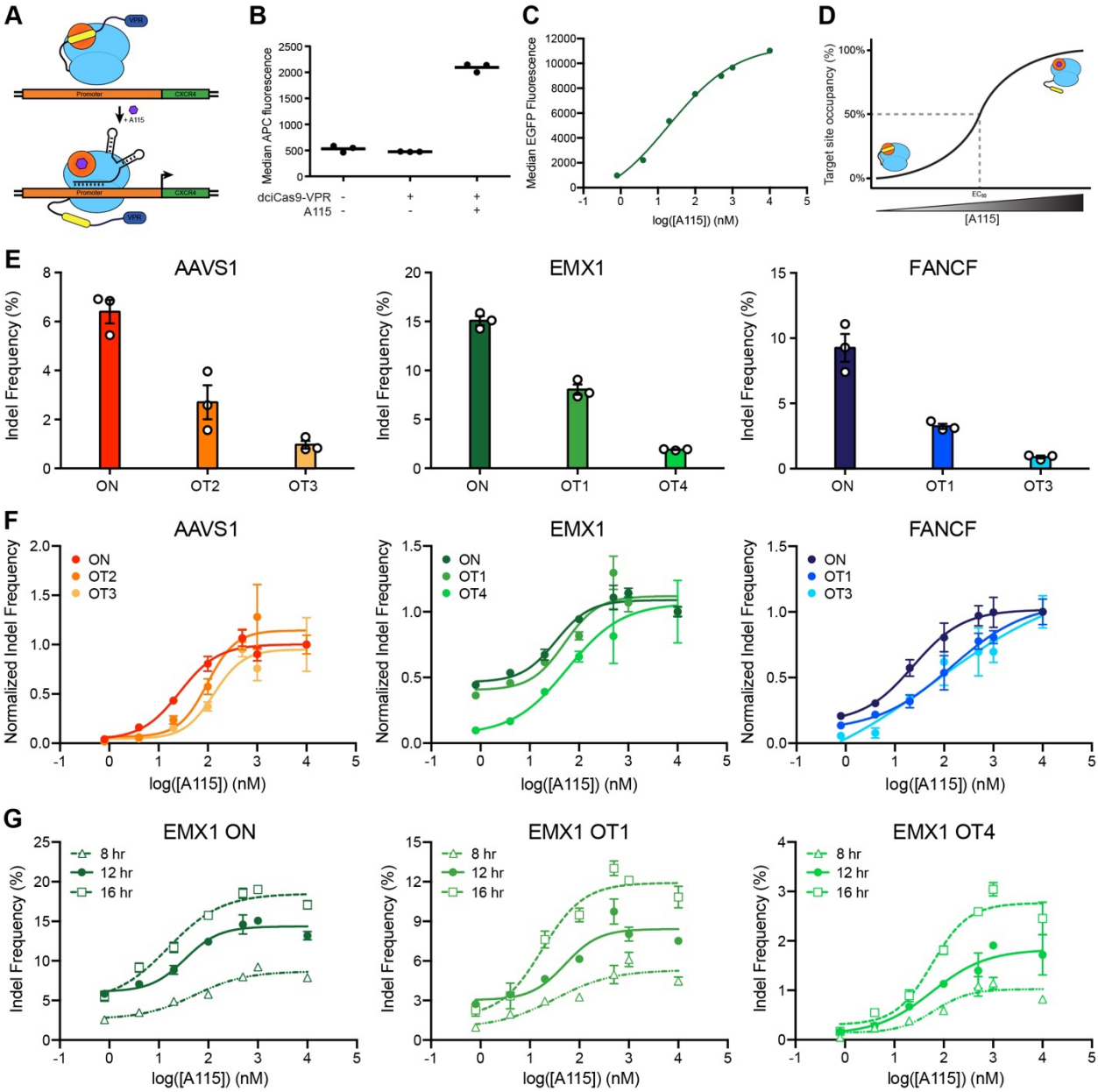

Figure 2.2. Cas9 binds to on- and off-target sites very similarly and is not the major determinant of off-target discrimination.

A) Schematic of dciCas9-VPR used in CRISPRa at the CXCR4 locus to study its ability to bind DNA in the presence and absence of A115.

B) Activation of CXCR4 expression by dciCas9-VPR in the presence and absence of A115. CXCR4 expression is measured by staining HEK293T cells with APC-conjugated anti-CXCR4 antibody and subsequent flow cytometry analysis.

C) Dose-response of dciCas9-VPR-induced EGFP expression in the EMX1-EGFP transcriptional synthetic locus cells at 48 hr after A115 addition. Data represented as mean ON:OT ± SEM of 3 cell culture replicates.

D) Schematic of tuning A115 concentration added to cell cultures to control occupancy of target sites by ciCas9 to measure EC_{50} of target occupancy.

Figure 2.2 cont.
E) Indel frequency of ciCas9-induced editing at 12 hr after A115 addition. Indels measured at on-target and most prominent off-target sites for AAVS1, EMX1, and FANCF sgRNA targets.
F) Normalized dose-response of ciCas9 editing at on- and off-target sites for AAVS1, EMX1, and FANCF sgRNA targets. Indel frequencies normalized to the mean of indels measured at 10 μM A115.
G) Dose-response of ciCas9 editing at EMX1 ON, EMX1 OT1, and EMX1 OT4 measured at 8, 12, and 16 hr after addition of A115. Data represented as mean ON:OT ± SEM of 3 cell culture replicates.

Table 2.1. EC_{50} values from ciCas9 dose-response measured by indel frequencies at the endogenous loci 12 hr after A115 addition.

Locus		EC_{50} (nM)	95% CI (nM)
AAVS1	ON	27	13 - 47
	OT2	97	28 - 232
	OT3	132	35 - 874
EMX1	ON	34	14 - 66
	OT1	51	11 - 141
	OT4	58	Too large
FANCF	ON	24	1 - 61
	OT1	142	Too large
	OT3	64	Too large

Table 2.2. EC_{50} values from ciCas9 dose-response measured by indel frequencies at the endogenous loci 8 hr after A115 addition.

Locus		EC_{50} (nM)	95% CI (nM)
AAVS1	ON	52	25 - 92
	OT2	152	Too large
	OT3	535	Too large
EMX1	ON	52	9 - 136
	OT1	32	Too large
	OT4	63	2 - 197
FANCF	ON	12	Too large
	OT1	52	26 - 89
	OT3	57	9 - 170

Table 2.3. EC_{50} values from ciCas9 dose-response measured by indel frequencies at the endogenous loci 16 hr after A115 addition.

Locus		EC_{50} (nM)	95% CI (nM)
AAVS1	ON	33	23 - 47
	OT2	68	45 - 99
	OT3	104	68 - 150
EMX1	ON	15	2 - 35
	OT1	17	5 - 39
	OT4	58	25 - 110
FANCF	ON	7	Too large
	OT1	42	Too large
	OT3	Ambiguous	Ambiguous

2.2.4 A library of target sites in a synthetic locus reveal how mismatches affect Cas9 binding versus cleavage efficiency

Investigating off-target Cas9 binding versus cleavage efficiency mechanisms at only a few previously characterized sites limits us to studying the effects of mismatches that are well tolerated and are still able to allow indel generation. Off-target sequences often have mismatches that are spread throughout the 20-nucleotide target sequence. Mismatches in the seed region 8-12 nucleotides immediately 5′ of the PAM sequence tend to disrupt Cas9 DNA binding[24–26]. Furthermore, mismatches in the PAM-distal region tend to disrupt Cas9 DNA cleavage[30]. Investigation of *in vitro* Cas9 DNA binding and cleavage of mismatched DNA:RNA sequences using a library of mismatched DNA target substrates reported similar findings[29]. Where mismatches in the seed region affect Cas9 DNA binding and mismatches in the PAM-distal region do not disrupt binding, but decrease DNA cleavage magnitude. Thus, Cas9 simply does not bind some off-target sites and, at others, bound Cas9 does not cleave effectively.

However, the effects of mismatches on binding and cleavage efficiency have not been studied systematically *in vivo.* We used a synthetic locus approach, where previously investigated on- and off-target sites were introduced into a genomic landing pad and EC50s were measured and compared to the endogenous on- and off-target sites (**Fig. 2.3A**)[133]. Binding of ciCas9 to on-

and off-target sequences in this synthetic locus closely resembled binding at the endogenous loci (**Figs. 2.3B-K; Table 2.4; Appendix B Fig. B.6**). Indel magnitudes at fully open ciCas9 (1 μM A115) were generally higher at the synthetic locus perhaps because the landing pad, integrated via lentiviral integration, is likely in the region of open chromatin.

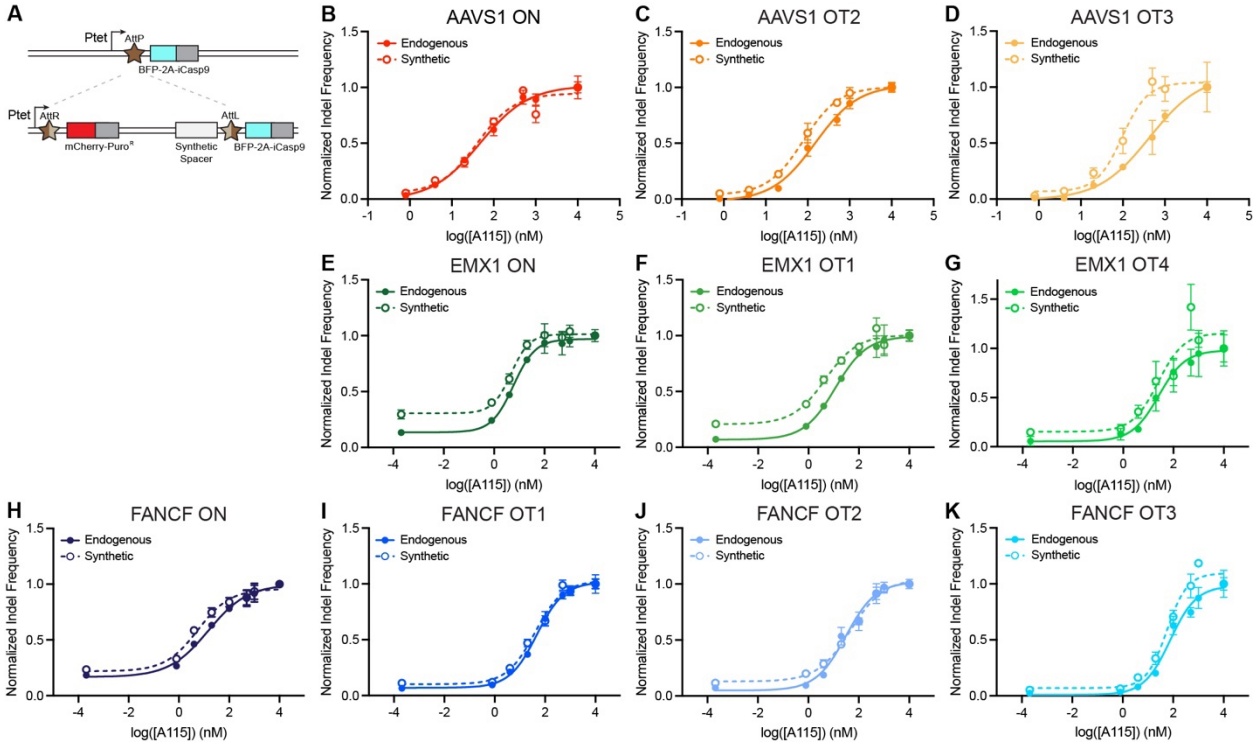

Figure 2.3. ciCas9 binds and edits synthetic locus target sites similar to endogenous target sites.

A) (Top) Schematic of HEK293T BFP iCasp9 landing pad system used to generate the synthetic loci. (Bottom) Schematic of recombined synthetic locus with an mCherry-Puro^R selection marker and the synthetic spacer target locus at the 3´ end of the recombined locus.

B-K) Normalized dose-response of ciCas9 editing at endogenous and synthetic loci for (**B**) AAVS1 ON, (**C**) AAVS1 OT2, (**D**) AAVS1 OT3, (**E**) EMX1 ON, (**F**) EMX1 OT1, (**G**) EMX1 OT4, (**H**) FANCF ON, (**I**) FANCF OT1, (**J**) FANCF OT2, and (**K**) FANCF OT3 synthetic locus HEK293T BFP iCasp9 landing pad cell lines. Indel frequencies normalized to the mean of indels measured at 10 μM A115. For all EMX1 and FANCF on- and off-target sites, indel frequency for DMSO was incorporated into EC_{50} calculations as 0.0002048 nM A115. Data represented as normalized mean indel frequency ± normalized SEM of 3 cell culture replicates.

Table 2.4. EC_{50} values from ciCas9 dose-response measured by indel frequencies at the endogenous and synthetic loci 12 hr after A115 addition.

Locus		Endogenous		Synthetic	
		EC_{50} (nM)	95% CI (nM)	EC_{50} (nM)	95% CI (nM)
AAVS1	ON	48.0	24.7-82.7	40.1	7.9-89.6
	OT2	151	94-247	78.1	51.9-117.9
	OT3	433	192-14769	96.0	31.6-209.8
EMX1	ON	5.82	3.42-9.93	4.68	2.57-8.30
	OT1	11.6	5.4-27.3	4.55	2.04-10.94
	OT4	25.8	6.5-314.6	24.5	3.9-133.4
FANCF	ON	14.8	7.1-38.3	5.61	2.78-14.27
	OT1	46.7	32.6-67.5	39.9	22.8-71.3
	OT2	30.8	15.9-64.5	47.6	27.0-87.1
	OT3	75.9	37.3-193.0	56.1	32.9-93.5

Next, we constructed a library of synthetic target sites containing single nucleotide mismatches throughout the entire 20 bp target sequence for each of the three sgRNAs investigated. We first validated the editing specificity of fully active WT-ciCas9 by comparing against WT-Cas9 specificity (**Appendix B Fig. B.7; Table 2.5**). The specificity of WT-ciCas9 is highly correlated with the specificity of WT-Cas9 in all three single mismatch libraries (AAVS1 Pearson r = 0.997, p<0.0001; EMX1 Pearson r = 0.978, p<0.0001; FANCF Pearson r = 0.928, p<0.0001; **Appendix B Fig. B.7B**). As previously observed, WT-Cas9 had greater editing magnitude compared to WT-ciCas9 at all single mismatch target sequences but was also highly correlated (**Appendix B Fig. B.7A; Table 2.5**). This reinforced the confidence in WT-ciCas9 to be able to discriminate against mismatched target sites similarly to WT-Cas9.

We then used the single mismatch libraries to measure editing specificity for every mismatch in the library, by computing the ratio of indel magnitudes at the perfectly matched target to the mismatched target sequence (PM:MM) (**Fig. 2.4; Appendix B Fig. B.8**). Mismatches within the ~8 bp seed region showed clear decreases in editing magnitude and higher specificity compared to mismatches in more PAM-distal positions (**Fig. 2.4A**). This is consistent with previous observations of mismatch tolerance and editing efficiency at off-target sequences[24–26]. Furthermore, position-specific patterns were observed that were consistent among all three

libraries. Interestingly, all three target sites show some mismatch tolerance in the first 2 nucleotides closest to the PAM, whereas mismatches in positions 3-6 upstream of the PAM were less tolerated and resulted in less editing and higher editing specificity (**Fig. 2.4B**). Mismatches at position 5 showed the lowest editing in all three libraries (**Fig. 2.4A**). In addition, the type of mismatches at position 5 that had the greatest effect were purine-purine mismatches in the AAVS1 and EMX1 libraries. FANCF contained a pyrimidine at position 5, and thus we could not measure the effects of a purine-purine mismatch. This suggests that purine-purine mismatches result in a base pairing that may be causing a steric clash with Cas9 that hinders the protein domain movements and affects its ability to cleave DNA. We further investigated the effects of mismatch type on the specificity of editing for all three libraries combined (**Appendix B Fig. B.8A**). We found no significant impact of the type of mismatch on the overall editing specificity. Although the individual single mismatch libraries showed significant differences in the effects of some mismatch types and indel frequency outcomes (**Appendix B Figs. B.8B-D**).

Next we explored whether single mismatch sequence discrimination occurred via altered Cas9 binding or cleavage efficiency mechanisms by applying our ciCas9 titration assay to select single mismatches that showed enhanced specificity. We found that single mismatches in the seed region decreased ciCas9 EC_{50} by up to 64-fold compared to the perfectly matched sequence (**Appendix B Figs. B.9-13; Table 2.6**). Thus, seed region mismatches impact Cas9 target site binding, resulting in decreased indel magnitudes. Mismatches in the PAM-distal region decreased indel magnitude with fully active ciCas9 but EC_{50} largely remained unchanged or only differed by 2- to 3-fold compared to the perfectly matched sequence. Thus, off-target discrimination is achieved by two synergistic mechanisms: loss of affinity for PAM-proximal mismatches and loss of cleavage efficiency for PAM-distal mismatches.

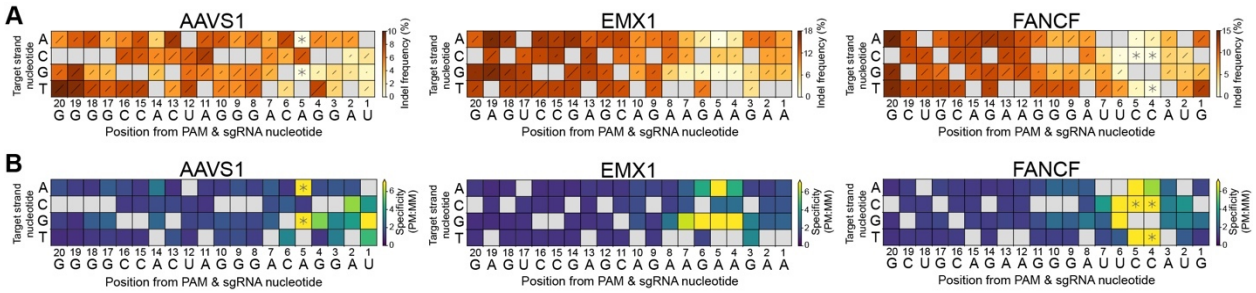

Figure 2.4. WT-ciCas9 editing in mismatched libraries.

A) WT-ciCas9 indel frequency at 10 μM A115 in the AAVS1, EMX1, and FANCF mismatch libraries. Lines in each box represent the relative SEM of 3 cell culture replicates. Mismatches where indel frequency < 0.3% with 10 μM A115 are denoted with *.

B) Specificity of WT-ciCas9 in the AAVS1, EMX1, and FANCF mismatch libraries. Specificity was calculated as a ratio of mean indel frequency with 10 μM A115 at the perfectly matched sequence to the mean indel frequency at each mismatched sequence. Mismatches where indel frequency < 0.3% with 10 μM A115 are denoted with *.

Table 2.5. Pearson r correlation between WT-ciCas9 and WT-Cas9 editing in mismatched libraries.

	AAVS1		EMX1		FANCF	
	Pearson r	**p**	**Pearson r**	**p**	**Pearson r**	**p**
Indel Frequency	0.885	<0.0001	0.938	<0.0001	0.942	<0.0001
Specificity (PM:MM)	0.997	<0.0001	0.978	<0.0001	0.928	<0.0001

Table 2.6. EC_{50} values from ciCas9 dose-response measured by indel frequencies at single mismatch synthetic target sites.

Locus		EC_{50} (nM)	95% CI (nM)
AAVS1	PM	31.3	10.1-276.2
	2C	283	197-471
	4G	137	79-311
	7C	59.5	Too large
	9T	100	Too large
	14C	578	Too large
EMX1	PM	2.28	0.43-38.11
	3G	50.5	Too large
	8G	146	Too large
	10G	3.57	1.49-9.85
	14G	6.66	2.26-18.89
	18A	2.79	0.96-8.77
FANCF	PM	11.1	6.7-19.6
	1A	33.7	18.7-74.1
	2C	50.1	33.9-72.5
	9G	23.1	15.1-36.3
	15G	12.3	7.4-22.2
	17G	8.43	5.47-13.26

2.3 Discussion

Here, we explored the mechanistic basis of on- versus off-target editing by Cas9 *in vivo* using our temporally and rheostatically controlled ciCas9 system. We found that Cas9 has similar overall editing kinetics at on- versus off-target sites and editing specificity does not change over time and that Cas9 binds to detected off-target sites with similar affinities as on-target sites. Therefore, lower magnitudes of editing at off-target sites *in vivo* is primarily due to Cas9's cleavage competency differences at on- versus off-target sites. DNA cleavage has previously been demonstrated to be a primary factor in Cas9 activity when bound to on- versus off-target DNA substrates *in vitro*[30–32]. Whereas mismatches in the seed region tend to disrupt stable DNA binding[29]. The endogenous off-target sites investigated in this study all contain at least one mismatch in the PAM-distal region with four of seven off-target sites also containing seed region

mismatches (**Appendix B Fig. B.4**). This suggests that some seed region mismatches can be tolerated in addition to PAM-distal mismatches with Cas9 binding at the off-target site with a similar affinity as the on-target site.

We further investigated the sequence determinants for *in vivo* off-target discrimination via binding or cleavage efficiency mechanisms using mismatch libraries containing every possible single mismatch for three different target sequences. We find that mismatches in positions 3-6 upstream of the PAM are the least tolerated. None of the endogenous off-target sites in this study contain mismatches in these positions. We coupled our mismatch library specificity analysis with *in vivo* binding affinity measurements at select single mismatches and showed the previously noted seed region versus PAM-distal region binding affinity differences[29,128]. PAM-distal mismatches do not disrupt Cas9 binding, but rather affect the cleavage competency of Cas9.

Although ciCas9 allows us to probe the kinetics and relative binding affinity of Cas9 at on- versus off-target sites *in vivo*, limitations remain to fully capture the mechanisms of Cas9 off-target discrimination *in vivo*. Our assays utilize indel frequency as a measurement of ciCas9 activity which is a result of both DSB generation and subsequent DNA repair. To directly measure ciCas9 activity, we would need to measure DSB outcomes. However, off-target DSB generation is often too low to be detected with our previously developed DSB-ddPCR assay. More sensitive quantitative DSB detection methods for a targeted locus need to be developed to enable more accurate measurements of ciCas9 activity, particularly at the low A115 doses in the *in vivo* binding affinity assay.

We have shown that ciCas9 offers a unique insight into the mechanisms of Cas9 off-target discrimination *in vivo*. The sequence determinants for off-target discrimination mechanisms *in vivo* largely match those that have been detailed in previous *in vitro* studies. The use of ciCas9-based *in vivo* assays to probe Cas9 mismatch tolerance mechanisms allows for better understanding of editing mechanisms in the context in which the systems are used. We envision this approach could better inform future engineering of high fidelity Cas9 variants.

2.4 Materials and methods

Expression plasmids

Mammalian expression of Cas9 with an N-terminal Flag sequence cloned into the pcDNA5/FRT/TO backbone (ThermoFisher) was previously reported[134]. Note that this study used the L22 BH3 variant of ciCas9 that we have previously reported[121]. The L22 variant was also used in all subsequent high-fidelity and transcriptional activation versions of ciCas9. Mammalian expression of ciCas9 cloned into the pcDNA5/FRT/TO backbone is available on Addgene (plasmid 100551). The ciCas9 pcDNA5/FRT/TO expression plasmid was used in the time course and titration studies at endogenous loci. For editing at the synthetic loci and the synthetic loci mismatch libraries, ciCas9 was subcloned into the pcDNA3.1 backbone (ThermoFisher) linearized using the BamHI and EcoRV restriction sites. Hypa-ciCas9 was cloned using site directed mutagenesis by primer extension and Gibson assembly into the pcDNA5/FRT/TO backbone linearized using the BamHI and EcoRV restriction sites for editing at endogenous loci. Hypa-ciCas9 was subcloned into the pcDNA3.1 backbone linearized using the BamHI and EcoRV restriction sites for editing in the synthetic loci mismatch libraries.

dciCas9-VPR for transcriptional activation was cloned in two steps. Gibson assembly was used to create the D10A and H935A (dCas9 H840A equivalent) mutations for dciCas9. Then dciCas9 was subsequently assembled with PCR-amplified VP64-p65-Rta from pEF045, a gift from Jesse Zalatan, using Gibson assembly into the pcDNA5/FRT/TO backbone linearized using the BamHI and EcoRV restriction sites. For transcriptional activation in the HEK293 TREx FlpIn EMX1-EGFP reporter, dciCas9-VPR was subcloned into the pcDNA3.1 backbone linearized using the BamHI and EcoRV restriction sites.

All sgRNAs were cloned into Addgene plasmid #41824 containing a U6 promoter and sgRNA scaffold, a gift from George Church. The AAVS1, EMX1, FANCF, and CXCR4 sgRNA plasmids have been previously reported[100,105,121]. Briefly, a single-stranded DNA (ssDNA) oligo

with overlap to the vector 5′ and 3′ of the 20 nt target sequence was purchased from Integrated DNA Technologies (IDT). The ssDNA oligo was Gibson assembled into the gRNA cloning vector linearized using the AflII restriction site.

Mammalian cell culture

HEK293T cells (ATCC, CRL-3216) and HEK293 TREx FlpIn cells (ThermoFisher) were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS, MilliporeSigma) and incubated at 37 °C, 5% CO_2. HEK293 TREx FlpIn EMX1-EGFP reporter cells were cultured in DMEM supplemented with 10% (v/v) FBS and 50 µg/mL hygromycin (Mirus) and incubated at 37 °C, 5% CO_2. HEK293T BFP iCasp9 LLP cells were cultured in DMEM supplemented with 10% (v/v) FBS and 2 µg/mL doxycycline (ThermoFisher) and incubated at 37 °C, 5% CO_2. Cells were tested and found negative for mycoplasma at least every 6 months.

Cas9 editing of endogenous genomic loci

For the 24 vs. 48 hr Cas9 editing, HEK293T cells were plated in 24-well plates at a density of 1.5×10^5 cells/well. The following day, cells were transfected with 250 ng ciCas9, 125 ng sgRNA, and 125 ng pMAX-GFP transfection control using 3 µL Turbofectin 8.0 (Origene) per well. 24 hr after transfection, cells were passaged 1:4 to allow for more growth. At the appropriate time points, cells were harvested by resuspending the cells in growth media. Resuspended cells were then pelleted at 1,500 × g for 10 min. Growth was removed and cell pellets were stored at -80 °C until genomic DNA extraction.

For the WT- vs. Hypa-Cas9 comparison, HEK293T cells were plated in 12-well plates at a density of 2×10^5 cells/well. The following day, cells were transfected with 450 ng WT- or Hypa-Cas9, 450 ng sgRNA, and 100 ng pMAX-GFP transfection control using 3 µL Turbofectin 8.0

(Origene) per well. 24 hr after transfection, cells were harvested after washing with 1 mL of ice-cold DPBS and then resuspending cells in 600 µL ice-cold DPBS. Resuspended cells were pelleted at 1,500 × g for 10 min at 4 °C. DPBS was removed and cell pellets were stored at -80 °C until genomic DNA extraction.

ciCas9 editing of endogenous genomic loci

For the ciCas9 time course and WT- vs. Hypa-ciCas9 comparison, HEK293T cells were plated in 12-well plates at a density of 2×10^5 cells/well. The following day, cells were transfected with 450 ng ciCas9, 450 ng sgRNA, and 100 ng pMAX-GFP transfection control using 3 µL Turbofectin 8.0 (Origene) per well. 24 hr following transfection, A-1155463 (ChemieTek) dissolved in DMSO was added to each well to a final concentration of 1 µM for the time course experiments, final [DMSO] of 0.1%. At the appropriate time points, cells were harvested after washing with 1 mL of ice-cold DPBS and then resuspending cells in 600 µL ice-cold DPBS. Resuspended cells were pelleted at 1,500 × g for 10 min at 4 °C. DPBS was removed and cell pellets were stored at -80 °C until genomic DNA extraction.

For the ciCas9 titrations at endogenous loci, the same steps were performed as the ciCas9 time course, except the cells were transfected with 225 ng ciCas9, 225 ng sgRNA, and 50 ng pMAX-GFP transfection control using 1.5 µL Turbofectin 8.0 (Origene) per well. A115 was diluted to the appropriate concentrations using DMSO and added to cells with a final [DMSO] of 0.1%.

Indel quantification using next generation sequencing

Genomic DNA isolation, sequencing, and indel frequency analysis for non-library loci were performed as previously described[121]. Briefly, genomic DNA was extracted from cells using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's protocol with an extended proteinase K digestion of 1 hr at 56 °C. The loci of interest was first amplified with 15 cycles of

PCR from 2 µL (~100-300 ng) of genomic DNA eluate using a 5 µL Kapa HiFi HotStart polymerase reaction (Roche). The first PCR was then diluted with 25 µL of DNAse-free water. Indexes and Illumina cluster generation sequences were then added with a secondary PCR reaction using 3 µL of the diluted primary PCR product with a 10 µL Kapa Robust HotStart polymerase reaction (Roche) for 20 cycles. The final amplicons were run on a 1% TBE-agarose gel and DNA was extracted using a Freeze and Squeeze column (BioRad) according to the manufacturer's protocol. Gel extracted amplicons were quantified using the Qubit dsDNA HS assay kit (Invitrogen). Up to 1000 indexed amplicons were pooled and sequenced on a NextSeq 550 using a NextSeq Mid 150 v2/v2.5 kit (Illumina).

For ciCas9 editing in the AAVS1 and FANCF mismatched libraries, cell pellets were split into 8 genomic DNA extractions. The synthetic locus was amplified using 5 cycles of PCR from 2.5 µg of genomic DNA using a 50 µL Q5 HotStart polymerase reaction (New England Biolabs). Each genomic DNA sample for the AAVS1 and FANCF libraries were amplified using 8 separate PCR reactions. The first PCR was then individually cleaned up using AMPure XP beads (Beckman Coulter) and PCR product from each reaction was eluted into 15 µL of DNAse-free water. Indexes and Illumina cluster generation sequences were then added with a secondary PCR reaction using all 15 µL of PCR product from the primary PCR. The secondary PCR reaction was also performed with 8 separate reactions, each with a 50 µL Q5 HotStart polymerase reaction tracked on a RT-qPCR machine using SYBR Gold (Invitrogen) and stopped when the RFU reached a level previously determined to be observable on an agarose gel. The 8 separate secondary PCR reactions were then pooled according to their final RFU values and run on a 2% TAE-agarose gel. DNA was extracted using a Freeze and Squeeze column according to the manufacturer's protocol. Gel extracted amplicons were quantified using the Qubit dsDNA HS assay kit. Up to 30 indexed amplicons were pooled and sequenced on a NextSeq 2000 using a NextSeq P3 100 cycle v3 kit (Illumina).

For the ciCas9 editing in the EMX1 mismatch library and all Cas9 editing in the mismatch

libraries, cell pellets were extracted using a single column with the DNeasy Blood and Tissue kit. The same amplification conditions were used as above except only 2 separate PCR reactions were performed per gDNA sample.

Quantification of indel frequency for all experiments were performed with custom scripts, available at github.com/cindytxwei/. Dose response fitting and measurement of EC_{50} were performed using a non-linear fit of log(agonist) vs. response - variable slope calculation in GraphPad Prism 9.

EGFP reporter construction for dciCas9-VPR transcriptional activation

The 20 bp EMX1 target and PAM sequence and 20 bp of endogenous gDNA sequence 5′ and 3′ of the target were cloned using Gibson assembly into a pcDNA5-FRT-TO backbone without a CMV promoter. 3′ of the target sequence is a synthetic minimal promoter followed by 100 bp of random DNA sequence and an EGFP reporter gene (**Appendix B Fig. B.3**)[135]. The plasmid containing this locus was then transfected into HEK293 TRex Flp-In cells (ThermoFisher) along with a pOG44 plasmid encoding a Flp-recombinase using Turbofectin according to manufacturer's protocols. Cells with successful integration of the reporter locus were selected using hygromycin (Mirus Bio).

dciCas9-VPR transcriptional activation

For CXCR4 transcriptional activation, 6×10^4 HEK293T cells were seeded in 12-well plates. ~20-24 hr after seeding, each well of cells was transfected with 1 µg total dciCas9-VPR and sgRNA plasmids targeting the CXCR4 locus (450 ng dciCas9-VPR, 450 ng CXCR4 sgRNA, 100 ng mCherry control) using Turbofectin (Origene) according to the manufacturer's protocols. ~24 hr after transfection, 1 µM of A115 was added to the cells, final [DMSO] of 0.1%. 48 hr after A115 addition, cells were harvested and incubated with APC anti-human CD184 (CXCR4) [12G5] (BioLegend) for 1 hr and then fluorescence was analyzed on the LSRII flow cytometer (BD

Biosciences). 30,000 single cell events were collected for each sample. The median APC fluorescence is reported for the brightest 25% of cells expressing mCherry transfection control. All flow cytometry data was analyzed using FlowJo. See **Appendix B Fig. B.12** for cell gating strategies.

For EGFP reporter transcriptional activation, a similar protocol as CXCR4 activation was used except dciCas9-VPR was targeted using an EMX1 sgRNA and no antibody incubation was performed, cells were directly analyzed for EGFP fluorescence by flow cytometry. See **Appendix B Fig. B.13** for cell gating strategies.

Synthetic loci of known on- and off-target sequences and select single mismatch target sequences

For the known on- and off-target sequences, a ssDNA oligo containing an on- or off-target sequence was annealed to a ssDNA oligo containing a 16 bp degenerate barcode. The annealed DNA was then made double-stranded using Klenow (exo-) polymerase (New England Biolabs). The locus was then cloned using Gibson assembly into a vector containing an AttB recombination sequence followed by mCherry-P2A-Puro^R used for selection of recombined cells. Plasmid DNA was extracted from 20 *E. coli* colonies and then pooled to create a plasmid pool with the same target locus corresponding to 20 different 16 bp barcode sequences.

For the single mismatch target sequences, a gBlock containing the target sequence and a random 16 bp barcode sequence was purchased from IDT. The gBlock was then cloned using Gibson assembly into a vector containing an AttB recombination sequence followed by mCherry-P2A-Puro^R. A single colony was selected for each single mismatch target sequence.

HEK293T BFP iCasp9 lenti-landing pad cells were seeded in no doxycycline media and transfected on the same day. The AttB plasmid containing the synthetic locus was co-transfected with pCAG-NLS-Bxb1 at a 15:1 (w/w) plasmid ratio, total 1.895 µg of plasmid DNA, for recombination into the landing pad site in the cells. 2 days after transfection, doxycycline was

added to induce expression of mCherry for evaluation of recombination efficiency. 2 days after adding doxycycline, cells were analyzed for mCherry expression using flow cytometry to evaluate recombination efficiency. On the same day, AP1903 (MedChem Express) was added to the cells to a final concentration of 10 nM to induce cell death in non-recombined cells. 1 day after adding AP1903, the cell culture media was changed to remove dead cells. 2 days after adding AP1903, cells were analyzed by flow cytometry to evaluate enrichment of recombined cells. If cells were <80% mCherry positive, cells were sorted using fluorescence activated cell sorting on the Aria III (BD Biosciences) to enrich for recombined cells. Recombined and enriched cells were then seeded for ciCas9 titrations following the ciCas9 editing of genomic loci method above.

ciCas9 editing in single synthetic loci experiments were performed following the same protocol as editing at endogenous loci.

Mismatch synthetic locus library construction

A ssDNA oligo containing the AAVS1, EMX1, or FANCF target sequence with a 6% hand-mixed non-WT nucleotide at each position of the 20 nt target sequence was purchased from Integrated DNA Technologies. This oligo library was annealed to another ssDNA oligo containing a 16 bp degenerate barcode. The annealed DNA was then made double-stranded using Klenow (exo-) polymerase (New England Biolabs). The locus was then cloned using Gibson assembly into a vector containing an AttB recombination sequence followed by mCherry-P2A-Puro^{R} used for selection of recombined cells. The library Gibson assemblies were electroporated into NEB10β electrocompetent *E. coli* (New England Biolabs). The library was bottlenecked at the transformation recovery to contain ~100,000 barcodes per library. Plasmid DNA was extracted using a GeneJET Plasmid Midiprep Kit (ThermoFisher). Plasmid libraries were then sequenced on a NextSeq 550 using a NextSeq Mid 300 v2/v2.5 kit (Illumina) to subassemble mismatched target sequences with their corresponding 16 bp barcode sequence.

To recombine the single mismatch libraries into HEK293T BFP iCasp9 lenti-landing pad

cells, 6.72×10^7 cells were seeded for the AAVS1 library, 7×10^6 cells were seeded for the EMX1 library, and 5.04×10^7 cells were seeded for the FANCF library. The AttB plasmids containing the mismatch synthetic locus libraries were co-transfected with pCAG-NLS-Bxb1 at a 15:1 (w/w) plasmid ratio for recombination, total 18.95 µg plasmid DNA for the AAVS1 and FANCF libraries and 7.58 µg plasmid DNA for the EMX1 library. 2 days after transfection, doxycycline was added to induce expression of mCherry for evaluation of recombination efficiency. 1 day after adding doxycycline, cells were analyzed for mCherry expression using flow cytometry to evaluate recombination efficiency. On the same day, AP1903 (MedChem Express) was added to the cells to a final concentration of 10 nM to induce cell death in non-recombined cells. 1 day after adding AP1903, the cell culture media was changed to remove dead cells. 2 days after adding AP1903, cells were analyzed by flow cytometry to evaluate enrichment of recombined cells. After selection of cells recombined with the single mismatch libraries with AP1903, cells were expanded until enough to seed a ciCas9 editing experiment.

ciCas9 editing of synthetic single mismatch libraries

For ciCas9 editing in the AAVS1 and FANCF libraries, 8.2×10^6 cells were seeded into 15 cm dishes in triplicate. 24 hr after seeding cells, 18.9 µg ciCas9 in pcDNA3.1(+), 18.9 µg sgRNA for the corresponding library, and 4.2 µg of pMAX-GFP were transfected into each 15 cm dish of cells. For the EMX1 library, 4.05×10^5 cells were seeded into 6-well plates in triplicate. 24 hr after seeding cells, 1.233 µg ciCas9 in pcDNA3.1(+), 1.233 µg EMX1 sgRNA, and 0.274 µg pMAX-GFP were transfected into each well of cells. Transfections were performed using Turbofectin following the manufacturer's protocols. 24 hr after transfection, A115 was added to each 15 cm dish to a final concentration of 10 µM. 12 hr after A115 addition, cells were harvested for gDNA extraction.

For Cas9 editing in all the single mismatch libraries, 4.05×10^5 cells were seeded into 6-well plates in triplicate. 24 hr after seeding cells, 1.233 µg ciCas9 in pcDNA3.1(+), 1.233 µg

sgRNA corresponding to each library, and 0.274 μg pMAX-GFP were transfected into each well of cells. Transfections were performed using Turbofectin following the manufacturer's protocols. 24 hr after transfection, cells were harvested for gDNA extraction.

Chapter 3: A versatile chemically-controlled DNA binding switch allows simple engineering of temporally-controlled Cas9-based effectors

A version of this chapter is in preparation for publication.

Wei, C.T., Maly, D.J., Fowler, D.M.

3.1 Introduction

CRISPR-Cas systems have revolutionized precise genome targeting. The CRISPR-Cas9 system consists of a Cas9 endonuclease that can be targeted to nearly any location within a genome using a single guide RNA (sgRNA) that encodes a 20 nucleotide sequence complementary to a specific genomic locus[16,126,136]. Targeting of the CRISPR-Cas9 system can be tuned by simply changing the 20 nucleotide sequence within the sgRNA. The CRISPR-Cas9 system is commonly used to create genomic double-strand breaks (DSBs) to facilitate incorporation of desired DNA edits at specific loci via homology directed repair (HDR) or to generate indels to knock out specific genetic elements via non-homologous end joining (NHEJ)[11].

