Published online 19 May 2016

Nucleic Acids Research, 2016, Vol. 44, No. 11 ***5365–5377***
doi: 10.1093/nar/gkw417

Distinct patterns of Cas9 mismatch tolerance *in vitro* and *in vivo*

Becky X.H. Fu[1,*], Robert P.St. Onge[2,3], Andrew Z. Fire[1,4,*] and Justin D. Smith[1,2,*]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA, [2]Stanford Genome Technology Center, Palo Alto, CA 94304, USA, [3]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA and [4]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA

Received March 31, 2015; Revised April 27, 2016; Accepted April 30, 2016

ABSTRACT

Cas9, a CRISPR-associated RNA-guided nuclease, has been rapidly adopted as a tool for biochemical and genetic manipulation of DNA. Although complexes between Cas9 and guide RNAs (gRNAs) offer remarkable specificity and versatility for genome manipulation, mis-targeted events occur. To extend the understanding of gRNA::target homology requirements, we compared mutational tolerance for a set of Cas9::gRNA complexes *in vitro* and *in vivo* (in *Saccharomyces cerevisiae*). A variety of gRNAs were tested with variant libraries based on four different targets (with varying GC content and sequence features). In each case, we challenged a mixture of matched and mismatched targets, evaluating cleavage activity on a wide variety of potential target sequences in parallel through high-throughput sequencing of the products retained after cleavage. These experiments evidenced notable and consistent differences between *in vitro* and *S. cerevisiae* (*in vivo*) Cas9 cleavage specificity profiles including (i) a greater tolerance for mismatches *in vitro* and (ii) a greater specificity increase *in vivo* with truncation of the gRNA homology regions.

INTRODUCTION

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats) immune system evolved naturally in archaea and bacteria and has been adapted as a tool for genome editing (*in vivo*) and biochemical manipulation (*in vitro*) in numerous applications. *Streptococcus pyogenes* Cas9 is the most intensively characterized nuclease used in CRISPR-Cas genome editing (1–3). Cas9 can be directed to specific genomic sites with a programmable RNA called the guide RNA (gRNA) (4), which includes 17–20 nucleotides of target sequence complementarity next to a protospacer adjacent motif (PAM) consisting of two G residues separated by a base (A,T, C or G) (5′-NGG-3′). Cas9 generates a blunt double-strand DNA break in the region of gRNA::target homology that can be repaired by non-homologous end-joining or homologous recombination (if a template for repair is provided) (2,5). Due to the broad range of research utilizing Cas9 as technology and its potential in clinical applications, much effort has been put into understanding the off-target cleavage activities (6–10) and requirements for gRNA::target complementarity for precise binding and cleavage (2,5). Functional studies have, in particular, identified mismatches in the PAM and 'seed' region (10-12 bp proximal to the PAM) as having strong effects on reducing Cas9's ability to bind and cleave its target (4,11–15).

Comparison of different studies reporting requirements for guide homology provides a number of paradoxes. For example, some studies have reported that the 'seed' region can tolerate mismatches (10,13) while some report little to no tolerance (16). These differences could be the result of varying target sequences, differences between conditions *in vitro* and *in vivo*, or other variations in the assays employed. We specifically sought to assay Cas9 target mismatch tolerance using the same gRNAs and assay system under *in vitro* (DNA biochemical manipulation) and *in vivo* editing conditions. In this work we assay mismatch tolerance in four different target sequences under a variety of *in vitro* and *in vivo* conditions.

MATERIALS AND METHODS

Plasmid and strain construction

Molecular cloning was done with Gibson Assembly as outlined by Gibson *et al*. (17). *Escherichia coli* DNA plasmid preps were performed with QIAprep Spin Minipreps (Qiagen). Preparation of competent *E. coli* DH5α and transformation used Zymo Mix & Go *E. coli* Transformation

*To whom correspondence should be addressed. Tel: +1 650 723 2885; Email: xuhua@stanford.edu
Correspondence may also be addressed to Andrew Z. Fire. Tel: +1 650 723 2885; Email: afire@stanford.edu
Correspondence may also be addressed to Justin D. Smith. Tel: +1 650 721 5597; Email: jdsmith4@stanford.edu

reagents and Zymo Broth (Zymo Research) or electrocompetent CloneCatcher DH5G cells purchased from Genlantis. Competent *Saccharomyces cerevisiae* (strain BY4741 and KU70 deletion strain from MATa collection (18)) were prepared either by standard lithium acetate transformation protocols or using Zymo Frozen-EZ Yeast Transformation II (19). Hifi Hotstart (Kapa Biosystems), Q5 High Fidelity polymerase (NEB) and Phusion Hot Start Flex (Thermo Scientific) were used for polymerase chain reactions (PCRs). Primers and gRNA oligonucleotides were ordered from Integrated DNA Technologies(IDT). DpnI treatment was used to remove template plasmids in PCRs that were followed by Gibson Assembly. Benchling.com and APE DNA editing software were used for plasmid design.

Cas9-gRNA plasmids were built in the yeast pRS416 Cen/ARS plasmid containing the Ura3 marker. First we cloned an engineered tetracycline inducible pRPR1 Pol III promoter (20,21), NotI site and gRNA sequence as well as the tetracycline repressor gene (TetR) under control of the GPM1 promoter and terminator into pRS416 at the PciI site adjacent to the ori using Gibson Assembly (17). This vector is referred to as pRS416gT. Next we digested pRS416gT at the multiple cloning site with KpnI and SacI. GalL-Cas9-Cyc1t was amplified from pRS415-GalL-Cas9-Cyc1t (Addgene 43804) (22) using M13 forward and reverse primers. This PCR product was Gibson assembled into cut pRS416gT and transformed into DH5α. The gRNA single stranded oligonucleotides were then cloned into the NotI cleaved site with Gibson Assembly.

For the initial experiments (Type I libraries), libraries *unc-22A* RVL-1 and RVL-2 (13) were amplified using Kapa HiFi Hotstart, followed by DpnI treatment to remove the template plasmid. This PCR product was then Gibson Assembled into SacII cut pRS416gT (with a specific guide sequence cloned into the gRNA locus as above) adjacent to the GPM1 terminator and transformed into DH5α. The plasmid design for the 20 nt complementarity *unc-22A* plasmid is available at https://benchling.com/s/O5VobNjd or as Supplementary Figure S1a. All oligonucleotide sequences used in this work are available in Supplementary Table S1.

For the Type II libraries, EGFP-1, EGFP-2, *rol-6* and new *unc-22A* libraries, oligos were synthesized by Custom Array, Inc. Plasmid designs can be found in Supplementary Figure S1b and c. These target libraries were designed to contain all possible single nucleotide changes, all possible unique single base deletions, and a set of adjacent double nucleotide changes tiling the entire target. Oligos were amplified from the Custom Array pool using 182-Fu-Library_fwd and 183-Fu-Library_rev, which also created overlaps for Gibson Assembly. This PCR product was cloned into pRS415 cut with XhoI and SacI by Gibson Assembly. Ethanol precipitation was used to purify and concentrate the Gibson product, which was then electroporated into CloneCatcher DH5G cells using a capacitance of 25μF, resistance of 400 ohms and voltage of 2.0 kV using an ice-cold 0.1 cm cuvette. The transformation was recovered for 90 min and then transferred to liquid LB + carbenicillin as well as a dilution on LB + carbenicillin agar plates to obtain colony counts. The resulting bacterial cells were spun down and plasmid DNA prepared (Qiagen). To create yeast strains for each of the gRNAs, we first transformed in pRS416gT plasmids containing Cas9 and each of the 18 different gRNAs into yeast strain BY4741. We then transformed the new pRS415 library into each of these strains and grew in liquid synthetic complete media lacking uracil and leucine to select for both plasmids.

Both Type I and II libraries are available upon request.

Cas9 *in vitro* cleavage specificity assay

In vitro Cas9 cleavage assays and gRNA transcription (17, 18, 20 and 21 nt of complementarity) were performed as described in (13). The *in vitro* assays were performed at 37°C, while the yeast *in vivo* incubation temperature is 30°C .To ensure temperature was not the cause of differences observed between *in vivo* and *in vitro* experiments, the *in vitro* assay with full-length gRNA for the Type I variant library was repeated at 30°C (Supplementary Figure S2). The temperature change slowed Cas9 cleavage activity *in vitro* but the general pattern of the cleavage profile did not vary. The cleavage assay was also performed with the addition of magnesium after incubation of Cas9::gRNA complex and DNA; similar results were observed (Supplementary Figure S3). In addition, the cleavage assay was performed with an independent source of Cas9 protein used previously (13). We observed comparable results with the independent source of Cas9 nuclease assayed with gRNAs of 18 and 20 nt complementarity (Supplementary Figure S4).

