Engineered transcription-associated Cas9 targeting in eukaryotic cells

Gregory W. Goldberg[¥*], Manjunatha Kogenaru[¥], Sarah Keegan[¥+], Max A. B. Haase[¥+], Larisa Kagermazova[¥], Mauricio A. Arias[#], Kenenna Onyebeke[¥], Samantha Adams[¥], David Fenyö[¥], Marcus B. Noyes[¥*], and Jef D. Boeke[¥&*]

[¥]Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

+Denotes equivalent contribution to the work

[#]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

[&]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn NY 11201

*Correspondence to Jef.Boeke@nyulangone.org, Gregory.Goldberg@nyulangone.org, and Marcus.Noyes@nyulangone.org

ABSTRACT

DNA targeting Class 2 CRISPR-Cas effector nucleases, including the well-studied Cas9 proteins, evolved protospacer-adjacent motif (PAM) and guide RNA interactions that sequentially license their binding and cleavage activities at protospacer target sites. Both interactions are nucleic acid sequence specific but function constitutively; thus, they provide intrinsic spatial control over DNA targeting activities but naturally lack temporal control. Here we show that engineered Cas9 fusion proteins which bind to nascent RNAs near a protospacer can facilitate spatiotemporal coupling between transcription and DNA targeting at that protospacer: Transcription-associated Cas9 Targeting (TraCT). Engineered TraCT is enabled when suboptimal PAM interactions limit basal activity in vivo and when one or more nascent RNA substrates are still tethered to the actively transcribing target DNA in *cis*. We further show that this phenomenon can be exploited for selective editing at one of two identical targets in distinct gene loci, or, in diploid allelic loci that are differentially transcribed. Our work demonstrates that temporal control over Cas9's targeting activity at specific DNA sites may be engineered without modifying Cas9's core domains and guide RNA components or their expression levels. More broadly, it establishes RNA binding in *cis* as a mechanism that can conditionally stimulate CRISPR-Cas DNA targeting in eukaryotes.

Introduction

Class 2 CRISPR-associated (Cas) effectors such as Cas9 are vastly simplifying (epi)genome editing applications in biomedical research, owing to their impressive DNA targeting specificity, facile guide RNA programmability, and tractable single-protein architecture[1-5]. In complex with guide RNAs, Cas9 proteins constitutively bind and cleave DNA at target sites known as protospacers[6-8]. Binding and cleavage occur sequentially through a coupled mechanism[9] initiated by direct protein-DNA interactions at protospacer-adjacent motifs (PAMs)[10,11], for example 'NGG' sequences[12] recognized by the model Cas9 protein from *Streptococcus pyogenes* SF370. Following PAM recognition, Cas9 forms an R-loop and cleaves its target once sufficient base-pairing is achieved between its guide RNA and protospacer DNA[8,9,13-17]. Mechanistic coupling between PAM and guide RNA interactions enforces sequence-specific spatial control over Cas9's DNA targeting activities; i.e., even perfectly complementary guides cannot license binding or cleavage at protospacers that lack a functional PAM. Furthermore, suboptimal PAM (subPAM) interactions may license intermediate activities, as seen with the NAG and NGA PAMs of *S. pyogenes* Cas9 that are recognized less strongly than NGG[12,18,19].

Beyond Cas9's broad utility as a tool for constitutive DNA binding and cleavage, certain applications demand dynamically modulated activity that can be controlled over time by an experimentalist or via an endogenous signal. Temporal control was previously engineered by inducing Cas9 expression[20], as well as by modifying Cas9's guide RNA or protein components to ensure that they are held inactive until post-transcriptionally or post-translationally induced by chemical[21-24], optical[25-28], or

enzymatic[29] inputs. Engineered Cas9 'pro-enzymes' were further shown to respond to endogenously expressed proteases via proteolytic activation[29]. Spatiotemporal coupling between endogenous signals and CRISPR-Cas DNA targeting has also been described in natural systems. The type III-A CRISPR-Cas systems encode multi-protein effector complexes[3] that naturally depend on co-transcriptional binding of nascent RNAs to license DNA degradation[30-32]. Nascent RNA binding simultaneously localizes the type III-A complex near target DNA and activates its DNase activity[31,32], thereby ensuring that transcribed DNA targets are degraded while identical but repressed protospacers may be conditionally tolerated in prokaryotes[33-36]. It is not known whether nascent RNA binding can be exploited to control DNA targeting by other CRISPR-Cas effectors. We investigated this by engineering Cas9 fusions that can bind to nascent RNA substrates transcribed near a protospacer in *cis*, thus partially emulating the type III-A targeting mechanism and forming the basis for Transcription-associated Cas9 Targeting (TraCT) in eukaryotes. subPAM interactions were employed to ensure that Cas9's basal DNA targeting activity in vivo did not mask the transcription-associated activity conferred by nascent RNA binding.