Engineered versions of the CRISPR-Cas9 system have enabled a plethora of applications beyond DSB generation for DNA editing[11]. For example, catalytically-inactive Cas9 (dCas9) has been fused to transcriptional activators and chromatin modifiers for targeted epigenome editing[13]. Nickase Cas9 (nCas9) has been fused to various DNA deaminase enzymes yielding DNA base editors[55]. DNA base editors rely on the DNA unwinding activity of nCas9 to expose a single-stranded DNA (ssDNA) substrate for the deaminases to convert cytidine to uracil or adenine to inosine[56,60]. DNA repair, induced by the nCas9-mediated DNA nick, then converts the uracil and inosine into thymine and guanine, respectively, thus resulting in a cytidine to thymine (C-to-T) or adenine to guanine (A-to-G) DNA base edit. Dual C-to-T and A-to-G base editors and C-to-G base editors have also been engineered[66–70]. Recently, prime editing has been developed to

introduce precise DNA edits using an RNA template encoding the desired edit[48]. Prime editing relies on an extended prime editing guide RNA (pegRNA) that encodes both the 20 nucleotide genome targeting sequence and a template for the desired edit. The pegRNA is coupled with nCas9 fused to a reverse transcriptase that uses the pegRNA template to incorporate the desired edit into the genome. Thus, the CRISPR-Cas9 system has proven to be extraordinarily versatile.

Precisely targeting and modifying specific loci is important to studying genotype-phenotype relationships, but the ability to temporally control DNA targeting is essential to study genetic elements that are temporally regulated. Engineered Cas9 variants have been developed to provide temporal control over targeted DSB generation and subsequent DNA editing[100,107,110–120]. Such temporally-controlled Cas9s have been used in a variety of applications, including studying the kinetics of CRISPR-Cas9 DNA editing, studying the kinetics of DNA repair after a DSB, and engineering systems that record biological events in cells[100–103]. Beyond temporally-controlled Cas9s for generating DSBs, temporally-controlled versions of other Cas9-based effectors have also been engineered. Examples include temporally-controlled dCas9-based DNA transcription and chromatin modifiers capable of turning on or off gene expression[104,105] and split-engineered base editors (seBEs) that allow for temporal control of C-to-T base editing[106]. Having precise temporally-controlled Cas9-based effectors would provide insight into the time scales on which the effectors act. For example, temporally-controlled base editors would provide insight into the formation of unwanted bystander editing and allow for greater predictability of edited DNA variant outcomes in pooled variant-function studies.

Although temporal control of some Cas9-based effectors has been achieved, existing systems comprise a patchwork of approaches that do not cover all important Cas9-based effector activities. Moreover, these existing systems are complicated by the fact that they use multiple protein components, often requiring screening of split enzymes to confer temporal control over specific Cas9-based effector domains. These current approaches can be laborious to engineer and are not easily applicable to all effectors used with Cas9. A generalizable system to engineer

temporal control for all Cas9-based effector activities would be based on a single component, would be rapidly activatable, and would allow precise tuning of activity. Since RNA-guided Cas9 binding to DNA is common to all Cas9-based effectors, we hypothesized that control of Cas9 DNA binding activity would enable engineering of all Cas9-based effector systems. We previously developed a single component temporally-controlled Cas9 protein, ciCas9, that contains a tightly autoinhibited switch that can be rapidly activated with a potent small molecule[100].

Here, we show that the ciCas9 switch can serve as a general platform for conferring temporal control over a wide range of Cas9-based effectors. We develop a chemically-controlled transcription factor, dciCas9-VPR, and use it to show that the ciCas9 switch functions as a chemically-controlled, two-state switch governing DNA target site binding. We then show that the ciCas9 switch can be used to engineer chemically-controlled base editors, allowing robust temporal control over C-to-T and A-to-G DNA editing. We use these chemically-controlled base editors to examine for the first time how nucleotide position within a target site affects the kinetics of base editing, and we reveal how early base editing kinetics affect the magnitude of editing outcomes. We also dissect the kinetics of multiple edits at several loci, elucidating the order in which nucleotides are edited and revealing how base editing at one nucleotide in the target site influences the appearance of bystander edits at other nucleotides within the same target. Finally, we highlight the versatility of the ciCas9 switch by engineering chemically-controlled dual A-to-G and C-to-T base editors and DNA prime editors whose activity can be controlled with high temporal precision.

3.2 Results

3.2.1 The ciCas9 switch acts as a chemically-controlled DNA target binding switch

We previously developed a chemically-controlled Cas9 variant (ciCas9) that uses an engineered protein-protein interaction to enable small-molecule control of Cas9's activity[100]. In

our system, the REC2 domain was replaced with Bcl-xL and a BH3 peptide was appended to the C-terminus of Cas9[100,121,131]. Bcl-xL and BH3 form a tight interaction that inhibits Cas9 activity (**Fig. 3.1A**). In the basal state, autoinhibited ciCas9 possesses low activity, but addition of a small molecule (A-1155463, hereafter A115) disrupts the interaction between Bcl-xL and the BH3 peptide resulting in dose-dependent generation of double-stranded breaks (DSBs) at target sites within minutes[100]. We reasoned that the single-protein architecture and rapid activation kinetics of ciCas9 could serve as a versatile platform for conferring chemical control over diverse Cas9 effector activities. Successful application of the ciCas9 switch to other Cas9-based effectors requires that the switch modulates DNA target site binding as opposed to another mechanism such as altering Cas9 enzymatic activity (**Fig. 3.1A**). To test whether ciCas9's autoinhibitory switch controls DNA target site binding *in vivo*, we measured transcriptional activation, which relies on Cas9 localization through DNA binding rather than Cas9 nuclease activity. Thus, we fused the transcriptional activator Vp64-p65-Rta (VPR) to the C-terminus of catalytically dead ciCas9 (dciCas9) and tested its ability to promote expression of CXCR4 (**Fig. 3.1B-C**). A115 treatment of HEK-293T cells expressing dciCas9-VPR resulted in induction of CXCR4 expression, supporting the DNA-blocking autoinhibition mechanism of dciCas9. Consistent with ciCas9 acting as a chemically-controlled DNA target site binding switch, we also found that unmodified dciCas9 functions in a multi-protein component transcriptional activation assay with previously reported scaffold RNAs (scRNA) (**Appendix C Fig. C.1**)[137]. Therefore, ciCas9 is unable to bind DNA in its autoinhibited state, and release of autoinhibition by A115 addition allows ciCas9 to bind. Finally, we demonstrated that dciCas9-VPR transcriptional activation could be dose-dependently tuned by targeting dciCas9-VPR to a synthetic EGFP reporter in the presence of different amounts of A115 (**Fig. 3.1D; Appendix C Fig. C.2**). Thus, the ciCas9 switch modulates DNA binding—a process common to all engineered Cas9-based effectors—which we can regulate to confer temporal control.

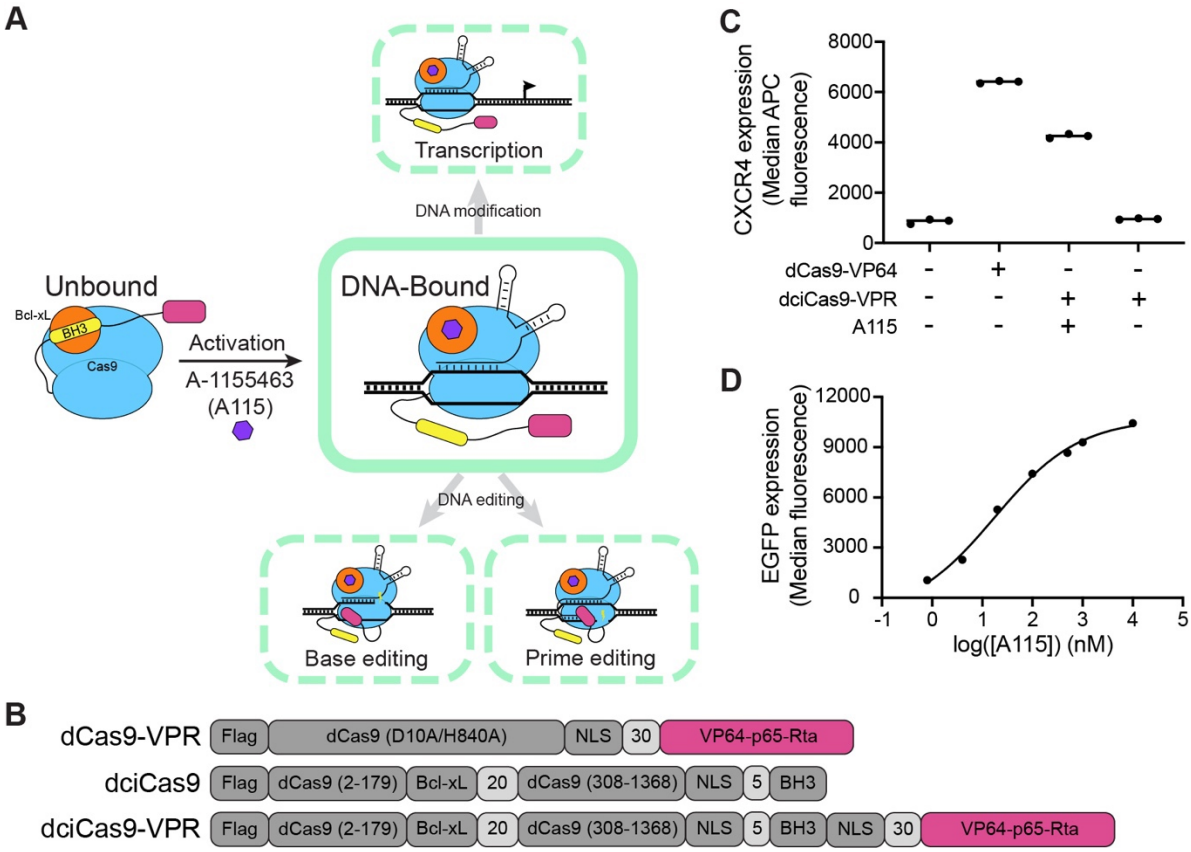

Figure 3.1. The ciCas9 switch can be used as a framework to create chemically-controlled Cas9-based effectors.

A) Schematic describing how the ciCas9 switch can be used to engineer different chemically-controlled, Cas9-based effectors and editors.

B) Schematic of the domains in catalytically-dead Cas9 (dCas9) and ciCas9 (dciCas9) fused to transcriptional activator VP64-p65-Rta (VPR).

C) Activation of CXCR4 expression using dCas9-VPR or dciCas9-VPR targeted to the promoter region in HEK293T cells in the presence or absence of 1 μM A115 activation for 48 hr. Cells were stained with CXCR4 antibody. Three cell culture replicates are shown, with a line indicating the mean.

D) Activation of EGFP expression using dciCas9-VPR and a range of A115 doses added to HEK293 TREx FlpIn cells that contain an EGFP reporter locus downstream of an EMX1 target sequence (EMX1-EGFP). Cells were treated with A115 for 48 hr prior to flow cytometry analysis. Points represent the mean of median EGFP fluorescence ± SEM of three cell culture replicates. Line shows fitting of dose-response data using a non-linear fit of log(agonist) vs. response - variable slope calculation in GraphPad Prism 9.

3.2.2 The ciCas9 switch can be used to create chemically-controlled DNA base editors

To further explore the utility of the ciCas9 switch, we created chemically-controlled cytidine base editors by fusing deaminases from the BE4max and AncBE4max base editors[138] to a ciCas9

nickase (nciCas9), preserving the original domain arrangements (**Fig. 3.2A**). We then transfected the chemically-controlled cytidine base editors into HEK-293T cells and fully activated them with a high concentration of A115 (1 μM) to test their maximum base editing output (**Fig. 3.2B-C; Appendix C Fig. C.3**). We quantified base editing outcomes using targeted-amplicon next-generation sequencing and observed robust A115-activated editing with modest background editing for both chemically-controlled cytidine base editors after 24 and 72 hr of activation with A115. The HEK3 site accumulated more background edits than the EMX1 target site. In an attempt to maximize overall editing and reduce background, we modified the epitope tag, nuclear localization sequence, peptide linker lengths, and codon optimization of the chemically-controlled cytidine base editors (**Appendix C Figs. C.3-5**). However, these factors did not have an appreciable impact.

We reasoned that background editing was due to the nciCas9 switch not being sufficiently closed, since transient dissociations of the BH3 from Bcl-xL complex would allow DNA binding and subsequent base editing. Thus, to minimize background, we tested if a higher affinity BH3 peptide variant, F22, would provide greater autoinhibition of nciCas9 activity (**Figs. 3.2D-E**)[121]. The higher affinity F22 variant did not appreciably reduce editing at 24 or 72 hr after A115 addition compared to the lower affinity L22 variant (**Figs. 3.2B-E**). However, the F22 variants of both chemically-controlled cytidine base editors demonstrated lower background editing (DMSO only) at both the EMX1 and HEK3 target sites. To verify F22 variant performance, we evaluated both editors at the ABE9 and HEK2 target sites, where we observed similarly low background (**Figs. 3.2F-G**). Thus, the dynamic range of chemically-controlled cytidine base editors can be increased by strengthening the autoinhibitory interaction between Bcl-xL and the BH3 peptide (**Appendix C Figs. C.4C, C.4E, C.5C, C.5E**). We used the F22 variants of both cytidine base editors, which we hereafter refer to as ciBE4max and ciAncBE4max.

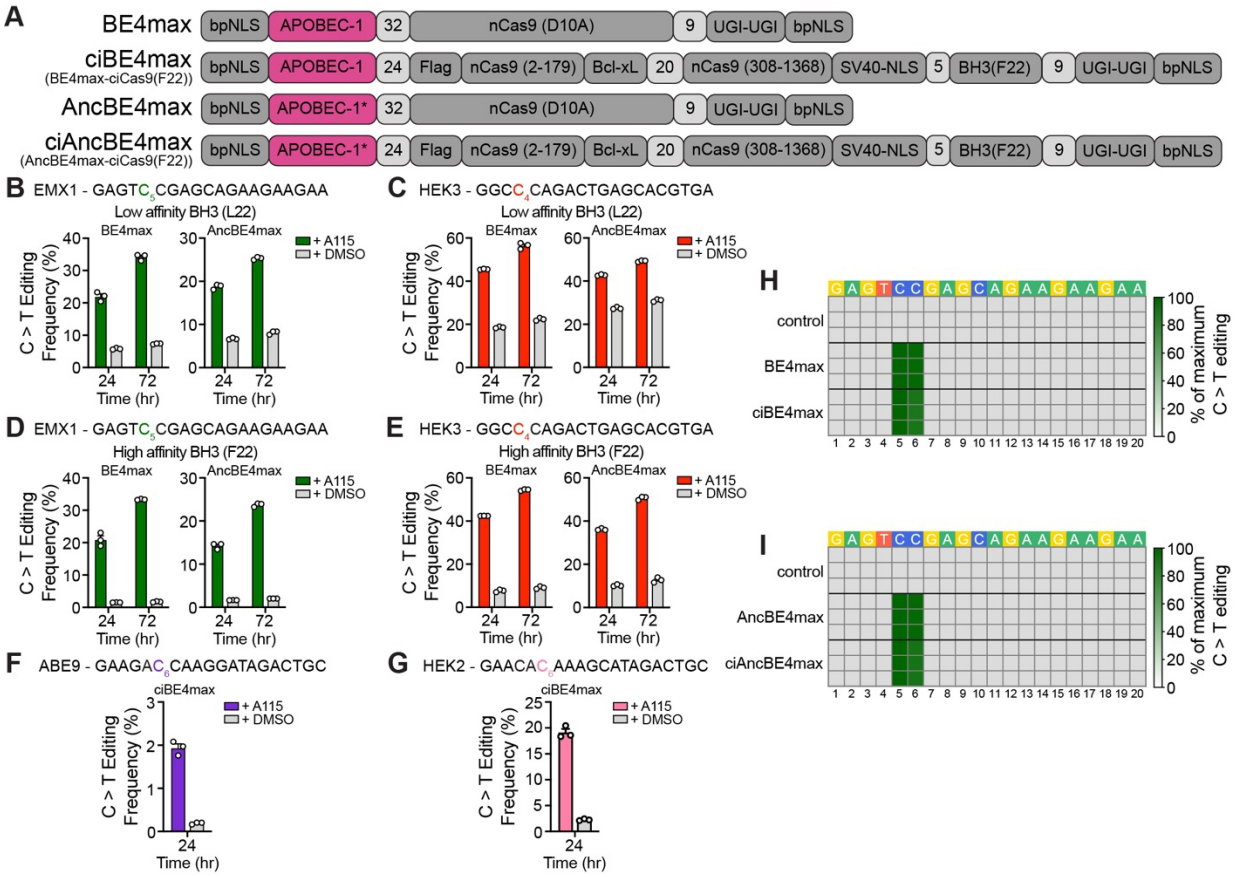

Figure 3.2. The ciCas9 switch can be used to create chemically-controlled cytidine base editors.

A) Schematic of the domain arrangements in the unmodified BE4max and AncBE4max base editors and chemically controlled versions that included codon optimization and the ciCas9(F22) variant.

B, C) C-to-T editing frequencies with the chemically-controlled BE4max and AncBE4max base editors with a low affinity BH3 peptide switch (L22) at the EMX1 **(B)** and HEK3 **(C)** target sites.

D, E) C-to-T editing frequencies with the chemically-controlled BE4max and AncBE4max base editors with a high affinity BH3 peptide switch (F22) at the EMX1 **(D)** and HEK3 **(E)** target sites.

F, G) C-to-T editing frequencies with the ciBE4max and ciAncBE4max base editors at the ABE9 **(F)** and HEK2 **(G)** target sites.

In **(B-G)** chemically-controlled BE4max and AncBE4max editing are quantified at the indicated time point after 1 μM A115 or DMSO addition to HEK293T cells. Editing is quantified at a single nucleotide within a target site that is colored in the target sequence. Bars show mean editing frequency ± SEM of three cell culture replicates, with white circles showing individual replicates.

H-I) Heatmaps of BE4max, ciBE4max **(H)** and AncBE4max, ciAncBE4max **(I)** editing as a percentage of the highest edited nucleotide for each editor throughout the entire EMX1 target site. Each row shows an individual cell culture replicate. BE4max and AncBE4max editing frequencies were quantified at 72 hr after transfection and ciBE4max and ciAncBE4max editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK293T cells. The control shows untransfected cells harvested at the same time as ciBE4max and ciAncBE4max. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.

We next engineered chemically-controlled adenine base editors by fusing either the ABEmax or ABE8e deaminases to the nciCas9 switch in the same domain arrangements as the unmodified ABEmax and ABE8e base editors (**Fig. 3.3A**)[57,138]. We observed robust editing for both chemically-controlled adenine base editors when activated with 1 μM A115 (**Figs. 3.3B-E**). Similar to the chemically-controlled cytidine base editors, the higher affinity F22 BH3 variant was able to improve the dynamic range of base editing (**Figs. 3.3B-E**) compared to the low affinity L22 variant and other protein modifications (**Appendix C Figs. C.6-8**). Comparison of the ratio of editing frequency in the presence of A115 to background editing at all target sites also reveals that the F22 variant demonstrates a greater dynamic range (**Appendix C Figs. C.7D, C.7G, C.8D, C.8G**). The F22 variants of both inducible adenine base editors demonstrated a suitable dynamic range of chemically-controlled base editing at the ABE16, ABE9, and HEK2 target sites, but limited chemical control was observed at the HEK3 locus due to high background in the absence of A115. This higher background editing was also observed at the HEK3 target site for cytidine chemically-controlled base editors. Thus, lack of chemical control at the HEK3 target site is locus-specific, not a property of the nciCas9 switch. As for the adenine base editors, we used the codon optimized chemically-controlled adenine base editors containing the F22 BH3 variant, ciABEmax and ciABE8e, for all subsequent experiments.

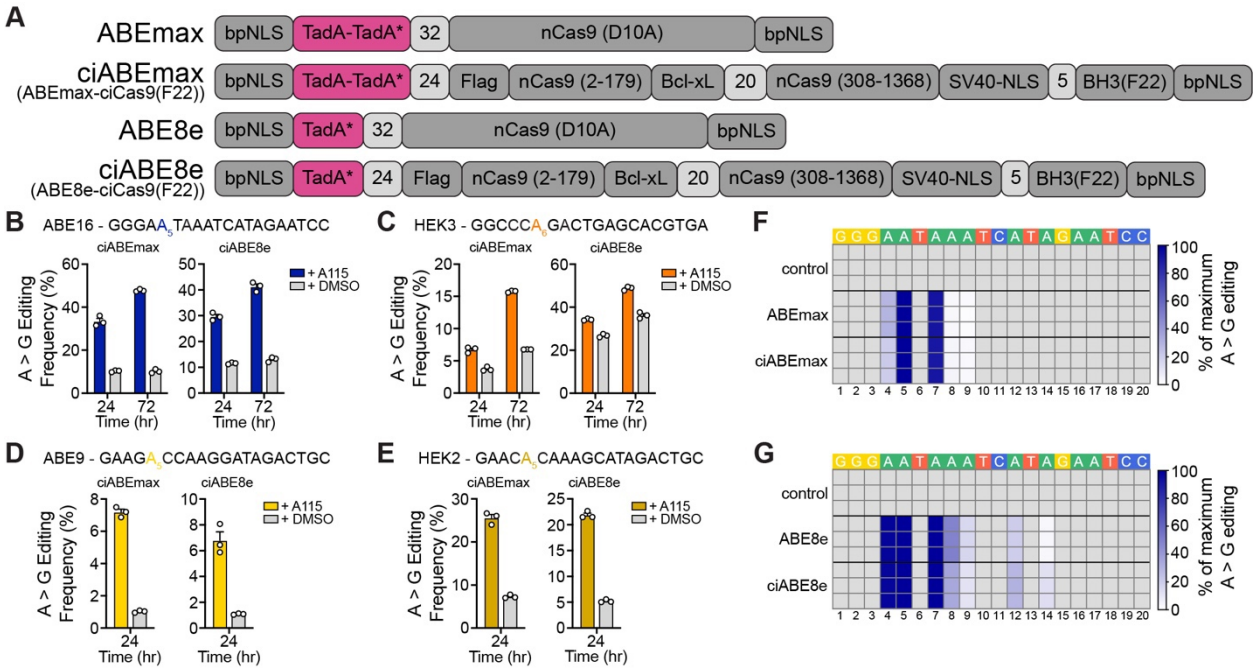

Figure 3.3. The ciCas9 switch can be used to create chemically-controlled adenine base editors.

A) Schematic of the domain arrangements in the unmodified ABEmax and ABE8e base editors and the chemically-controlled ciABEmax and ciABE8e base editors with codon optimization and using the ciCas9(F22) variant.

B-E) A-to-G editing frequencies with ciABEmax and ciABE8e base editors at the ABE16 **(B)**, HEK3 **(C)**, ABE9 **(D)**, and HEK2 **(E)** target sites. ciABEmax and ciABE8e editing are quantified at the indicated time point after 1 μM A115 or DMSO addition to HEK293T cells. Editing is quantified at a single nucleotide within a target site that is colored in the target sequence. Bars show mean editing frequency ± SEM of three cell culture replicates, with white circles showing individual replicates.

F, G) Heatmaps of ABEmax, ciABEmax **(F)** and ABE8e, ciABE8e **(G)** editing as a percentage of the highest edited nucleotide for each editor throughout the entire ABE16 target site. Each row shows an individual cell culture replicate. ABEmax and ABE8e editing frequencies were quantified at 72 hr after transfection and ciABEmax and ciABE8e editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK293T cells. The control shows untransfected cells harvested at the same time as ciABEmax and ciABE8e. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.

Having engineered a set of robust, chemically-controlled base editors, we validated that they edit DNA with similar efficiency to their parent base editors by exploring their targeting precision. All four chemically-controlled base editors were only able to edit the intended on-target site in the presence of an sgRNA (**Appendix C Fig. C.9**). Furthermore, all four edit the same nucleotides within a target site to a similar degree as the parental versions (**Figs. 3.2H-I, 3.3F-G;**

Appendix C Fig. C.10), with minimal indel formation at the target site (**Appendix C Fig. C.11**). Finally, off-target DNA base editing occurs at similar or lower magnitudes and at the same nucleotide positions compared to the parental base editors at all off-target sites investigated (**Appendix C Fig. C.12**). Thus, our chemically-controlled base editors do not appear to appreciably impact R-loop formation, positioning and dynamics of the DNA deaminase enzymes, or unwanted off-target DNA base editing activities by the chemically-controlled base editors. Moreover, nciCas9 can be used as a direct replacement of nCas9, and simply appending the deaminase components is able to confer chemical control over base editing activities.

3.2.3 Chemically-controlled base editing reveals nucleotide position effects on the kinetics of base editing

A key application of chemically-controlled enzymes, including Cas9, is exploring kinetics of enzyme activity and downstream cellular processes using time course experiments. For example, chemically- and light-controlled CRISPR/Cas9 systems have been used to study the kinetics of DNA repair after DSB formation by Cas9[100–102]. For base editors, time courses could provide insight into the relative kinetics of different DNA deaminase enzymes, reveal positional effects on base editing kinetics at target sites that contain multiple editable nucleotides within the base editing window, and shed light on the relationship between deamination, repair and editing. All four chemically-controlled base editors yielded appreciable editing within 24 hr of A115 addition, suggesting that activity is induced rapidly. For the chemically-controlled cytidine base editors, we decided to only move forward with ciBE4max due to the similar editing window and magnitudes of editing achieved as ciAncBE4max. Thus, for ciBE4max, ciABEmax and ciABE8e, we quantified editing using next-generation sequencing at 1, 2, 4, 8, 12, and 24 hrs after A115 addition at 4 different target sites (**Fig. 3.4; Appendix C Figs. C.13-17**)[56]. For ciBE4max, the fastest-edited nucleotides within each target site began accumulating edits within 2-4 hr after A115 addition (**Fig. 3.4A; Appendix C Figs. C.13A**). Based upon our previous study of DSB

generation kinetics, we know that the open ciCas9 switch is able to bind target sites within minutes of A115 addition[100]. Thus, within 2-4 hr of ciBE4max localization to a target site, a sufficient degree of deamination, DNA nicking by nciCas9, and subsequent DNA repair occurs to accumulate measurable base editing. Moreover, once a detectable level of editing was observed at a target site, C-to-T base edits by ciBE4max accumulated nearly linearly for the first 12 hr at all four target sites (**Appendix C Figs. C.14-15**). The rate of C-to-T base editing at different nucleotides within each target site correlate with the eventual magnitude of editing observed after 24 and 72 hr, with nucleotides in positions 5-7 (counting the PAM as positions 21–23) edited the earliest, fastest, and to the eventual greatest magnitude (**Fig. 3.2H; Appendix C Figs. C.10A, C.14A, C.15**). Nucleotides outside the position 5-7 window demonstrated slower editing kinetics, which correlated with less overall editing at 24 and 72 hr. Thus, we found that the early editing kinetics of ciBE4max at each nucleotide position dictates the eventual editing magnitudes observed at later time points.

The chemically-controlled adenine base editors showed similar positional effects as ciBE4max in terms of early base editing kinetics and subsequent editing magnitudes (**Figs. 3.3F-G; Appendix C Figs. C.10C-D, C.14B-C, C.15-16**). For ciABEmax, the fastest edited nucleotide within each target site began accumulating base edits 1-2 hr after A115 addition (**Fig. 3.4B; Appendix C Fig. C.13B**). The HEK3 target site with greater background activity showed accumulation of edits at later time points (**Appendix C Figs. C.13B**). The fastest edited nucleotide by ciABEmax was at position 5 of all three target sites that contained a substrate adenine in this position, with adenines to the 5´ and 3´ accumulating edits more slowly (**Fig. 3.4B, Appendix C Figs. C.13B, C.14B**). The ciABEmax editing kinetics also correlated to the different magnitudes of editing achieved at different positions within a target site, with positions 5-7 edited the fastest and to the greatest magnitude (**Figs. 3.3F; Appendix C Figs. C.10C, C.14B, C.15**). Compared to ciABEmax, ciABE8e has similar initial accumulation of A-to-G base edits at the nucleotide that achieves the greatest eventual magnitude (**Fig. 3.4C; Appendix C Figs. C.13C**). Outside of the

fastest edited nucleotide within a target site, ciABE8e accumulated edits faster at 15 of 17 nucleotides across all target sites studied than ciABEmax, mostly at nucleotides where editing is detected at very low magnitudes with ciABEmax (**Appendix C Figs. C.14B-C**). This also explains the greater number of positions edited by ciABE8e (**Fig. 3.3G; Appendix C Fig. C.10D**). Here, the faster kinetics of ciABE8e at more positions within a target site results in lower selectivity and more bystander editing. Thus, ciBE4max, ciABEmax, and ciABE8e enable *in vivo* kinetic studies, revealing that early editing kinetics correlate with the magnitude of editing later and highlighting the kinetic differences between deaminase enzymes which have previously only been explored *in vitro*[57,139].

To further investigate positional effects on the kinetics of base editing, we normalized editing frequency to the nucleotide edited at the greatest magnitude within each target site to facilitate comparisons across target sites (**Figs. 3.4D-F**). At nucleotide positions where ciBE4max edits, positional effects on the kinetics of base editing were readily apparent (**Fig. 3.4D**). Across all target sites studied, C-to-T editing by ciBE4max occurs fastest at positions 5-7 (**Fig. 3.4D**). As previously noted, this is also reflected in the greater magnitudes of editing achieved at positions 5-7 with ciBE4max at 72 hr (**Fig. 3.2H; Appendix C Fig. C.10A**). ciABEmax and ciABE8e show similar positional effects on editing rate and magnitude as ciBE4max where adenines at positions 5 and 6 are edited the fastest (**Figs. 3.4E-F**). ciABE8e shows more rapid editing at positions 4, 7, and 8 compared to ciABEmax, emphasizing the greater number of nucleotides edited within a target site by ciABE8e. The rates of editing at positions 4 and 7 by ciABE8e are almost as fast as editing at positions 5 and 6. Furthermore, ciABE8e shows editing at positions 9 and 12 across multiple target sites, whereas this is not observed with ciABEmax. The editing magnitude differences within a single target site commonly observed with base editing are readily explained by the position of the substrate nucleotide. At every target site we tested, nucleotides that are edited to high magnitudes due to rapid early editing kinetics were at positions 5-7. Whereas

nucleotides both 3´ and 5´ of these rapidly edited positions were edited more slowly and thus to a lower final magnitude.

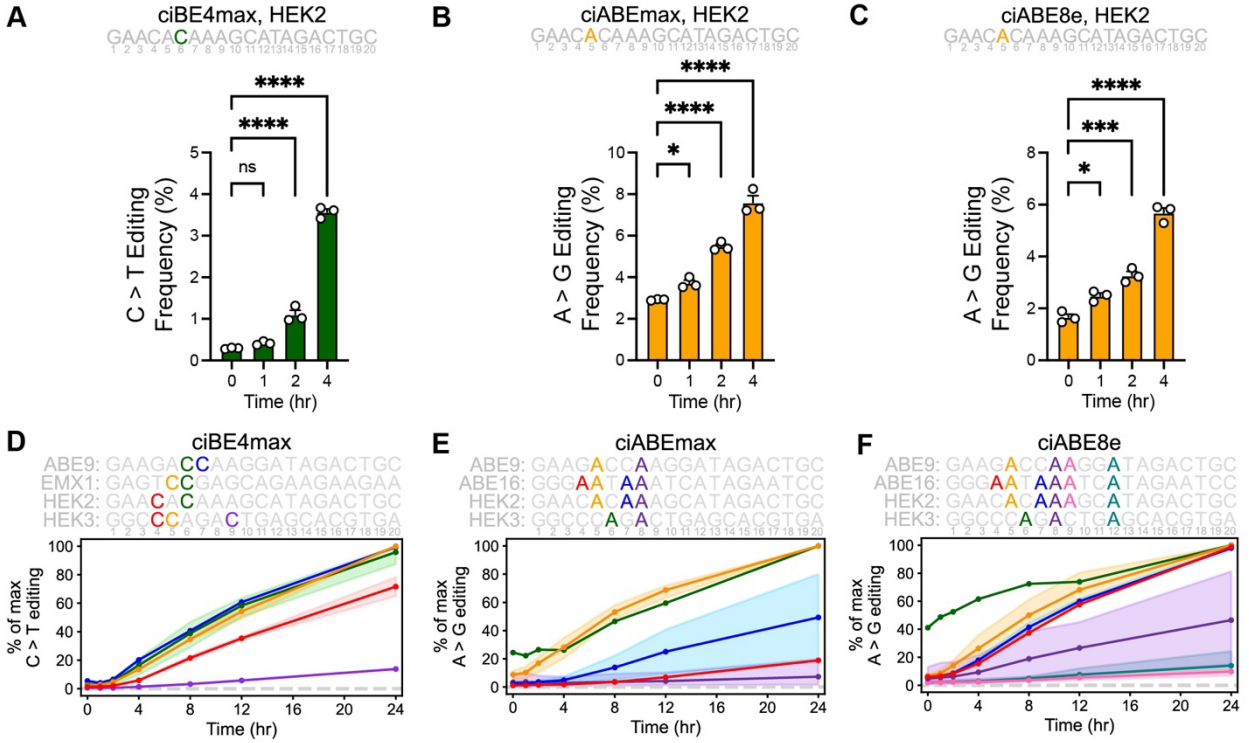

Figure 3.4. Chemically-controlled base editing reveals nucleotide position effects on the kinetics of base editing.

A-C) Early time courses of editing at the HEK2 target site with the ciBE4max **(A)**, ciABEmax **(B)**, and ciABE8e **(C)** base editors activated using 1 μM A115 at the indicated target sites. Time courses shown for the nucleotide colored in the target sequences shown. Numbers underneath the target sequence show the position of the nucleotide from the most PAM-distal nucleotide. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates. Significance of editing at different time points were compared to editing frequency at 0 hr using a One-way ANOVA, statistical values shown in **Appendix C Table C.1**.

D-F) Normalized time courses of ciBE4max **(A)**, ciABEmax **(B)**, and ciABE8e **(C)** base editing. Time courses were normalized within each target sequence where the highest edited nucleotide within each target site at 24 hr after A115 addition was set to 100%. Lines show the mean normalized editing frequency at that position of all target sites listed, and shading shows the range between maximum and minimum normalized editing frequency at that position across all target sites. Color corresponds to the nucleotide positions in each target sequence, shown above each plot. Numbers underneath the target sequences show the position of the nucleotide from the most PAM-distal nucleotide.

3.2.4 Chemically-controlled base editors provide insight into the kinetics of multiply-edited allele formation and nucleotide editing dependency

Analyzing the kinetics of cumulative base edits at individual nucleotides within a target site provides insight into positional effects, but this approach masks the heterogeneity of base editing outcomes within individual cells. In particular, target sites with multiple A or C nucleotides are able to acquire different combinations of multiple edits, resulting in the accumulation of different alleles. The longer the period of active base editing, the greater the accumulation of these multiply-edited alleles[140]. Using our chemically-controlled base editors, we can, for the first time, dissect the kinetics of multiply-edited allele formation to better understand the order in which nucleotides are edited and the impact of initial edits on subsequent ones in the formation of multiply-edited alleles.

To determine the order in which nucleotides are edited within a target site, we identified all distinct combinations of edits comprising alleles of the target site and tracked the frequencies of each allele we observed over time (**Fig. 3.5A-C; Appendix C Figs. C.18-20**). As expected, we observed early accumulation of singly-edited alleles and later accumulation of multiply-edited alleles. Singly-edited alleles begin to accumulate 1-2 hr after A115 activation, for example with ciBE4max at HEK3 and ciABEmax at ABE16, similar to the time frame observed in the cumulative nucleotide editing analysis (**Figs. 3.5A-B; Appendix C Figs. C.14A-B**). Generally, all alleles accumulate linearly for 2-6 hr after they are first detected. However, some alleles eventually show a decrease in accumulation rate or even a decrease in frequency. We hypothesized that these decreases reflect consumption of these alleles to form more highly edited alleles. For example, the singly-edited A5G allele created by ciABEmax at ABE16 shows a decrease in accumulation rate between 4 and 8 hr and then a decrease in frequency after 8 hr (**Fig. 3.5B**). Presumably, this A5G allele is being consumed to form the doubly-edited alleles A5G/A7G or A4G/A5G which both begin to accumulate 4 hr after activation. Finally, the triply-edited A4G/A5G/A7G allele is formed later, first appearing at 8 hr following a third edit of the A4G/A5G or A5G/A7G alleles. We executed this analysis at four target sites each for ciBE4max, ciABEmax, and ciABE8e base editors (**Fig.**

3.5A-C; Appendix C Figs. C.18-20). We found similar kinetic patterns for lower-order and higher-order edited alleles for all three chemically-controlled base editors.

The faster editing kinetics and larger range of positions edited by ciABE8e compared to ciABEmax is also reflected in the greater diversity and frequency of higher order edited alleles detected at all four target sites (**Figs. 3.5B-C; Appendix C Figs. C.19-20**). For example, the A4G/A5G/A7G allele at ABE16 accumulated linearly with ciABEmax to a maximum frequency of 3.2% at 24 hr (**Fig. 3.5B**). Whereas with ciABE8e, the A4G/A5G/A7G allele accumulated linearly increasing to a frequency of 4.5% as early as 8 hr, reflecting the faster editing by ciABE8e (**Fig. 3.5C**). Similarly, after 8 hr at ABE16, we also observed a reduction in accumulation of the A4G/A5G/A7G allele due to consumption of this allele to form higher order alleles. The appearance of alleles with up to six edits reflects the ability of ciABE8e to edit more positions within a target site due to its faster deaminase kinetics. Furthermore, for ciABE8e, we rarely detect singly-edited alleles suggesting that ciABE8e generates singly-edited alleles which are very quickly consumed to form higher order alleles. Thus, the appearance of multiply-edited alleles earlier and the greater detection of higher order alleles with ciABE8e provides clear evidence of its faster *in vivo* editing kinetics compared to ciABEmax.

The formation of multiply-edited alleles can be achieved through two possible kinetic models. An independent model of base editing posits that multiply-edited alleles form by separate independent nucleotide editing events without editing at one position impacting the rate of editing at another position within a target site. The dependent model of editing posits that base editing at one nucleotide position affects the rate of base editing at other positions. Under the independent model, the frequency of a higher order allele at any given time should be the product of the frequency of the individual edits it contains. For each base editor at each target site, we computed single edit frequencies from our cumulative editing analysis (**Appendix C Fig. C.14**). Then, we computed the predicted frequency of each higher order edit by multiplying the frequencies of the constituent single edits. We compared the predicted and observed initial rates of accumulation

for each allele at the earliest time point when the allele was detected. If the rate of accumulation predicted under an assumption of independence was equal to or greater than the measured rate, we classified that allele as “independent.” If the predicted rate was less than the measured rate, we classified that allele as “dependent.” Across all base editors and target sites, 28 of 31 higher order alleles showed a predicted rate that was less than the measured rate (**Figs. 3.5D; Appendix C Figs. C.21-23**). For example, the C4T/C5T allele created by ciBE4max at the HEK3 target site showed 5.9-fold greater measured allele frequency compared to the predicted allele frequency (**Fig. 3.5D**). Thus, these 28 alleles are classified as dependent, suggesting that editing of the first nucleotide increases the rate of editing at all subsequent nucleotides. The remaining three alleles that were not clearly classified as dependent show a slow down in allele accumulation at later time points compared to the expected allele accumulation (**Fig. 3.5E**). We speculate that these three ambiguous alleles also form following a dependent model of editing, but the slower measured accumulation compared to the expected accumulation is a result of these alleles being consumed as substrates to form higher order multiply-edited alleles. Our speculation is supported by the increase in accumulation of higher order alleles that can be derived from these three ambiguous alleles (**Figs. 3.5A-B; Appendix C Fig. C.20**). Thus, the formation of multiply-edited alleles appears to follow a dependent model, where editing at the first nucleotide, often at positions 5-7 within a target site, increases the rate of editing at subsequent nucleotides in other positions within the same target site.