Cas9 *in vivo* cleavage specificity assays

For the initial Type I library experiments, starting cultures were grown overnight at 30°C in synthetic complete media –Ura. These cultures were then used to inoculate experimental cultures to OD 0.1 in YP Galactose media with 250 ng/ml anhydrotetracycline (ATc), the gRNA inducing agent. To control for possible effects of non-homologous end joining (NHEJ) events, we did our initial experiments in both wild-type BY4741 strain and a KU70 null, NHEJ-deficient strain. Strains expressing gRNAs with L + CAG + 17, 18, 20 or 21 nt (G + 20) nt of target complementarity were grown for 12 h in inducing conditions. No detectible difference in the cleavage pattern was observed (Supplementary Figure S5, (19)), and wild-type BY4741 was used for all further experiments. Approximately 3000–4500 unique target variants were assayed *in vivo*. Further experiments with libraries *unc-22A* RVL-1 and RVL-2 were done with 18 and 20 nt complementarity versions of the gRNAs in BY4741 (Supplementary Figure S6, (8)). For the Type II libraries (EGFP-1, EGFP-2, *rol-6* and new *unc-22A* libraries), experiments were inoculated as above with 17, 18 and 20 nt complementarity versions of the gRNAs, but starting cultures were grown overnight at 30°C in synthetic complete media –Ura –Leu +dextrose, and growth experiments were conducted in synthetic complete media –Ura –Leu +galactose with 250 ng/ml anhydrotetracycline (ATc). For EGFP-1 target site, there were three different orientations of guide RNAs tested (EGFP-1, EGFP-1-R, EGFP-1-OS).

Yeast culturing and sample collection was performed using a cell-screening platform that integrates temperature-controlled absorbance plate readers, plate coolers and a liquid handling robot. Briefly, 700 ul yeast cultures were grown

Table 1. Table of all gRNAs used in experiments and their corresponding targets and sequences

gRNA name	gRNA length (nt)	conditions:	gRNA sequence:	DNA target sequence	gRNA target
EGFP-1 L+CAG+20	20	in *vivo*	L-CAGGGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-1 L+CA+19	19	in *vivo*	L-GAGGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-1 L+CA+18	18	in *vivo*	L-CAGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-1-R L+CAG+20	20	in *vivo*	L-CAGCCGGCAAGCTGCCCGTGCCC	CCGGCAAGCTGCCCGTGCCC	EGFP-1
EGFP-1-R L+CAG+18	18	in *vivo*	L-CAGGGCAAGCTGCCCGTGCCC	CCGGCAAGCTGCCCGTGCCC	EGFP-1
EGFP-1-R L+CA+18	18*	in *vivo*	L-CAGGCAAGCTGCCCGTGCCC	CCGGCAAGCTGCCCGTGCCC	EGFP-1
EGFP-2 L+CAG+20	20	in *vivo*	L-CAGGATGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
EGFP-2 L+CAG+18	18	in *vivo*	L-CAGTGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
EGFP-2 L+CAG+17	17*	in *vivo*	L-CAGGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
rol-6 L+CAG+20	20	in *vivo*	L-CAGGTGAGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
rol-6 L+CAG+18	18	in *vivo*	L-CAGGAGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
rol-6 L+CA+18	18	in *vivo*	L-CAGAGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
EGFP-1-OS L+CAG+20	20	in *vivo*	L-CAGCCAGGGCACGGGCAGCTTGC	CCAGGGCACGGGCAGCTTGC	EGFP-1
EGFP-1-OS L+CAG+18	18	in *vivo*	L-CAGAGGGCACGGGCAGCTTGC	CCAGGGCACGGGCAGCTTGC	EGFP-1
EGFP-1-OS L+CAG+17	17*	in *vivo*	L-CAGGGCACGGGCAGCTTGC	CCAGGGCACGGGCAGCTTGC	EGFP-1
unc-22A L+CAG+20	20	in *vivo*	L-CAGGCACCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
unc-22A L+CAG+G+20	20	in *vivo*	L-CAGGGCACCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
unc-22A L+CAG+18	18	in *vivo*	L-CAGACCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
unc-22A L+CAG+17	17	in *vivo*	L-CAGCCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
EGFP-1 G+20	20	in *vitro*	G+GGGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-1 19	19	in *vitro*	G+GCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-1 18	18	in *vitro*	G+CACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGG	EGFP-1
EGFP-2 G+20	20	in *vitro*	G+GATGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
EGFP-2 G+18	18	in *vitro*	G+TGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
EGFP-2 G+17	17	in *vitro*	G+GCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGT	EGFP-2
rol-6 G+20	20	in *vitro*	G+GTGAGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
rol-6 G+18	18	in *vitro*	G+GAGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
rol-6 18	18	in *vitro*	G+AGACGTCAACAATATGG	GTGAGACGTCAACAATATGG	rol-6
unc-22A 21	21	in *vitro*	G+GCACCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
unc-22A G+18	18	in *vitro*	G+ACCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A
unc-22A G+17	17	in *vitro*	G+CCACCGTCGCCGGCATT	GCACCACCGTCGCCGGCATT	unc-22A

*Extra upstream base matches

'L' denotes the remainder of the uncleaved leader sequence. The three bases of the leader sequence next to the gRNA sequence are 'CAG'. To clarify potential homology of the unprocessed/partially processed gRNA and target, if any bases of the 'CAG' match the target sequence then the guide is named L (leader)+ 'CAG' mismatched base + gRNA name. Bases that match the target sequence but are separated from the core target complementarity are colored as blue in the gRNA name. If the flanking 'CAG' bases match the target then the only mismatched bases will be in the gRNA name. The bases that represent mismatches to the target are shown in pink in the diagram pairing the gRNA with it is target. All *invitro* transcribed gRNAs have an additional G at the 5′ end and are named 'G'+gRNA name. If this G is not indicated, the extra G extended the 5′ homology.

in 48-well plates at 30°C with orbital shaking in Infinite plate readers (Tecan). To maintain cultures in log phase over 10 doublings, 80 μl of the culture was removed when it reached an OD of 0.76, added to a well containing 620 μl of media, and then allowed to grow further. After two such dilutions, 600 μl of the culture was collected and saved to a 4°C cooling station (Torrey Pines) when it reached an OD of 0.76. This amounted to ∼10 culture doublings from the beginning of the experiment. Pipetting events were triggered automatically by Pegasus Software and performed by a Freedom EVO workstation (Tecan). After sample collection, yeast plasmids were purified using the Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research).

Formation of gRNA 5′ ends

We note that all synthetic gRNAs were produced with T7 polymerase and therefore have an additional G on the 5′ end (see Table 1). *In vivo* production of gRNAs utilized an engineered Tet inducible RPR1 Pol III promoter (TetO-RPR1). The RPR1 promoter transcribes an 84 nt 5′ leader that is cleaved *in vivo* in it's native context (RPR1 gene) to give the mature RPR1 transcript (23). We and others have used this promoter (20) and other Pol III promoters (SNR52) (22,24) that produce leader sequences as a standard tool to produce effective gRNAs *in vivo* in yeast; nonetheless, it is unclear if this 5′ leader is present, partially cleaved or has been removed at the time that the Cas9::gRNA complexes act *in vivo*. Supporting the potential for decreased activity in the presence of the full-length leader, *in vitro* assays using gRNA molecules synthesized with the full 84 nt RPR1-TetO leader showed variable degrees of decreased efficacy (data not shown).

Cas9 *in vitro* and *in vivo* sequence retention calculation

Cas9 *in vitro* and *in vivo* substrate cleavage efficiency was calculated as described in Fu *et al*. (13). A log retention score (effectively the inverse of cleavage efficiency) for each sequence in each experiment was calculated by quantifying the representation of each sequence before and after either the addition of Cas9 (*in vitro*) or induction of Cas9 (*in vivo*). A more negative retention score indicates more cleavage while a more positive retention score indicates less cleav-

age. Only sequences with $n \geq 50$ counts in the non-cleaved control were considered for all experiments.

A list of all Cas9 *in vitro* and *in vivo* experiments, experimental conditions and sequencing run IDs are reported in Supplementary Table S2.

RESULTS/DISCUSSION

To assess the effects of variants on Cas9-mediated cleavage *in vivo,* we adapted a previous high-throughput sequencing approach that had been used to characterize Cas9 specificity *in vitro* (13). The approach makes use of a defined pool of perfectly matched and imperfect target sequences (a variant library). Sequencing of the target pool before and after Cas9 action yields a detailed picture of retained (and removed) species, providing a profile of specificity as a function of position and mutational character. Two methods of library construction were used for this analysis, the first (Type I) involving chemically degenerate oligonucleotide synthesis (13) (Figure 1B), with the second (Type II) involving ordered synthesis of designed variants using a microarray-based synthesis technology (25) (Figure 1C).

We used two Type I libraries. These had previously been examined *in vitro* (13), with the canonical target for both libraries comprising a segment (*'unc-22A'*) from the *Caenorhabditis elegans unc-22* gene (Figure 1A) with adjacent PAM. These libraries derive from oligonucleotide pools synthesized with mixtures of bases designed to have a 10% variation rate in the gRNA homology region (13). Adding to the diversity of sequences, each potential target + PAM sequence is flanked upstream and downstream by 6 nt regions of random sequence (used as a barcode). Parallel experiments employed the two independently prepared *unc-22A* Type I libraries, each containing several thousand of the $>10^9$ possible barcoded variants. Individual single base variants within the target region are each represented by multiple distinct barcoded clones in each library, with consistency for a common guide + PAM sequence between libraries and between distinct clones supporting the robustness of the assay (13). An extensive analysis of the assay's noise and reproducibility can be found in (13).

To extend the study to a broader set of target molecules, a second type of variant library (Type II) was constructed using 2056 rationally-designed oligonucleotides for four different target sites: EGFP-1 (EGFP site 1), EGFP-2 (EGFP site 2), *rol-6* and *unc-22A*. In addition to the original sequences for each target, the ordered library included all single base variants, all single base deletions and a set of double adjacent variants (Figure 1A). The four canonical target sequences represented a range of GC contents (80, 50, 45 and 70% respectively). The EGFP-1 and EGFP-2 targets derived from the EGFP coding sequence characterized previously for their gRNA cleavage requirements in mammalian cells (26). The *rol-6* sequence is a highly efficient target used for co-conversion in *C. elegans* genome editing (27,28). The *unc-22A* target was the same site as previously characterized *in vitro* and used for the Type I library above (13). Both the EGFP-1 and *rol-6* gRNAs contain a 3′ GG motif within the region of target homology just upstream of the PAM (5′ GGNGG 3′) (28), which has been shown to increase cleavage efficiency (28). A list of sequences used for the Type II library can be found in Supplementary Table S3.

For the Type I libraries, assays were carried out with gRNAs of different target complementarity lengths for *unc-22A* (17, 18 and 20 nt). In addition (related to the operational inclusion of an additional 5′ G in vectors to produce gRNAs in T7 synthesis reactions), we examined effects from assays that used a lengthened version of the 20 nt guide (designated G + 20 nt). For the Type II libraries, assays were carried out for the EGFP-1, EGFP-2, *rol-6* and *unc-22A* targets with gRNAs lengths of 17, 18 and 20 nt.

For *in vivo* assays, gRNAs were transcribed using a TetO modified RPR1 promoter (21). The modified TetO RPR1 promoter has been previously used for gRNA production in other *S. cerevisiae* studies (20,29). The RPR1 gene produces a precursor RNA with an 84 nt 5′ leader sequence. In the context of the RPR1 gene, the leader sequence is excised via self cleavage (23). There is evidence that gRNAs exist in a mixed population of cleaved, partially cleaved and uncleaved versions (see 'Materials and Methods' section for details) (30), and as with any such *in vivo* experiment it is not clear to what extent original, partially processed and fully processed versions may contribute to the observed activity. To convey the composition of the initial *in vivo* transcript, we use a gRNA nomenclature that describes the extent of target-matching sequences as well as potential upstream leader sequences retained from transcription (see Table 1).

In each experiment, we challenged a library of potential target sequences with a single Cas9::gRNA species and examined sequences that remained intact (indicating failure to cleave). Each sequence variant is represented by several distinctly 'barcoded' sequence species in the initial library, with robustness of results supported by consistent retention fractions among the barcoded sub-pools for any single target variant. As previously described (13), a log retention is calculated for each sequence. A negative value for log (retention) suggests cleavage, while a zero value suggests lack of cleavage. Positive log(retention) values are likewise indicative of lack of cleavage. For the *S. cerevisiae in vivo* assay, the library of target plasmids was maintained continually in yeast, during expression of both Cas9 with a nuclear localization signal (22) and the relevant gRNAs (Figure 1A–C). Following growth under inducing conditions (galactose + anhydrotetracycline (ATc)), the resulting change in target representation was assessed as described above.

The use of an internal control population (a set of sequences in each library not predicted to be cleaved by the relevant Cas9::gRNA complex) provides an essential normalization for each experiment. For the Type I variant library we used two normalizations (13) (i) target-related sequences with 4–7 mismatches to the original trigger and (ii) a series of distinct sequences spiked in from a different target library ('protospacer 4' a.k.a 'PS4'). Comparable results were seen from the two normalization methods. For the Type II library design, the target of interest was normalized against all target sequences not related to the gRNA in question. For example, if the gRNA being assayed targeted EGFP-1, then all the target sequence read counts for EGFP-2, *rol-6* and *unc-22A* were aggregated for normalization.

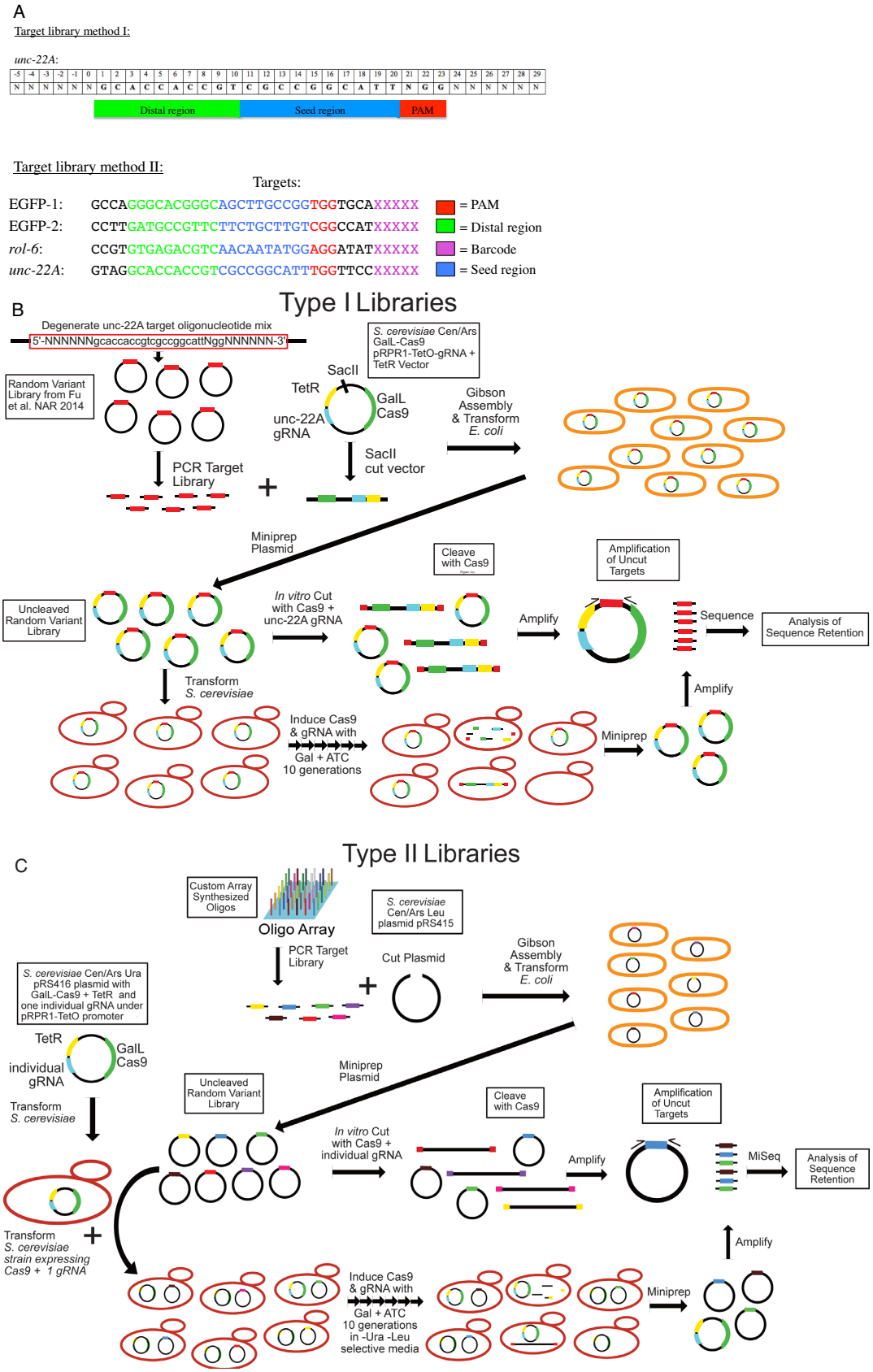

Figure 1. (**A**) The target sequences used for Type I and II variant libraries. Type I libraries are constructed using a synthetic pool of oligonucleotides with random degeneracy at each guide-homologous and PAM base (90% of molecules carry the specified base at any position, with the remaining 10%