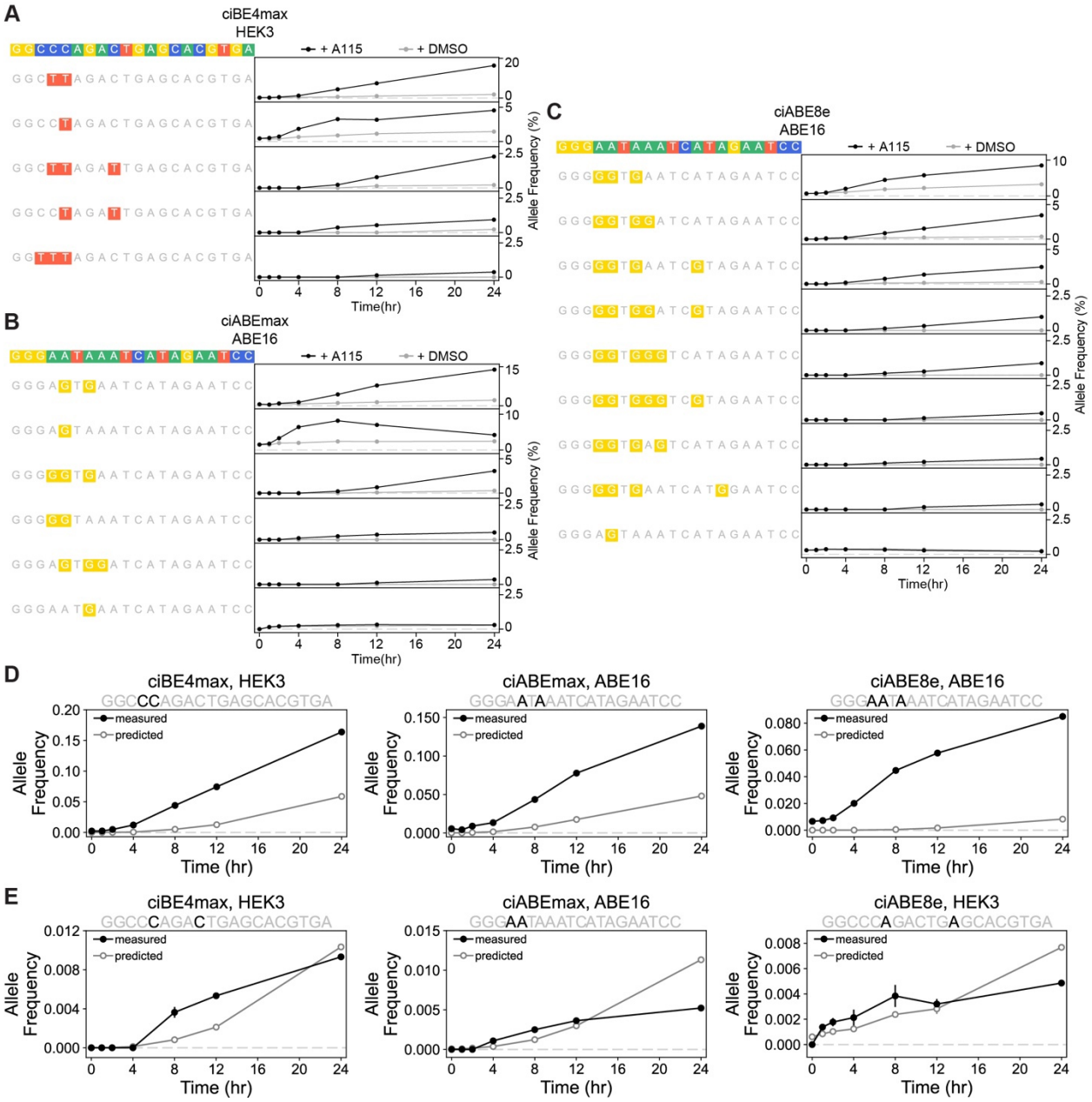

Figure 3.5. Chemically-controlled base editors provide insight into the kinetics of multiply-edited allele formation and nucleotide editing dependency.

A-C) Time course of allele formation by chemically-controlled base editing after activation with 1 μM A115 or DMSO. ciBE4max was targeted to the HEK3 site **(A)**, ciABEmax was targeted to the ABE16 site **(B)**, and ciABE8e was targeted to the ABE16 site **(C)**. Cells were harvested and editing was quantified at specified time points after activation. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.

D) Examples of measured and predicted allele frequencies over time created by ciBE4max (left), ciABEmax (center), ciABE8e (right) that show a dependent model of base editing for multiply-edited alleles.

Figure 3.5 cont.
E) Measured and predicted allele frequencies over time created by ciBE4max (left), ciABEmax (center), ciABE8e (right) that show an ambiguous model of base editing for multiply-edited alleles. In **(D)** and **(E)**, black lines and solid circles show measured allele frequencies, gray lines and open circles show predicted allele frequencies. Measured data represented as mean editing frequency ± SEM of 3 cell culture replicates. Predicted editing frequency represented as mean predicted editing frequency ± relative error. Calculations for predicted frequency and relative error described in the methods.

3.2.5 The ciCas9 switch can also be used to engineer chemically-controlled dual A-to-T and C-to-G base editors and prime editors

Given that the ciCas9 switch is able to provide chemical control of transcriptional activation and DNA base editing, we wondered whether the switch could also be applied to the recently developed dual A-to-T and C-to-G base editors and to prime editors[48,66]. One of the reported dual A-to-T and C-to-G base editors, SPACE, utilizes two deaminase domains fused to nCas9, with the adenine deaminase at the N-terminus of nCas9 and the cytidine deaminase at the C-terminus[66]. When SPACE binds to a target site through the nCas9 domain, it is able to create both A-to-T and C-to-G base edits within a single target site. We generated a chemically-controlled dual A-to-T and C-to-G base editor using the nciCas9 switch, ciSPACE, constructed with the same domain architecture as the unmodified version (**Fig. 3.6A**). We found that ciSPACE was able to introduce both C-to-T and A-to-G edits upon A115 addition in cells transiently expressing ciSPACE and sgRNAs targeting specific loci, with minimal background editing in the absence of A115 (**Fig. 3.6C, E**). Moreover, ciSPACE edits at the exact same positions within target sites as SPACE (**Fig. 3.6F**). ciSPACE also forms minimal indels and off-target base edits, at magnitudes similar or lower to those of SPACE (**Appendix C Figs. C.24-25**). We next explored the kinetics of the two deaminase domains with time course experiments (**Fig. 3.6G; Appendix C Fig. C.26**). At all three target sites investigated, C-to-T edits appeared to accumulate faster than A-to-G edits. Furthermore, at the ABE18 and ABE19 target sites, we did not observe any A-

to-G edits with magnitudes above the detection limit (**Appendix C Fig. C.26**). Thus, at least at these target sites, cytidine deamination and repair appears much faster than adenine deamination and repair.

Finally, we applied the nciCas9 switch to engineer temporal control of the PE3b prime editor[48]. PE3b consists of a PE2 enzyme, which is a nCas9(H840A) fused to M-MLV reverse transcriptase, coupled with two gRNAs. PE3b utilizes a 3´ extended pegRNA encoding the desired edit that is cotransfected with an additional sgRNA that targets the edited sequence to introduce a DNA nick to induce DNA repair and incorporation of the edit. We constructed a chemically-controlled PE2 enzyme, ciPE2, in the same domain architecture as the unmodified version (**Fig. 3.6A**). We tested two sets of previously reported pegRNA/sgRNA pairs with ciPE2 and observed efficient incorporation of the desired edit above DMSO background (**Fig. 3.6H**). Moreover, we observe minimal indel formation at the prime editing target site, similar to that of the PE2 editor (**Appendix C Fig. C.27**). Thus, the ciCas9 switch can be applied to Cas9-based effectors with diverse architectures by simply replacing Cas9 with ciCas9, including to control dual base editing and prime editing with minimal unwanted editing.

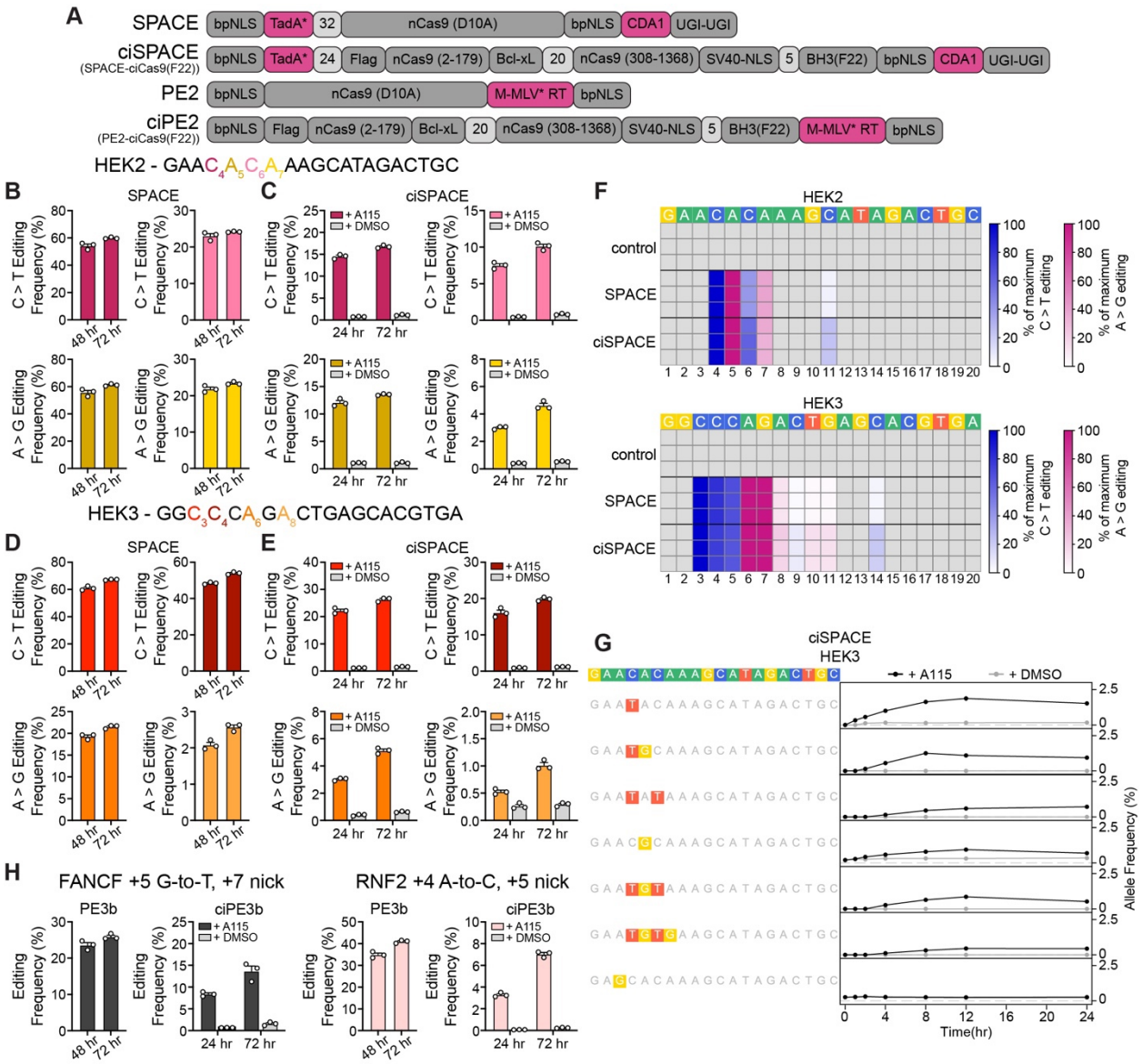

Figure 3.6. The ciCas9 switch can also be used to engineer chemically-controlled dual A-to-T and C-to-G base editors and prime editors.

A) Schematic of the domain arrangements in the unmodified SPACE base editor and PE2 prime editor and the chemically-controlled ciSPACE and ciPE2 editors using the codon optimized ciCas9(F22) variant. ciSPACE is codon optimized SPACE-ciCas9(F22) and ciPE2b is codon optimized PE2b-ciCas9(F22).

B-E) Dual A-to-T and C-to-G base editing by SPACE **(B,D)** and ciSPACE **(C,E)** at the HEK2 **(B,C)** and HEK3 **(D,E)** target sites. SPACE/ciSPACE base editing is shown at the 2 adenine and 2 cytidine nucleotides in each target site with the highest editing frequency with the Cas9 version of SPACE. The 4 different nucleotides in each target site are indicated by color in the target sequence. SPACE editing is quantified at 48 and 72 hr after cotransfection of base editor and sgRNA into HEK-293T cells. ciSPACE editing is quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK-293T cells. Bars show mean editing frequency ± SEM of 3 cell culture replicates with white circles showing individual replicates.

Figure 3.6 cont.
F) Heatmaps of SPACE and ciSPACE editing through the entire HEK2 (top) and HEK3 (bottom) target sites. A-to-G base editing is shown in pink, C-to-T base editing is shown in blue. Each row shows an individual cell culture replicate. SPACE editing frequencies were quantified at 72 hr after transfection and ciSPACE editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK-293T cells. The control shows untransfected cells harvested at the same time as ciSPACE. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.
G) Time course of allele formation by ciSPACE base editing at the HEK3 target site after activation with 1 μM A115 or DMSO. Cells were harvested and editing was quantified at specified time points after activation. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.
H) Prime editing using the PE3b system which consists of PE2 editor cotransfected with a pegRNA targeting FANCF encoding a +5 G-to-T edit (black) or pegRNA targeting RNF2 encoding a +4 A-to-C edit (pink). PE3b editing frequencies were quantified at 48 and 72 hr after transfection and ciPE3b editing frequencies were quantified at 24 and 72 hr after 1 μM A115 addition to HEK-293T cells.

3.3 Discussion

We demonstrate a general method for gaining precise chemical control over different Cas9-based effectors by modulating DNA target site binding using our recently-developed ciCas9 switch. Because the ciCas9 switch consists only of the replacement of the Cas9 REC2 domain with Bcl-xL and appendage of a BH3 peptide, it can be installed while preserving nearly any desired architecture of Cas9-based effector systems. Thus, use of the ciCas9 switch to engineer chemically-controlled transcriptional activators and base editors is simple compared to currently available multi-protein systems that require screening for functional split proteins and careful co-expression of multiple protein components[104–106]. Rapid activation kinetics mean the ciCas9 switch is also more temporally precise than other chemically-controlled Cas9 systems that rely on relocalization of protein to the nucleus or shutoff of degradation—processes that have much slower kinetics[112–114,116–118,120]. As a result of using domain replacement to confer chemical control over Cas9 activity, the overall size of the ciCas9 switch is similar to that of Cas9 itself.

Furthermore, many different Bcl-xL/BH3 disruptors used to activate the ciCas9 switch are available and are compatible with a variety of organisms[141,142]. Thus, ciCas9 can easily replace Cas9 in any Cas9-based effector system to confer chemical control over effector activities. We used the ciCas9 switch to enable chemical control of transcriptional activation, base editing, and prime editing, demonstrating the versatility and simplicity of the switch.

We used three chemically controlled base editors, ciBE4max, ciABEmax and ciABE8e, to elucidate the early *in vivo* kinetics of base editing. We found that rapid early editing 5-7 nucleotides from the 5´ end of the target site generally led to higher editing at later time points. Investigation of the kinetics of alleles with multiple edits revealed that these alleles do not form following an independent model of editing. Instead, bystander edits form at an increased rate following a first edit. We hypothesize that this editing dependency arises from one of two different mechanisms depending on whether the base editor remains bound to the target site during the entire editing process or undergoes cycles of binding and dissociation. If the base editor remains bound, then the first base edit within a target site must increase the affinity of the deaminase for bystander nucleotides within the same target site, yielding faster bystander editing. If the base editor undergoes cycles of binding and dissociation, the first deamination event would create a mismatch in the DNA double helix, thus favoring subsequent cycles of base editor binding and deamination. Discriminating between these possibilities requires investigation of the *in vivo* kinetics of binding and dissociation as well as direct measurement of deamination rather than editing to dissect deaminase activity and subsequent DNA repair. Insights into the kinetics of base editing, and, especially, multiply-edited allele formation could inform future efforts to engineer more efficient and selective base editors, which are sorely needed. For example, correction of pathogenic mutations using base editing often means the only available target site contains additional C or A bases that can also be unintentionally edited[72]. Base editing enables pooled screening of DNA variant effects of genes at their endogenous locus[143]but the unpredictable and

partially specific nature of base editing complicates assessment of DNA variants without sequencing the edited locus.

Despite the utility and power of the ciCas9 switch, challenges remain. Installing the ciCas9 switch often modestly decreases the efficiency of Cas9-based effector by diminishing expression. We also observe appreciable DMSO background for the chemically-controlled base editors at select target sites, which could be mitigated by further engineering of the BH3 peptide to increase its affinity to Bcl-xL[144]. While some background editing is observed and could be problematic for therapeutic applications, low background editing is observed prior to ciCas9 switch activation in our time course experiments and provides enough dynamic range to allow insight into editing kinetics. Replacement of the REC2 domain could affect the proofreading mechanisms that rely on the conformational movement of this domain[30]. However, these proofreading mechanisms are not critical for applications involving dciCas9 or nciCas9 since they have a negligible ability to drive indel formation.

Beyond limitations of the ciCas9 switch itself, challenges remain to fully capture the kinetics of all processes involved in base editing. The ciCas9 base editors enabled chemical control base editing activity and read out activity at early time points after activation, but it does not allow measurement of each step that is involved in the base editing process. Furthermore, cellular processes such as DNA repair and cell proliferation can impact the kinetic measurements at later time points of base editing activity. Development of assays to directly measure deaminase activity and DNA repair *in vivo* coupled with computational modeling of the data could provide a more accurate picture of base editing mechanisms and the timing of different allele outcomes. For example, ciCas9 base editors could generate data to help improve the recently reported model of bystander base editing[145]. This theoretical framework to design kinetically different base editors with reduced bystander activity could be used in conjunction with the ciCas9 switch to screen for new base editor enzymes with the desired kinetics and reduced bystander editing activity. However, from our study, we have found that changing the rate of editing is not enough

to develop more selective base editors as bystander editing is a result of a dependent process. We need a deeper understanding of the underlying mechanisms of editing dependence to be able to engineer more selective base editors.

We have shown that the ciCas9 switch offers a general approach to engineering chemical control of Cas9-based effectors. For example, the ciCas9 switch could be used to temporally control the expression of specific genes during different stages of development or cell differentiation. Precise definition of a starting time for lineage tracing is also achievable with the ciCas9 switch. Temporal control over base or prime editing of clinically relevant loci could also be beneficial to better control editing efficiency and specificity. Finally, other Cas9-based effector proteins could be temporally controlled using the ciCas9 switch such as directly reading or writing chromatin marks and colocalizing genetic elements within the genome to study the effects of 3D genome architecture or localization to organelles within the nucleus. We envision that the ciCas9 switch can be applied to confer temporal control over a plethora of Cas9-based effector proteins that are currently available or will be engineered in the future.

3.4 Methods

Expression plasmids

Mammalian expression plasmids of dCas9, dciCas9(L22), and dciCas9(L22)-VPR were expressed using the pcDNA5/FRT/TO backbone (ThermoFisher). dciCas9(L22) was constructed by introducing the D10A and H935A (dCas9 H840A equivalent) mutations into previously reported ciCas9(L22)[100]. To create dciCas9(L22)-VPR, PCR-amplified VP64-p65-Rta from pEF045, a gift from Jesse Zalatan, was assembled with dciCas9(L22) into the pcDNA5/FRT/TO backbone linearized using the BamHI and EcoRV restriction digest sites. dCas9-VPR was expressed in a pEF backbone and was a gift from David Shechner. The expression of MCP-VPR and CXCR4 scRNA were from a bicistronic vector, a gift from Jesse Zalatan.

All ciCas9 base editors and prime editors were expressed using the pcDNA3.1(+) backbone. nciCas9(D10A) and nciCas9(H840A) were constructed using PCR amplification products from previously constructed ciCas9(L22) and dciCas9(L22). Codon optimized ciCas9(L22) and dciCas9(L22) were purchased from Twist Biosciences. The Bcl-xL and BH3 components were codon optimized using the Genscript codon optimization tool. nciCas9(F22) constructs were made by introducing a point mutation into the L22 constructs using Gibson assembly. The deaminase and UGI components in ciBE4max and ciAncBE4max were PCR-amplified from pCMV_BE4max (Addgene #112093) and pCMV_AncBE4max (Addgene #112094), respectively, both gifts from David Liu. The deaminase components in ciABEmax and ciABE8e were PCR-amplified from pCMV_ABEmax (Addgene #112095) and pCMV_ABE8e (Addgene #138489), respectively, both gifts from David Liu. The M-MLV* reverse transcriptase in ciPE2 was PCR-amplified from pCMV_PE2 (Addgene #132775), a gift from David Liu. The base and prime editing components were assembled with the nciCas9(L22/F22) component using Gibson assembly into the pcDNA3.1(+) backbone linearized using the BamHI and EcoRI restriction digest sites. For a full list of constructs and corresponding amino acid sequences, see Supplemental Table 1.

All sgRNAs were cloned into the gRNA cloning vector (Addgene #41824), a gift from George Church. The CXCR4 sgRNA plasmid has been previously reported[105]. Briefly, a single-stranded DNA (ssDNA) oligo with overlap to the gRNA cloning vector 5´ and 3´ of the 20 nt target sequence was purchased from Integrated DNA Technologies (IDT). The ssDNA oligo was assembled into the gRNA cloning vector linearized using the AflII site by Gibson assembly.

Mammalian cell culture

HEK293T cells (ATCC, CRL-3216) and HEK293 TREx FlpIn cells (ThermoFisher) were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS, MilliporeSigma). Cells were all incubated at 37 °C, 5% CO_2

and found to be free from mycoplasma at least every 6 months. HEK293 TREx FlpIn EMX1-EGFP reporter cells were cultured in DMEM supplemented with 10% (v/v) FBS and 50 µg/mL hygromycin (Mirus).

EGFP reporter construction for dciCas9-VPR transcriptional activation

A 20 bp EMX1 target and PAM sequence and 20 bp of endogenous gDNA sequence 5´ and 3´ of the target were cloned using Gibson assembly into a pcDNA5-FRT-TO backbone lacking the CMV promoter. 3´ of the target sequence is a minimal promoter followed by 100 bp of random DNA sequence and an EGFP reporter gene (**Appendix C Fig. C.3**)[135]. The plasmid containing this locus was then transfected into HEK293 TRex Flp-In cells (ThermoFisher) along with a pOG44 plasmid encoding a Flp-recombinase using Turbofectin according to manufacturer's protocols. Cells with successful integration of the reporter locus were selected using hygromycin (Mirus Bio).

Transcriptional activation

For CXCR4 transcriptional activation with dciCas9-VPR, 6×10^4 HEK293T cells were seeded in 12-well plates. ~20-24 hr after seeding cells, each well was transfected with 1 µg total dciCas9-VPR and CXCR4 sgRNA plasmids (450 ng dciCas9-VPR, 450 ng CXCR4 sgRNA, 100 ng mCherry control) using Turbofectin (Origene) according to the manufacturer's protocol. ~24 hr after transfection, a final concentration of 1 µM of A-1155463 (ChemieTek) was added to the cells, final [DMSO] of 0.1%. 48 hr after A115 addition, cells were harvested and incubated with APC anti-human CD184 (CXCR4) [12G5] (BioLegend) for 1 hr and then fluorescence was analyzed on the LSRII flow cytometer (BD Biosciences). 30,000 single cell events were collected for each sample. The median APC fluorescence is reported for the brightest 25% of cells expressing mCherry transfection control. A similar protocol was followed for dciCas9 + scRNA transcriptional activation of CXCR4 with the exception of 450 ng dciCas9, 450 ng CXCR4 scRNA/MCP-VPR-

IRES-mCherry, and 100 ng empty pcDNA5/FRT/TO plasmid was transfected into HEK293T cells. All flow cytometry data was analyzed using FlowJo. See **Appendix C Fig. C.28** for example cell gating strategies.

For EGFP reporter transcriptional activation, a similar protocol as CXCR4 activation was used except an EMX1 sgRNA was used to target dciCas9-VPR and no antibody incubation was performed, cells were directly analyzed for EGFP fluorescence by flow cytometry. A115 was diluted to the indicated concentrations using DMSO and added to cells with a final [DMSO] of 0.1%. See **Appendix C Fig. C.29** for example cell gating strategies.

Base editing and prime editing with Cas9 and ciCas9 base editors

For both base editing and prime editing experiments, HEK293T cells were seeded at 1.8-2.0 × 10^4 cells per well in a 12-well plate. ~20-24 hr after seeding cells, cells were transfected with 1 µg total plasmid DNA of base/prime editor, sgRNA, and a pMAX-GFP transfection control using Turbofectin according to the manufacturer's protocol. For base editing, 690 ng base editor, 230 ng sgRNA, and 80 ng of pMAX-GFP were cotransfected into each well. No sgRNA control experiments were conducted with 690 ng base editor and 310 ng of pMAX-GFP. For prime editing, 630 ng prime editor, 210 ng pegRNA, 70 ng sgRNA, and 90 ng of pMAX-GFP were cotransfected into each well. ~24 hr after transfection, a final concentration of 1 µM of A115 was added to the wells containing ciCas9 base or prime editor, final [DMSO] of 0.1%. Cas9 base and prime editor conditions were harvested at the indicated time points after transfection. ciCas9 base and prime editor conditions were harvested at the indicated time points after A115 addition.

Library preparation for targeted amplicon DNA sequencing

Genomic DNA isolation, sequencing, and indel frequency analysis for non-library loci were performed as previously described[121]. Briefly, genomic DNA was extracted from cells using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's protocol with an extended

proteinase K digestion of 1 hr at 56 °C. The loci of interest was first amplified with 15 cycles of PCR from 2 µL (~100-300 ng) of genomic DNA eluate using a 5 µL Kapa HiFi HotStart polymerase reaction (Roche). The first PCR was then diluted with 25 µL of DNAse-free water. Indexes and Illumina cluster generation sequences were then added with a secondary PCR reaction using 3 µL of the diluted primary PCR product with a 10 µL Kapa Robust HotStart polymerase reaction (Roche) for 20 cycles. The final amplicons were run on a 1% TBE-agarose gel and DNA was extracted using a Freeze and Squeeze column according to the manufacturer's protocol (BioRad). Gel extracted amplicons were quantified using the Qubit dsDNA HS assay kit (Invitrogen). Up to 1000 indexed amplicons were pooled and sequenced on a NextSeq 550 using a NextSeq Mid 150 v2/v2.5 kit (Illumina). A minimum of 1,000 reads was acquired for each sample except for replicate 1 of ciABE8e at ABE16 with 24 hr DMSO treatment in the time course experiments (**Fig. 3.5C; Appendix C Fig. C.15**).

Editing quantification and analysis

Editing was quantified using the CRISPResso2 package, version 2.0.45[146]. All editing was quantified using batch analysis. Base editing was quantified using the additional flags "-wc -10 -w 10 -q 30". Cumulative base edits at each nucleotide within the target site were extracted from the output table "Nucleotide_percentage_summary." Normalized base editing in **Fig. 3.4C** was calculated by setting the mean of the highest edited nucleotide within each target site at 24 hr to 100% editing. Normalized editing frequencies at other nucleotide positions and at other time points within the same target site were calculated as a percentage of the maximum editing at that highest edited nucleotide using the mean editing frequency of a triplicate of cell culture replicates.

Allele frequencies were extracted from the output table "Alleles_frequency_table_around_sgRNA." Allele frequencies were also determined by only looking at base changes within the 20 nucleotide target sequence, any changes outside the target sequence were trimmed and allele frequencies were summed using the custom script

"allele_frequency_merge_v1.py." Plotting of allele frequency time courses (**Figs. 3.5, 3.6G; Appendix C Figs. C.18-23, C.26**) were filtered for alleles that were detected at >0.3% at any time point and for alleles that showed only A-to-G or C-to-T base edits corresponding to the base editor studied.

Indel frequencies from the base editors were calculated using the output table "CRISPRessoBatch_quantification_of_editing_frequency" by calculating:

$$(1) \qquad Indel\ frequency\ = \frac{"Insertions" + "Deletions" - "Insertions\ and\ Deletions"}{"Reads\_aligned"} \times 100$$

from the table columns. Heatmaps showing base editing frequencies were filtered to only show base conversion frequencies at A or C nucleotides within the target site. Editing frequencies in **Figs. 3.2H-I, 3.3F-G, 3.6F, and Appendix C Fig. C.10** were filtered for positions with base conversion greater than 0.7%. Heatmaps showing off-target base editing in **Appendix C Fig. C.12** were filtered for positions with base conversion greater than 0.1%.

Prime editing was quantified using the additional flag "-q 30". Prime editing frequencies were calculated using the output table "CRISPRessoBatch_quantification_of_editing_frequency" using the "Prime-edited" row for each sample and calculating:

$$(2) \qquad Prime\ editing\ frequency\ = \frac{"Unmodified"}{"Reads\_aligned\_all\_amplicons"} \times 100$$

from the table columns. Indel frequencies from prime editing were analyzed using standard NHEJ CRISPResso2 settings. Indel frequencies were then calculated using the output table "CRISPRessoBatch_quantification_of_editing_frequency" using the equation (1) and the same table columns.

For analysis of editing at early time points in **Figs. 3.4A-C and Appendix C Fig. C.13**, statistical comparison of editing at 0 hr to 1, 2, and 4 hr after A115 addition to editing was completed using a One-way ANOVA using the Graphpad Prism 9 software. Results from the One-way ANOVA are reported in **Appendix C Table C.1**.

Calculation of expected allele frequencies

The expected frequency for an allele was calculated as a product of the frequency of edits at each nucleotide position that make up the allele:

$$(3) \qquad x_{allele} = x_1 \times x_2 \times \ldots \times x_i$$

Where x_i represents the edited nucleotide frequency at nucleotide position i. The relative error was calculated using the standard error of the mean for each nucleotide position:

$$(4) \qquad \sigma_{allele} = x_{allele} \times \sqrt{\left(\frac{\sigma_1}{x_1}\right)^2 + \left(\frac{\sigma_2}{x_2}\right)^2 + \ldots + \left(\frac{\sigma_i}{x_i}\right)^2}$$

Where σ_i represents the standard error of the mean at nucleotide position i.

To determine the dependent vs. independent models of editing for each allele, we chose the earliest time points where each allele was detected as these time points would have minimal impact from cellular factors such as cell division and subsequent plasmid dilution.

Chapter 4: Conclusions and the path forward with the ciCas9 system

The ciCas9 system has enabled exploration into CRISPR/Cas9 editing mechanisms *in vivo* and has provided a generalizable method to engineer future temporally-controlled Cas9-based effectors. While these studies have demonstrated the functionality of ciCas9 and the versatility of the system, other applications of the ciCas9 system are apparent. For example, ciCas9 could be used as a platform to screen and investigate the *in vivo* mechanisms of future high fidelity Cas9 variants. ciCas9-based effector systems could also be used in a variety of research applications, such as genetic screens in induced pluripotent stem cell (iPSC) differentiation and cell lineage tracing. Moreover, further engineering to create a temporally-controlled off-switch to complement the ciCas9 system would enable a reversible CRISPR/Cas9 system to study the effects of pulsed Cas9-based effector activities. Here, I explore the future possibilities and applications of the ciCas9 system.

4.1 *In vivo* mechanistic studies to better inform future engineering of CRISPR/Cas9 DNA editing technologies

While *in vitro* studies are important to understand the biochemical mechanisms underlying Cas9 targeting and double-strand break (DSB) formation, *in vitro* studies lack consideration of cellular factors such as chromatin state and 3D genome architecture. Expansion of the mismatch library-based studies in chapter 2 could provide a more complete picture of Cas9 on- versus off-target discrimination mechanisms *in vivo*. Development of efficient *in vivo* DSB capture techniques that are compatible with DNA target sequence libraries would allow us to directly measure the outcome of Cas9 activity rather than measuring indel frequency. Finally, ciCas9 can also be leveraged to understand how DNA target sequence, chromatin state, and transcriptional machinery may impact Cas9 editing kinetics.

4.1.1 Expanding ciCas9 and mismatched target sequence library studies to further elucidate CRISPR/Cas9 *in vivo* off-target discrimination mechanisms

In chapter 2, I demonstrated the use of ciCas9 to investigate the sequence-based effects on Cas9 binding versus cleavage using a mismatched target sequence library. While measuring relative binding affinity at select single mismatched target sites showed some sequence-based effects on Cas9 off-target discrimination, this did not provide a comprehensive view. A ciCas9 dose-response assay using cells recombined with the mismatched libraries, rather than a single mismatched target sequence, would allow for a more in-depth analysis of sequence-based effects on Cas9 off-target discrimination. Furthermore, as the mismatched libraries were constructed using a doped oligo approach, most double mismatches and some triple mismatched target sequences are also included in the library. Scaling up the experiment size and deeper sequence coverage would allow investigation into target sequences that more resemble endogenous off-target sites where often more than one mismatch is present.

To perform ciCas9 dose-response experiments in the mismatched libraries, scale up optimizations must be performed to allow sufficient recombination coverage of all mismatched target sites and to achieve maximum editing efficiency. I attempted to scale up mismatched library recombination and subsequent ciCas9 dose-response experiments. However, I found that the magnitude of editing with fully active ciCas9, at high [A115], in the recombined mismatched libraries was lower than the magnitudes of editing achieved in recombined cells with one mismatched target sequence. Interestingly, the background editing observed with DMSO was similar in both the library and single target experiments. Thus, the dynamic range of editing was limited in the library-based experiments, and accurate EC_{50} values could not be calculated from the dose-response curves. This could be due to incomplete enrichment of cells recombined with the mismatched libraries and inefficient lipid-based transfection of ciCas9 and sgRNA into recombined cells. The inefficiencies from both the recombination enrichment and transfection of ciCas9/sgRNA could be compounded and result in cells with a successfully recombined

mismatched library synthetic locus but no ciCas9/sgRNA transfection. Thus, resulting in lower indel frequencies compared to small scale experiments where recombination and transfection efficiencies are much higher. Fluorescence-activated cell sorting (FACS) can be used to enrich recombined cells to >95% and electroporation methods to more efficiently introduce ciCas9 and sgRNA plasmids should also be considered. Another potential explanation of lowered maximum editing magnitudes is the greater formation of non-productive ciCas9:sgRNA:DNA complexes when transfected at larger scales[19]. However, the exact cause of lowered editing magnitudes in the mismatched libraries remains to be identified.

Future experiments using ciCas9 in the mismatched libraries should also be reconstructed using DNA synthesized to have defined mismatched sequences rather than using a doped oligo approach. The doped oligo approach resulted in a large proportion of perfectly matched target sequences and a large proportion of double and triple mismatched sequences that were not relevant in our studies. This resulted in a large number of cells that needed to be recombined to reach sufficient coverage to not lose any single mismatched target sequences that we were interested in pursuing. Using a user-defined approach to select specific DNA target sequences and the proportions of types of mismatches would enable more efficient recombination of the libraries into cells. Improving both library construction and the protocols to recombine and enrich recombined cells would enable better measurements of ciCas9 dose-responses in the mismatched libraries.

Information from expanded *in vivo* mismatch library experiments to capture binding affinity and cleavage efficiency effects on Cas9 off-target discrimination could potentially facilitate the development of better high fidelity Cas9 variants or predictive models for off-target editing. The information generated from an *in vivo* investigation into Cas9 off-target mechanisms could also be used to validate the *in vitro* findings of Cas9 binding and cleavage activities performed with mismatched DNA substrates[29,128]. We have also previously demonstrated the compatibility of ciCas9 with high fidelity mutations that confer greater editing specificity in WT-Cas9[100]. ciCas9

could be used in conjunction with the mismatched libraries as a tool to screen for future high fidelity Cas9 variants that have altered binding affinity or cleavage efficiency at mismatched target sites *in vivo*.

4.1.2 Using unbiased double-strand break (DSB) capture techniques with DNA target sequence libraries to elucidate sequence effects on CRISPR/Cas9 *in vivo* editing kinetics

Investigations into the kinetics of DSB generation and subsequent DNA repair have revealed the timing between these two processes in Cas9-mediated DNA editing[100–102,121]. These studies have been enabled by techniques that directly measure DSB formation and DNA repair at the target site after activation of an inducible Cas9 system. While these studies have revealed important mechanisms in Cas9-mediated DNA editing, investigating these responses at a limited number of target sequences do not fully capture the effects of DNA sequence on Cas9 DSB generation kinetics and subsequent DNA repair kinetics. In particular, it has been reported that DNA repair outcomes after a Cas9-mediated DSB is dependent on the target sequence itself more than other factors such as cell type[147]. Investigating DSB generation and DNA repair kinetics at thousands of on- and off-target sequences would enable a better understanding of the factors that govern Cas9-mediated DNA editing efficiency. To study editing kinetics at thousands of on- and off-target sequences, the architecture of the synthetic target sequence library needs to be modified and an unbiased DSB capture technique at the synthetic locus is needed.

An unbiased method to detect DSBs would need to be rapid and contain a unique primer binding sequence to allow for targeted amplicon sequencing and identification of the original target sequence. Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq) is one unbiased method that is currently available that can detect genome-wide DSBs[27]. However, GUIDE-seq relies on the rapid incorporation of an exogenously introduced dsODN through end-joining prior to indel formation. Since end-joining requires DNA repair machinery, GUIDE-seq often does not capture all of the DSBs that occur and indel formation occurs. Breaks

Labeling *In Situ* and Sequencing (BLISS) is another method that can capture DSBs in cells but without the need for DNA repair machinery[148]. BLISS relies on fixation of cells prior to *in situ* ligation of DNA adapters at DSBs. Since cell fixation is required, DSBs can be quantified in a time-resolved manner by fixing and harvesting cells at different time points after ciCas9 activation. However, blunt end ligation of DNA is inefficient, thus DSB capture by BLISS may not provide accurate quantification. BLISS would have to be coupled with an enrichment technique for the synthetic target site prior to blunt end ligation to increase DSB capture efficiency.

The architecture of the synthetic target sequence library would also need to be modified to be able to study thousands of on- and off-target sequences in parallel. In the mismatched target sequence library, the sgRNA could be transfected with ciCas9 after stable integration of the library as the library consisted of only mismatched sequences corresponding to the one sgRNA. In this case, we would need a separate sgRNA for each on-target sequence in the library. In this target sequence library, the locus would contain an sgRNA expressed from a U6 promoter that corresponds to the target sequence in the same locus. The U6-sgRNA element would be positioned for transcription to proceed away from the target sequence and also insulated from the target sequence using a 500 base pair (bp) random DNA sequence to ensure no interference of RNA polymerase III activity with CRISPR/Cas9 editing. This would allow simultaneous introduction of the target sequence and its sgRNA into the same cell in a library-based format. ciCas9 can then be transfected and activated to edit at the synthetic target sequence following successful recombination of the sgRNA and target sequence library.

Simultaneous detection of DSB and indel formation in a time course experiment with ciCas9 editing in the target sequence library would enable exploration into the sequence-based effects of Cas9 DNA editing kinetics. We can measure the timing between DSB and indel formation for each target sequence. We can then use kinetic modeling to approximate the rates of DSB generation and indel formation for a given target sequence. Comparing the magnitudes of DSB generation and indel formation could also provide insight into the level of perfect repair

that happens at different target sequences. Elucidating sequence-based effects on Cas9 editing kinetics could lead to better selection of target sites where high editing efficiency is needed.

4.1.3 A modular DNA target sequence library to investigate the effects of chromatin state and DNA transcription on CRISPR/Cas9 *in vivo* editing kinetics

Cellular factors have been shown to play a role in CRISPR/Cas9 editing outcomes, particularly chromatin state and active DNA transcription[23,130]. Heterochromatin has been shown to lower CRISPR/Cas9 DNA editing efficiency by blocking accessibility of Cas9 to DNA that is tightly associated with the histones[149]. In contrast, active DNA transcription by RNA polymerase has been shown to increase CRISPR/Cas9 DNA editing efficiency by dislodging Cas9 from the DSB to allow access to the DNA repair machinery[23]. While these studies have revealed some of the factors that impact DNA editing efficiency, investigations into the kinetics of CRISPR/Cas9 editing under the absence or presence of these factors at a library of DNA target sequences has not been reported. Using a combination of ciCas9 for temporally controlled Cas9 activity, an unbiased DSB capture technique, and a modular synthetic locus DNA target sequence library where chromatin state and DNA transcription can be controlled, we can systematically study how each of these factors impact CRISPR/Cas9 *in vivo* editing kinetics.