Effects of single base variants on Cas9 *in vivo* and *in vitro* cleavage specificity

From these data, we see a range of retention values, varying from sequences that are removed completely from the pool, to those that are unaffected by Cas9 nuclease activity. For Type II libraries, all four targets cleaved with full-length gRNAs exhibited intolerance *in vivo* to variation in the 'seed' region (positions 11–20), but tolerated variation in the seed-distal region (Figure 2). Results for the Type I variant libraries showed similar patterns for the *unc-22A* gRNAs (Supplementary Figures S6 and 8). Figure 2A–D depicts the retention values of all transversion variants of the four targets following expression of gRNAs. For graphical simplicity only the results for transversion variants (purine to pyrimidine or pyrimidine to purine) are shown in Figure 2, with transition variants showing comparable retention values to transversions (Supplementary Figure S9). There is an evident dip in the distal region (positions 1–10) indicating the relaxation of sequence requirements outside the seed region.

Truncated gRNAs have been reported to increase the specificity of Cas9 *in vivo* in mammalian cells (26,31). Although the yeast pol III polymerases produce ambiguous gRNA products (30), we tested the ability of gRNAs with shortened guide::target homology to increase specificity in yeast. Examining gRNAs with 18 and 17 nt complementarity in the yeast *in vivo* Cas9 cleavage assay, we observed each target had a striking increase in specificity (over the gRNA with 20 nt complementarity) with at least one of homology-truncated gRNA lengths. In particular, transversions throughout both seed and distal regions led to a dramatic drop in cleavage by the 18 nt complementarity Cas9::gRNA complex for the EGFP-1, EGPF-2, *rol-6* and *unc-22A* targets. A graph single base transversion effects for all four targets with the full-length gRNAs can be found in Supplementary Figure S10. For both Type I and Type II libraries, a shorter *unc-22A* gRNA (17 nt of complementarity) showed no cleavage (Supplementary Figure S6).

The experimental design of the EGFP-1 target (Type II library) allowed us to address the question of whether detailed homology requirements reflect specific base identities or positional effects within the gRNA. This flexibility derives from the fact that the EGFP-1 target matches three different PAM-adjacent 20-base segments, each of which can be targeted by a unique gRNA. To augment the analyses with the single gRNA in Supplementary Figures S11–12, we carried out assays with the two additional gRNAs *in vivo*, one offset from the original gRNA by 3 bases (EGFP-1-OS) and one the precise reverse-complement of the original guide (EGFP-1-R) for the EGFP-1 target. As shown in Supplementary Figures S11–12, we see comparable mismatch tolerance patterns as a function of position within the gRNA, arguing (at least for this particular site) that some of the observed positional variance in mismatch tolerance through the length of the guide is based on proximity to the PAM site rather than on some shared 5′ or 3′ base composition.

The observed *in vivo* cleavage profiles of the gRNAs are markedly different in a number of aspects from previously described *in vitro* cleavage profiles with the *unc-22A* parent plasmid library and the same gRNA. We evaluated the robustness of this difference by assessing the *in vitro* cleavage preferences with the same plasmid libraries used in this work, including the recloned *unc-22A* Type I variant library and all four targets with the Type II variant libraries (Figure 3, Supplementary Figure S7). Short (3 min) and long (180 min) incubation times were used for both Type I and II variant libraries (Figure 3, Supplementary Figures S7 and 14). The effects of transitions on cleavage *in vitro* are shown in Supplementary Figure S13. For all gRNAs tested *in vitro*, the early time points show greater tolerance for the seed region, with some loss of specificity at longer incubation times (Supplementary Figures S7 and 14). The homology-truncated gRNAs (17 and 18 nt of complementarity) and the full-length gRNA (20 nt of complementarity) have similar cleavage specificity profiles *in vitro* for all time points for all targets, lacking the dramatic improvement in specificity observed with homology-truncated gRNAs *in vivo*.

Effects of adjacent double variants and deletions on Cas9 *in vivo* and *in vitro* cleavage specificity cleavage specificity

The Type II variant library was constructed to include all single base deletions and a set of double variants (consecutive single nucleotide variants) throughout the EGFP-1, EGFP-2, *rol-6* and *unc-22A* targets, allowing the effects of these variants to be measured *in vivo* and *in vitro*. Adjacent double variants *in vivo* resulted in a general decrease of cleavage efficiency for all targets and all lengths of gR-

←

having an equimolar mixture of the other three bases; 'N' bases were from an equimolar mixture of the four individual bases). Type II libraries were constructed using dedicated synthesis on a microarray platform (Custom Array, Inc) of a series of barcoded variants, followed by insertion of these into the relevant target vector. (**B**) To create a yeast target library, we first inserted the relevant *unc-22A* gRNA sequences into the ATc-inducible guide RNA cassette in a yeast Cen/Ars vector also containing inducible Cas9. The resulting plasmids were linearized (SacII) for insertion of the amplified target library of Fu *etal*. (13). Insertion of target libraries employed Gibson Assembly (17), using primers giving appropriate overlaps. Following transformation into *Escherichia coli,* we plasmid prepped each library and used the library for both *in vivo* and *in vitro* experiments. For the *in vivo* experiments the library was transformed into haploid *Saccharomyces cerevisiae*, followed by Cas9 and gRNA induction with galactose and ATc, respectively, for 10 generations (experiments using different numbers of generations are presented in Supplementary Figure S19). Cas9/gRNA cut plasmids should then be lost, while retained plasmids were quantified by PCR (from yeast plasmid minipreps) and sequencing. For the *in vitro* experiments, the *E. coli* minipreped plasmid libraries were cut in reactions with a mixture of purified Cas9 enzyme and *in vitro* transcribed *unc-22A* gRNA. For both methods the retained plasmids were isolated and amplified and sequenced and retention profiles were determined. (**C**) As in 1b except: Type II target libraries were created from microarray oligo pools designed to four different target sequences. These pools were PCR amplified and inserted into the multiple cloning site of a leucine selectable Cen/Ars plasmid using Gibson Assembly, and transformed into *E. coli* as in 1b. Pooled plasmid DNA from this library was prepared and used for *in vitro* experiments exactly as in 1b. For the *in vivo* experiments, the Cas9 and ATc-inducible gRNA genes were on a separate uracil selectable Cen/Ars plasmid from the target library. First, we transformed the Cas9/gRNA plasmid into haploid *S. cerevisiae.* Next, a second round of transformation was used to introduce the library was transformed into yeast containing the Cas9/gRNA plasmid. During *in vivo* cutting, selection was maintained for both plasmids. Otherwise, this protocol was identical to that used for the Type I libraries.