A modular DNA target sequence library would allow systematic and simple modification of the DNA context at the target site. This modular DNA target sequence library can be built using the DNA target sequence library that simultaneously introduces the corresponding sgRNA. The 500 bp random DNA sequence insulating the U6-sgRNA element from the target sequence can be replaced with DNA context modifiers such as promoters that recruit DNA transcription machinery or contain recognition sequences for recruiting DNA chromatin modifiers. Recombination of these target sequence libraries with different DNA context modifiers will then allow systematic combinatorial investigations into how each of these factors affect CRISPR/Cas9 editing.

Using ciCas9 with unbiased DSB capture and targeted amplicon sequencing to quantify indels in this modular DNA target sequence library will provide an in-depth combinatorial analysis into sequence and DNA context factors that influence DNA editing kinetics. Knowledge of chromatin state and DNA transcription activity at different target sequences would also enable development of better predictive models for CRISPR/Cas9 DNA editing efficiency.

4.2 Engineering rapid CRISPR/Cas9 off-switches for tight regulation of genome engineering activities

Although an on-switch for CRISPR/Cas9 activity is useful for temporal studies of Cas9-based activities, a rapidly-inducible off-switch would enable greater control of overall Cas9 activity. Having on- and off-switch capabilities of Cas9 would allow us to study the downstream effects of a pulse of Cas9 activity rather than prolonged and cycles of Cas9 activity that can complicate kinetic analysis. Furthermore, if prolonged Cas9 activity does result in deleterious mutation accumulation corresponding to a therapeutic target, having an off-switch would create a safety mechanism to minimize unwanted DNA editing. Ideally, a general Cas9 off-switch would also allow tuning of Cas9-based effector activities beyond DSB generation. Some studies have shown reversibility of their inducible-Cas9 systems by washing out the small molecule used to turn on Cas9 activity, or using a different wavelength of light to turn off Cas9 activity[120,123]. However, these methods are slow and do not fully eliminate Cas9 activity. Here, I explore possible methods to engineer rapidly-inducible CRISPR/Cas9 inhibitors that can be coupled with ciCas9 to enable “on” and “off” control of Cas9 activity.

For a system that can turn off Cas9 activity, we need to inhibit the DNA binding or DNA cleavage activities. Nature’s evolutionary arms race has resulted in anti-CRISPR proteins (Acrs) that are able to directly inhibit the activities of CRISPR/Cas systems[150,151]. Acrs have been shown in nature to be able to block PAM recognition, block guide RNA binding, or inhibit DNA cleavage. Computational design of binding interfaces on Acrs has also been used to expand the inhibition

mechanisms[152]. We can use these Acr proteins to create rapidly-inducible systems to inhibit CRISPR/Cas9 activity. For the *Streptococcus pyogenes* CRISPR/Cas9 system, the AcrIIA2 and AcrIIA4 proteins have been shown to be the most potent Cas9 inhibitors that directly block PAM recognition[153].

To create inducible Cas9 inhibition, we can use chemically-inducible dimerization systems to recruit AcrIIA2 or AcrIIA4 to Cas9 in a temporally-controlled manner. Using our recently developed hepatitis C viral protease-based inducible dimerization system[105], we could fuse NS3a to AcrIIA2/AcrIIA4, forming NS3a-AcrIIA2/AcrIIA4, and sequester it outside of the nucleus using the ANR peptide (**Fig. 4.1**). The ANR peptide can be anchored on the cytoplasmic side of an organelle outside of the nucleus. NS3a-AcrIIA2/AcrIIA4 would then be held outside of the nucleus upon protein expression due to the binding of NS3a to ANR in the absence of small molecules that disrupt this interaction. Addition of danoprevir or grazoprevir can then disrupt the NS3a-ANR interaction and induce translocation and dimerization of NS3a-AcrIIA2/AcrIIA4 to nuclear localized DNCR2 or GNCR1, respectively. As the AcrIIA2/AcrIIA4 proteins are very small (<20 kDa), nuclear translocation by diffusion is possible. This method of inducible Cas9 inhibition would require ANR anchoring of NS3a-AcrIIA2/AcrIIA4 outside of the nucleus to be very high affinity to maintain high editing efficiency of ciCas9 prior to inhibition. Tandem ANR peptides and NS3a fusions can be created to provide avidity and prevent Cas9 inhibition prior to induction. Furthermore, the stoichiometry of ANR anchor to NS3a-AcrIIA2/AcrIIA4 has to be considered to ensure sufficient sequestration outside of the nucleus.

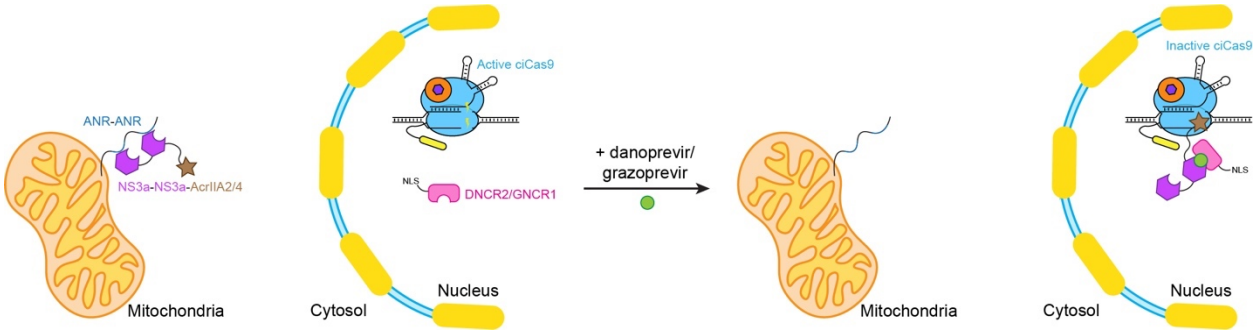

Figure 4.1. Schematic of the AcrIIA2/AcrIIA4 sequestration strategy for inducible ciCas9 inhibition.
Upon protein expression, NS3a-AcrIIA2/AcrIIA4 is sequestered outside the nucleus using tandem ANR peptides anchored on the cytosolic side of an organelle, such as the mitochondria. The other half of the dimerization system, DNCR2/GNCR1, is expressed with an NLS in the nucleus. Upon addition of danoprevier or grazoprevir to bind NS3a, NS3a-AcrIIA2/AcrIIa4 will dimerize with DNCR2 or GNCR1, respectively. This will then cause translocation of NS3a-AcrIIA2/AcrIIA4 into the nucleus where it can inhibit ciCas9 activity.

The latching orthogonal cage-key proteins (LOCKR) system could also be used as an alternative approach to engineer inducible off-switches for CRISPR/Cas9 activity[154]. In this system, computational protein design could inform grafting of the AcrIIA2/AcrIIA4 interface that binds Cas9 onto the cage-latch switch component of LOCKR (**Fig. 4.2**). This Acr-based Cas9 binding interface would be hidden in the cage-latch interface prior to key activation of the switch. The latch would also contain a small interface of Bcl-xL that can bind to the BH3 peptide in ciCas9, forming a BH3-AcrSwitch. Activation of ciCas9 with A115 will then release the BH3 peptide which can then act as a key to activate and recruit the BH3-AcrSwitch and cause inhibition of ciCas9 activity through the Acr interface on the latch. In this LOCKR-based system, the affinities of the BH3 binding and Cas9 binding interfaces in the latch need to be carefully considered to allow enough ciCas9 editing to occur prior to inhibition. Rapid recruitment and inhibition of ciCas9 activity immediately after activation would decrease editing magnitudes. An alternative approach to allow prolonged ciCas9 activity is to use a dox-inducible promoter for key expression instead of using the ciCas9 BH3 peptide. Although dox-induced expression is slow, our time course studies show ciCas9 takes at least 16 hrs to reach maximum editing which is enough time to allow transcription and translation of a key to unlatch an AcrSwitch.

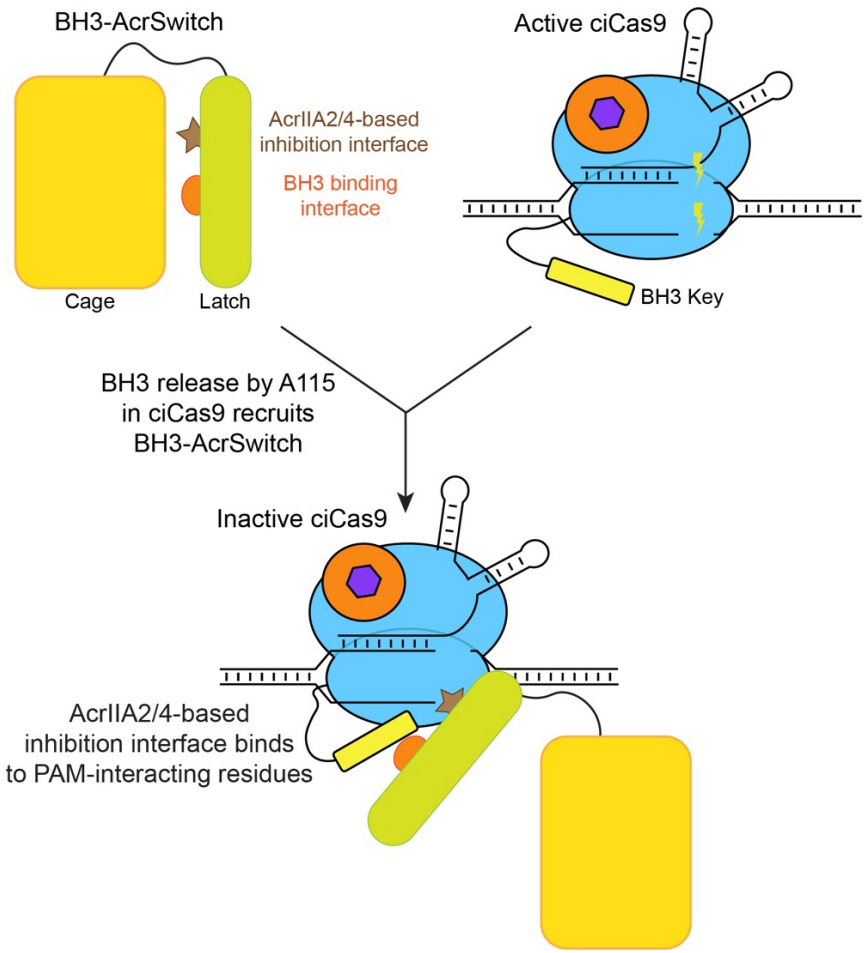

Figure 4.2. Schematic of a BH3-AcrSwitch LOCKR system for inducible ciCas9 inhibition. The BH3-AcrSwitch contains a latch with a BH3 binding surface and an AcrIIA2/AcrIIA4-based inhibition surface to inducibly turn off ciCas9 activity. Activation of ciCas9 with A115 to release the BH3 peptide will be detected by the BH3 binding surface of the latch and recruit the AcrIIA2/AcrIIA4-based inhibition residues also present in the latch to interact with ciCas9 and turn off activity.

Both hypothetical frameworks explored here to create an inducible off-switch for ciCas9 requires careful tuning of protein interfaces to balance the magnitude of editing and the speed of inhibition. Structural studies of AcrIIA2 and AcrIIA4 have also revealed the exact residues involved in Cas9 inhibition[153]. Mutation of these binding residues could alter the potency of AcrIIA2/AcrIIA4 to allow for more ciCas9 editing prior to inhibition. Having an off-switch to complement the ciCas9 system that specifically inhibits DNA binding of Cas9 would enable tighter temporal control and pulses of ciCas9-based effector activities in cells.

4.3 Using ciCas9-based systems to elucidate biological processes

While ciCas9 is able to reveal *in vivo* mechanisms and the factors that impact CRISPR/Cas9 DNA editing, in chapter 3 I showed the potential of using ciCas9 to create chemically-controlled Cas9-based effector systems. These ciCas9-based effector systems could be used to elucidate a plethora of biological processes. For example, dciCas9-VPR could be used to elucidate the kinetics of transcriptional activation. Chemically-controlled base editors could inform the possible DNA variants for a given sgRNA in a base editing screen. ciCas9-based systems could also be used for more precise cell lineage tracing or CRISPR screens in different stages of iPSC differentiation. Here, I highlight the versatility of the ciCas9 system by exploring possible applications.

4.3.1 Understanding transcriptional activation kinetics with dciCas9-VPR

Transcriptional activation using dCas9-VPR has been used extensively to study the effects of expressing specific genes and also in genome-wide CRISPRa functional screens [155]. However, the kinetics of transcriptional activation is not well understood. The timing of DNA binding of dCas9-VPR and subsequent transcriptional machinery recruitment is unknown. As activation of ciCas9 is known to be rapid and DSBs are detected within 10 min of activation, dciCas9 could be used to precisely control the timing of DNA binding[100]. The use of dciCas9-VPR with time course experiments at a fluorescent reporter locus could then elucidate the timing of DNA transcription and translation after DNA binding. The transcriptional activation kinetics of a specific gene could also be investigated by creating a reporter locus with the gene of interest and detecting DNA transcription using qRT-PCR. Understanding the timing of transcriptional activation by dCas9 would be important in applications such as CRISPRa screens for processes that are temporally regulated.

4.3.2 ciCas9 base editing for more predictable DNA variant formation in base editing screens

Generating a DNA variant library at the endogenous locus of a gene for a multiplex assay of variant effects (MAVE) is highly desirable as the gene is still subject to endogenous regulation mechanisms. Base editing has been demonstrated to be able to generate DNA variant libraries at endogenous loci without using saturation genome editing which requires generation of a DSB[143]. However, current base editing systems often have bystander editing effects, thus complicating the identification of DNA variants after functional assaying. The entire gene of interest must be sequenced to identify the variant generated by base editing.

The chemically-controlled base editors could be used in conjunction with models of base editing activity to predict editing outcomes. Recently, a kinetic model of base editing was reported that dissected the possible factors that influence bystander editing[145]. Generation of a predictive model of base editing outcomes could inform the variants present in a MAVE library at different time points after active editing. Therefore, only sequencing of the sgRNA used for base editing would need to be performed to identify the DNA variants after selection using a functional assay. This would simplify MAVE investigations of genetic elements at their endogenous genomic locus.

4.3.3 CRISPR screens at different stages of iPSC differentiation with the ciCas9 system

The ability to temporally control CRISPR/Cas9-based activities with ciCas9 allows the investigation of genetic elements involved in temporally-regulated processes. For example, precisely timed chromatin modification and expression of different genes is required during iPSC differentiation[156]. ciCas9 could be used in CRISPR-based genetic screens to understand the genes that are involved in each stage of iPSC differentiation to different cell lineages. ciCas9 and the sgRNA library would be stably introduced into the genome prior to differentiation. Then, DNA editing or ciCas9-based CRISPRa/i activities can be activated at different steps of differentiation to evaluate the contribution of genes targeted in the library. Using ciCas9 also provides an advantage over using constitutively active Cas9 systems for screening. Genes that may be

important in multiple stages of differentiation could be missed at the later differentiation stages if they disrupted earlier stages. Furthermore, the precise timing and rapid activation of ciCas9-based activities enables investigation into the precisely timed steps in iPSC differentiation. Cas9-based screens using constitutively active systems could also be performed at different stages of differentiation, but it would require slow activation of a dox-induced Cas9-based effector or multiple lentiviral transductions. The ciCas9-based systems offer a simpler and more rapid method to perform CRISPR screens in the precisely timed stages of iPSC differentiation.

4.3.4 More precise cell lineage tracing using ciCas9-based genetic tracing tools

CRISPR/Cas9-based methods have been developed to aid in lineage tracing to understand the cellular relationships during organism development and regeneration[157]. The single cell ScarTrace (scScarTrace), single cell genome editing of synthetic target arrays for lineage tracing (scGESTALT), and lineage tracing by nuclease-activated editing of ubiquitous sequences (LINNAEUS) methods all use CRISPR/Cas9 to generate indel-based barcodes for cell lineage tracing[158–160]. All of these methods rely on stably integrated DNA target barcode sequences into the genome of single cells. Cas9:sgRNA is then introduced into the cells at an early developmental stage and that targets these DNA barcodes. Accumulation of indels at these DNA barcodes become genetic scars that are permanently recorded and inherited by the descendents of that cell. Tracing of cell lineage is then possible by looking at the indel sequences that have formed in each cell. Although the CRISPR/Cas9 system is able to generate cellular barcodes that can then be used to unravel cell lineages, Cas9:sgRNA is only active in the cells for a short period of time. Thus, use of these CRISPR/Cas9-based lineage tracing systems is limited to early time points in development and regeneration.

ciCas9 could provide an alternative approach to CRISPR/Cas9-based genetic lineage tracing methods by allowing activation of DNA barcode indel recording at later time points during development. In a ciCas9-based system, the DNA barcode target sequences and ciCas9:sgRNA

would be introduced into the cells at an early developmental stage. ciCas9 activity can then be induced at different time points in separate populations of cells to study different stages of development. The rapid activation kinetics of ciCas9 DNA editing allows precise timing of genetic recording activation. One limitation of the ciCas9 system is the background editing activity that is observed for some target sequences prior to activation with A115. Background indel formation could complicate construction of lineage trees if indels were already present at the DNA barcode loci prior to ciCas9 activation. This complication could be mitigated by using higher affinity BH3 variants in ciCas9 that have lower background activity[144]. Furthermore, identification of the exact DNA barcode sequences and indels that have already accumulated at the beginning of lineage tracing prior to activating ciCas9 would be needed. Using ciCas9 instead of Cas9 in scScarTrace, scGESTALT, or LINNAEUS expands the possible time frames during development in which genetic cell lineage tracing can be performed.

4.4 Summary

The discovery of CRISPR/Cas systems has revolutionized biological science. Using RNA to encode the target sequence and guide the Cas9 endonuclease has greatly simplified and improved the outcomes of targeted genome editing. The adaptability of the CRISPR/Cas9 system to recruit other effector proteins to targeted sites in the genome has also vastly expanded the genome engineering toolbox. My dissertation projects discussed here involved using a chemically-controlled CRISPR/Cas9 system to understand and develop new CRISPR/Cas9-based genome engineering technologies. I elucidated the mechanisms of CRISPR/Cas9 off-target editing by performing *in vivo* biochemical measurements of Cas9 activity. Furthermore, I expanded the temporally-controlled genome engineering toolbox by adapting ciCas9 into a versatile DNA binding switch that can be used to control various Cas9-based effectors. I envision the results presented here could be used to inform engineering of future Cas9-based systems to be more efficient and accurate.

Appendix A: Methods for using ciCas9 in mammalian cell lines and detection of DNA double-strand breaks (DSBs) using DSB-ddPCR

Appendix A has been adapted from:

Wei, C.T., Maly, D.J., Fowler, D.M. Temporal and rheostatic control of genome editing with a chemically-inducible Cas9. *Methods Enzymol.* **633**, 119-141 (2020).

This appendix outlines the protocol of performing ciCas9 experiments in HEK-293T cells, genomic DNA isolation, and DNA double-strand break (DSB) quantification using DSB-ddPCR.

A.1 Temporal and rheostatic control of genome editing with ciCas9

The ciCas9 system allows for kinetic studies of Cas9 activity through rapid activation of nuclease activity upon addition of the small molecule activator, A115. We commonly introduce ciCas9 into cells using transient transfection along with an sgRNA targeting a locus of choice. The kinetics of Cas9 editing at this locus can then be monitored using a time course experiment. Here, cells are harvested at various time points after A115 addition; DSBs are assessed using DSB-ddPCR and indels are assessed using high-throughput DNA sequencing. We have used this system to study DNA cleavage and indel kinetics at different loci and have found that CRISPR/Cas9 genome editing kinetics can vary at different target sites[100].

This protocol is written based on usage in HEK-293T cells, which allow for rapid cell harvesting for temporal precision. We allow around 24 hours between seeding of cells and transient transfection of ciCas9/sgRNA and then another 24 hours before A115 addition. In HEK-293T cells, this timing allows the cells to reach suitable confluency and stable Cas9 expression following transient transfection. We note that in other cell lines, timing should be adjusted to optimize confluency and expression.

A.1.1 Seeding cells for *in vivo* kinetics studies of Cas9 activity

Equipment

- CO_2 incubator
- Biosafety cabinet
- Centrifuge
- Hemocytometer or automated cell counter

Reagents

- Human embryonic kidney 293T cells (ATCC)
- Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific), DMEM
- Fetal Bovine Serum (Millipore Sigma), FBS
- Dulbecco's Phosphate-Buffered Saline (Thermo Fisher Scientific), DPBS
- 0.25% Trypsin-EDTA (Thermo Fisher Scientific)

Protocol

1) HEK293T cells are maintained in DMEM supplemented with 10% FBS (D10) and kept at 37 °C with 5% CO_2 until the cells are 80-90% confluent before seeding for an experiment.
2) The day of seeding cells, remove the media and wash the cells with DPBS.
3) Detach the cells from the vessel by adding 0.25% Trypsin-EDTA, just enough to cover the surface area of the cells. Incubate at room temperature until the cells detach. Add D10 to quench the trypsin. Move all the cells into a conical tube.
4) Centrifuge the cells at 100 ×g for 3 min and remove the media. Then resuspend the cells in new D10, making sure to disrupt all cell clumps. Count the number of cells using a hemocytometer or an automated cell counter.
5) Dilute the cells to 1.8-2 × 10^5 cells/mL using new D10.
6) Add 1 mL of cells into each well of a 12-well plate. This should result in 1.8-2 × 10^5 cells per well. Gently rock the plate to ensure even distribution of cells in each well.

7) Incubate the plates at 37 °C with 5% CO_2 for 24 hours, reaching 40-50% confluency.

Notes

- Entire wells of cells are harvested at each time point, so separate 12-well plates should be used for each time point. This will allow for minimal disruption to the cells that are to be collected at later time points.
- 12-well plates are used in this protocol as this provides enough cells for downstream analysis by high-throughput sequencing and DSB-ddPCR.
- Other cell lines can also be used to study genome editing kinetics with ciCas9. We often use HEK293T cells due to their high transfection efficiency and the ease of harvest at early time points due to the weak adherence of these cells. Cells should be seeded to 40-50% confluency at time of transfection.

A.1.2 Transient transfection of ciCas9 into human cells

Equipment

- CO_2 incubator
- Biosafety cabinet

Reagents

- Turbofectin 8.0 (OriGene)
- Opti-MEM (Thermo Fisher Scientific)

Plasmids

- ciCas9 in mammalian cell expression backbone (e.g. pcDNA5-FRT-TO or pcDNA3.1(+))
- sgRNA cloned into Addgene #41824 backbone
- pMAX-GFP transfection control

Protocol

1) Calculate the volume of transfection mixture required for 100 µL of transfection mix per well. Plasmids should be mixed in a ratio of 450 ng ciCas9, 450 ng sgRNA, and 100 ng pMAX-GFP per well. The remaining volume should be made up with Opti-MEM.
2) To create the transfection mix, add Opti-MEM to the tube first, followed by adding the plasmids. Mix by inverting the tube 5-6 times.
3) Add 3 µL of Turbofectin 8.0 reagent per 1 mg of plasmid DNA to the transfection mix. Try not to touch the side of the tube when adding Turbofectin. Invert the tube 5-6 times then incubate the transfection mix with Turbofectin at room temperature for 20 min.
4) Gently add 100 µL of transfection mix to each well drop-wise and gently rock the plate to mix. Be sure to not pipette the transfection mix onto the cells too quickly, this may result in cells detaching from the plate.
5) Incubate the plates at 37 °C with 5% CO_2 for 24 hours.

Notes

- If DSB-ddPCR will be used to analyze kinetics of DSB formation, half the amount of DNA should be transfected for each well; resulting in 225 ng ciCas9, 225 ng sgRNA, and 50 ng pMAX-GFP.
- Other transfection reagents can also be used to deliver plasmids. We have found that Turbofectin 8.0 works efficiently for delivering large plasmids, such as ciCas9, and also has minimal effects on cell viability.
- We have found that 1 mg of DNA with 3 µL of Turbofectin works well in HEK293, HEK293T, HCT116, and U2OS cell lines for delivering ciCas9 and sgRNA. Transfection has been observed to be most efficient in cells at 40-50% confluency at the time of transfection.

- 1 mg of DNA is suitable for the number of cells plated per well in a 12-well plate, as described in section A1.1. DNA amounts can be adjusted for the number of cells and well size.

A.1.3 Activation of ciCas9 to study *in vivo* editing kinetics and dose-dependence

Equipment

- CO_2 incubator
- Biosafety cabinet

Reagents

- A-1155463 dissolved in DMSO at 10 mM stock concentration, aliquoted and stored at -20 °C (ChemieTek, Catalog #: CT-A115), A115
- DMSO (Thermo Fisher Scientific)

Protocol

1) Dilute the 10 mM stock of A115 to 1 mM using DMSO.
2) For time course studies, add 1 µL of 1 mM A115 stock to each well where ciCas9 will be activated, reaching a final drug concentration of 1 µM and 0.1% DMSO. For dose-response studies, create serial dilutions of A115 to desired concentrations using DMSO. Add the drug to the side of the well, being sure to not detach any of the cells by forceful pipetting.
3) Gently rock the plate to mix the drug with the media.
4) Incubate the plates at 37 °C with 5% CO_2 for the amount of time for each time point.

Notes

- We have observed that 1 µM of A115 can fully activate ciCas9 for editing without affecting cellular viability. In dose-response studies, 1 µM A115 is usually sufficient as the highest dose.
- At any concentration of A115 added, the final DMSO percentage should be 0.1% or less.

- If there are short time points (e.g. under 1 hour), it is advised to add A115 to the later time points first to ensure accurate harvesting times without delay for the earlier time points.

A.1.4 Harvesting of HEK293T cells after ciCas9 activation

Equipment

- Biosafety cabinet
- Centrifuge at 4 °C

Reagents

- Ice cold Dulbecco's Phosphate Buffered Saline (Thermo Fisher Scientific), DPBS

Protocol

1) Label all 1.5 mL microcentrifuge tubes that the cells will be collected into and pre-chill the tubes on ice for 10 min prior to harvesting cells.
2) At appropriate time points, move the plates onto ice and immediately remove the growth medium. Wash the cells in each well with 1 mL of ice cold DPBS.
3) Add 600 µL of ice cold DPBS to each well and pipette up and down to disrupt the cells off the surface of the plate. Move the cells into a pre-chilled 1.5 mL microcentrifuge tube.
4) Centrifuge the cells at 1,500 ×g for 10 min at 4 °C. Remove DPBS and immediately store cell pellets at -80 °C until isolation of genomic DNA.

Notes

- To ensure temporal precision, all steps must be performed quickly and the cells must be kept cold at all times.
- Cell pellets can be kept at -80 °C indefinitely until isolation of genomic DNA.

A.1.5 Harvesting of other adherent cell lines after ciCas9 activation

This section describes harvesting of adherent cell lines that are more strongly adherent than HEK-293T cells and will require trypsin cleavage to detach cells.

Equipment

- CO_2 incubator
- Biosafety cabinet
- Centrifuge at 4 °C

Reagents

- Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific) supplemented with 10% Fetal Bovine Serum (Millipore Sigma), D10
- Warmed Dulbecco's Phosphate-Buffered Saline (Thermo Fisher Scientific), warmed DPBS
- Ice cold Dulbecco's Phosphate Buffered Saline (Thermo Fisher Scientific), ice cold DPBS
- 0.25% Trypsin-EDTA (Thermo Fisher Scientific)

Protocol

1) Same as section A1.4, label all 1.5 mL microcentrifuge tubes that the cells will be collected into and pre-chill the tubes on ice for 10 min prior to harvesting cells.
2) At appropriate time points remove the growth medium and wash the cells in each well with 1 mL of warmed DBPS.
3) Add 100 µL of 0.25% Trypsin-EDTA and incubate at room temperature until cells detach. Quench trypsin using 1 mL of D10 media.
4) Resuspend the cells and move into a pre-chilled 1.5 mL microcentrifuge tube and place the tube on ice.
5) Centrifuge the cells at 1,500 ×g for 10 min at 4 °C. Remove the media.

6) Resuspend the cell pellet with 600 µL of ice cold DPBS to wash the cells.
7) Centrifuge the cells again at 1,500 ×g for 10 min at 4 °C. Remove DPBS and immediately store cell pellets at -80 °C until isolation of genomic DNA.

Notes

- The trypsin incubation for detaching cells can occur at 37 °C if the cells are strongly adherent. However, incubation time should be minimized to ensure temporal precision.

A.2 DSB-ddPCR to measure ciCas9-mediated DSB formation

The outcomes of Cas9 activity *in vivo* are often measured by quantifying the indels formed at the target locus. Quantification of indels by sequencing does not directly measure the outcomes of Cas9 activity as indel formation involves repair of Cas9-mediated DSBs. Cas9 is only involved in the generation of DSBs prior to DNA repair. Thus, to directly measure Cas9 activity, methods for quantifying DSBs are needed. We have developed DSB-ddPCR to directly quantify the frequency of DSBs created by Cas9[100].

DSB-ddPCR involves using two sets of probes and primers to quantify the percentage of a target site that contains DSBs. One set of primers span the sgRNA target locus and the other set forms a control template amplicon adjacent to the target locus that does not span the site targeted by the sgRNA used (**Appendix A Fig. A.1**). Amplification by these primers are detected by using fluorescence-quenched qPCR probes that anneal within these amplicons. Droplets that contain fluorescent signals for both the template and target probes represent uncleaved or repaired DNA targets. Droplets that contain fluorescent signal for only the template probe represent DNA target sites containing DSBs. By counting the number of droplets with these different fluorescent outcomes, DSB frequencies are quantified.

The use of DSB-ddPCR coupled with ciCas9-based time course experiments allows for the direct quantification of Cas9-mediated DSBs. Additional insight into DNA repair following DSB

generation can be measured with indel formation using high-throughput sequencing in parallel. Protocols for measuring indels with targeted high-throughput sequencing have been described[100]. We have previously coupled these techniques using the original ciCas9(G22) variant and A3 activation to study the kinetics of various loci. Here, we describe methods for the processing and DSB-ddPCR analysis of frozen cell pellets that have been collected from ciCas9 time course experiments. These methods also include guidelines for designing target and template probe and primer sets for DSB-ddPCR and the creation of standard curves for quantifying DSBs.

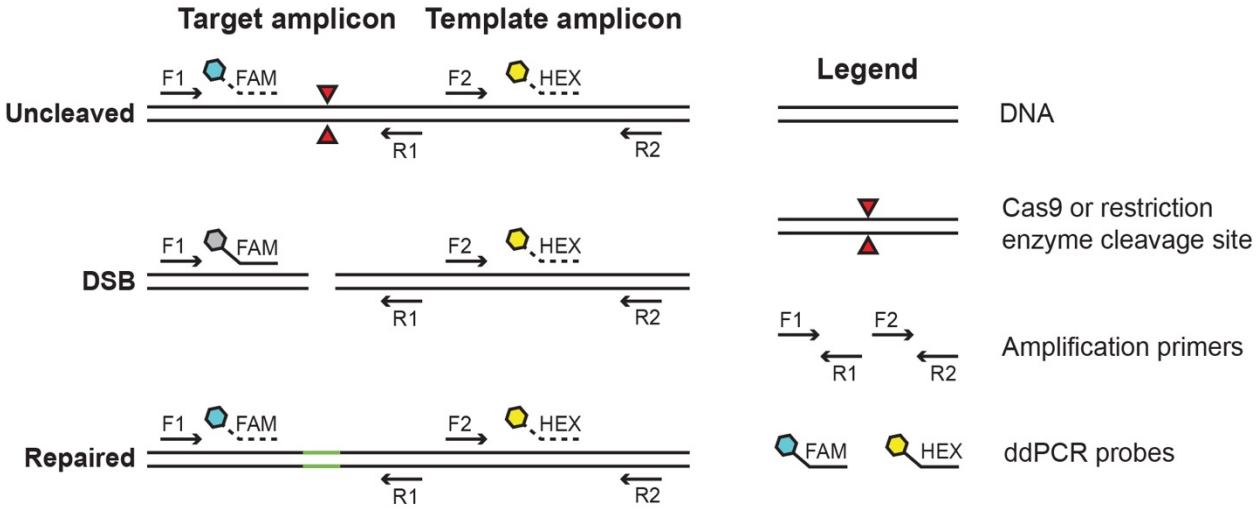

Figure A.1. Schematic of the DNA double-strand break droplet digital PCR (DSB-ddPCR) assay.

A.2.1 Design of primers and probes

Retrieval of genomic DNA sequence using UCSC Genome Browser

- Use the BLAST-like alignment tool (BLAT) from UCSC Genome Browser (https://genome.ucsc.edu/) to retrieve the genomic DNA sequence flanking the sgRNA and PAM sequence of interest.
- 1,000 base pairs of DNA sequence should be retrieved 5´ and 3´ of the sgRNA target and PAM sequenced.
- Use of a DNA sequence visualizer is helpful in identifying the sgRNA target site and where designed probes and primers will anneal, such as SnapGene.

Setting parameters in IDT PrimerQuest to design primers and probes

- Amplification primers and double-quenched qPCR probes can be designed and ordered using the Integrated DNA Technologies PrimerQuest tool. Use the qPCR (2 Primers + Probe) option with the Custom Design Parameters.
    - https://www.idtdna.com/PrimerQuest/Home/Index
- Helpful parameters for designing probes and primers can be found in the BioRad ddPCR experiment guidelines.
    - https://www.bio-rad.com/en-us/applications-technologies/planning-droplet-digital-pcr-experiments?ID=MDV33OKG4
- General parameters for designing probes and primers for both target and template amplicons can be found in **Appendix A Table A.1**. These parameters assume 1,000 bp of sequence flanking the sgRNA target and PAM sequence has been entered into PrimerQuest.

Considerations for probe and primer selection

- Probes should anneal distal to the sgRNA target and PAM site due to formation of indels after DNA repair.
- Template amplicons can be located anywhere 5´ or 3´ from the target amplicon as long as the amplicons do not overlap. The parameters suggested for "*Excluded region list*" and "*Target region list*" in **Appendix A Table A.1** is based on locating the template amplicon approximately 300 bp 3´ of the PAM.
- The target amplicon should contain a restriction enzyme site that is not present between the target and template amplicons. This restriction enzyme site will be used to create the standard curve.
- Probes with multiple primer sets should be selected so many alternative amplicons can be tested for efficiency in detecting cut and uncut DNA when troubleshooting.

- Use of Black Hole quenchers with probes are recommended. The FAM dye is used with the target probe and the HEX dye is used with the template probe.

A.2.2 Extraction of genomic DNA

Equipment

- Dry block incubator at 56 °C
- Centrifuge

Reagents

- Qiagen DNeasy Blood and Tissue kit
- 96-100% ethanol
- Nuclease free water or elution buffer without EDTA
- Dulbecco's Phosphate-Buffered Saline (Thermo Fisher Scientific), DPBS

Protocol

1) Thaw frozen cell pellets at room temperature for 3 min.
2) Resuspend cell pellet in 200 µL of DPBS.
3) Add 20 µL of Proteinase K and then 200 µL buffer AL. Invert tube to mix.
4) Incubate on a dry block incubator at 56 °C for 1 hour.
5) Add 200 µL of 96-100% ethanol. Invert tube to mix.
6) Transfer mixture into a DNeasy spin column with a collection tube. Centrifuge at 8,000 ×g for 1 min. Discard flow through and exchange for a new collection tube.
7) Add 300 µL of buffer AW1. Centrifuge at 8,000 ×g for 1 min. Discard flow through and exchange for a new collection tube.
8) Add 300 µL of buffer AW2. Centrifuge at 17,000 ×g for 1 min. Discard flow through.
9) Move the DNeasy spin column into a clean 1.5 mL microcentrifuge tube.

10) Add 150 µL of nuclease free water or elution buffer without EDTA. Incubate at room temperature for 1 min. Centrifuge at 8,000 ×g for 1 min. Expected concentration ranges are from 30-150 ng/µL.

Notes

- DPBS, Proteinase K, and buffer AL can be made together as a master mix. 400 µL of master mix should be added to each cell pellet followed by quick vortexing to disrupt the cell pellet.
- The extended 1 hour incubation at 56 °C ensures Proteinase K digestion of ciCas9 from the target sites that could disrupt DSB-ddPCR analysis.
- Buffer AE from the Qiagen DNeasy Blood and Tissue kit should not be used for elution.
- Extracted genomic DNA can be stored at -20 °C until downstream analysis by DSB-ddPCR or high-throughput sequencing.

A.2.3 Standard curve sample creation

To create a standard curve, extracted genomic DNA is cut with a restriction enzyme within the target amplicon. Select a restriction enzyme that should not cut between the target and template amplicons or within the template amplicon. DNA from an untreated sample should be used to create the standard curve. Cut DNA is mixed with uncut DNA at different ratios to create the standard curve (**Appendix A Fig. A.2**).

Equipment

- Incubators at 37 °C and 80 °C
- Magnetic microcentrifuge tube rack
- Fume hood

Reagents

- Restriction enzyme with associated buffer
- 70% ethanol
- AMPure XP beads (Beckman Coulter)
- Nuclease free water

Protocol

1) To create the cut DNA reaction, make a 35 µL digest reaction containing: 3.5 µL 10x buffer, 2 µL restriction enzyme, and 29.5 µL DNA. Restriction enzyme and DNA amounts can be adjusted according to restriction enzyme efficiency.
2) To create the uncut DNA reaction, make a 70 µL reaction containing: 7 µL 10x buffer, 63 µL DNA. No restriction enzyme should be added.
3) Incubate both the cut and uncut reactions separately at 37 °C for 20 min. The length of digestion can be adjusted according to restriction enzyme efficiency.
4) Add 70 µL of AMPure XP beads to the cut DNA reaction and 140 µL of AMPure XP beads to the uncut DNA reaction. Place reactions on the magnetic microcentrifuge tube rack.
5) Once all of the beads have settled onto the magnet, aspirate and discard the liquid.
6) Wash both the cut and uncut DNA reactions with 140 µL of 70% ethanol. Repeat.
7) Place the open tubes in the fume hood for 10 min to allow any residual ethanol to evaporate.
8) Add 35 µL of nuclease free water to the cut DNA sample and 70 µL of nuclease free water to the uncut DNA sample.
9) Mix the cut DNA and uncut DNA samples to create 50%, 33%, 25%, 10%, 5%, and 1% cut DNA samples for the standard curve. A total volume of 12.5 µL of each standard is adequate for several DSB-ddPCR runs.

Notes

- Standard curve samples can be frozen at -20 °C for future use.
- Restriction digest volumes suggested here create enough for 5 standard curve assays.
- See section A.2.5 for calculating and plotting the standard curve after droplet analysis.

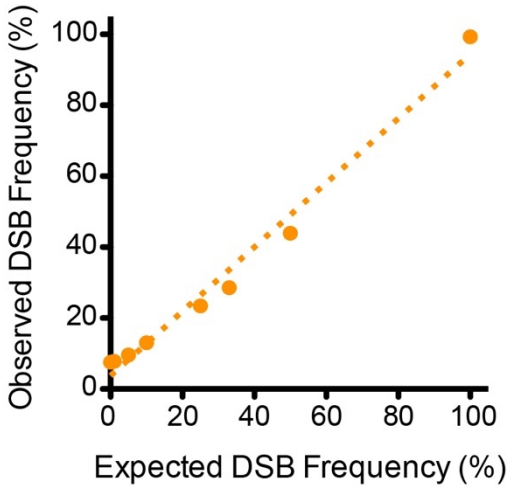

Figure A.2. Example of a standard curve created using restriction enzyme digested gDNA.