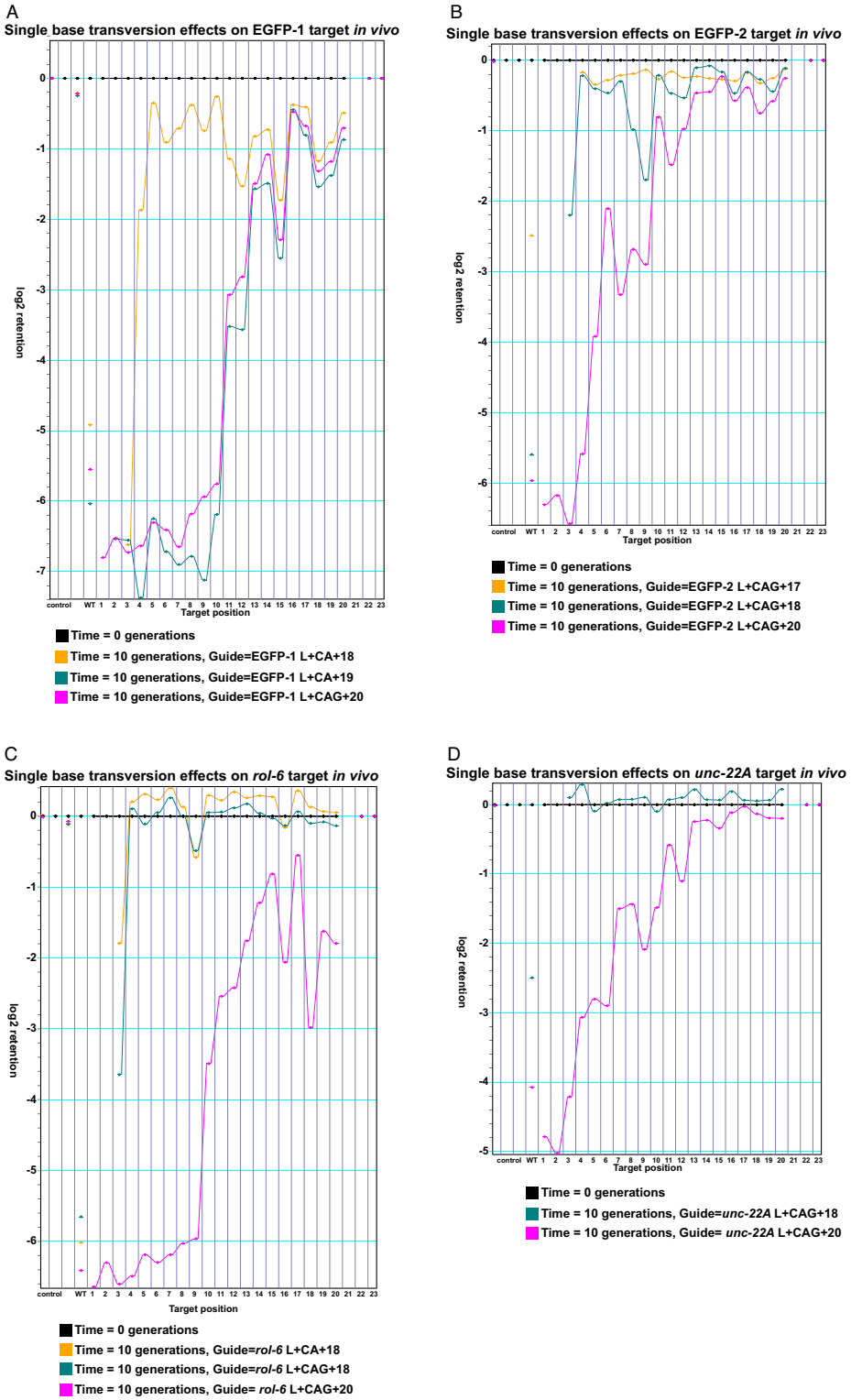

Figure 2. *In vivo* cleavage effects of single base variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs over 10 generations. A median of effects for single base transversion variants is indicated for each target position by a dot. Coloring of lines is used to indicate the length of the gRNA. These graphs show transversion variant retention for the target region (i.e. the gRNA binding site) and PAM nucleotides (positions 1–20 and 22–23). Only sequences with ≥50 reads in the control (uncut) library were considered in calculating this median. Targets that were not complementary to the gRNA of interest (i.e. for the other three targets in the library) were used as negative controls (labeled 'control'). Only positions with gRNA::DNA targeting have a reported median retention value. Wild-type ('WT') targets are included for each gRNA. A negative retention score indicates sequence cleavage, while a retention score of zero corresponds to a lack of cleavage, comparable to the pool of non-complementary sites. A slightly positive retention score seen in some cases is likewise indicative of a lack of Cas9 cleavage (such modest anomalies reflect the limitations of consistent normalization, with small differences in PCR recovery between gRNA pools leading to the apparently positive retention (13)). (AF_SOL_672, 673,676).

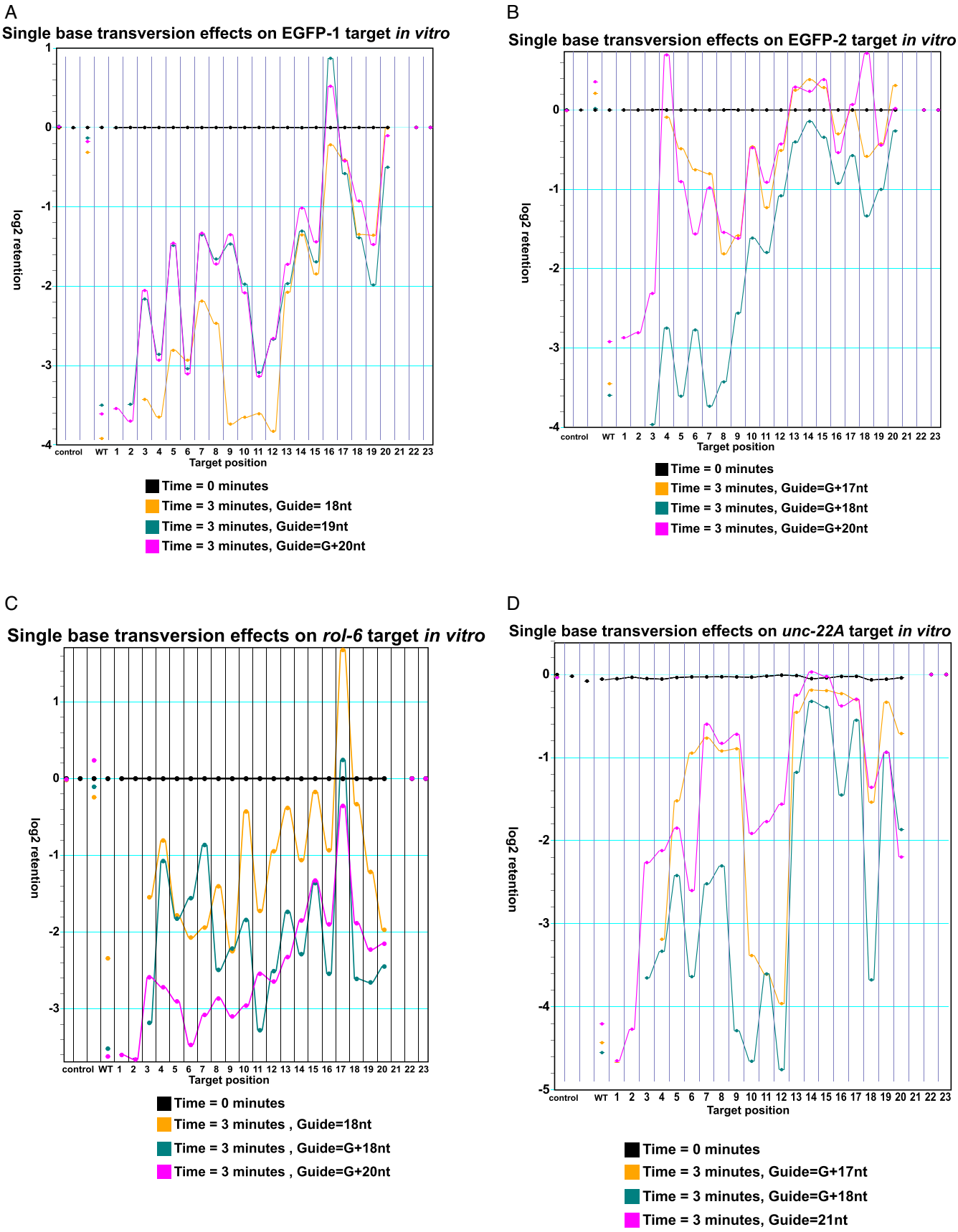

Figure 3. *In vitro* cleavage effects of single base variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs with 3 min incubation (analytical methods for generating these graphs are as in Figure 2). (AF_SOL_636, 637, 643, 664).