A.2.4 DSB-ddPCR to quantify DSBs at a Cas9/ciCas9 targeted locus

Equipment

- Thermal cycler for 96-well plates
- PX1™ PCR plate sealer (Bio-Rad)
- QX200™ Droplet Generator (Bio-Rad)
- QX200™ Droplet Reader (Bio-Rad)
- DG8™ Cartridge Holder (Bio-Rad)

Consumables

- PCR plate heat seal, foil, pierceable (Bio-Rad)
- ddPCR™ 96-well plates (Bio-Rad)
- DG8™ Cartridges for QX200™ Droplet Generator (Bio-Rad)
- DG8™ Gaskets for QX200™ Droplet Generator (Bio-Rad)

Reagents

- ddPCR™ Supermix for Probes (Bio-Rad)
- Droplet generation oil for probes (Bio-Rad)
- ddPCR™ droplet reader oil (Bio-Rad)
- 10 mM forward and reverse primers mix with both target and template primers
- 5 mM target and template probes mix
- Nuclease free water

Mix for a 1x ddPCR reaction

10 µL 2x ddPCR™ Supermix for Probes

1.8 µL Primers mix (10 mM)

1 µL Probes mix (5 mM)

30-150 ng DNA (acceptable range)

Bring to final volume of 20 µL with nuclease free water

Protocol for generating droplets and PCR cycling

1) Place a DG8™ cartridge into the DG8™ cartridge holder.
2) Add 50 µL of droplet generation oil for probes into each oil well in the cartridge.
3) Transfer samples into the sample wells in the cartridge. Make sure there are no bubbles in the sample at the bottom of the well. If there are less than eight samples, fill the remaining wells with 20 µL of water.
4) Place a DG8™ gasket onto the cartridge and holder.
5) Place the cartridge and holder into the QX200™ droplet generator. Close the lid to begin droplet generation.
6) Pipette droplets carefully from cartridge into a ddPCR™ 96-well plate.
7) Repeat for all samples.

8) Seal the plate using a PCR plate pierceable foil heat seal in the PX1™ PCR plate sealer. Make sure to take out the plate support block when heating up the plate sealer to avoid heating samples when placing the plate into the support block.
9) Load the 96-well plate into a thermal cycler.
10) Run thermal cycler with program:

    10 min at 95 °C

    40 cycles of:

        30 sec at 94 °C

        60 sec at 60 °C

    10 min at 98 °C

    Hold at 4 °C

Protocol for loading QX200™ droplet reader

1) Take the sealed 96-well plate from the thermal cycler and place it in the metal plate holder for the QX200™ droplet reader.
2) Open the QuantaSoft™ software and create a new plate layout template.
3) Select wells with droplets to be analyzed and input the sample name, experiment type (choose “Rare Event Detection”), master mix used (choose “ddPCR Supermix for Probes”), target name(s), target type(s) (choose “Unknown” for both channels), and fluorescent channels (choose Ch1 for FAM and Ch2 for HEX).
4) Click “Run” to start the droplet reading process.

Notes

- Probe and primer sets should first be tested using uncut and 100% cut DNA samples to ensure adequate detection. Different primer sets should be tested for a single probe.
- Restriction enzymes that cut distal to both amplicons can be used to improve amplification efficiency. Ideally, selected restriction enzyme cut sites should create a fragment approximately 3000 bp in length and neither enzyme should cut within or between the

target and template amplicons. If restriction enzymes are used for this purpose, 0.25 µL of each enzyme should be added to each 1x ddPCR reaction. No additional incubation steps are required.

- Droplets can be stored at 4 °C for 24 hours after PCR cycling. Avoid freezing droplets and storing at 4 °C for longer than 24 hours.

  More information about the use and setup of the QX200™ droplet reader can be found on the BioRad website.

  - https://www.bio-rad.com/en-us/product/qx200-droplet-digital-pcr-system?ID=MPOQQE4VY

A.2.5 Data analysis and DSB quantification

Processing of raw data and gating of droplets

- Droplets are analyzed using the QuantaSoft™ Analysis Pro software. Open the experiment and select wells and fluorescent channels to be analyzed.
- If there is good separation between positive and negative droplets, the fluorescence thresholds are automatically called and gates are automatically drawn. For probes that do not have good separation between positive and negative droplets, use the uncut and 100% cut samples from the standard curve to draw gates (**Appendix A Fig. A.3**).
- Data can be exported as a .csv file for DSB calculations.

Calculation and quantification of DSB frequency

- Raw DSB frequency for all samples and a standard curve can be calculated from the number of droplets using the following equation:

$$\frac{[target-, target+]}{[target-, target+] + [target+, target+]}$$

- The standard curve can be created by plotting raw DSB frequencies to expected DSB frequencies and performing a linear regression.

- Absolute DSB frequencies can then be calculated by fitting the raw DSB frequencies to the standard curve.

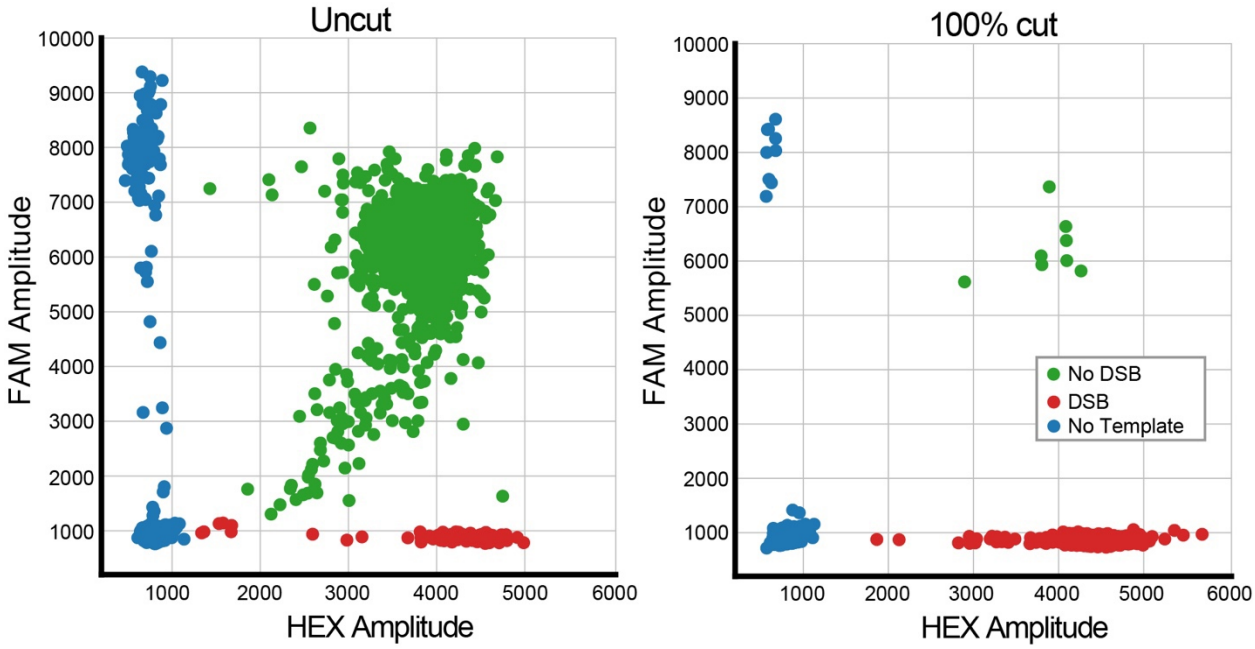

Figure A.3. Example of droplet gating to quantify cut versus. uncut gDNA.
The blue colored droplets represent those without the desired gDNA locus. The green colored droplets represent the uncut or repaired gDNA. The red colored droplets represent the cut gDNA.

A.3 Summary

We have engineered a rapidly-inducible, single-component Cas9 that can be temporally and rheostatically controlled to study Cas9 editing kinetics and dose-dependence *in vivo*. We have found that ciCas9 rapidly generates DSBs and subsequent indels upon activation with the small molecule A115. In addition, we have developed DSB-ddPCR to directly quantify the outcomes of Cas9 activity by measuring DSB generation *in vivo*. The coupling of these methods with high-throughput sequencing allows the kinetics and dose-dependence of CRISPR/Cas9 genome editing at specific loci of interest to be investigated. DSB-ddPCR can also be readily adapted to measure DSB generation in other contexts.

Appendix B: Supplementary material for chapter 2

A version of appendix B is in preparation for publication with chapter 2.

Wei, C.T., Stephany, J.J., Popp, N.A., Fowler, D.M., Maly, D.J.

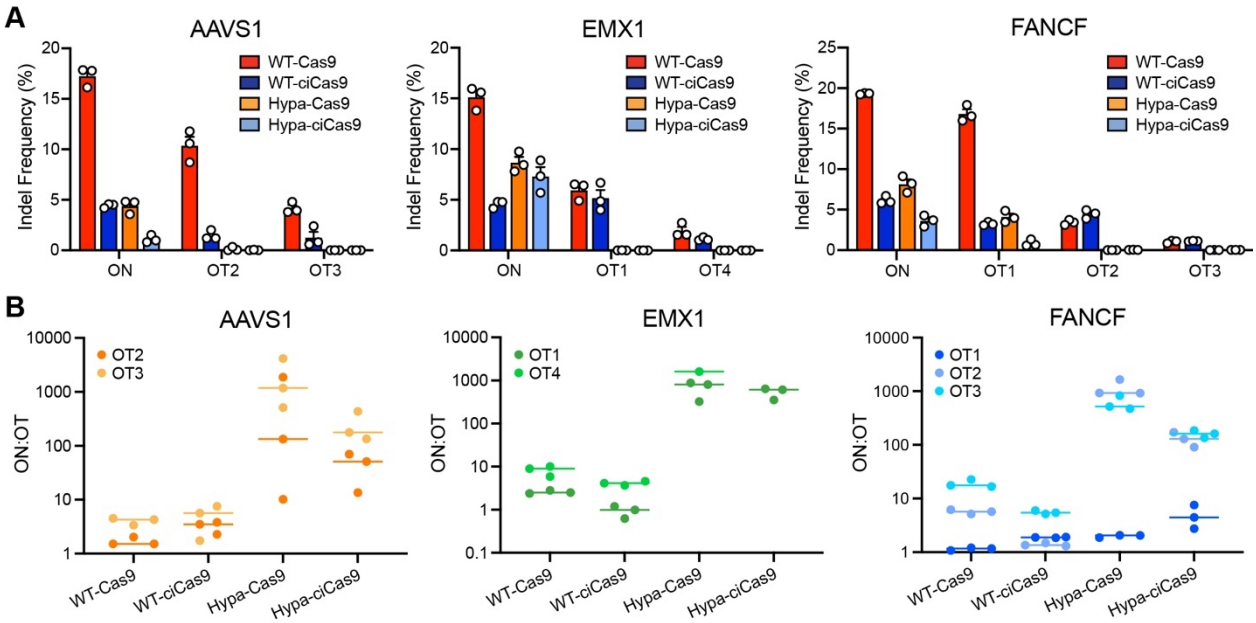

Figure B.1. Comparison of WT and Hypa variants of Cas9 and ciCas9.
A) Indel frequency of WT-Cas9, Hypa-Cas9, WT-ciCas9, and Hypa-ciCas9 at ON and OT with sgRNAs targeting AAVS1, EMX1, and FANCF 24 hr after transfection for WT- and Hypa-Cas9 and 24 hr after A115 addition for WT- and Hypa-ciCas9.
B) Specificity (ON:OT of indel frequencies) of WT-Cas9, WT-ciCas9, Hypa-Cas9, and Hypa-ciCas9 at off-target sites with sgRNAs targeting AAVS1, EMX1, and FANCF. Cas9 indels measured at 24 hr after transfection, ciCas9 indels measured at 24 hr after A115 addition. Some replicates of Hypa-ciCas9 editing at EMX1 OT1 and EMX1 OT4, had indel frequencies of 0% where we were unable to calculate specificity. Mann-Whitney U-test $p > 0.05$ for all comparisons of WT-Cas9 with WT-ciCas9 and Hypa-Cas9 with Hypa-ciCas9 ON:OT specificity where indel frequency could be measured.

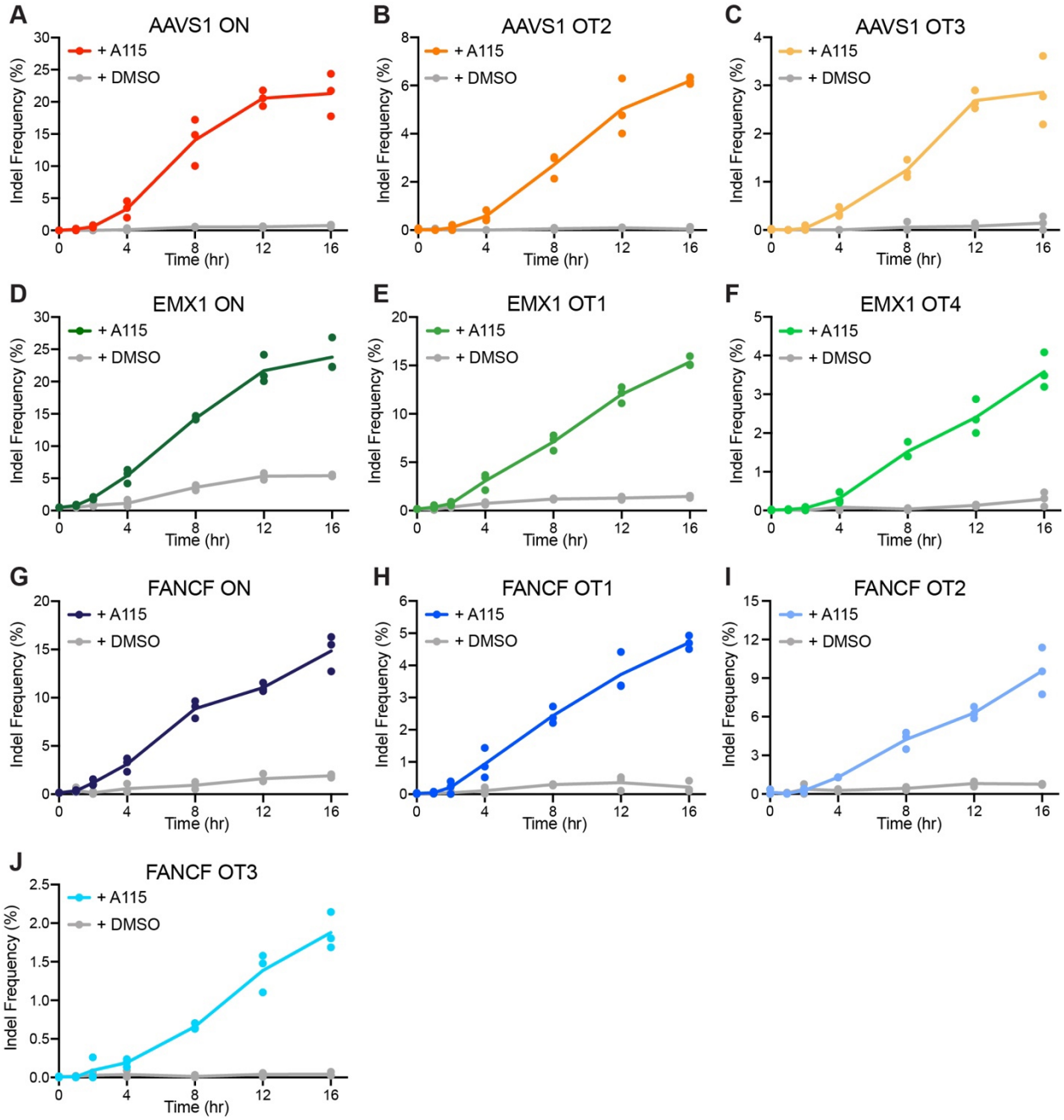

Figure B.2. ciCas9 time course of background editing with DMSO.

Indel frequency of ciCas9-induced editing over a 16 hr time course. A115 or DMSO is added to the cells at t=0 hr. Indel frequency is measured at **(A)** AAVS1 ON, **(B)** AAVS1 OT2, **(C)** AAVS1 OT3, **(D)** EMX1 ON, **(E)** EMX1 OT1, **(F)** EMX1 OT4, **(G)** FANCF ON, **(H)** FANCF OT1, **(I)** FANCF OT2 and **(J)** FANCF OT3 sites.

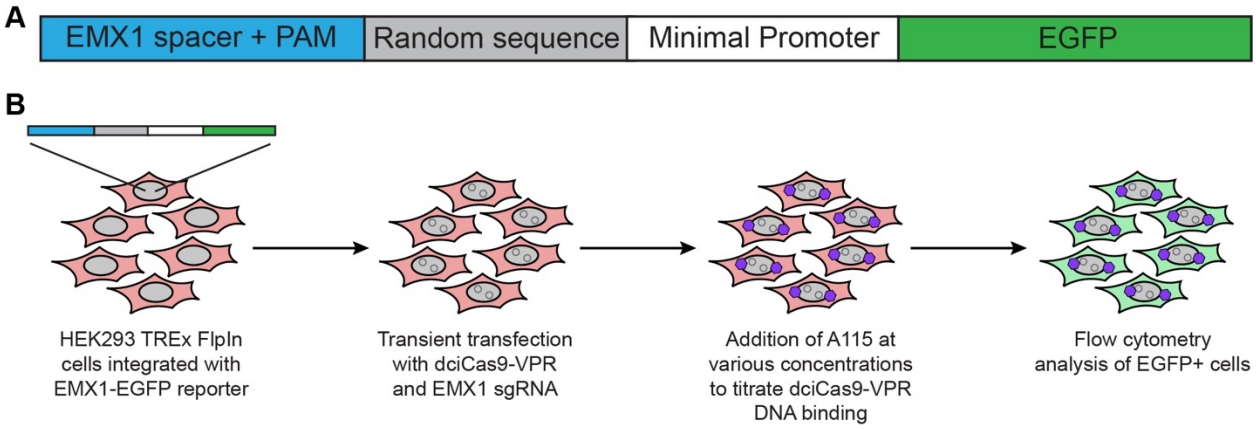

Figure B.3. Schematic of EGFP transcriptional synthetic locus

A) Schematic of the EMX1-EGFP transcriptional synthetic locus integrated into HEK293 TREx FlpIn cells.

B) Workflow of using EMX1-EGFP transcriptional synthetic locus cells with dciCas9-VPR + EMX1 sgRNA for a dose-response of EGFP expression.

	Guide	PAM
AAVS1 ON	GGGGCCACTAGGGACAGGAT	TGG
OT2	GGGACCATCAGGGACAGGAT	GGG
OT3	GGGGCCAATTAGGACAGGAT	GGG
EMX1 ON	GAGTCCGAGCAGAAGAAGAA	GGG
OT1	GAGTTAGAGCAGAAGAAGAA	AGG
OT4	GAGTCTAAGCAGAAGAAGAA	GAG
FANCF ON	GCTGCAGAAGGGATTCCATG	AGG
OT1	GCTGCAGAAGGGATTCCAAG	GGG
OT2	GACGCAGAAGGGACTCCATG	GGG
OT3	GGTACAGAAGGGCTTCCATG	AGG

Figure B.4. Sequences of on- and off-target sites used with ciCas9.

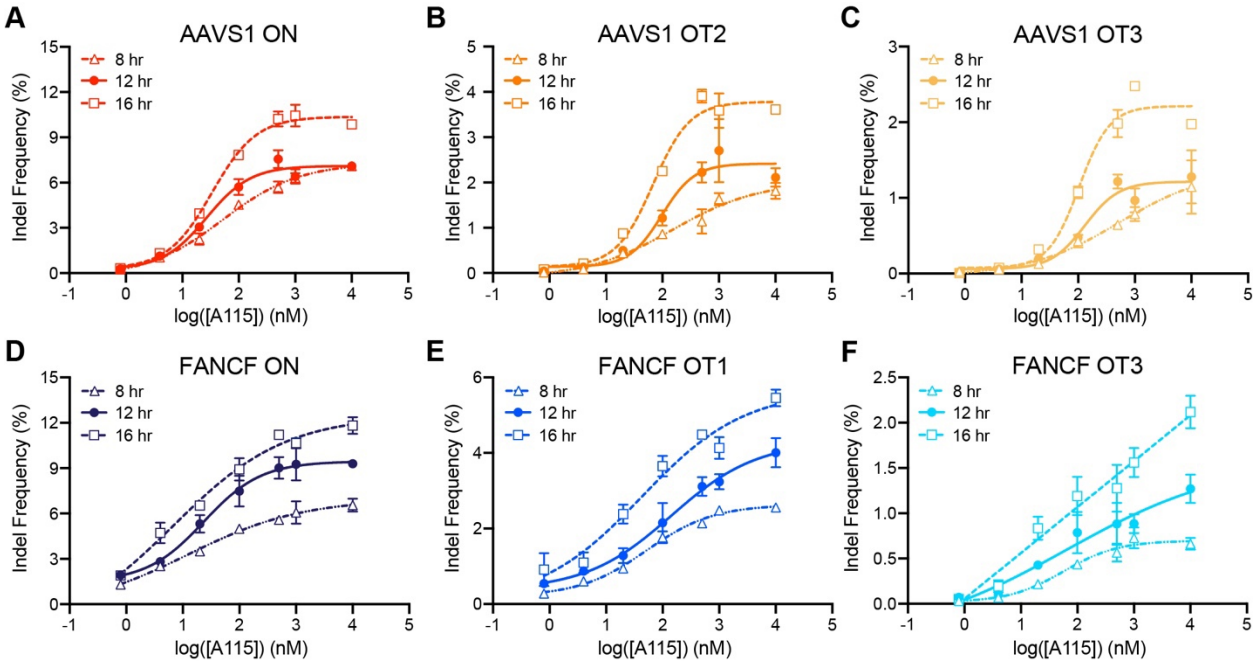

Figure B.5. ciCas9 dose-response over time for on- and off-target sites.
Dose-response of ciCas9 editing at **(A)** AAVS1 ON, **(B)** AAVS1 OT2, **(C)** AAVS1 OT3, **(D)** FANCF ON, **(E)** FANCF OT1, and **(F)** FANCF OT3 at 8, 12, and 16 hr after A115 addition. Data represented as mean ON:OT ± SEM of 3 cell culture replicates.

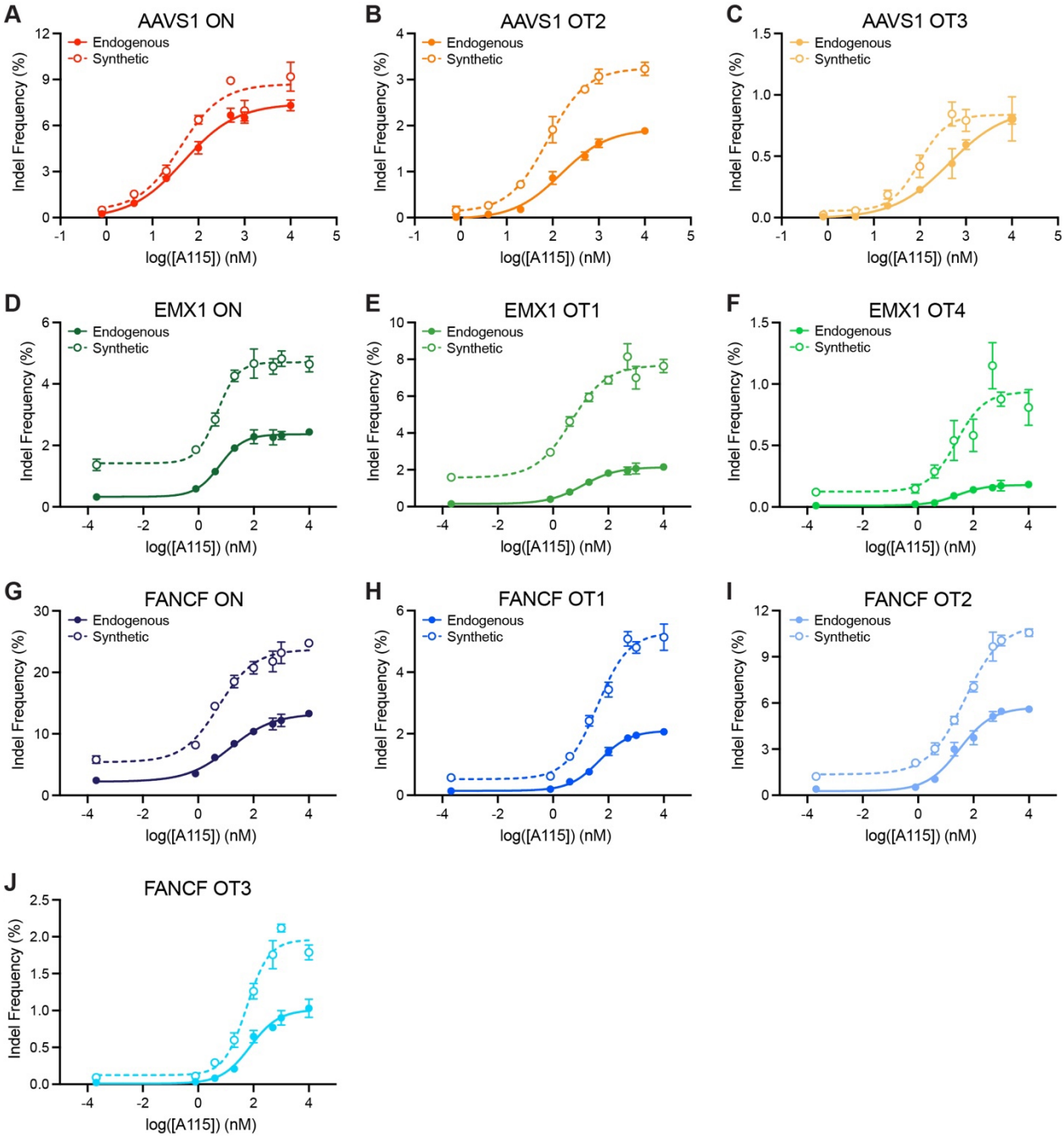

Figure B.6. ciCas9 dose-response at endogenous and synthetic loci.

Dose-response of ciCas9 editing at endogenous and synthetic loci for (**A**) AAVS1 ON, (**B**) AAVS1 OT2, (**C**) AAVS1 OT3, (**D**) EMX1 ON, (**E**) EMX1 OT1, (**F**) EMX1 OT4, (**G**) FANCF ON, (**H**) FANCF OT1, (**I**) FANCF OT2, and (**J**) FANCF OT3 synthetic locus HEK293T BFP iCasp9 landing pad cell lines. For all EMX1 and FANCF on- and off-target sites, indel frequency for DMSO was incorporated into EC_{50} calculations as 0.0002048 nM A115. Data represented as indel frequency ± SEM of 3 cell culture replicates.

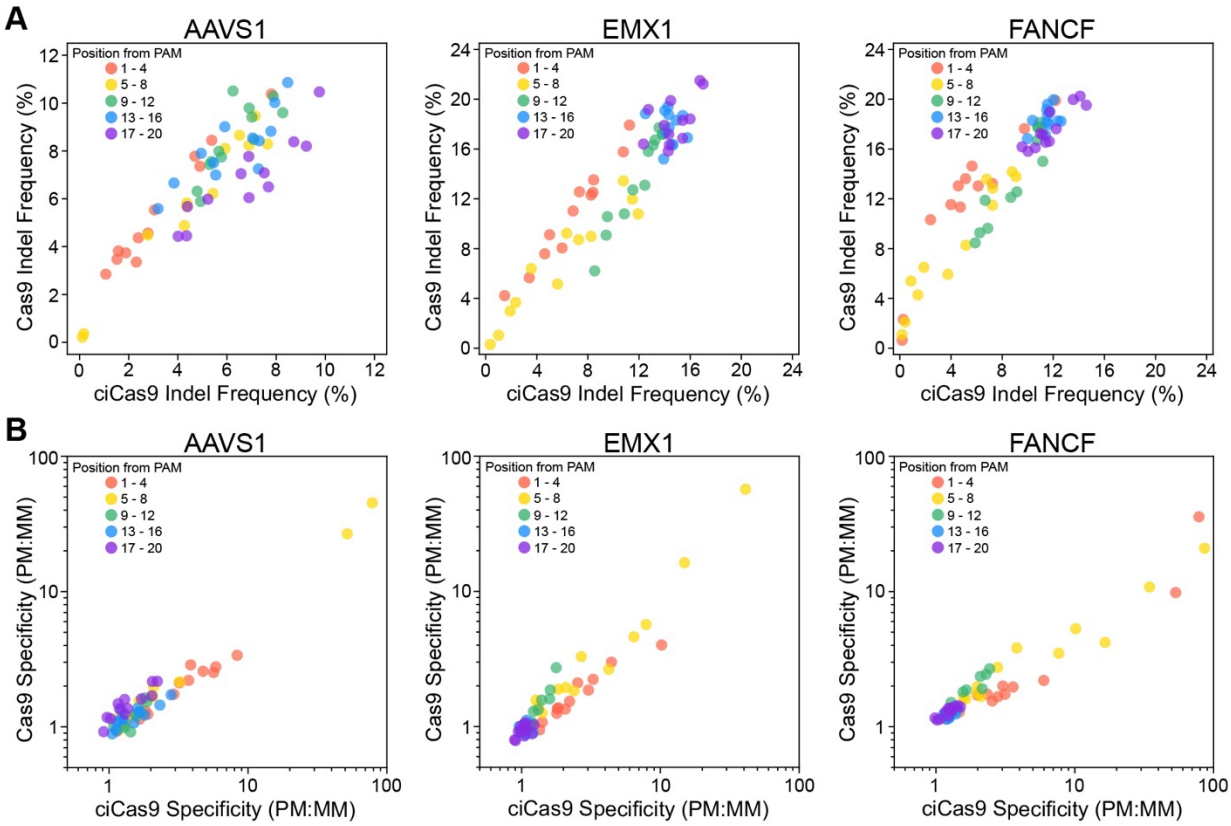

Figure B.7. WT-ciCas9 vs. WT-Cas9 in mismatched libraries.

A) Comparison of WT-Cas9 and WT-ciCas9 indel frequencies in the AAVS1, EMX1, and FANCF mismatch libraries. Each point represents the mean indel frequency for a single mismatch in the library and the color corresponds to the position of the mismatch in the target sequence. For WT-ciCas9 editing, indel frequencies were measured at 10 μM A115.

B) Comparison of WT-Cas9 and WT-ciCas9 specificities in the AAVS1, EMX1, and FANCF mismatch libraries. Specificity was calculated by taking the ratio of the mean indel frequency at the perfectly matched target sequence to the mean indel frequency at the mismatched target sequence. Each point represents the mean specificity for a single mismatch in the library and the color corresponds to the position of the mismatch in the target sequence. For WT-ciCas9 editing, indel frequencies were measured at 10 μM A115.

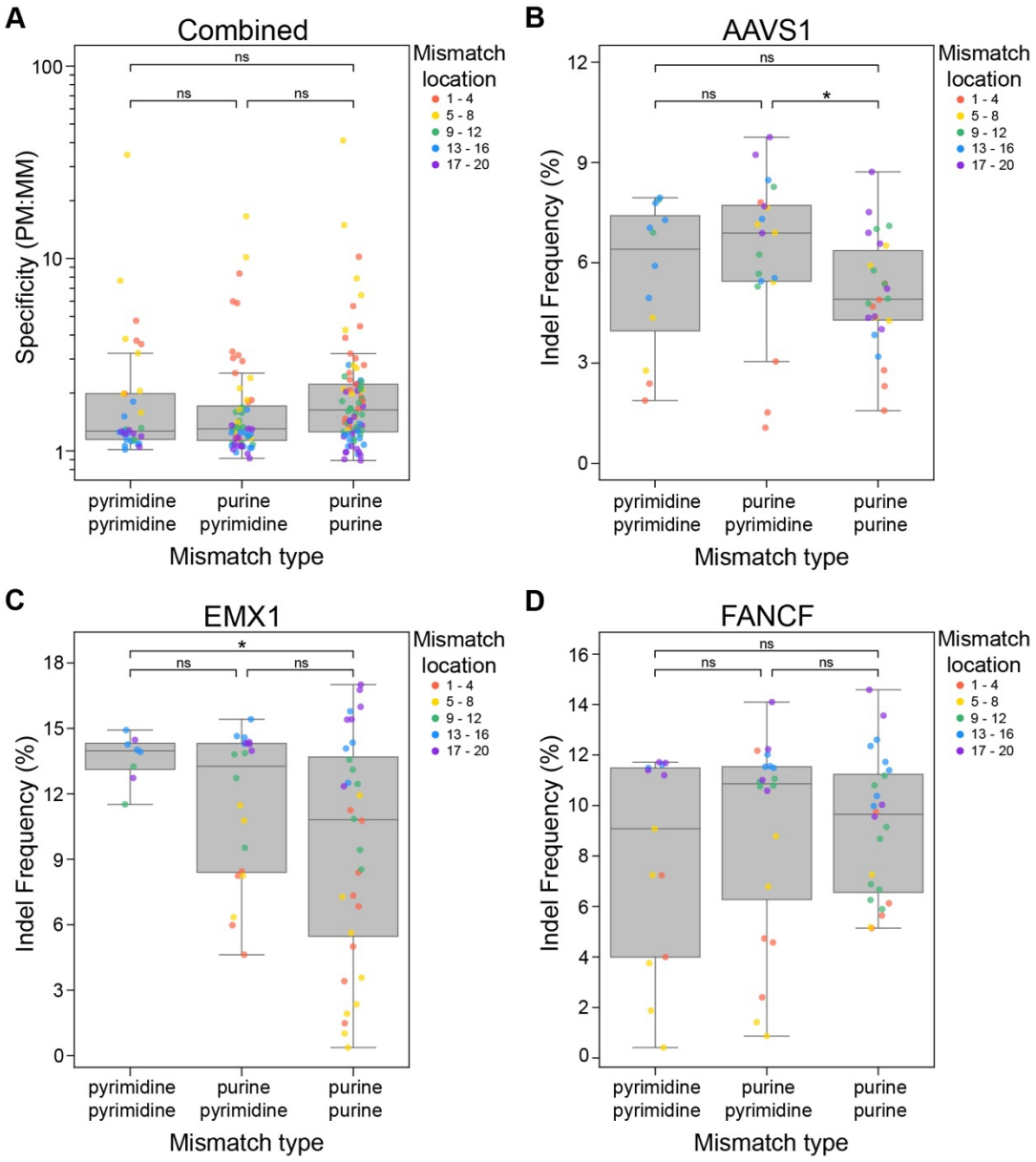

Figure B.8. Mismatch type (i.e. purine-purine, purine-pyrimidine, pyrimidine-pyrimidine) vs. indel frequency

A) Boxplots showing the distribution of calculated editing specificity for different types of mismatches combined from all three single mismatch libraries. Specificity was calculated as a ratio of mean indel frequency with 10 μM A115 at the perfectly matched sequence to the mean indel frequency at each mismatched sequence. Mismatches were filtered for those that showed >0.3% indel frequency. Each point represents the specificity for a single mismatch in the library and the color corresponds to the position of the mismatch in the target sequence.

B-D) Boxplots showing the distribution of indel frequency for different types of mismatches for the (**B**) AAVS1, (**C**) EMX1, and (**D**) FANCF libraries. Mismatches were filtered for those that showed >0.3% indel frequency. Each point represents the mean indel frequency for a single mismatch in the library and the color corresponds to the position of the mismatch in the target sequence.

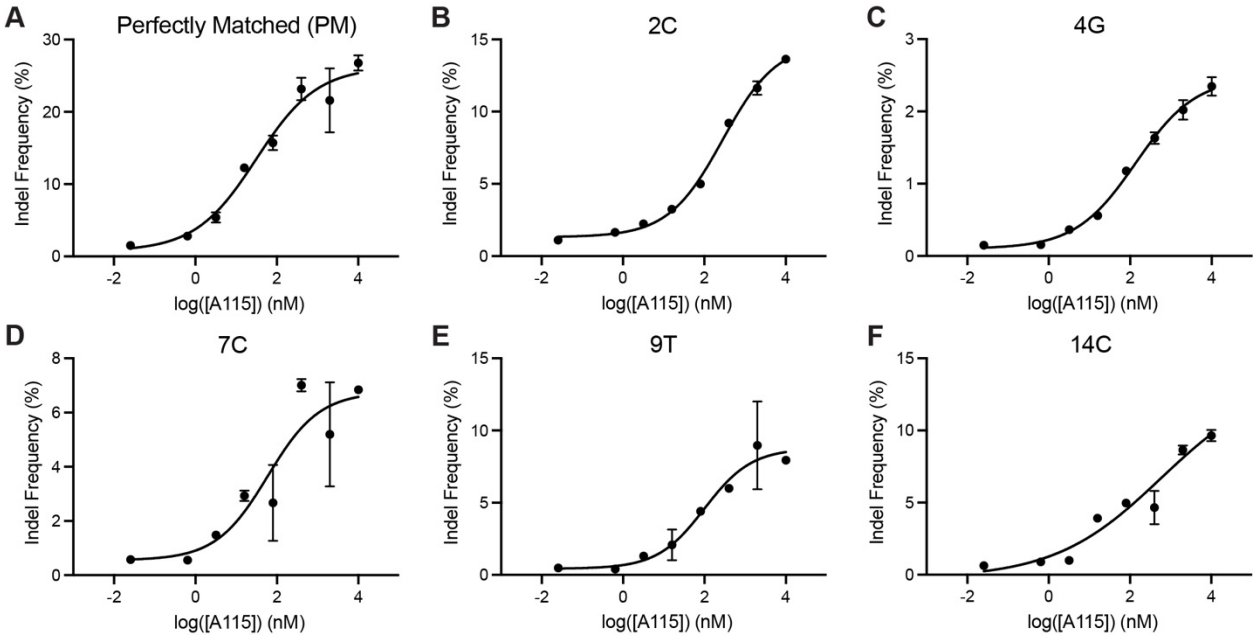

Figure B.9. Validation mismatch titrations, AAVS1.

Dose-response of ciCas9 editing at single synthetic target sites with the indicated single mismatches in the AAVS1 target sequence. Data represented as indel frequency ± SEM of 3 cell culture replicates.

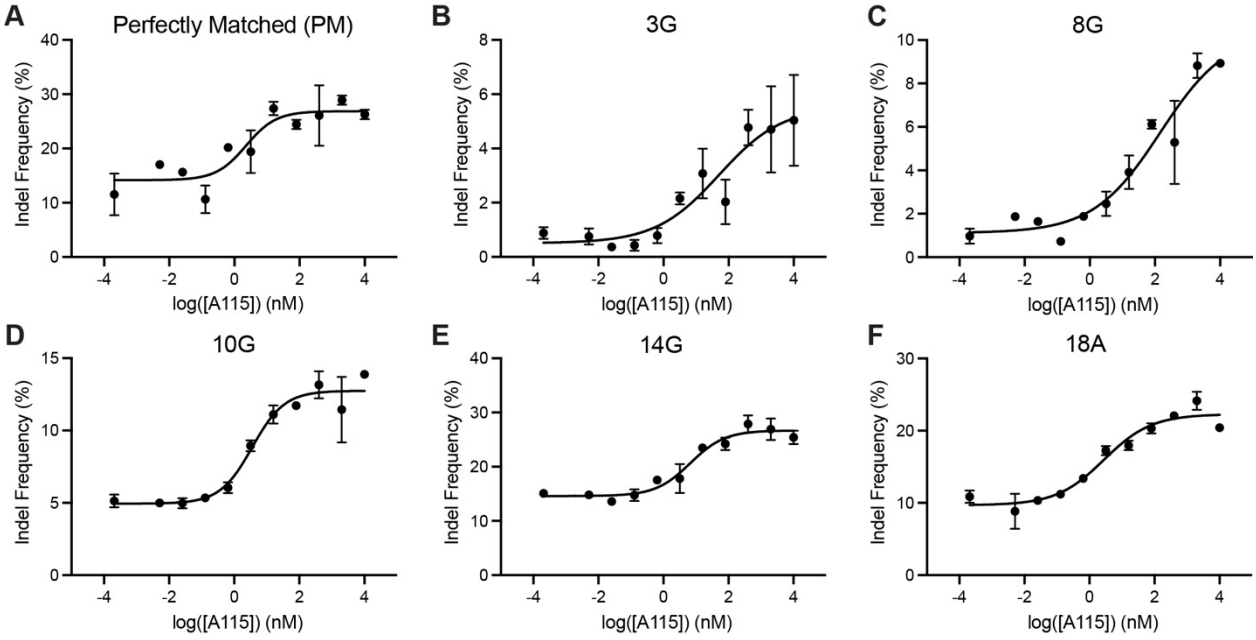

Figure B.10. Validation mismatch titrations, EMX1.