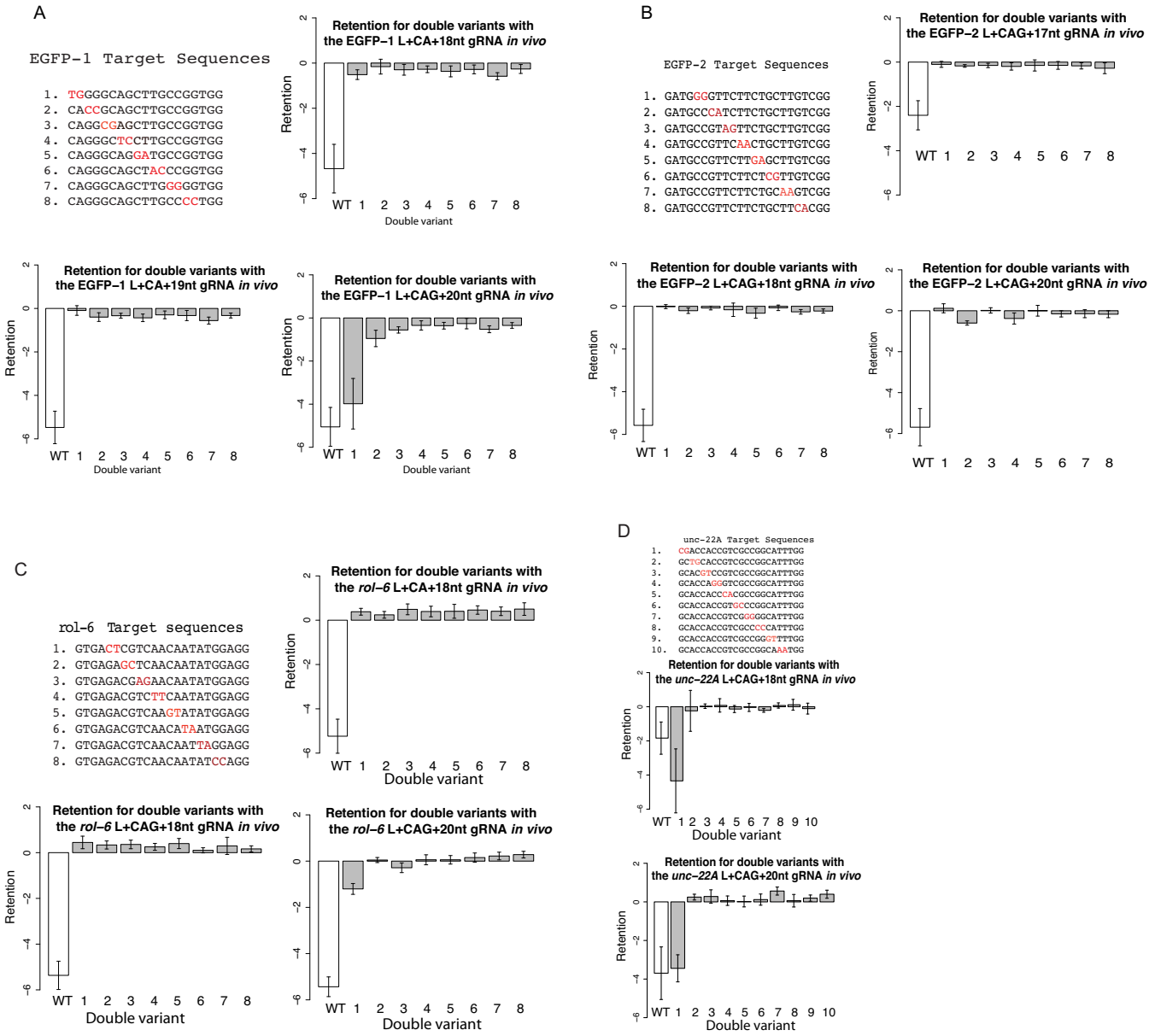

Figure 4. *In vivo* cleavage effects of adjacent double variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs. This graph depicts the *in vivo* median retention scores (Figure 2) of each double variant assayed in the Type II libraries. The error bars represent the standard deviation. The key to the target variants assayed is shown in the top panel. The bases in red are the variants in the given target of interest.

NAs (17, 18 and 20 nt complementarity) (Figure 4A–D). Although the double variants *in vivo* caused striking decreases in cleavage with full and homology-truncated gRNAs, some deletion variants were tolerated or even modestly increased cleavage efficiency (Figure 6A–D). For example, for the full-length guide for EGFP-1 (EGFP-1 L + CAG + 20) deletions toward middle and the 5′ end of the target caused cleavage comparable to the wild-type sequences (deletions denoted by 1, 2, 4 and 5) (Figure 6A). Results for adjacent double variants and deletions for the EGFP-1-R and EGFP-OS gRNAs follow similar trends and are in Supplementary Figures S11–12. The Type I library double variant results for the *unc-22A* target for 17, 18 and 20 nt complementarity gRNAs *in vivo* can be found in Supplementary Figure S15.

Figure 5 shows results of cleavage of the corresponding lengths of gRNAs tested in Figure 4 *in vitro* for the 3-min time point. The double variants at the early time points have similar decrease in cleavage as the *in vivo* results. However, with longer incubation time (180 min) double variants in the distal region of all targets are cleaved while most variants in the seed remain intolerant (Supplementary Figure S16). Figure 7 shows the 3-min time point *in vitro* results comparable to the *in vivo* results from Figure 6. For all lengths of gRNAs tested, most of the deletion variants are not cleaved at the early time points. However, with longer incubation time (180 min) almost all deletion variants in the distal region are cleaved, with some deletion variants in the seed region remaining uncleaved (Supplementary Figure S17). The

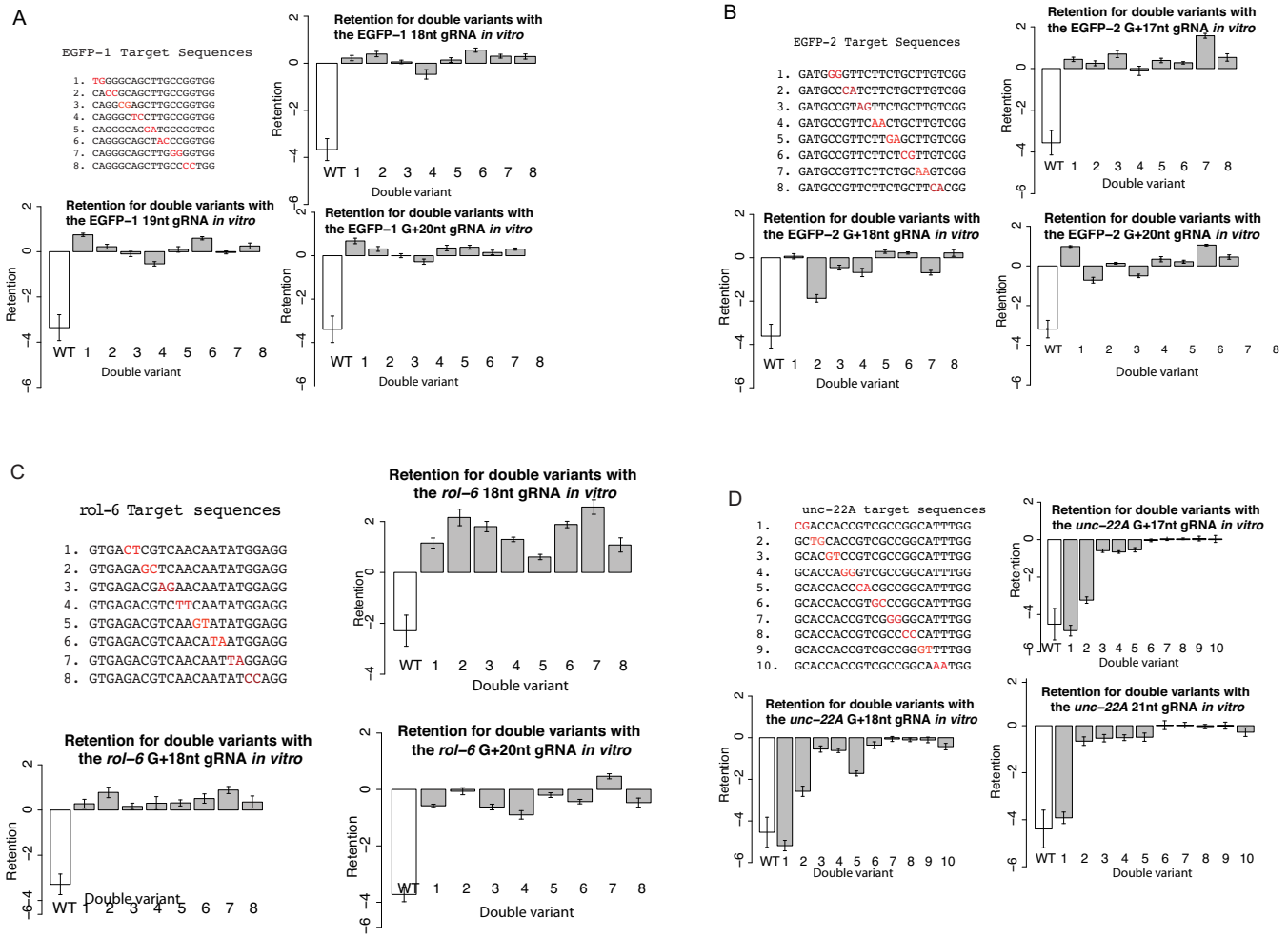

Figure 5. *In vitro* cleavage effects of adjacent double variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs. This graph depicts the *in vitro* median retention scores (Figure 3) of each double variant assayed in the Type II libraries.

trend toward decreased specificity at longer time points is of substantial relevance to *in vitro* applications of Cas9 (e.g. utilization of Cas9::guide RNA complexes as surrogate restriction enzymes may require care in terms of incubation times that is generally not afforded to conventional restriction systems). The Type I library double variant results for the *unc-22A* target for 17, 18 and 20 nt complementarity gRNAs *in vitro* can be found in Supplementary Figure S18.

Comparison of mismatch tolerance in yeast and mammalian cell culture

Two of the four targets characterized in this study had previously been characterized in mammalian cells, EGFP-1 and EGFP-2 (26). The cleavage efficiency profiles for these targets in the mammalian and yeast systems show notable similarities including a strong requirement for continuous homology between target and guide for successful cleavage using the homology-shortened (18 nt) gRNAs. However, there were some small differences. The 18 nt gRNA to EGFP-1 in yeast did not tolerate any double (i.e. consecutive) mismatches (Figure 4A) whereas in the mammalian cell line used, double mismatches were tolerated in the middle of the target sequence. Indeed, full-length gRNAs in mammalian cells also appeared to be more tolerant of double mismatches than those in the yeast system. These differences are potentially due to a number of experimental factors. For example, our assay measures cleavage through the loss of the target sequence, rather than reading downstream expression via fluorescence (26). In addition, unlike the mammalian cell experiments, our method varied the target and not the gRNA sequence. Our yeast system also uses a yeast pol III promoter that transcribes a leader sequence whereas the mammalian U6 promoter transcribes a G to the 5′ end of gRNAs. Finally, as neither EGFP in mammalian cells, nor our plasmid library represent native host genes, chromatin structure is largely ignored and has been shown to be an important factor in predicting effective guides (29,32) and may impact the two systems differently. Overall, our results show more consistencies than differences with the mammalian data, suggesting that the determinants of Cas9 specificity are generally shared across a diversity of cellular environments.

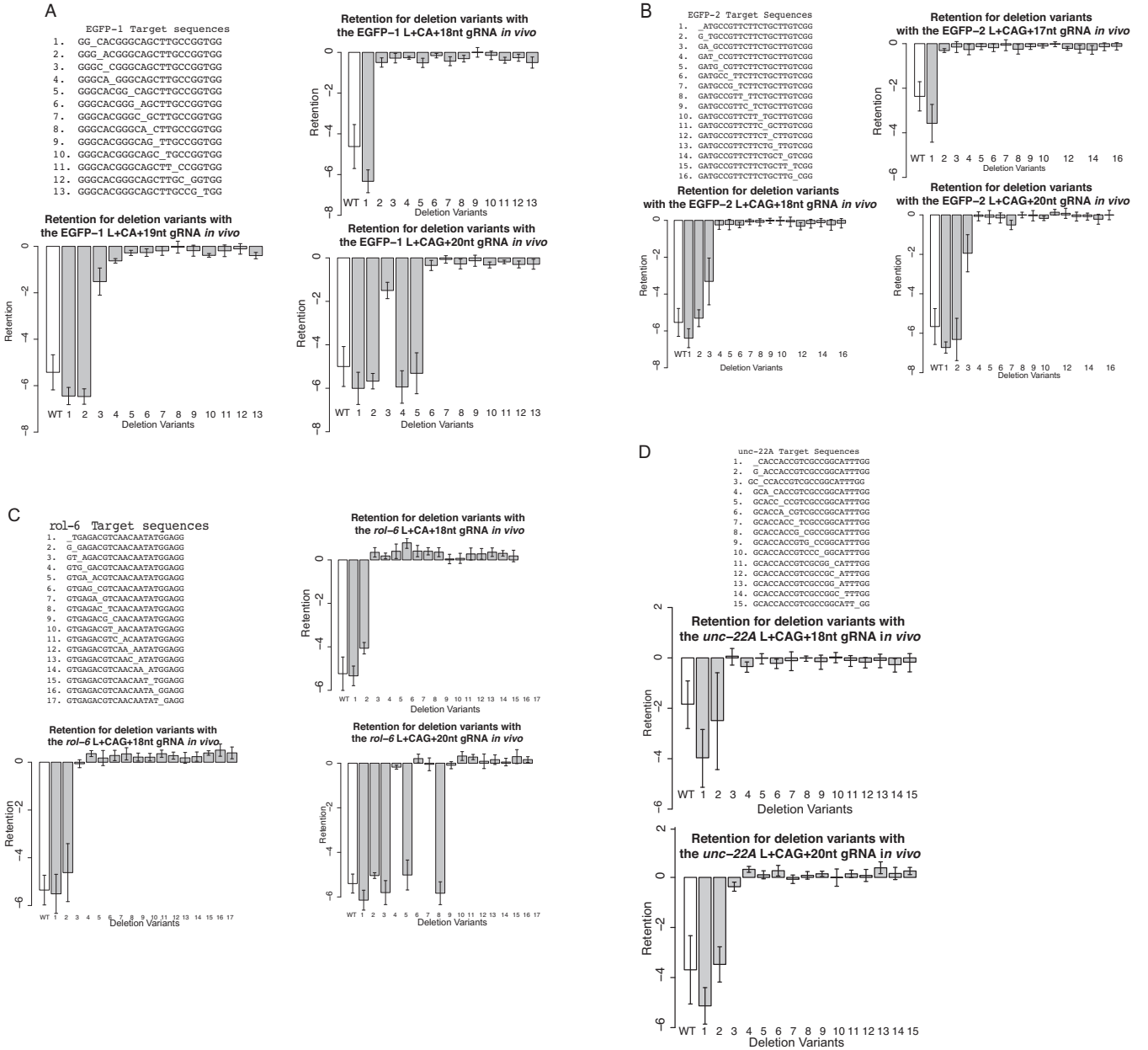

Figure 6. *In vivo* cleavage effects of deletion variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs. This graph depicts the *in vivo* median retention scores (Figure 2) of each deletion variant assayed in the Type II libraries. The key to the target variants assayed is shown in the top panel. The '_' indicate the deleted base in the given target of interest.

CONCLUSIONS

Cas9 has become a dominant platform for both biochemical manipulation of DNA and *in vivo* genome editing, thus warranting extensive research to explore the gRNA::DNA interaction influence on Cas9 cleavage specificity. This study provides a detailed map of Cas9 specificity and cleavage for gRNAs of various lengths in the parallel contexts of *in vitro* and *in vivo* conditions. We found that the cleavage efficiency profiles mammalian and yeast systems show notable similarities including reduced tolerance of mismatches with homology-shortened (18 nt) gRNAs, with a few minor differences that could be due to differences between the organisms or experimental setup. As the use of Cas9 purified protein and of Cas9 expression constructs expands to routine assays and larger screens, knowledge of which mismatched and off-target cleavages are possible becomes an underpinning of experimental design, particularly in cases where allele-specific Cas9 targeting is either a possibility or a goal. These results identify differences in Cas9 activity in different contexts, highlighting the caution that must be taken in the use of Cas9 in different systems *in vitro* versus *in vivo*.