Dose-response of ciCas9 editing at single synthetic target sites with the indicated single mismatches in the EMX1 target sequence. Indel frequency measured for DMSO was incorporated into EC_{50} calculations as 0.0002048 nM A115. Data represented as indel frequency ± SEM of 3 cell culture replicates.

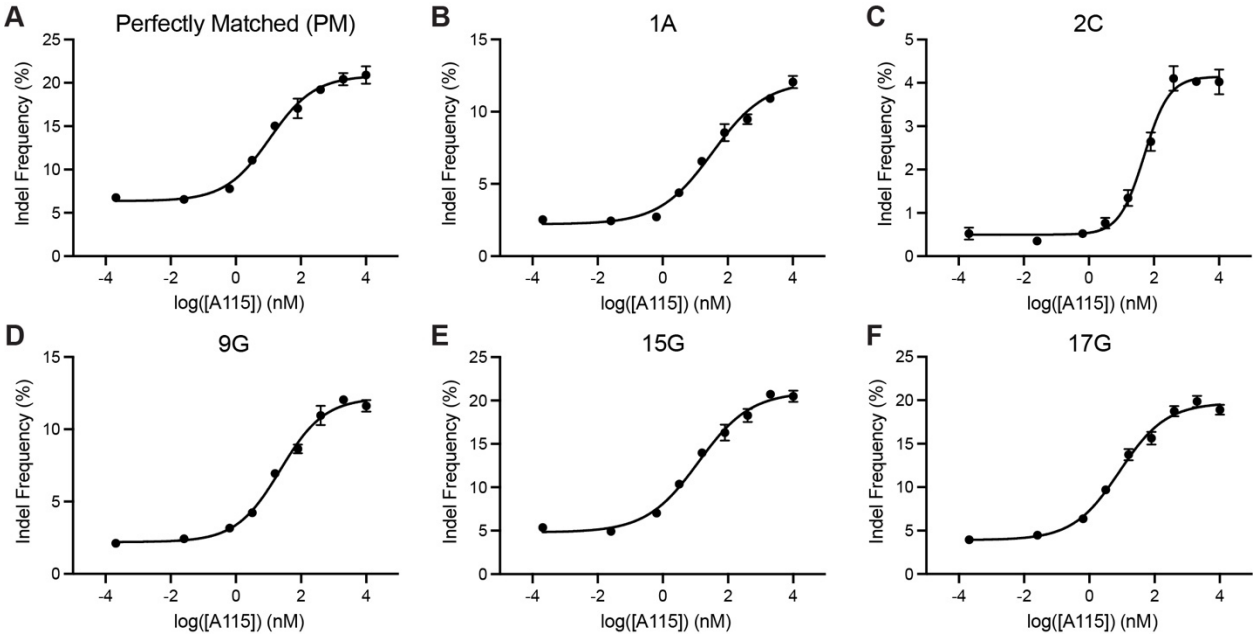

Figure B.11. Validation mismatch titrations, FANCF.

Dose-response of ciCas9 editing at single synthetic target sites with the indicated single mismatches in the FANCF target sequence. Indel frequency measured for DMSO was incorporated into EC_{50} calculations as 0.0002048 nM A115. Data represented as indel frequency ± SEM of 3 cell culture replicates.

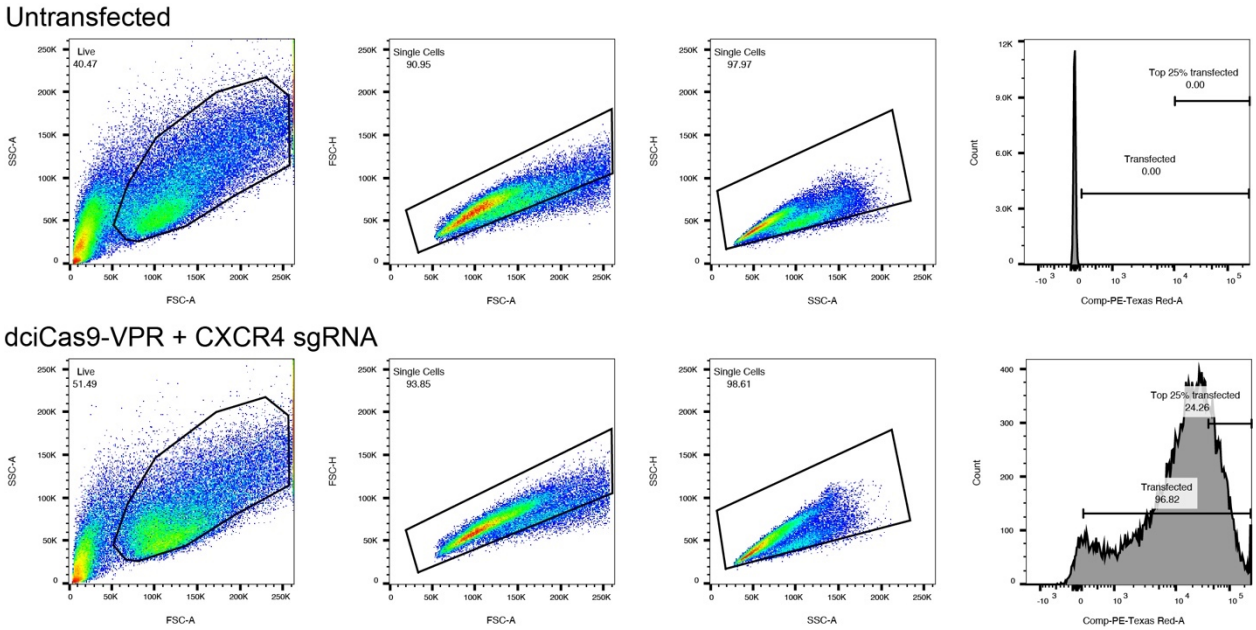

Figure B.12. Cell gating strategies for CXCR4 transcriptional activation.

Example gates set for determining CXCR4 transcriptional activation measured by median APC fluorescence of an anti-CXCR4 antibody, corresponding to **Fig. 2.2B**. HEK293T cells were transfected with the indicated plasmids, outlined in the methods. Untransfected cells were stained with anti-CXCR4 antibody to determine background CXCR4 expression on HEK293T cells.

dciCas9-VPR + EMX1 sgRNA

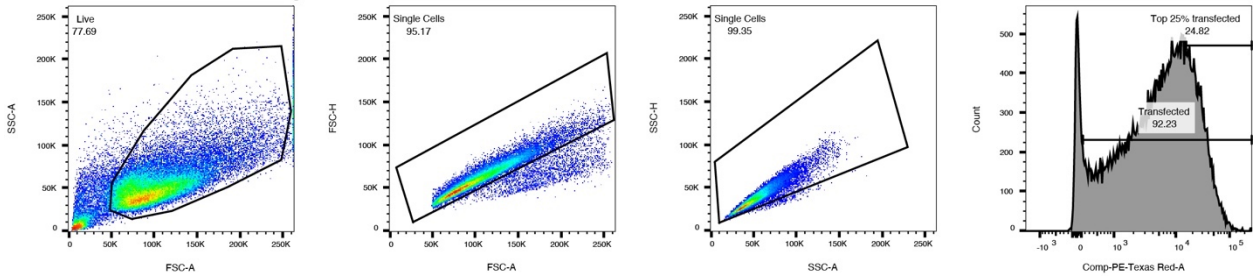

Figure B.13. Cell gating strategies for EGFP transcriptional activation.

Example gates set for determining EGFP transcriptional activation measured by EGFP fluorescence, corresponding to **Fig. 2.2C**. HEK293 TREx FlpIn EMX1-EGFP cells were transfected with the indicated plasmids, outlined in the methods.

Table B.1. Mann-Whitney U-test statistical analysis results.

Figure	Condition 1	Condition 2	P-value
Fig. 2.1D.	FANCF, ON:OT, 24 hr	FANCF ON:OT, 48 hr	0.2000
	HBB, ON:OT, 24 hr	HBB ON:OT, 48 hr	0.7000
	VEGFA, ON:OT, 24 hr	VEGFA ON:OT, 48 hr	0.7000
Appendix B Fig. B.1B.	AAVS1 OT2, ON:OT, WT-Cas9	AAVS1 OT2, ON:OT, WT-ciCas9	0.1000
	AAVS1 OT3, ON:OT, WT-Cas9	AAVS1 OT3, ON:OT, WT-ciCas9	0.7000
	AAVS1 OT2, ON:OT, Hypa-Cas9	AAVS1 OT2, ON:OT, Hypa-ciCas9	0.7000
	AAVS1 OT3, ON:OT, Hypa-Cas9	AAVS1 OT3, ON:OT, Hypa-ciCas9	0.7000
	EMX1 OT1, ON:OT, WT-Cas9	EMX1 OT1, ON:OT, WT-ciCas9	0.1000
	EMX1 OT4, ON:OT, WT-Cas9	EMX1 OT4, ON:OT, WT-ciCas9	0.1000
	EMX1 OT1, ON:OT, Hypa-Cas9	EMX1 OT1, ON:OT, Hypa-ciCas9	0.7000
	FANCF OT1, ON:OT, WT-Cas9	FANCF OT1, ON:OT, WT-ciCas9	0.1000
	FANCF OT2, ON:OT, WT-Cas9	FANCF OT2, ON:OT, WT-ciCas9	0.1000
	FANCF OT3, ON:OT, WT-Cas9	FANCF OT3, ON:OT, WT-ciCas9	0.1000
	FANCF OT1, ON:OT, Hypa-Cas9	FANCF OT1, ON:OT, Hypa-ciCas9	0.1000
	FANCF OT2, ON:OT, Hypa-Cas9	FANCF OT2, ON:OT, Hypa-ciCas9	0.1000
	FANCF OT3, ON:OT, Hypa-Cas9	FANCF OT3, ON:OT, Hypa-ciCas9	0.1000

Appendix C: Supplementary material for chapter 3

A version of appendix C is in preparation for publication with chapter 3.

Wei, C.T., Maly, D.J., Fowler, D.M.

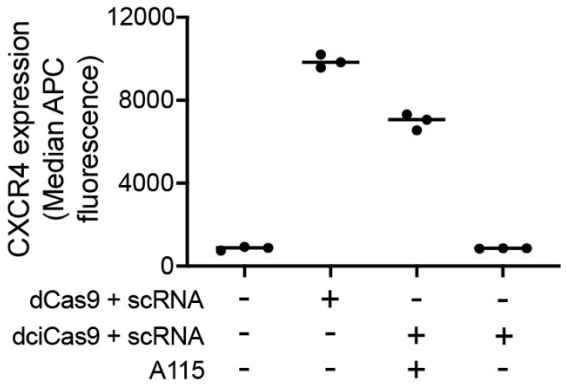

Figure C.1. dciCas9 transcriptional activation using a scRNA
Activation of CXCR4 expression using dCas9 or dciCas9 with a scRNA targeted to the promoter region in HEK-293T cells that recruits MCP-VPR in the presence or absence of 1 μM A115 activation for 48 hr. Data represented as 3 cell culture replicates shown with a line showing the mean.

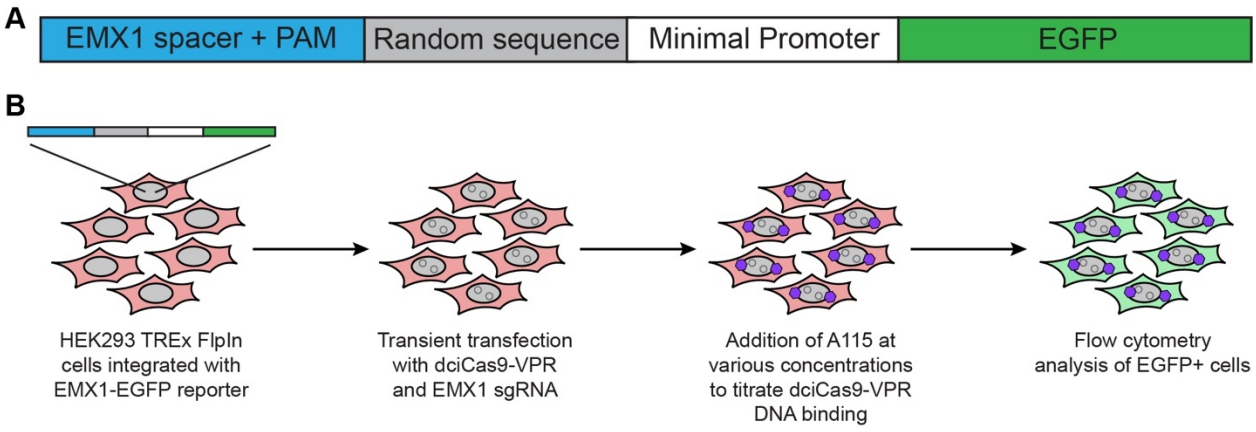

Figure C.2. Schematic of EMX1-EGFP reporter locus in HEK293 TREx FlpIn cells.
A) Schematic of the EMX1-EGFP transcriptional synthetic locus integrated into HEK293 TREx FlpIn cells.
B) Workflow of using EMX1-EGFP transcriptional synthetic locus cells with dciCas9-VPR + EMX1 sgRNA for a dose-response of EGFP expression.

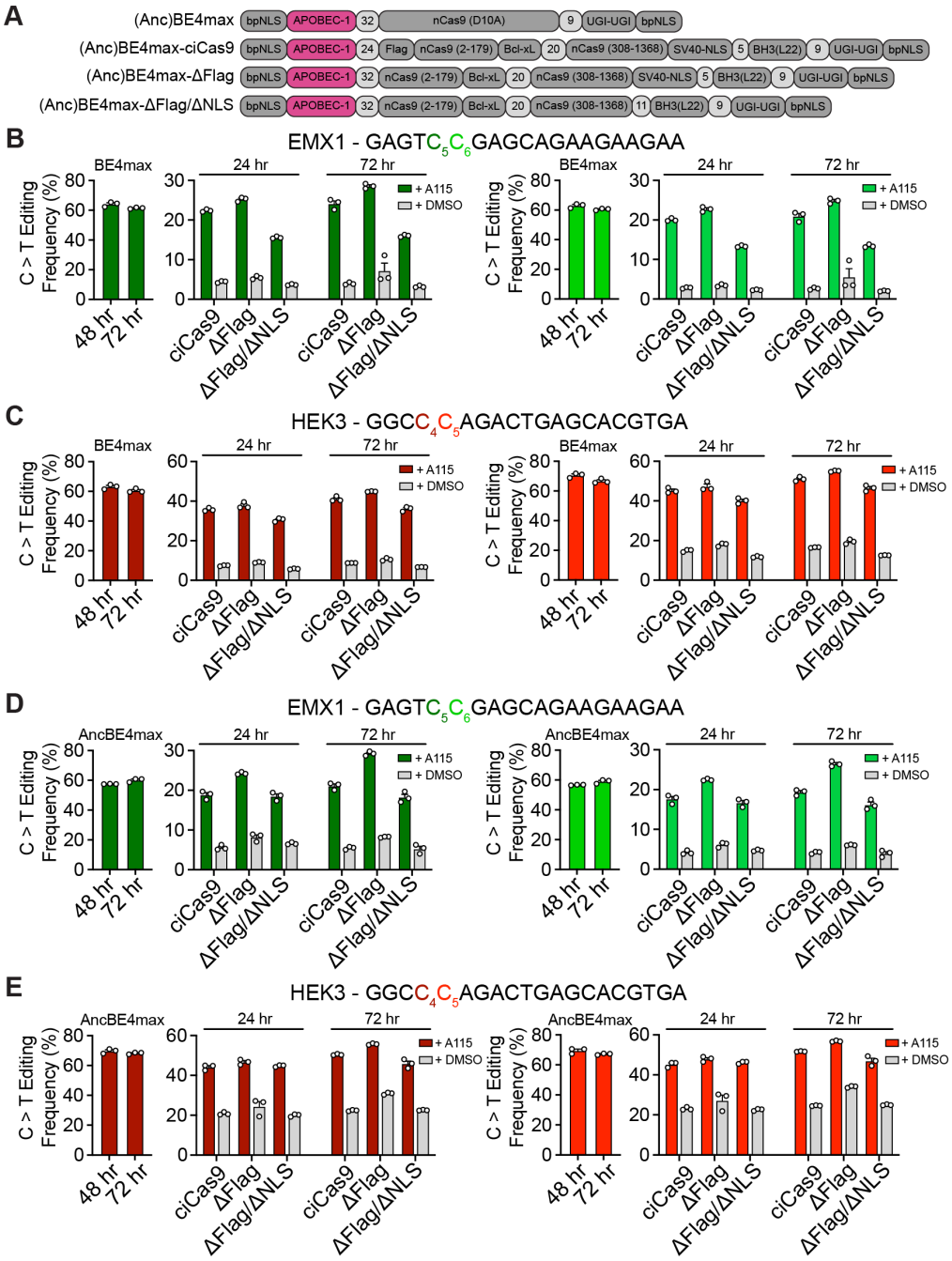

Figure C.3. Chemically-controlled cytidine base editors without codon optimization.

A) Schematic of the domain arrangements in the unmodified BE4max and AncBE4max base editors and the chemically-controlled ciBE4max and ciAncBE4max base editors without codon optimization and using the ciCas9(L22) variant. 3 different versions of ciCas9 were used, ciCas9(L22), ciCas9(L22) without a Flag-tag (ΔFlag), and ciCas9(L22) without a Flag-tag and additional SV40-NLS (ΔFlag/ΔNLS).

B-C) C-to-T editing frequency with BE4max and BE4max-ciCas9 at the EMX1 **(B)** and HEK3 **(C)** target sites.

D-E) C-to-T editing frequency with AncBE4max and AncBE4max-ciCas9 at the EMX1 **(D)** and HEK3 **(E)** target sites.

Figure C.3 cont.
BE4max and AncBE4max editing were measured at 48 and 72 hr after co-transfection of BE4max and sgRNA. BE4max-ciCas9 and AncBE4max-ciCas9 editing were measured at 24 and 72 hr after 1 μM A115 addition. C-to-T editing is shown at the 2 nucleotides in each target site with highest editing frequency with the Cas9 version of base editors (BE4max or AncBE4max). The 2 different nucleotides are indicated by color in the target sequence. Bars show mean editing frequency ± SEM of 3 cell culture replicates with white circles showing individual replicates.

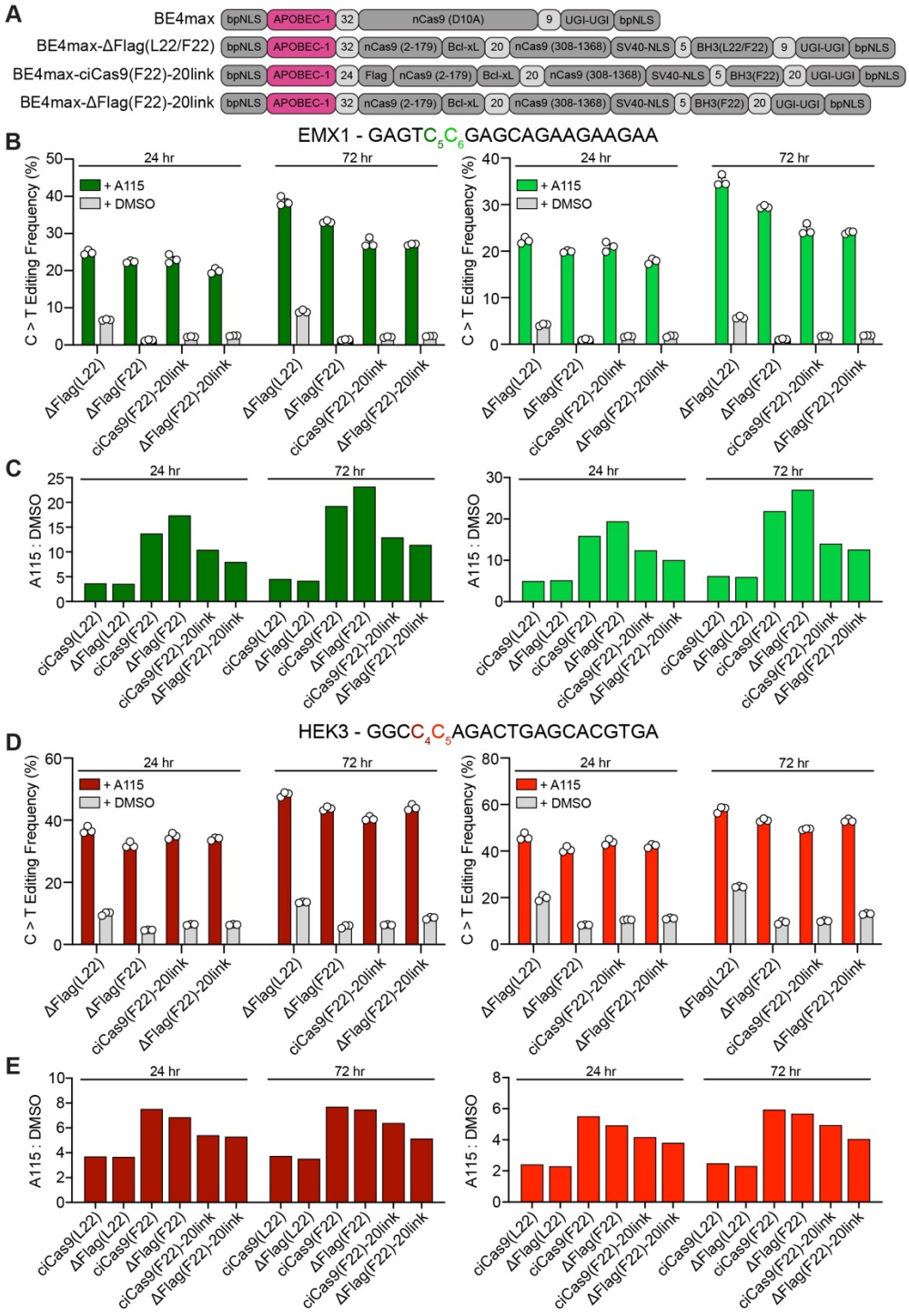

Figure C.4. Additional constructs of codon optimized chemically-controlled BE4max editors.

A) Schematic of domains in the unmodified BE4max base editor and additional constructs of the codon optimized BE4max-ciCas9 base editors tested. 4 different versions of ciCas9 were additionally tested: No-Flag-ciCas9(L22) (ΔFlag(L22)), No-Flag-ciCas9(F22) (ΔFlag(F22)), ciCas9(F22)-20linker-2xUGI (ciCas9(F22)-20link), and No-Flag-ciCas9(F22)-20linker-2xUGI (ΔFlag(F22)-20link).

Figure C.4 cont.
B,D) C-to-T editing frequencies of the BE4max-ciCas9 constructs at the EMX1 **(B)** and HEK3 **(D)** target sites. C-to-T editing is shown at the 2 nucleotides in each target site with highest editing frequency with BE4max. The 2 different nucleotides are indicated by color in the target sequence. Editing by all BE4max-ciCas9 constructs are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing frequency ± SEM of 3 cell culture replicates with white circles showing individual replicates.
C,E) Ratio of the mean C-to-T editing frequency with 1 μM A115 to the mean C-to-T editing frequency with DMSO (A115:DMSO) for all tested BE4max-ciCas9 base editors in Fig. 2 and Supplemental Fig. 4 at the EMX1 **(C)** and HEK3 **(E)** target sites. Bars show the ratios of editing at the 2 nucleotides at each target site with highest editing frequency with BE4max. Editing frequencies used to calculate the ratio were measured at 24 and 72 hr after A115 addition to HEK293T cells.

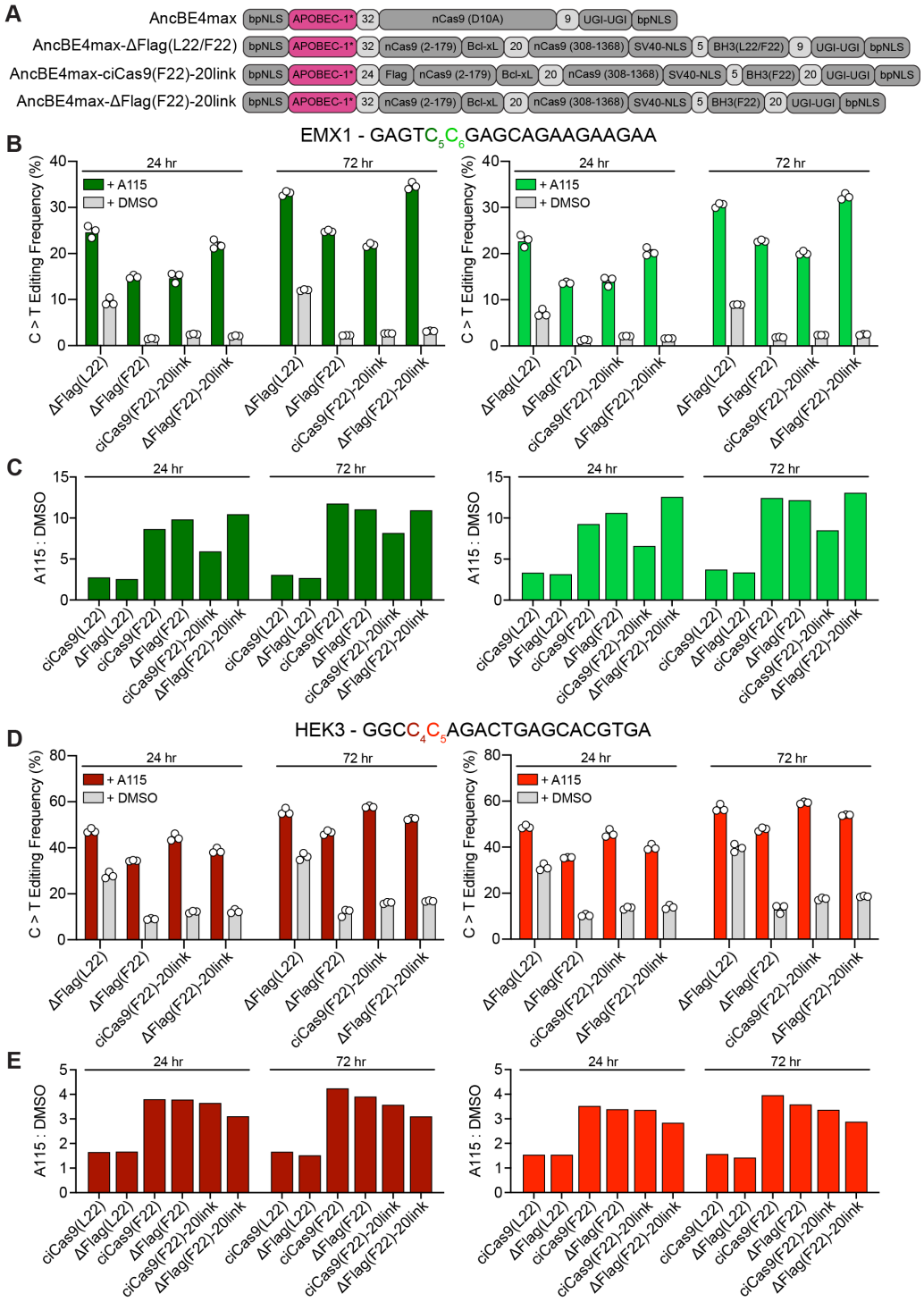

Figure C.5. Additional constructs of codon optimized chemically-controlled AncBE4max editors.

A) Schematic of domains in the unmodified AncBE4max base editor and additional constructs of the codon optimized AncBE4max-ciCas9 base editors tested. 4 different versions of ciCas9 were additionally tested: No-Flag-ciCas9(L22) (ΔFlag(L22)), No-Flag-ciCas9(F22) (ΔFlag(F22)), ciCas9(F22)-20linker-2xUGI (ciCas9(F22)-20link), and No-Flag-ciCas9(F22)-20linker-2xUGI (ΔFlag(F22)-20link).

Figure C.5 cont.
B,D) C-to-T editing frequencies of the AncBE4max ciCas9 constructs at the EMX1 **(B)** and HEK3 **(D)** target sites. C-to-T editing is shown at the 2 nucleotides at each target site with highest editing frequency with AncBE4max. The 2 different nucleotides are indicated by color in the target sequence. Editing by all AncBE4max-ciCas9 constructs are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing frequency ± SEM of 3 cell culture replicates with white circles showing individual replicates.
C,E) Ratio of the mean C-to-T editing frequency with 1 μM A115 to the mean C-to-T editing frequency with DMSO (A115:DMSO) for all tested AncBE4max-ciCas9 base editors in **Fig. 3.2** and **Appendix C Fig. C.5** at the EMX1 **(C)** and HEK3 **(E)** target sites. Bars show the ratios of editing at the 2 nucleotides in each target site with highest editing frequency with AncBE4max. Editing frequencies used to calculate the ratio were measured at 24 and 72 hr after A115 addition to HEK293T cells.

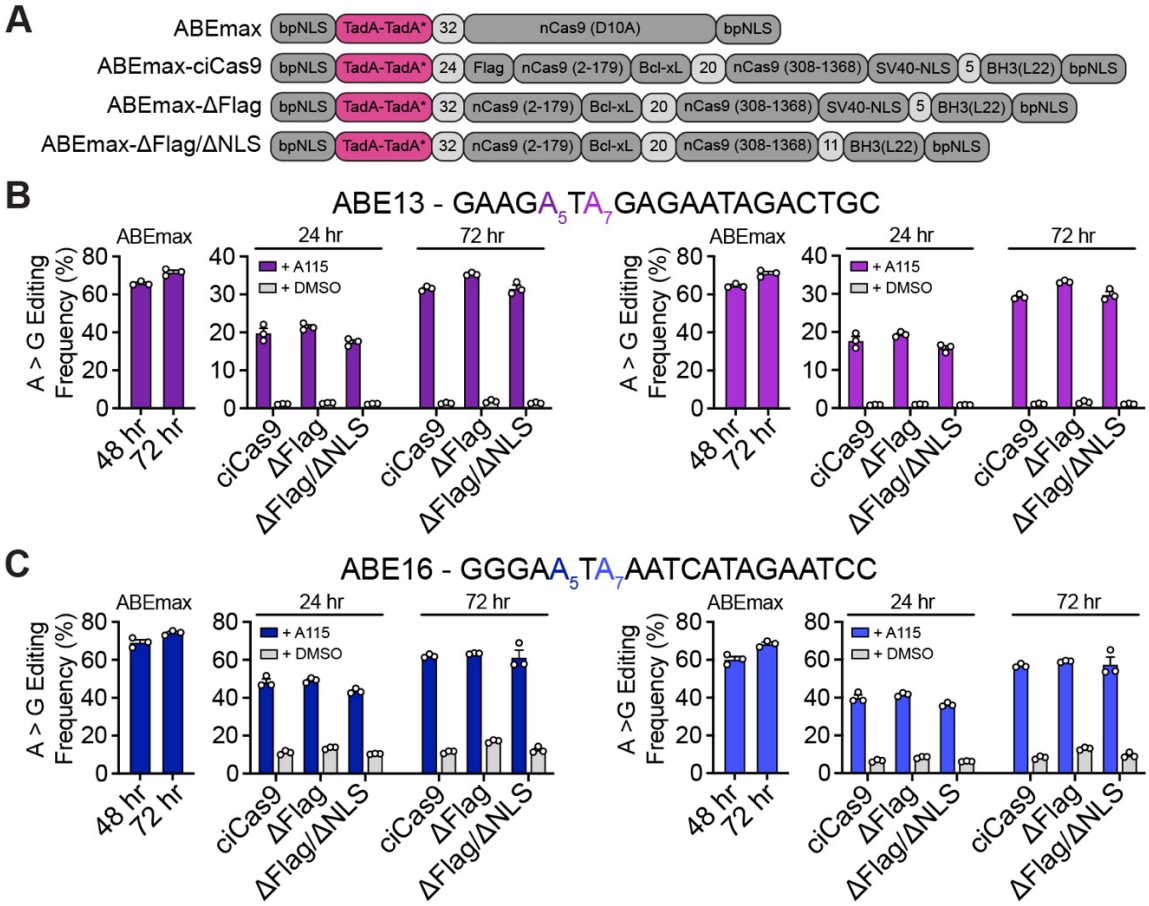

Figure C.6. Chemically-controlled adenine base editors without codon optimization.
A) Schematic of the domain arrangements in the unmodified ABEmax base editor and the chemically-controlled ciABEmax base editor without codon optimization and using the ciCas9(L22) variant. 3 different versions of ciCas9 were used, ciCas9(L22), ciCas9(L22) without a Flag-tag (ΔFlag), and ciCas9(L22) without a Flag-tag and additional SV40-NLS (ΔFlag/ΔNLS).
B-C) A-to-G editing frequency with ABEmax and ABEmax-ciCas9 at the ABE13 **(B)** and ABE16 **(C)** target sites.

Figure C.6 cont.
ABEmax editing was measured at 48 and 72 hr after co-transfection of ABEmax and sgRNA. ABEmax-ciCas9 editing was measured at 24 and 72 hr after 1 μM A115 addition. A-to-G editing is shown at the 2 nucleotides in each target site with highest editing frequency with the Cas9 version of the ABEmax base editor. The 2 different nucleotides are indicated by color in the target sequence. Bars show mean editing frequency ± SEM of 3 cell culture replicates with white circles showing individual replicates.

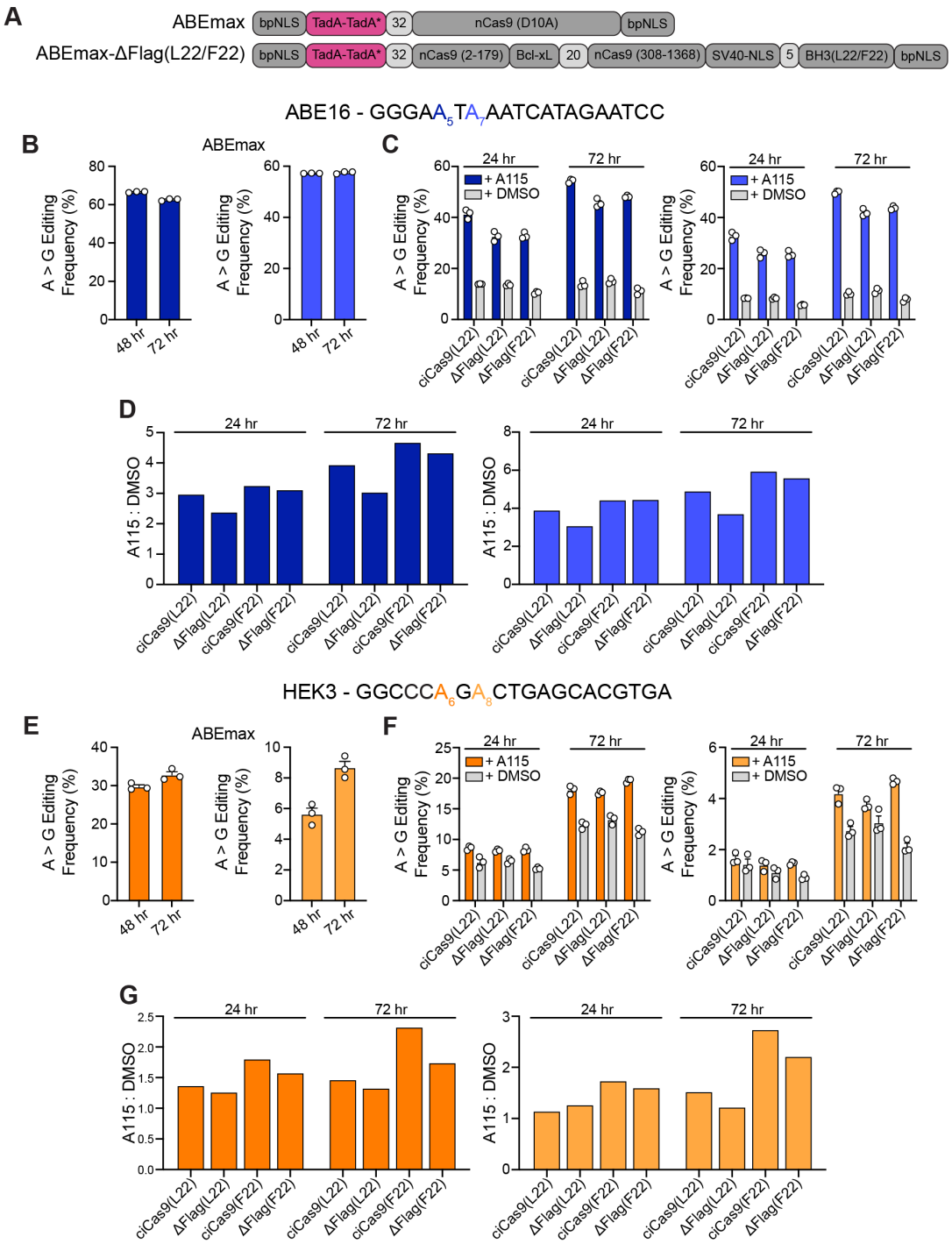

Figure C.7. Additional constructs of codon optimized chemically-controlled ABEmax editors.

A) Schematic of domains in the unmodified ABEmax base editor and additional constructs of the codon optimized ABEmax-ciCas9 base editors tested. 3 different versions of ciCas9 were additionally tested: ciCas9(L22), No-Flag-ciCas9(L22) (ΔFlag(L22)), and No-Flag-ciCas9(F22) (ΔFlag(F22)).

B,E) A-to-G editing frequencies of the unmodified ABEmax base editor at the ABE16 **(B)** and HEK3 **(E)** target sites.

Figure C.7 cont.
C,F) A-to-G editing frequencies of the ABEmax-ciCas9 constructs at the ABE16 **(C)** and HEK3 **(F)** target sites.
In (**B-C, E-F**) A-to-G editing is shown at the 2 nucleotides at each target site with highest editing frequency with ABEmax. The 2 different nucleotides are indicated by color in the target sequence. Editing by all ABEmax-ciCas9 constructs are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.
D,G) Ratio of the mean A-to-G editing frequency with 1 μM A115 to the mean A-to-G editing frequency with DMSO (A115:DMSO) for all tested ABEmax-ciCas9 base editors in **Fig. 3.3** and **Appendix C Fig. C.7** at the ABE16 **(D)** and HEK3 **(G)** target sites. Bars show the ratios of editing at the 2 nucleotides at each target site with highest editing frequency with ABEmax. Editing frequencies used to calculate the ratio were measured at 24 and 72 hr after A115 addition to HEK293T cells.

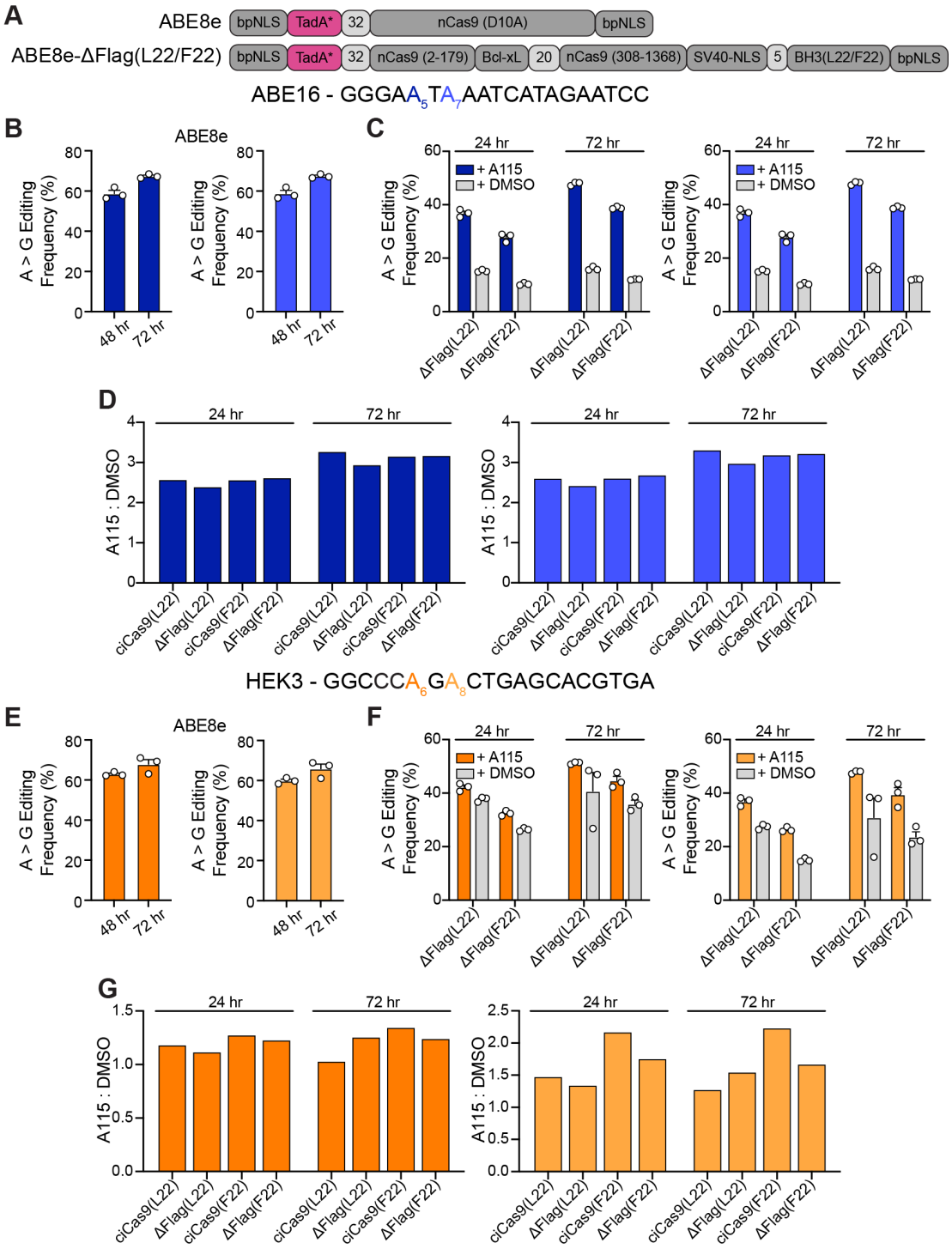

Figure C.8. Additional constructs of codon optimized chemically-controlled ABE8e editors. **A)** Schematic of domains in the unmodified ABE8e base editor and additional constructs of the codon optimized ABE8e-ciCas9 base editors tested. 3 different versions of ciCas9 were additionally tested: ciCas9(L22), No-Flag-ciCas9(L22) (ΔFlag(L22)), and No-Flag-ciCas9(F22) (ΔFlag(F22)).

B,E) A-to-G editing frequencies of the unmodified ABE8e base editor at the ABE16 **(B)** and HEK3 **(E)** target sites.

Figure C.8 cont.
C,F) A-to-G editing frequencies of the ABE8e-ciCas9 constructs at the ABE16 **(C)** and HEK3 **(F)** target sites.
In (**B-C, E-F**) A-to-G editing is shown at the 2 nucleotides at each target site with highest editing frequency with ABE8e. The 2 different nucleotides are indicated by color in the target sequence. Editing by all ABE8e-ciCas9 constructs are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.
D,G) Ratio of the mean A-to-G editing frequency with 1 μM A115 to the mean A-to-G editing frequency with DMSO (A115:DMSO) for all tested ABE8e-ciCas9 base editors in **Fig. 3.3** and **Appendix C Fig. C.8** at the ABE16 **(D)** and HEK3 **(G)** target sites. Bars show the ratios of editing at the 2 nucleotides at each target site with highest editing frequency with ABE8e. Editing frequencies used to calculate the ratio were measured at 24 and 72 hr after A115 addition to HEK293T cells.

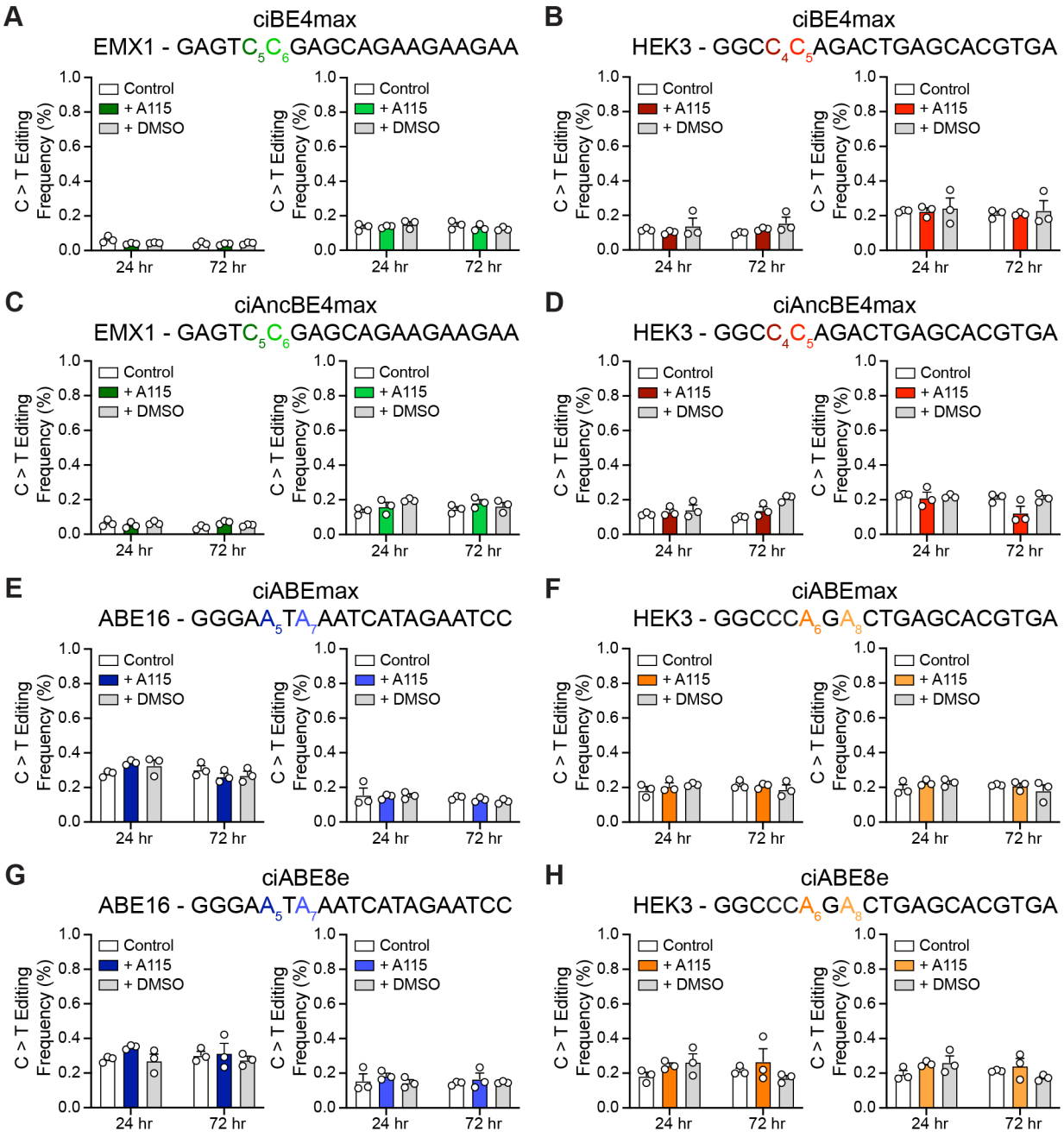

Figure C.9. No sgRNA control for chemically-controlled base editors.

A-B) ciBE4max base editing without sgRNA transfected at the EMX1 **(A)** and HEK3 **(B)** target sites.

C-D) ciAncBE4max base editing without sgRNA transfected at the EMX1 **(C)** and HEK3 **(D)** target sites.

E-F) ciABEmax base editing without sgRNA transfected at the ABE16 **(E)** and HEK3 **(F)** target sites.

G-H) ciABE8e base editing without sgRNA transfected at the ABE16 **(G)** and HEK3 **(H)** target sites.

Figure C.9 cont.
More transfection control plasmid, pMAX-GFP, was used to replace the sgRNA plasmid in the cotransfection with base editor. C-to-T editing **(A-D)** and A-to-G editing **(E-H)** is shown at the 2 nucleotides at each target site with highest editing frequency with the Cas9 version of the base editor. Editing by all chemically-inducible base editor constructs are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.

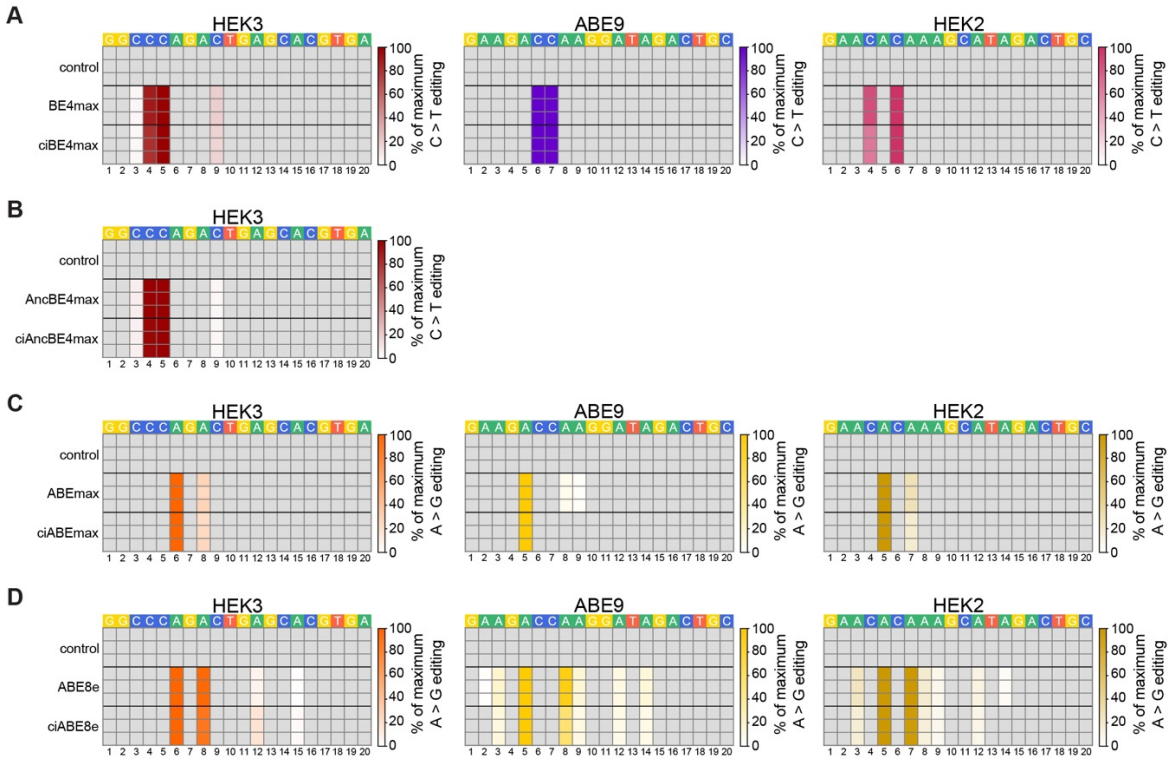

Figure C.10. Heatmaps of base editing by chemically-controlled base editors compared to unmodified base editors.
A-B) Heatmaps of BE4max, ciBE4max **(A)** and AncBE4max, ciAncBE4max **(B)** C-to-T base editing as a percentage of the highest edited nucleotide for each editor throughout the entire indicated target sites.
C-D) Heatmaps of ABEmax, ciABEmax **(C)** and ABE8e, ciABE8e **(D)** A-to-G base editing as a percentage of the highest edited nucleotide for each editor throughout the entire indicated target sites.
Each row shows an individual cell culture replicate. Editing frequencies of the unmodified base editors were quantified at 72 hr after transfection for the HEK3 target site and 48 hr after transfection for the ABE9 and HEK2 target sites. Chemically-controlled base editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK293T cells for the HEK3 target site and 24 hr after 1 μM A115 addition to HEK293T cells for the ABE9 and HEK2 target sites. The control shows untransfected cells harvested at the same time as the chemically-controlled base editors. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.

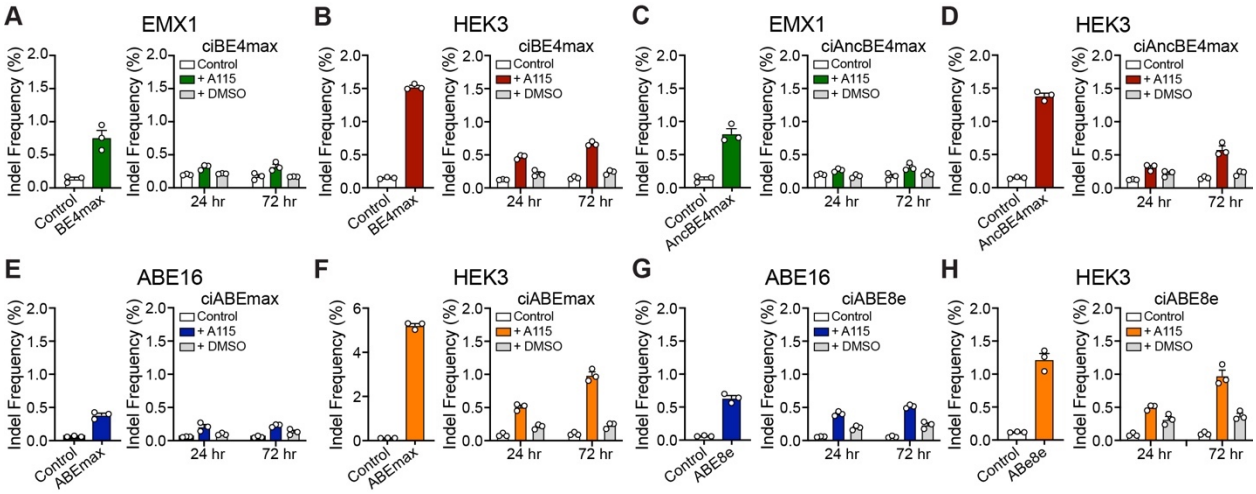

Figure C.11. Indel formation by chemically-controlled base editors.

A-B) BE4max (left) and ciBE4max (right) induced indel formation at the EMX1 **(A)** and HEK3 **(B)** target sites.

C-D) AncBE4max (left) and ciAncBE4max (right) induced indel formation at the EMX1 **(C)** and HEK3 **(D)** target sites.

E-F) ABEmax (left) and ciABEmax (right) induced indel formation at the ABE16 **(E)** and HEK3 **(F)** target sites.

G-H) ABE8e (left) and ciABE8e (right) induced indel formation at the ABE16 **(G)** and HEK3 **(H)** target sites.

Control samples were untransfected HEK293T cells harvested at the same time as transfected cells. Editing by all unmodified base editors are quantified at 72 hr after transfection. Editing by all chemically-controlled base editors are quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.

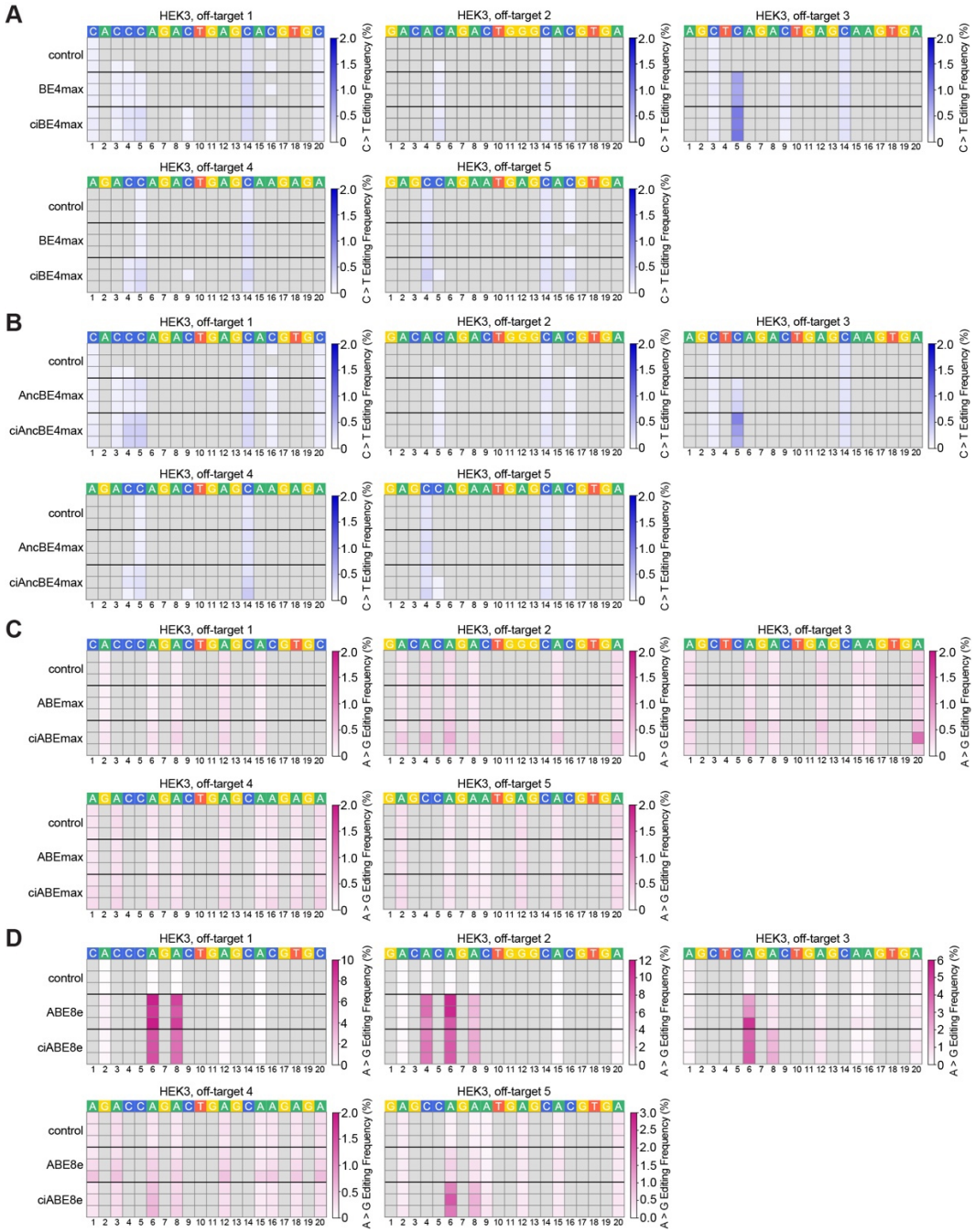

Figure C.12. Off-target base editing by chemically-controlled base editors.

Heatmaps of off-target base editing by ciBE4max **(A)**, ciAncBE4max **(B)**, ciABEmax **(C)**, and ciABE8e **(D)** with untransfected control and unmodified base editors. Each row shows an individual cell culture replicate. Unmodified base editor editing frequencies were quantified at 72 hr after transfection and chemically-controlled base editor editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK293T cells. Untransfected control cells were harvested at the same time as chemically-controlled base editing cells. C-to-T and A-to-G base editing frequencies have been filtered to only include C or A nucleotides in the target site where >0.1% of base conversion is observed. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.

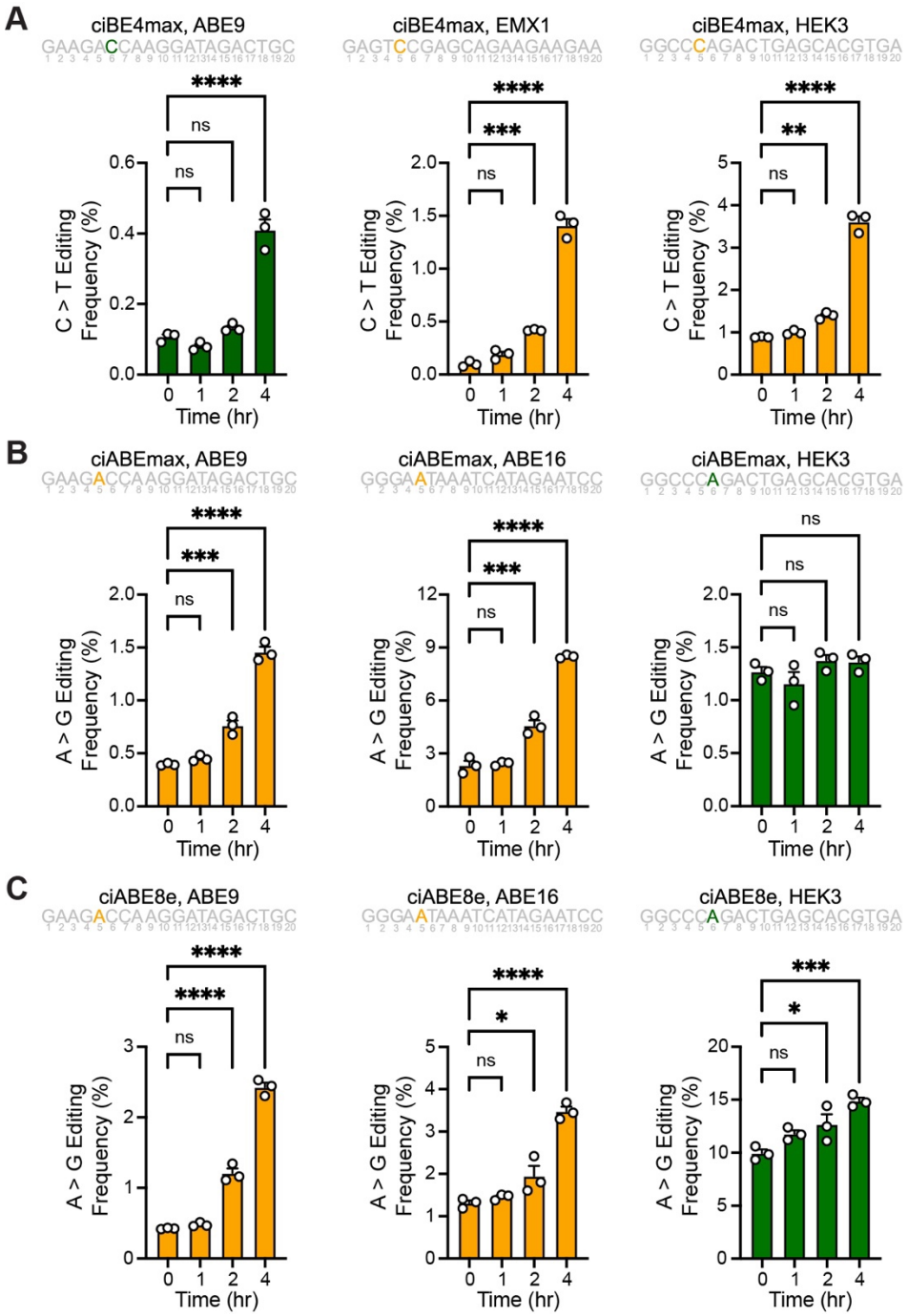

Figure C.13. Early time points in time courses of base editing with the chemically-controlled base editors.

Early time courses of chemically-controlled base editing using ciBE4max (**A**), ciABEmax (**B**), and ciABE8e (**C**) activated using 1 μM A115 at the indicated target sites. Time courses shown for the nucleotide colored in the target sequences shown. Numbers underneath the target sequence show the position of the nucleotide from the most PAM-distal nucleotide. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates. Significance of editing at different time points were compared to editing frequency at 0 hr using a One-way ANOVA, statistical values shown in **Appendix C Table C.1**.

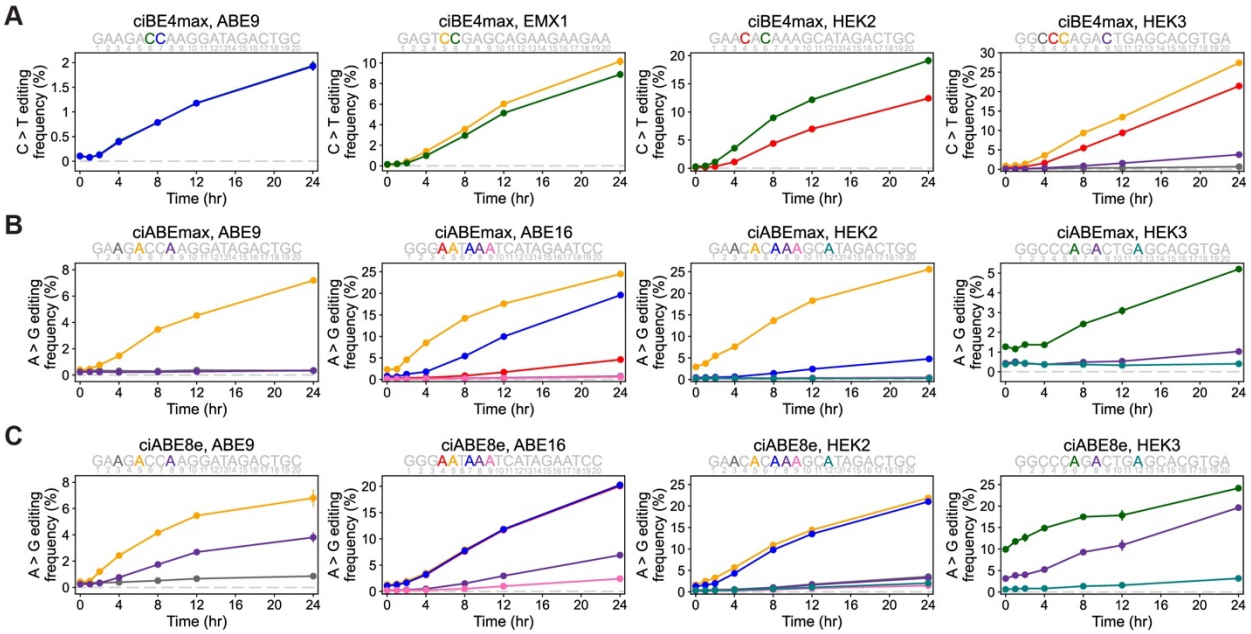

Figure C.14. Time courses of base editing with the chemically-controlled base editors.
A) Time course of chemically-controlled cytidine base editing by ciBE4max at the ABE9, EMX1, HEK2, and HEK3 target sites. ciBE4max was activated with 1 μM A115. Cells were harvested and editing was quantified at specified time points after activation. Colors of lines represent the corresponding nucleotide within the target site. Numbers underneath the target sequence show the position of the nucleotide from the most PAM-distal nucleotide.
B, C) Time course of chemically-controlled adenine base editing by ciABEmax **(B)** and ciABE8e **(C)** at the ABE9, ABE16, HEK2, and HEK3 target sites. ciABEmax and ciABE8e were activated with 1 μM A115. Cells were harvested and editing was quantified at specified time points after activation. Colors of lines represent the corresponding nucleotide within the target site. Numbers underneath the target sequence show the position of the nucleotide from the most PAM-distal nucleotide.
Data represented as mean editing ± SEM of 3 cell culture replicates. Time courses shown for all nucleotides where base editing frequency was greater than 0.5% at 24 hr after A115 addition.

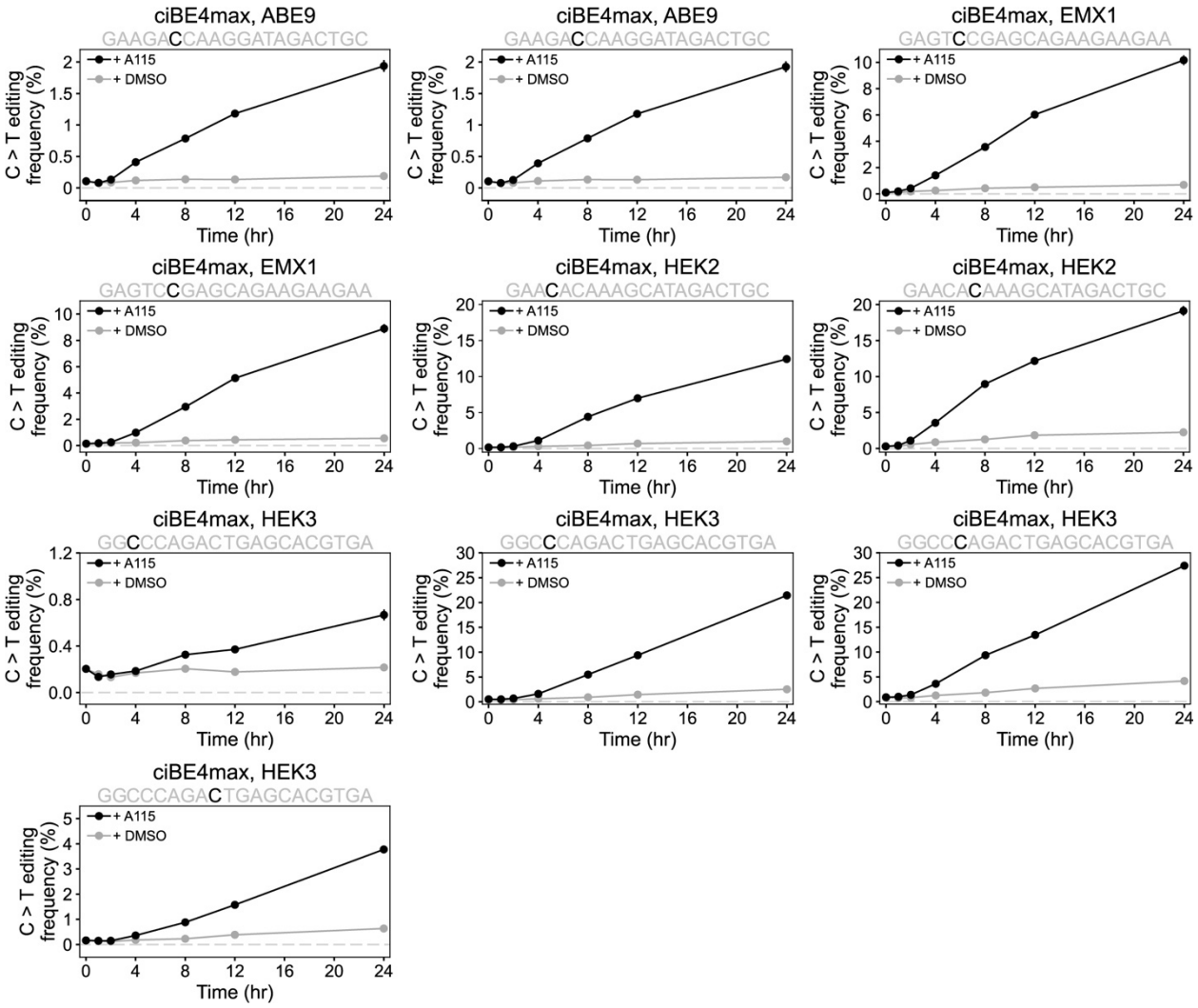

Figure C.15. Time courses of ciBE4max base editing at individual nucleotides with A115 or DMSO.

Time courses of ciBE4max C-to-T base editing. ciBE4max was activated with 1 μM A115 or DMSO. Cells were harvested and editing was quantified at the specified time points after activation. Black lines and circles show ciBE4max editing with 1 μM A115, gray lines and circles show ciBE4max editing with DMSO. Data represented as mean editing frequency ± SEM of 3 cell culture replicates.

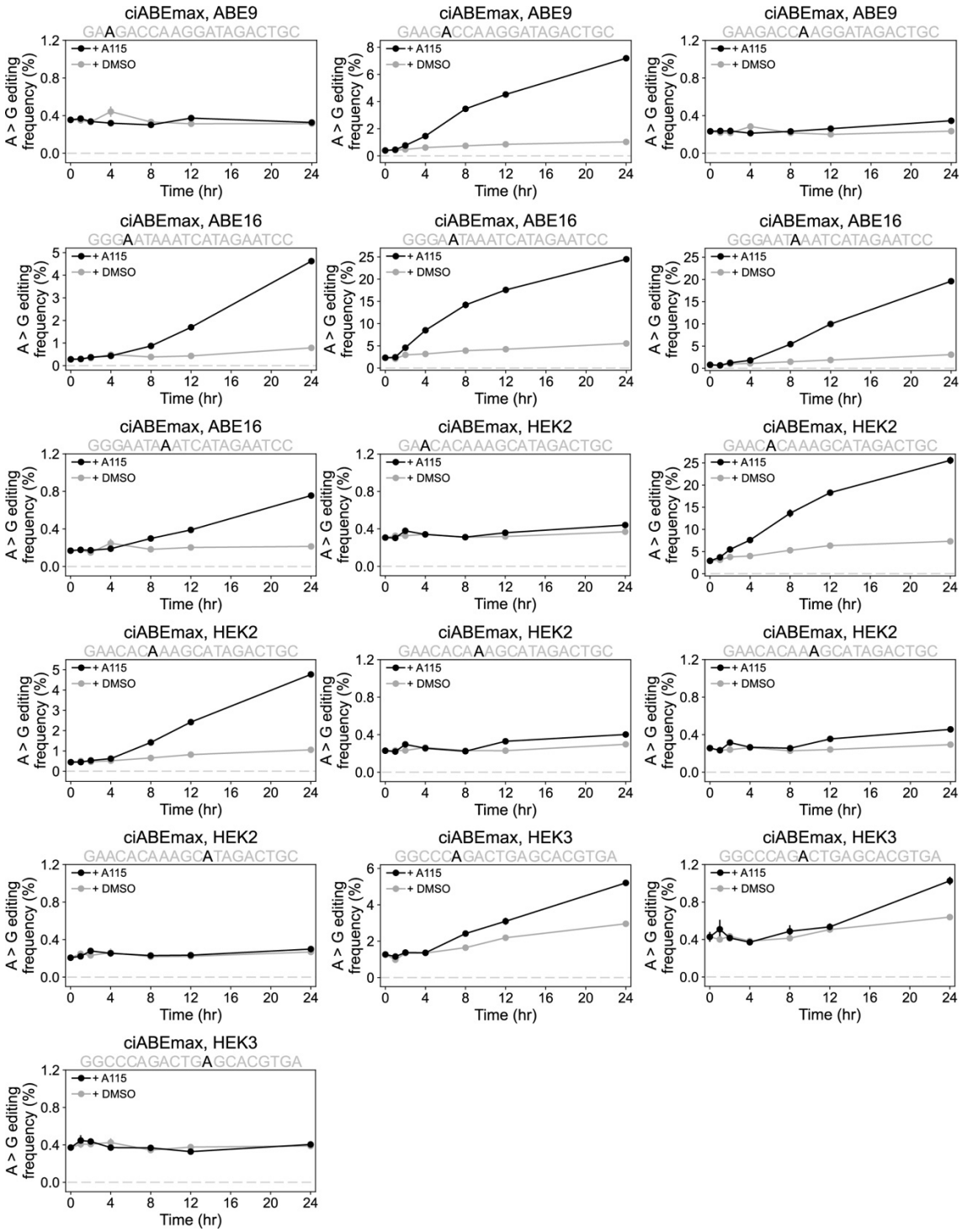

Figure C.16. Time courses of ciABEmax base editing at individual nucleotides with A115 or DMSO.

Time courses of ciABEmax C-to-T base editing. ciABEmax was activated with 1 μM A115 or DMSO. Cells were harvested and editing was quantified at the specified time points after activation. Black lines and circles show ciABEmax editing with 1 μM A115, gray lines and circles show ciABEmax editing with DMSO. Data represented as mean editing frequency ± SEM of 3 cell culture replicates.

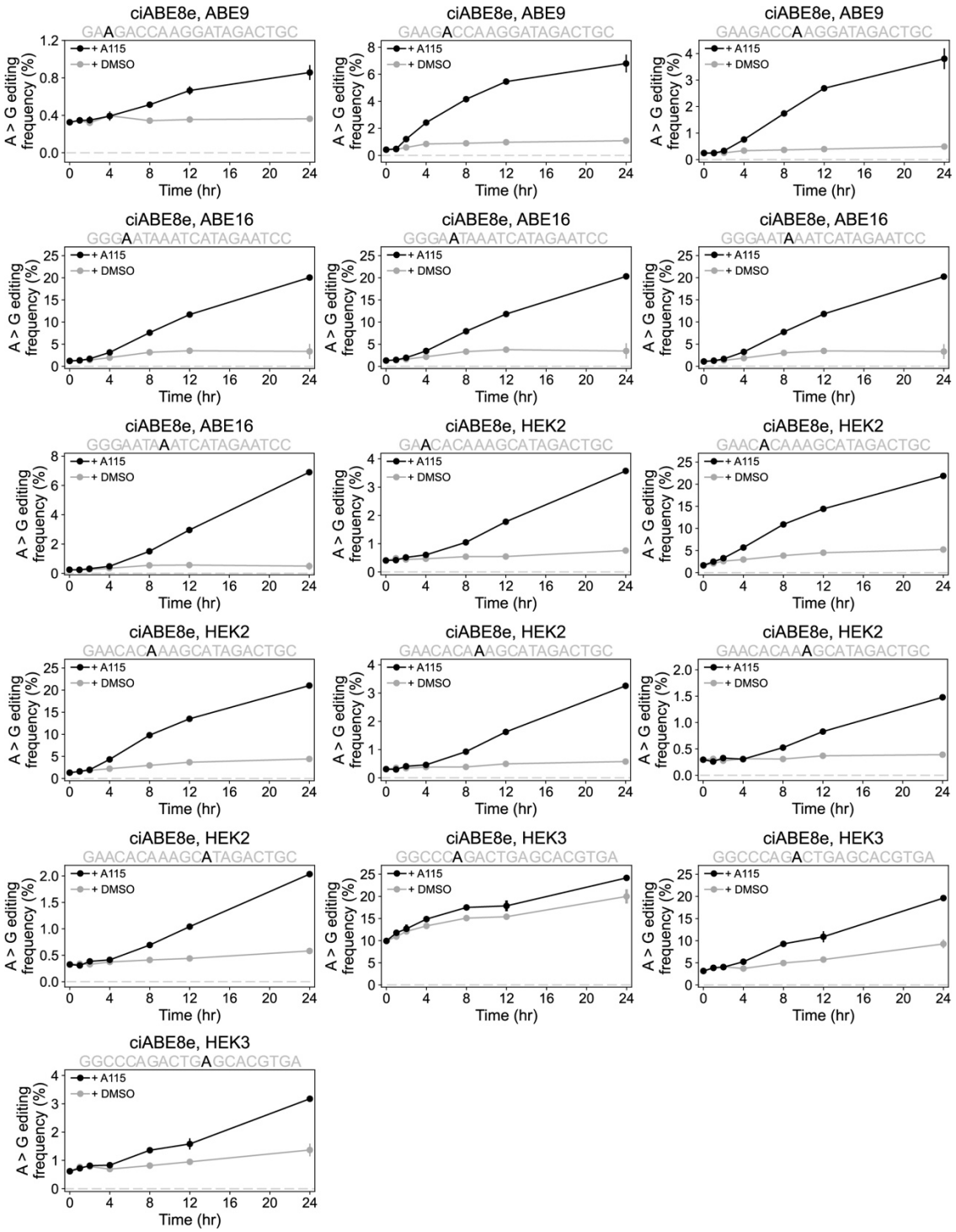

Figure C.17. Time courses of ciABE8e base editing at individual nucleotides with A115 or DMSO.

Time courses of ciABE8e C-to-T base editing. ciABE8e was activated with 1 μM A115 or DMSO. Cells were harvested and editing was quantified at the specified time points after activation. Black lines and circles show ciABE8e editing with 1 μM A115, gray lines and circles show ciABE8e editing with DMSO. Data represented as mean editing frequency ± SEM of 3 cell culture replicates.