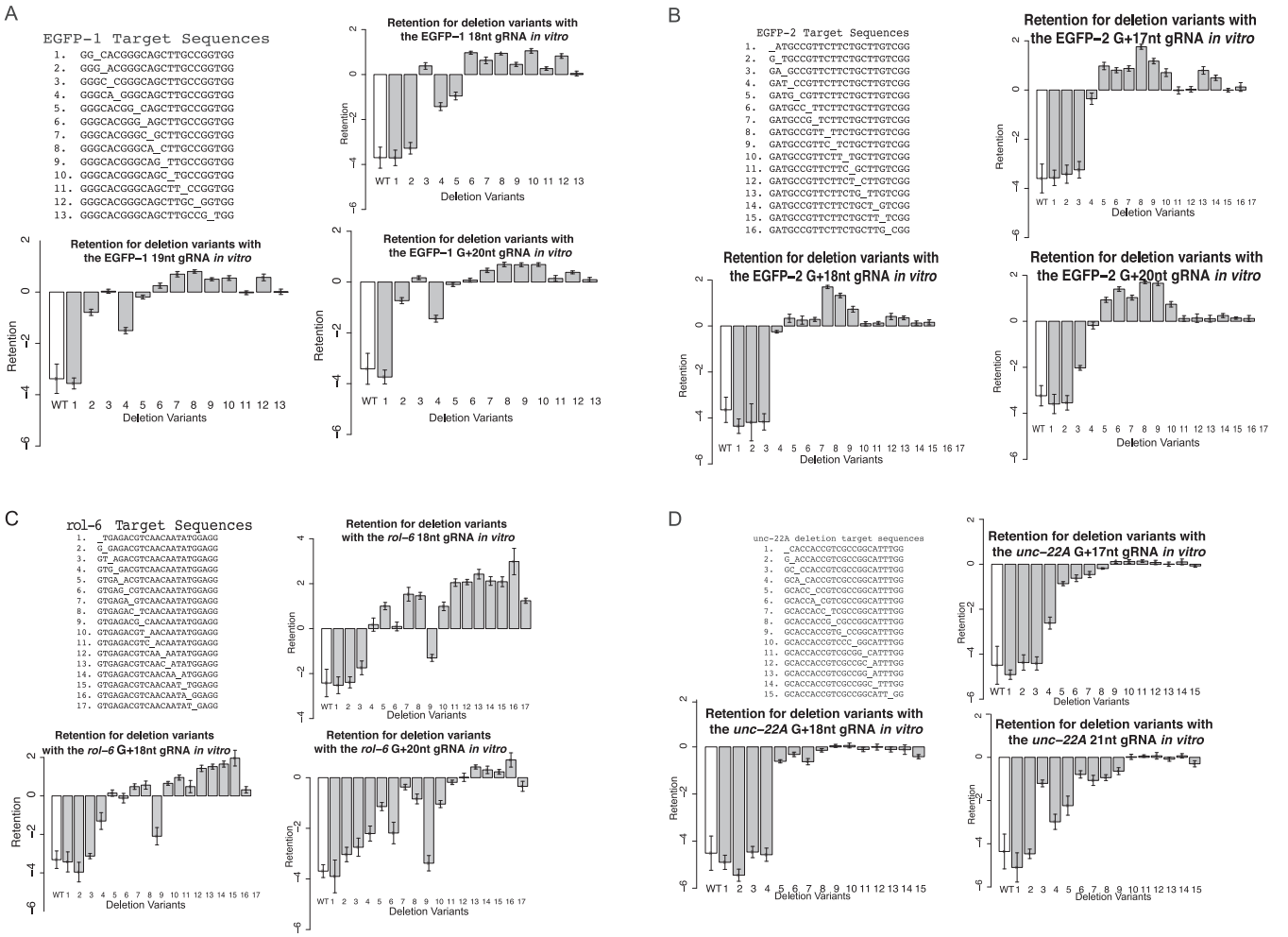

Figure 7. *In vitro* cleavage effects of deletion variants with (**A**) EGFP-1, (**B**) EGFP-2, (**C**) *rol-6* and (**D**) *unc-22A* gRNAs. This graph depicts the *in vitro* median retention scores (Figure 3) of each deletion variant assayed in the Type II libraries.

ACCESSION NUMBER

GEO ID: GSE79667

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Kevin RJ Roy, Daniel Herschlag, Michael Nonet, Ronald W Davis, Karen Artiles and Gavin Sherlock for their help and suggestions.

FUNDING

[NIH R01GM37706 to A.Z.F]; [NIH T32GM00779 to B.X.H.F]; NSF Graduate Fellowship (to B.X.H.F); [NIH P01HG000205 to R.P.S, J.D.S.]. Funding for open access charge: [NIH R01GM37706 to A.Z.F].
Conflict of interest statement. None declared.

REFERENCES

1. Mali,P., Esvelt,K.M. and Church,G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods*, **10**, 957–963.
2. Terns,R.M. and Terns,M.P. (2014) CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet.*, **30**, 111–118.
3. Bhaya,D., Davison,M. and Barrangou,R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.*, **45**, 273–297.
4. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
5. Hendel,A., Fine,E.J., Bao,G. and Porteus,M.H. (2015) Quantifying on- and off-target genome editing. *Trends Biotechnol.*, **33**, 132–140.
6. Tsai,S.Q., Zheng,Z., Nguyen,N.T., Liebers,M., Topkar,V.V, Thapar,V., Wyvekens,N., Khayter,C., Iafrate,A.J., Le,L.P. *et al.* (2014) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.*, **33**, 187–197.
7. Kim,D., Bae,S., Park,J., Kim,E., Kim,S., Yu,H.R., Hwang,J., Kim,J.-I. and Kim,J.-S. (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods*, **12**, 237–243.
8. Pattanayak,V., Lin,S., Guilinger,J.P., Ma,E., Doudna,J.A. and Liu,D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839–843.

9. Cho,S.W., Kim,S., Kim,Y., Kweon,J., Kim,H.S., Bae,S. and Kim,J.-S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.*, **24**, 132–141.
10. Wang,X., Wang,Y., Wu,X., Wang,J., Wang,Y., Qiu,Z., Chang,T., Huang,H., Lin,R.-J. and Yee,J.-K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
11. Wu,X., Scott,D.A., Kriz,A.J., Chiu,A.C., Hsu,P.D., Dadon,D.B., Cheng,A.W., Trevino,A.E., Konermann,S., Chen,S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
12. Hwang,W.Y., Fu,Y., Reyon,D., Maeder,M.L., Tsai,S.Q., Sander,J.D., Peterson,R.T., Yeh,J.-R.J. and Joung,J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.*, **31**, 227–229.
13. Fu,B.X.H., Hansen,L.L., Artiles,K.L., Nonet,M.L. and Fire,A.Z. (2014) Landscape of target: guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res.*, **42**, 13778–13787.
14. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Keith Joung,J. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
15. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2015) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.
16. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
17. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobas1. *Nat. Methods.*, **6**, 343–345.
18. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Véronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., André,B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387–391.
19. Amberg,D.C., Burke,D.J. and Strathern,J.N. (2006) High-efficiency transformation of yeast. *CSH Protoc.*, **2006**, doi:10.1101/pdb.prot4145.
20. Farzadfard,F., Perli,S.D. and Lu,T.K. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.*, **2**, 604–613.
21. Bak,G., Hwang,S.W., Ko,Y., Lee,J., Kim,Y., Kim,K., Hong,S.K. and Lee,Y. (2010) On-off controllable RNA hybrid expression vector for yeast three-hybrid system. *BMB Rep.*, **43**, 110–114.
22. DiCarlo,JE, Norville,JE, Mali,P, Rios,X, Aach,J and Church,GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.*, **41**, 4336–4343.
23. Lee,J.Y., Rohlman,C.E., Molony,L.A. and Engelke,D.R. (1991) Characterization of RPR1, an essential gene encoding the RNA component of Saccharomyces cerevisiae nuclear RNase P. *Mol. Cell. Biol.*, **11**, 721–730.
24. Harismendy,O., Gendrel,C.-G., Soularue,P., Gidrol,X., Sentenac,A., Werner,M. and Lefebvre,O. (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J.*, **22**, 4738–4747.
25. Kosuri,S. and Church,G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499–507.
26. Fu,Y., Sander,J.D., Reyon,D., Cascio,V.M. and Joung,J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279–284.
27. Arribere,J.A., Bell,R.T., Fu,B.X., Artiles,K.L., Hartman,P.S. and Fire,A.Z. (2014) Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics*, **198**, 837–846.
28. Farboud,B. and Meyer,B.J. (2015) Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics*, **199**, 959–971.
29. Smith,J.D., Suresh,S., Schlecht,U., Wu,M., Wagih,O., Peltz,G., Davis,R.W., Steinmetz,L.M., Parts,L. and St.Onge,R.P. (2016) Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol.*, **17**, 45.
30. Camier,S., Dechampesme,A.M. and Sentenac,A. (1995) The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 9338–9342.
31. Fu,Y., Reyon,D. and Joung,J.K. (2014) Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs. *Methods Enzymol.*, **546**, 21–45.
32. Horlbeck,M.A., Witkowsky,L.B., Guglielmi,B., Replogle,J.M., Gilbert,L.A., Villalta,J.E., Torigoe,S.E., Tijan,R. and Weissman,J.S. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife*, **5**, e12677.