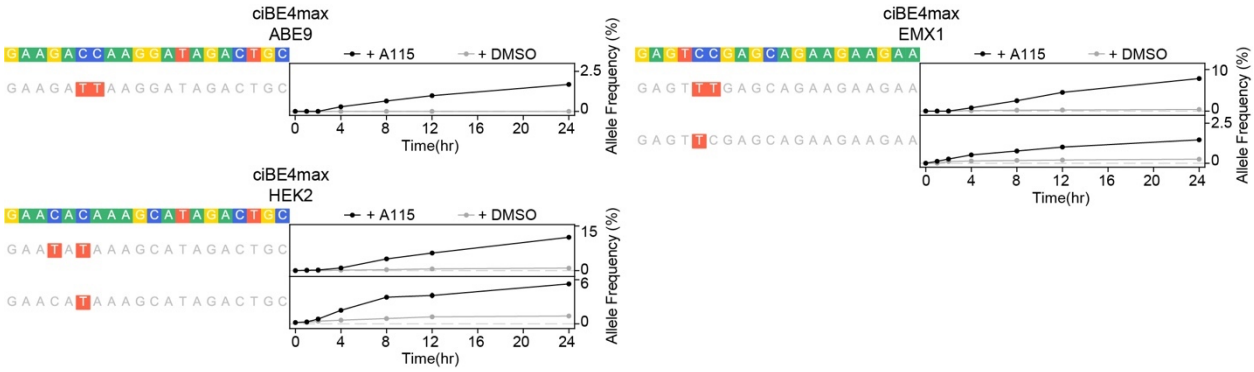

Figure C.18. Time courses of ciBE4max base editing allele outcomes.
Time course of allele formation by ciBE4max after activation with 1 μM A115 or DMSO. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.

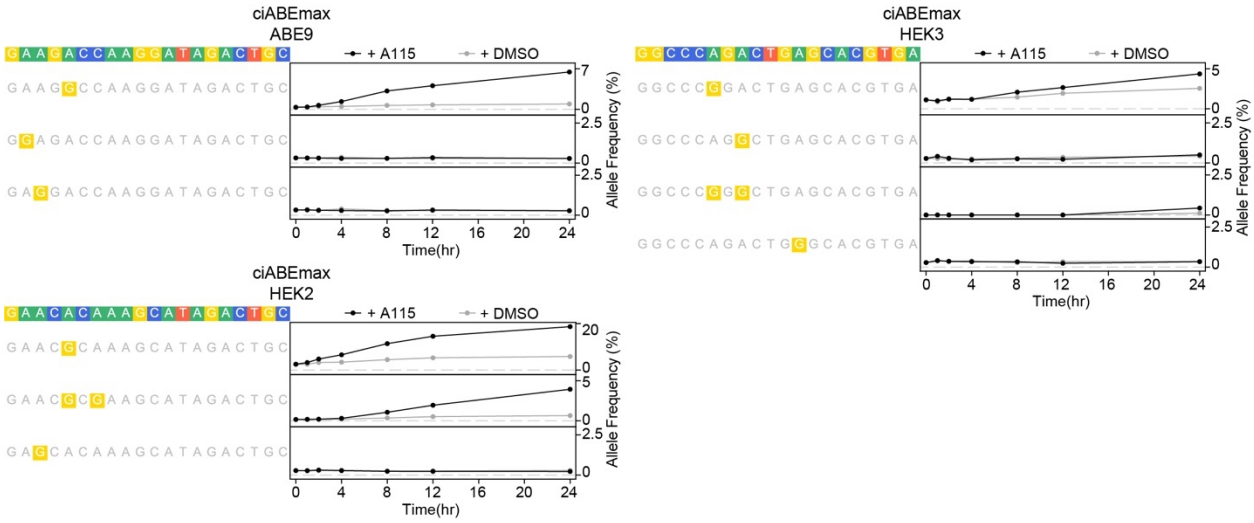

Figure C.19. Time courses of ciABEmax base editing allele outcomes.
Time course of allele formation by ciABEmax after activation with 1 μM A115 or DMSO. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.

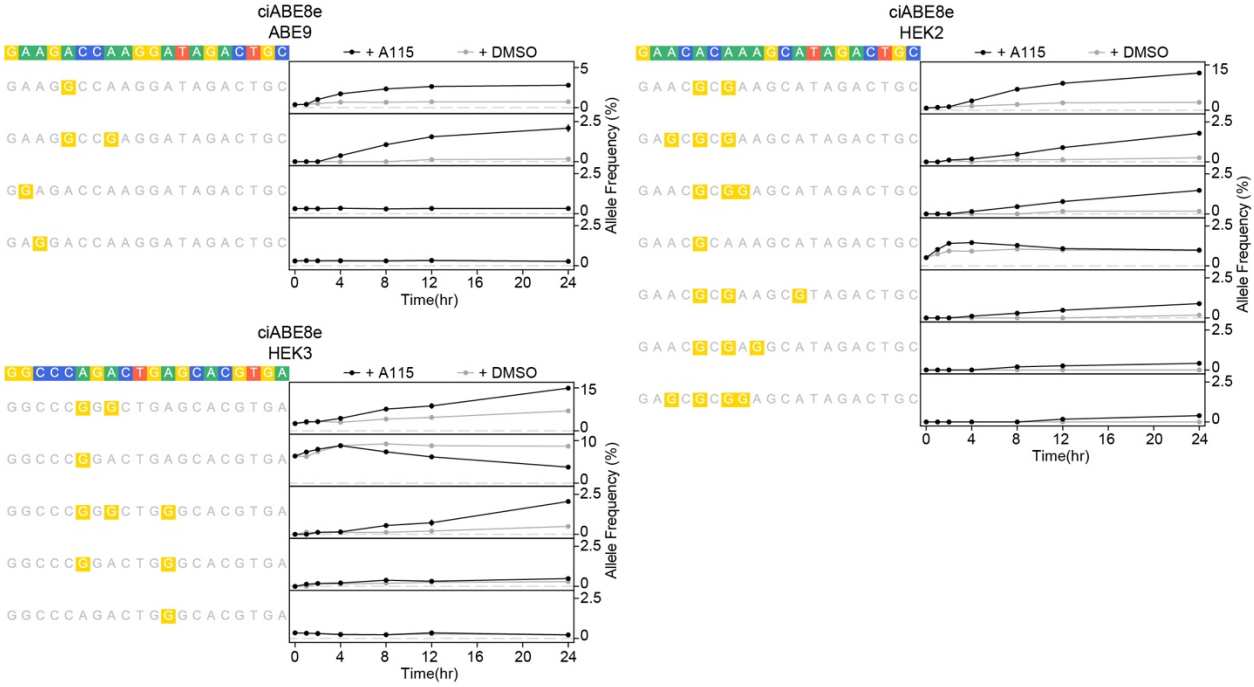

Figure C.20. Time courses of ciABE8e base editing allele outcomes.
Time course of allele formation by ciABE8e after activation with 1 μM A115 or DMSO. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.

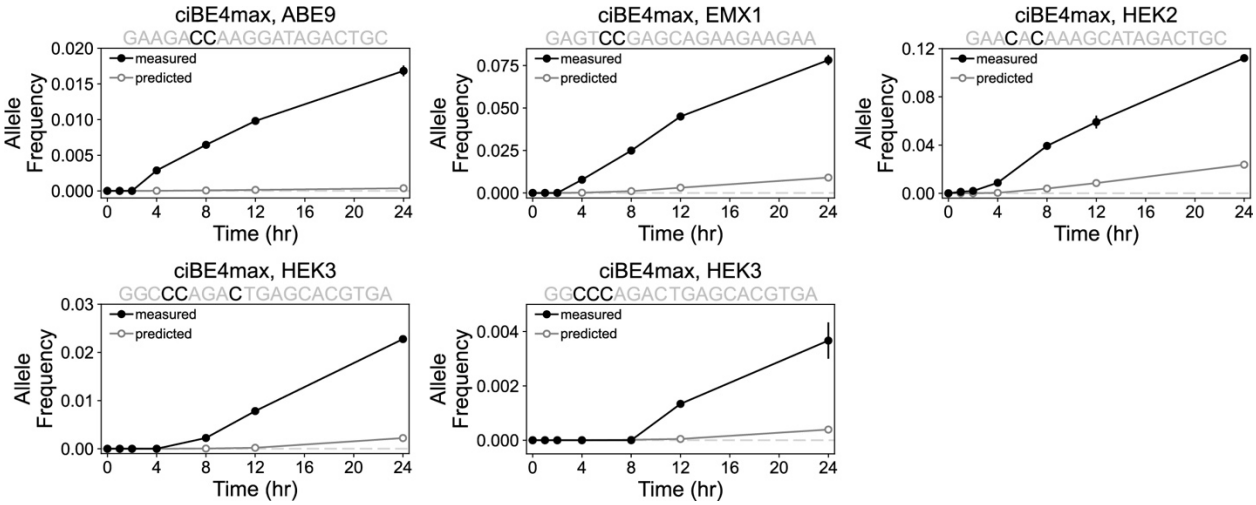

Figure C.21. Time course of measured and predicted allele frequencies by ciBE4max.
Measured and predicted allele frequencies over time created by ciBE4max that show a dependent model of base editing for multiply-edited alleles. Black lines and solid circles show measured allele frequencies, gray lines and open circles show predicted allele frequencies. Measured data represented as mean editing frequency ± SEM of 3 cell culture replicates. Predicted editing frequency represented as mean predicted editing frequency ± relative error. Calculations for predicted frequency and relative error described in the methods.

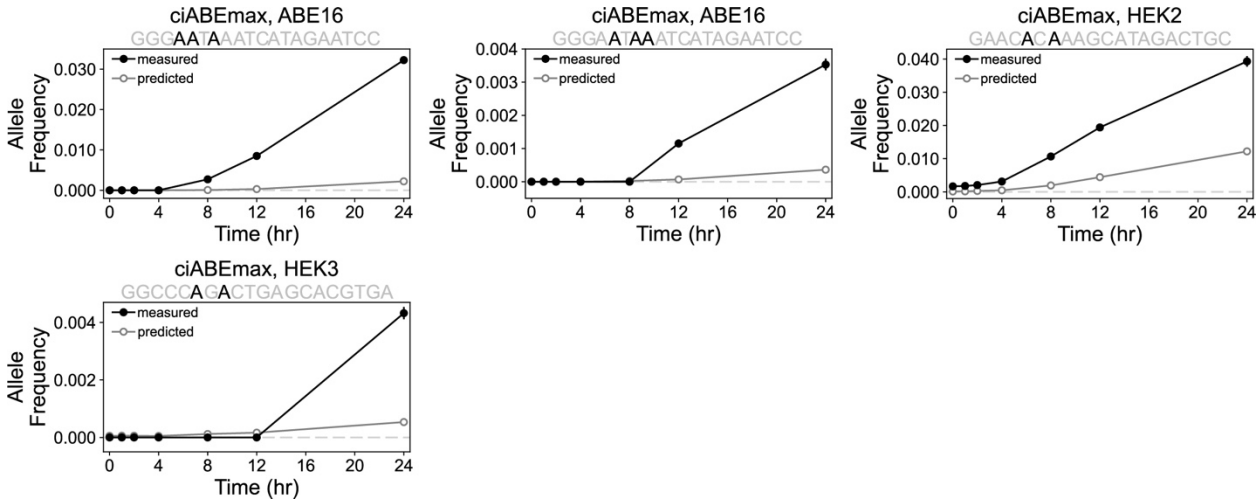

Figure C.22. Time course of measured and predicted allele frequencies by ciABEmax.
Measured and predicted allele frequencies over time created by ciABEmax that show a dependent model of base editing for multiply-edited alleles. Black lines and solid circles show measured allele frequencies, gray lines and open circles show predicted allele frequencies. Measured data represented as mean editing frequency ± SEM of 3 cell culture replicates. Predicted editing frequency represented as mean predicted editing frequency ± relative error. Calculations for predicted frequency and relative error described in the methods.

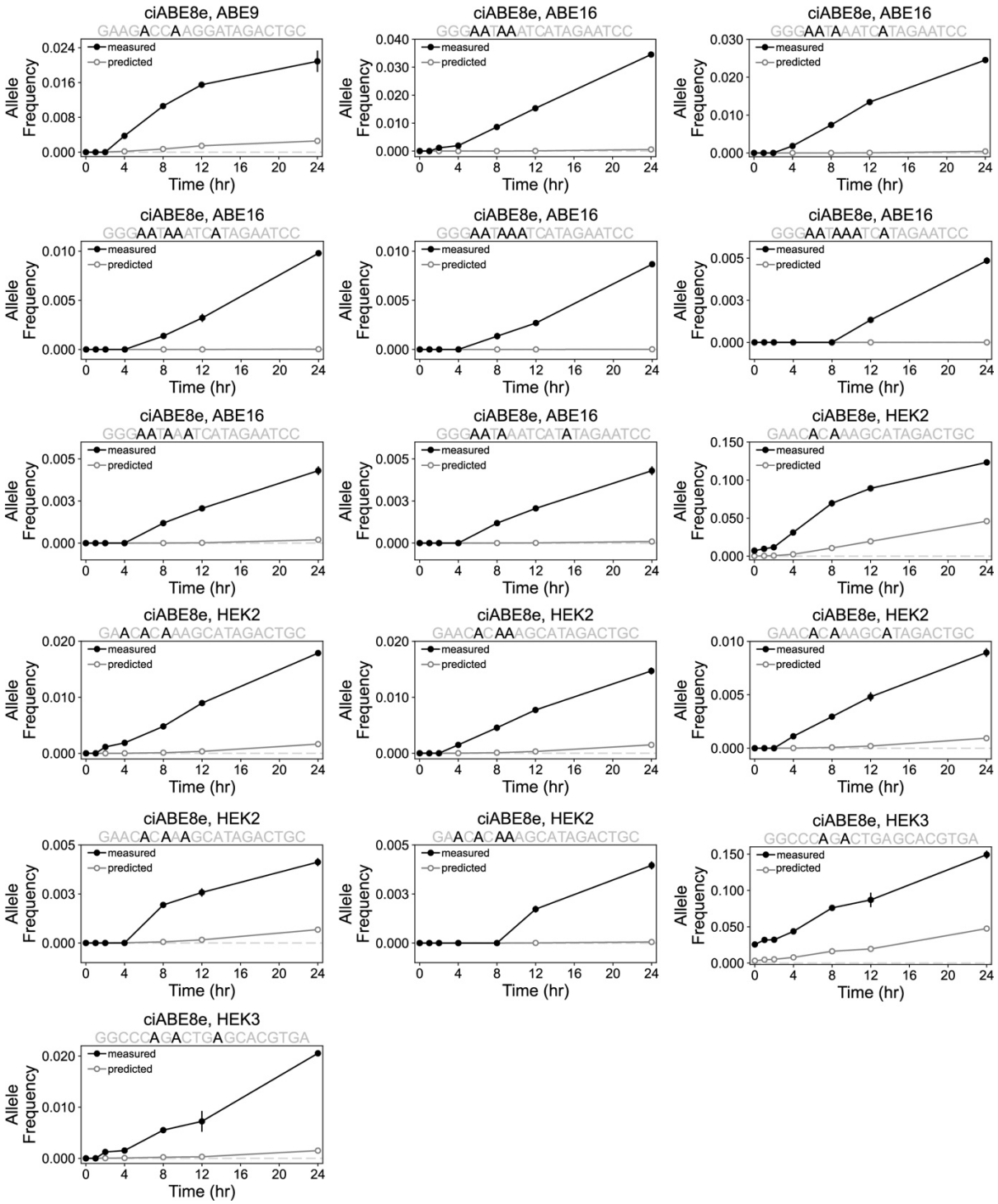

Figure C.23. Time course of measured and predicted allele frequencies by ciABE8e.
Measured and predicted allele frequencies over time created by ciABE8e that show a dependent model of base editing for multiply-edited alleles. Black lines and solid circles show measured allele frequencies, gray lines and open circles show predicted allele frequencies. Measured data represented as mean editing frequency ± SEM of 3 cell culture replicates. Predicted editing frequency represented as mean predicted editing frequency ± relative error. Calculations for predicted frequency and relative error described in the methods.

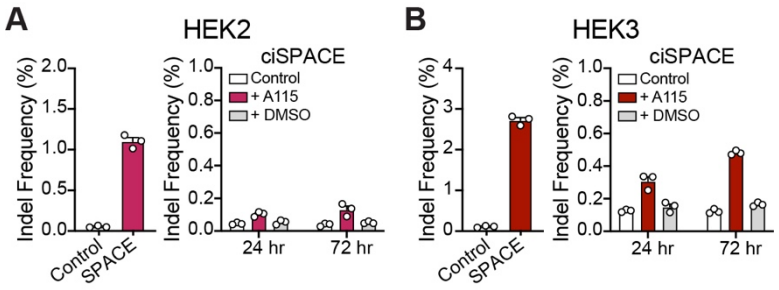

Figure C.24. Indel formation by ciSPACE.

SPACE (left) and ciSPACE (right) induced indel formation at the HEK2 **(A)** and HEK3 **(B)** target sites. Control samples were untransfected HEK-293T cells harvested at the same time as transfected cells. Editing by SPACE is quantified at 72 hr after transfection. Editing by ciSPACE is quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK-293T cells. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.

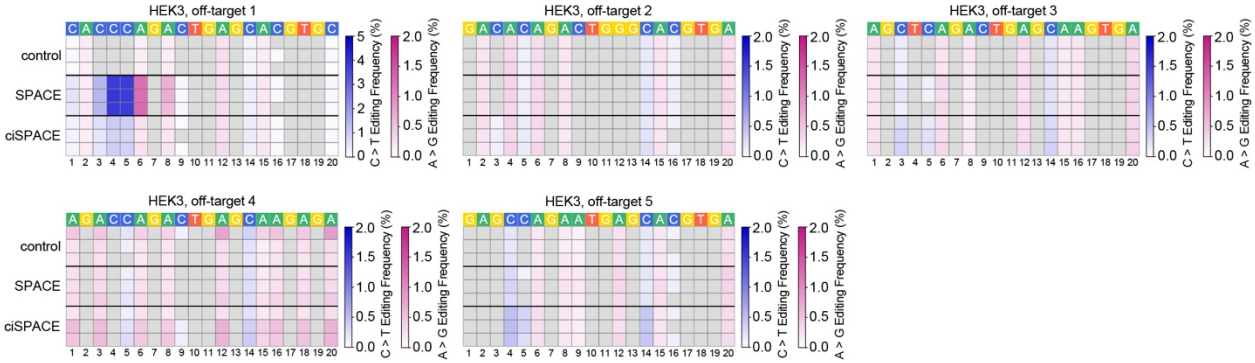

Figure C.25. Off-target base editing by ciSPACE.

Heatmaps of off-target base editing by ciSPACE with untransfected control and SPACE. Each row shows an individual cell culture replicate. SPACE editing frequencies were quantified at 72 hr after transfection and ciSPACE editing frequencies were quantified at 72 hr after 1 μM A115 addition to HEK-293T cells. Untransfected control cells were harvested at the same time as ciSPACE. C-to-T and A-to-G base editing frequencies have been filtered to only include C or A nucleotides in the target site where >0.1% of base conversion is observed. The numbers below the heatmaps show the position of the nucleotide from the most PAM-distal nucleotide.

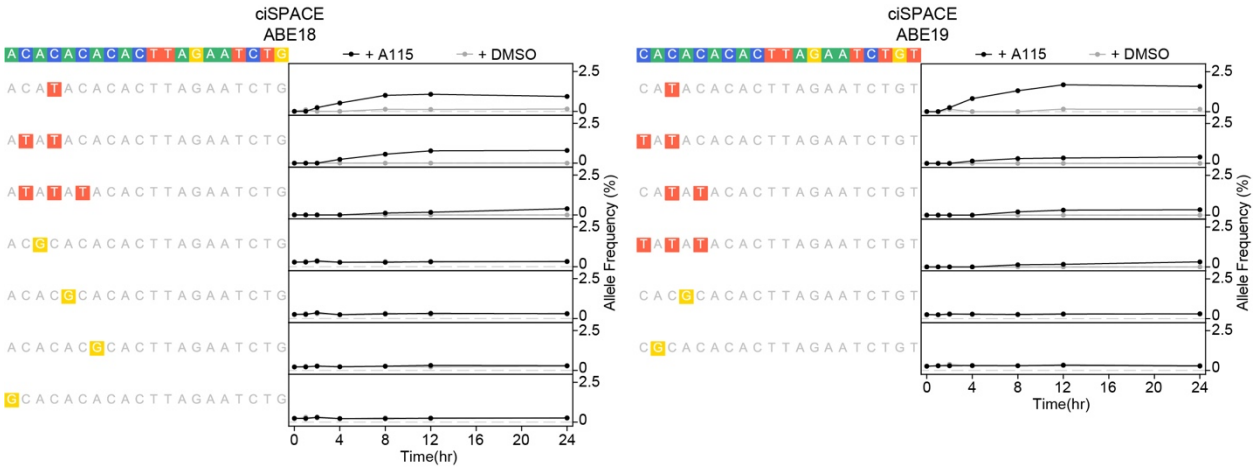

Figure C.26. Allele frequency time courses by ciSPACE.
Time course of allele formation by ciSPACE after activation with 1 μM A115 or DMSO. Black lines and circles show editing with 1 μM A115, gray lines and circles show editing with DMSO. Data represented as mean allele frequency ± SEM of 3 cell culture replicates.

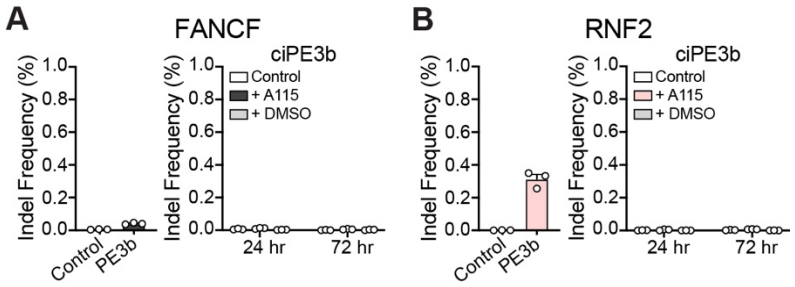

Figure C.27. Indel formation by ciPE3b.
PE3b (left) and ciPE3b (right) induced indel formation at the FANCF **(A)** and RNF2 **(B)** target sites corresponding to **Fig. 3.6H**. Control samples were untransfected HEK293T cells harvested at the same time as transfected cells. Editing by PE3b is quantified at 72 hr after transfection. Editing by ciPE3b is quantified at 24 and 72 hr after 1 μM A115 or DMSO addition to HEK293T cells. Bars show mean editing ± SEM of 3 cell culture replicates with white circles showing individual replicates.

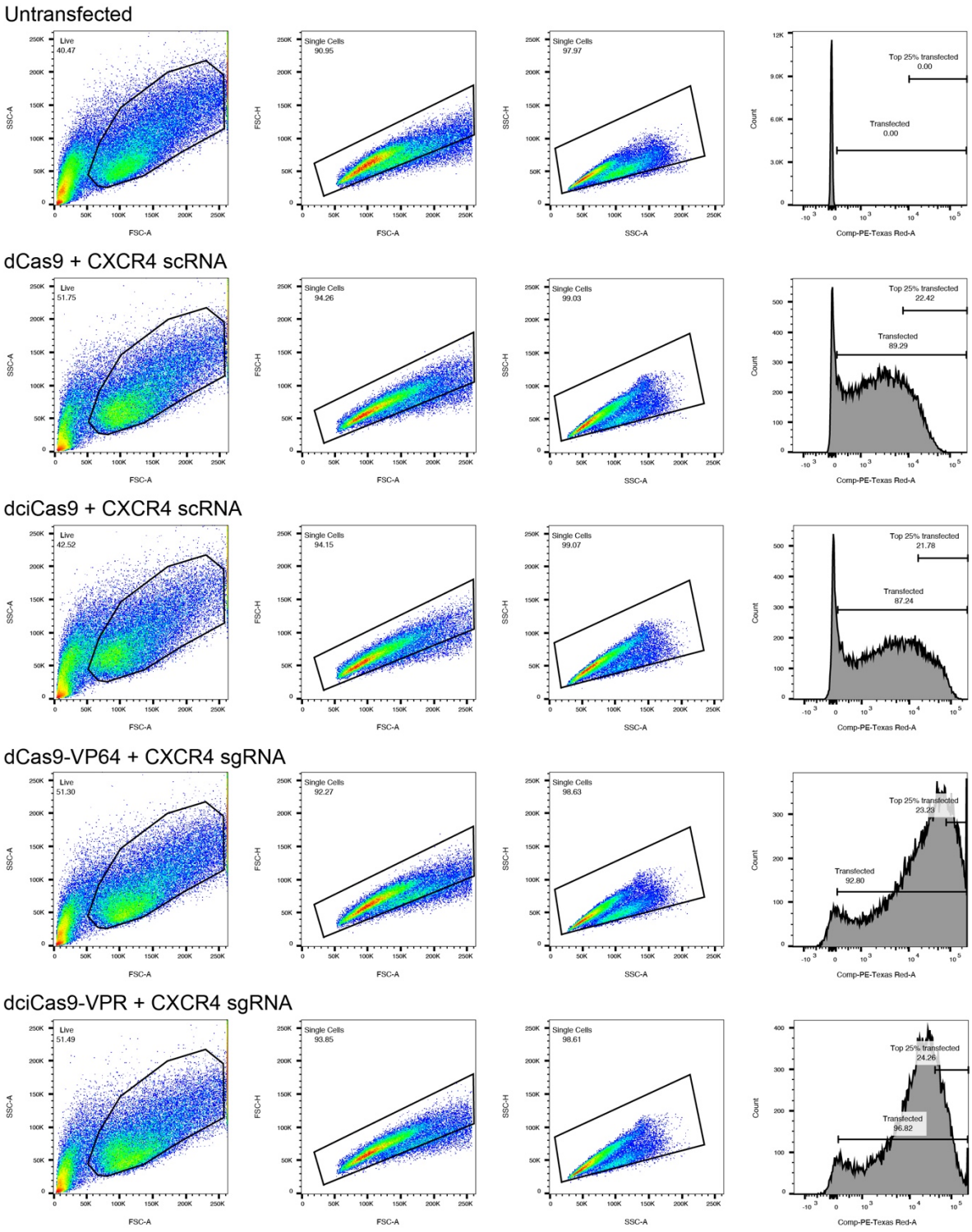

Figure C.28. Example gating of transcriptional activation at the CXCR4 locus.
Example gates set for determining CXCR4 transcriptional activation measured by median APC fluorescence of an anti-CXCR4 antibody, corresponding to **Fig. 3.1C** and **Appendix C Fig. C.1**. HEK293T cells were transfected with the indicated plasmids, outlined in the methods. Untransfected cells were stained with anti-CXCR4 antibody to determine background CXCR4 expression on HEK293T cells.

dciCas9-VPR + EMX1 sgRNA + 10 μM A115

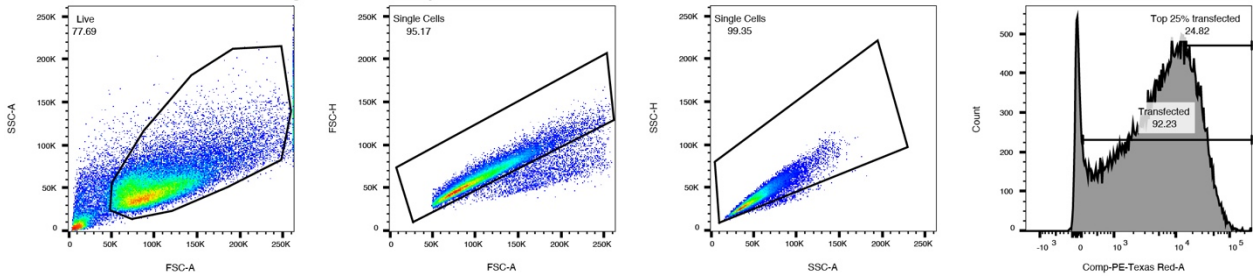

Figure C.29. Example gating of transcriptional activation at the EMX1-EGFP synthetic reporter locus.

Example gates set for determining EGFP transcriptional activation measured by EGFP fluorescence, corresponding to **Fig. 3.1D**. HEK293 TREx FlpIn EMX1-EGFP cells were transfected with the indicated plasmids, outlined in the methods.

Table C.1. One-way ANOVA results for comparison of early time points in chemically-controlled base editing.

Base editor	Target site	Figure	One-way ANOVA comparison to 0 hr time point					
			1 hr		2 hr		4 hr	
			Mean difference	P-value	Mean difference	P-value	Mean difference	P-value
ciBE4max	ABE9	Fig. C.13A	0.02648	0.5536	-0.02574	0.5730	-0.3029	<0.0001
	EMX1	Fig. C.13A	-0.09143	0.2287	-0.3205	0.0005	-1.309	<0.0001
	HEK2	Fig. 3.4A	-0.1230	0.4713	-0.8084	<0.0001	-3.274	<0.0001
	HEK3	Fig. C.13A	-0.1094	0.6052	-0.5002	0.0033	-2.723	<0.0001
ciABEmax	ABE9	Fig. C.13B	-0.05634	0.5825	-0.3672	0.0003	-1.061	<0.0001
	ABE16	Fig. C.13B	-0.1103	0.9623	-2.260	0.0002	-6.178	<0.0001
	HEK2	Fig. 3.4B	-0.7965	0.0476	-2.565	<0.0001	-4.681	<0.0001
	HEK3	Fig. C.13B	0.1135	0.5533	-0.1069	0.5943	-0.09362	0.6781
ciABE8e	ABE9	Fig. C.13C	-0.05563	0.7647	-0.7806	<0.0001	-2.001	<0.0001
	ABE16	Fig. C.13C	-0.1471	0.8027	-0.6336	0.0315	-2.167	<0.0001
	HEK2	Fig. 3.4C	-0.8219	0.0121	-1.609	0.0002	-4.018	<0.0001
	HEK3	Fig. C.13C	-1.836	0.1199	-2.740	0.0228	-4.936	0.0007

References

1. Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet.* **40**, 363–383 (2006).
2. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* **5**, a012740 (2013).
3. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
4. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell* **47**, 497–510 (2012).
5. Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
6. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
7. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
8. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
9. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
10. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
11. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **337**, 816–821 (2020).
12. Rees, H. A., Minella, A. C., Burnett, C. A., Komor, A. C. & Gaudelli, N. M. CRISPR-derived genome editing therapies: Progress from bench to bedside. *Mol. Ther.* **29**, 3125–3139

(2021).

13. Nakamura, M., Gao, Y., Dominguez, A. A. & Qi, L. S. CRISPR technologies for precise epigenome editing. *Nat. Cell Biol.* **23**, 11–22 (2021).
14. Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
15. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
16. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
17. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
18. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
19. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **22**, 359–371 (2018).
20. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
21. Jiang, F. & Doudna, J. Structure of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 863–867 (2016).
22. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).
23. Clarke, R. *et al.* Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. *Mol. Cell* **71**, 42–55.e8 (2018).
24. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).

25. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
26. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–837 (2013).
27. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2014).
28. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).
29. Boyle, E. A. *et al.* Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. *Sci Adv* **7**, (2021).
30. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
31. Liu, M.-S. *et al.* Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
32. Singh, D. *et al.* Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **25**, 347–354 (2018).
33. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
34. Sreekanth, V. *et al.* Chemogenetic System Demonstrates That Cas9 Longevity Impacts Genome Editing Outcomes. *ACS Cent Sci* **6**, 2228–2237 (2020).
35. Shin, J. *et al.* Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci. Adv.* **3**, (2017).
36. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
37. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).

38. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
39. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
40. Aprilyanto, V., Aditama, R., Tanjung, Z. A., Utomo, C. & Liwang, T. CROP: a CRISPR/Cas9 guide selection program based on mapping guide variants. *Sci. Rep.* **11**, 1504 (2021).
41. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
42. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
43. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
44. Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 12778 (2016).
45. 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.* **24**, 1–16 (2018).
46. Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
47. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 7 (2018).
48. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

49. Kim, H. K. *et al.* High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nature biomedical engineering* **4**, 111–124 (2020).

50. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

51. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).

52. Chatterjee, P. *et al.* A Cas9 with PAM recognition for adenine dinucleotides. *Nat. Commun.* **11**, 2474 (2020).

53. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

54. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* 38, 471–481 (2020).

55. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).

56. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

57. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).

58. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

59. Huang, T. P. *et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37**, 626–631 (2019).

60. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without

DNA cleavage. *Nature Publishing Group* **551**, 464–471 (2017).

61. Wood, R. D. DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167 (1996).
62. Wang, L. *et al.* Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* **27**, 1289–1292 (2017).
63. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **533**, 420–410 (2018).
64. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, (2016).
65. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).
66. Newald, J. G. X. *et al.* A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* **38**, 861–864 (2020).
67. Sakata, R. C. *et al.* Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* **38**, 865–869 (2020).
68. Zhang, X. *et al.* Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* **38**, 856–860 (2020).
69. Zhao, D. *et al.* New base editors change C to A in bacteria and C to G in mammalian cells. *Nat. Biotechnol.* **39**, 35–40 (2021).
70. Chen, L. *et al.* Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat. Commun.* **12**, 1384 (2021).
71. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).
72. Newby, G. A. *et al.* Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295–302 (2021).
73. Villiger, L. *et al.* Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).

74. Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).

75. Liang, P. *et al.* Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).

76. Kim, D., Kim, D.-E., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).

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

78. Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **36**, 292–295 (2019).

79. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **148**, 289–292 (2019).

80. Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 2052 (2020).

81. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).

82. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, (2019).

83. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).

84. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* **73**, 589–615 (2004).

85. Keijzers, G., Bohr, V. A. & Rasmussen, L. J. Human exonuclease 1 (EXO1) activity characterization and its function on flap structures. *Biosci. Rep.* **35**, (2015).

86. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*

12, 326–328 (2015).

87. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
88. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
89. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
90. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
91. Black, J. B. & Gersbach, C. A. Synthetic transcription factors for cell fate reprogramming. *Curr. Opin. Genet. Dev.* **52**, 13–21 (2018).
92. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
93. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
94. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13**, 127–137 (2016).
95. Kim, J.-M. *et al.* Cooperation between SMYD3 and PC4 drives a distinct transcriptional program in cancer cells. *Nucleic Acids Res.* **43**, 8868–8883 (2015).
96. Kwon, D. Y., Zhao, Y.-T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* **8**, 15315 (2017).
97. Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**, 46545–46556 (2016).
98. Kuscu, C. *et al.* Temporal and Spatial Epigenome Editing Allows Precise Gene Regulation

in Mammalian Cells. *J. Mol. Biol.* **431**, 111–121 (2019).

99. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
100. Rose, J. C. *et al.* Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat. Methods* **14**, 891–896 (2017).
101. Brinkman, E. K. *et al.* Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol. Cell* **70**, 801–813.e6 (2018).
102. Liu, Y. *et al.* Very fast CRISPR on demand. *Science* **368**, 1265–1269 (2020).
103. Park, J. *et al.* Recording of elapsed time and temporal information about biological events using Cas9. *Cell* **184**, 1047–1063.e23 (2021).
104. Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat. Methods* **13**, 1043–1049 (2016).
105. Foight, G. W. *et al.* Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nat. Biotechnol.* **37**, 1209–1216 (2019).
106. Berríos, K. N. *et al.* Controllable genome editing with split-engineered base editors. *Nat. Chem. Biol.* **17**, 1262–1270 (2021).
107. Gangopadhyay, S. A. *et al.* Precision Control of CRISPR-Cas9 Using Small Molecules and Light. *Biochemistry* **58**, 234–244 (2019).
108. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).
109. Lu, J. *et al.* Multimode drug inducible CRISPR/Cas9 devices for transcriptional activation and genome editing. *Nucleic Acids Res.* **46**, e25 (2017).
110. González, F. *et al.* An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Stem Cells* **15**, 215–226 (2014).
111. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**,

390–394 (2015).

112. Maji, B. *et al.* Multidimensional chemical control of CRISPR-Cas9. *Nat. Chem. Biol.* **13**, 9–11 (2017).

113. Senturk, S. *et al.* Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat. Commun.* **8**, 14370 (2017).

114. Kleinjan, D. A., Wardrope, C., Nga Sou, S. & Rosser, S. J. Drug-tunable multidimensional synthetic gene control using inducible degron-tagged dCas9 effectors. *Nat. Commun.* **8**, 1191 (2017).

115. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).

116. Liu, K. I. *et al.* A chemical-inducible CRISPR–Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12**, 980–987 (2016).

117. Nguyen, D. P. *et al.* Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat. Commun.* **7**, 12009 (2016).

118. Zhao, C. *et al.* HIT-Cas9: A CRISPR/Cas9 Genome-Editing Device under Tight and Effective Drug Control. *Molecular Therapy - Nucleic Acids* **13**, 208–219 (2018).

119. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).

120. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646–651 (2016).

121. Rose, J. C., Stephany, J. J., Wei, C. T., Fowler, D. M. & Maly, D. J. Rheostatic Control of Cas9-Mediated DNA Double Strand Break (DSB) Generation and Genome Editing. *ACS Chem. Biol.* **13**, 438–442 (2018).

122. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).

123. Bubeck, F. *et al.* Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9. *Nat. Methods* **15**, 924–927 (2018).

124. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).

125. Kundert, K. *et al.* Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat. Commun.* **10**, 2127 (2019).

126. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

127. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1–11 (2019).

128. Boyle, E. A. *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5461–5466 (2017).

129. Lee, C. M., Cradick, T. J., Fine, E. J. & Bao, G. Nuclease Target Site Selection for Maximizing On-target Activity and Minimizing Off-target Effects in Genome Editing. *Mol. Ther.* **24**, 475–487 (2016).

130. Verkuijl, S. A. N. & Rots, M. G. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Curr. Opin. Biotechnol.* **55**, 68–73 (2018).

131. Wei, C. T., Maly, D. J. & Fowler, D. M. Temporal and rheostatic control of genome editing with a chemically-inducible Cas9. *Methods Enzymol.* **633**, 119–141 (2020).

132. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).

133. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **11**, 1782–1712 (2019).

134. Rose, J. C. *et al.* Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. *Nat. Commun.* **11**, 2697 (2020).

135. Ede, C., Chen, X., Lin, M.-Y. & Chen, Y. Y. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth. Biol.* **5**, 395–404 (2016).

136. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).

137. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339–350 (2015).

138. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).

139. Lapinaite, A. *et al.* DNA capture by a CRISPR-Cas9-guided adenine base editor. *Science* **369**, 566–571 (2020).

140. Jang, H.-K. *et al.* High-purity production and precise editing of DNA base editing ribonucleoproteins. *Sci Adv* **7**, (2021).

141. Tao, Z.-F. *et al.* Discovery of a Potent and Selective BCL-XL Inhibitor with in Vivo Activity. *ACS Med. Chem. Lett.* **5**, 1088–1093 (2014).

142. Wilson, W. H. *et al.* Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid malignancies: a phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics, and antitumour activity. *Lancet Oncol.* **11**, 1149–1159 (2010).

143. Hanna, R. E. *et al.* Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021).

144. Boersma, M. D., Sadowsky, J. D., Tomita, Y. A. & Gellman, S. H. Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein-protein recognition: application to the Bim BH3 domain. *Protein Sci.* **17**, 1232–1240 (2008).

145. Wang, Q. *et al.* A general theoretical framework to design base editors with reduced

bystander effects. *Nat. Commun.* **12**, 6529 (2021).

146. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

147. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016).

148. Zhang, F. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).

149. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci.* **54**, 9351–9368 (2018).

150. Marino, N. D., Pinilla-Redondo, R., x00151, B. X. L. C. X. R. & Bondy-Denomy, J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nat. Methods* **17**, 471–479 (2020).

151. Li, Y. & Bondy-Denomy, J. Anti-CRISPRs go viral: the infection biology of CRISPR-Cas inhibitors. *Cell Host Microbe* **29**, 704-714 (2020)

152. Mathony, J. *et al.* Computational design of anti-CRISPR proteins with improved inhibition potency. *Nat. Chem. Biol.* **16**, 725–730 (2020).

153. Dong, D. *et al.* Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**, 436–439 (2017).

154. Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).

155. Chen, M. & Qi, L. S. Repurposing CRISPR System for Transcriptional Activation. *Adv. Exp. Med. Biol.* **983**, 147–157 (2017).

156. Serrano, L., Vazquez, B. N. & Tischfield, J. Chromatin structure, pluripotency and differentiation. *Exp. Biol. Med.* **238**, 259–270 (2013).

157. Baron, C. S. & van Oudenaarden, A. Unravelling cellular relationships during development

and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* **20**, 753–765 (2019).

158. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).

159. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).

160. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).