Results

Dynamic ligand-controlled rescue of subPAM interactions

Prior to investigating TraCT we first sought to demonstrate that Cas9 targeting activity can be dynamically controlled without disrupting its core protein domains or single guide RNA (sgRNA) components. To this end, we reasoned we could engineer Cas9 fusions that conditionally interact with *cis* elements in proximity to protospacers and thereby

stimulate targeting when subPAM interactions limit basal activity. Target site editing was assayed using a single-strand annealing[37,38] (SSA) based reporter system[19] wherein protospacer cleavage drives scarless reconstitution of a disrupted *LYS2* open reading frame (ORF) in *Saccharomyces cerevisiae* yeast (Fig. 1a). Throughout the study, this system enabled isogenic testing of various protospacer-proximal *cis* elements inserted between direct repeats without requiring exogenous repair templates. It was previously shown that fusion to strong zinc finger (ZF) DNA binding domains can constitutively increase Cas9 editing at protospacers with subPAMs when a ZF binding site is located nearby[39]. It was furthermore shown, in the same work[39], that ZF-DNA interactions can compensate for suboptimal activity on NGG PAMs when Cas9's PAM-interacting domain[40] (PID) was genetically attenuated with a single substitution known as 'MT3' ($Cas9^{MT3}$). Although ZFs can now be designed to recognize nearly any DNA sequence[41], they do not confer temporal control over Cas9's activity. We therefore initially tested a single *tetO2* DNA element, the target substrate for wild-type (Wt) TetR and reverse TetR (RevTetR) DNA binding proteins that can be dynamically induced with doxycycline (Dox) to dissociate from or bind *tetO2*, respectively[42]. The (Rev)TetR proteins function as homodimers so we fused tandem monomers to the C-terminus of Cas9 in accordance with a previously described tandem single-chain (sc) design[43], generating Cas9-scTetR and Cas9-scRevTetR fusions. These were constructed from an isogenic scaffold that included N- and C-terminal nuclear localization signal (NLS) tags and also served as our 'native' Cas9 control throughout the work (Extended Data Fig. 1a; Supplementary Note 1). Our heterologous protospacer insertion contained an NGG PAM that allowed us to compare editing activities for three PID variants which

recognize NGG at different strengths: Wt Cas9 ($Cas9^{Wt}$), $Cas9^{MT3}$, and SpCas9-NG with multiple PID substitutions[44] (henceforth, 'NG' or $Cas9^{NG}$). As expected, editing with native Cas9 was unaffected by Dox and consistently lower with the MT3 and NG variants compared to Wt (Fig. 1b). In contrast, Cas9-scTetR fusions achieved similarly high levels of editing with all three PID variants in the *absence* of Dox, though editing with MT3 and NG was reduced to background levels when Dox was present (Fig. 1b). This phenotype was qualitatively reversed when the scTetR domain was replaced with either V10 or V16 variants[42] of scRevTetR: in this context, maximal editing by MT3 and NG was achieved in the *presence* of Dox (Fig. 1c). All Dox-controllable phenotypes strictly required the nearby *tetO2* element (Extended Data Fig. 1b).

To confirm that cleavage by $Cas9^{NG}$ fusions is stimulated on physiological timescales we set up a liquid induction assay that allowed us to monitor growth while controlling the scRevTetR(V16) domain with Dox. We used modified yeast strains lacking a downstream SSA repeat or other endogenous repair templates that would rescue cells from persistent protospacer cleavage, thus ensuring a prolonged DNA damage response that slows growth upon targeting. Dox-dependent growth reductions required both the scRevTetR domain and the *tetO2* element (Extended Data Figs. 2a–e). Importantly, when compared with non-inducible controls, this Dox-dependent growth reduction was significantly detectable 260 minutes post-induction (Extended Data Fig. 2f). Assuming our apparent plate reader doubling time of ~90 min (Extended Data Fig. 2e), this indicates that growth was indeed reduced for a substantial portion of the population in under three generations.

mRNA binding by Cas9 fusions can facilitate TraCT

Having found that Cas9-sc(Rev)TetR fusions allow for dynamic control over DNA targeting activity when combined with subPAM interactions and a nearby *tetO2* element, we next sought to test whether binding DNA-tethered nascent RNA in *cis* could supplant direct DNA binding at *tetO2*. This was first accomplished with well-described MS2 phage parts: RNA hairpins containing either Uracil ($MS2^U$) or Cytosine ($MS2^C$) at a crucial position in their loop[45-47], and MS2 coat protein (MCP) variants that can bind them co-transcriptionally[47-49]. MCP binds to individual MS2 hairpins as a homodimer[46,50,51] so we again fused single-chain tandem monomers, generating Cas9-scMCP (Fig. 1d; Extended Data Fig. 1a). We first tested reporter strains with two Hairpin Templates (HTs) either upstream or downstream of the protospacer (Supplementary Note 2). It was unclear whether the higher-affinity $MS2^C$ loops[45-47] would be preferred over Wt $MS2^U$ loops for our purposes, so we used one of each. This yielded a small but significant increase in editing by PID mutants of Cas9-scMCP compared to native Cas9, but only with hairpins placed upstream (Fig. 1e). To probe whether alternative upstream configurations could improve activity we tested single and double $MS2^U$ loops, or $MS2^C$ loops. Editing activities were comparable but we obtained lower *p* values with $MS2^U$ loops (Extended Data Fig. 3a) so proceeded with one or two *$MS2^U$* HTs in all subsequent experiments. $Cas9^{Wt}$ and $Cas9^{Wt}$-scMCP controls displayed indistinguishable editing in all strains as expected (Extended Data Fig. 3b).

We then tested the phage lambda (λ) N-boxB system[52] which is orthogonal to MS2[53]. Individual boxB RNA hairpins are bound co-transcriptionally by λ N protein monomers[54-57] or a minimal 21-residue tract (N^{2-22}) from its N-terminus[58-60]. We fused

two N^{2-22} peptides (2xN) to the C-terminus of Cas9, generating Cas9-2xN, and assayed editing with one, two, or three *boxB* HTs upstream of the protospacer (Supplementary Note 2). Considerably higher editing by PID mutants was observed with Cas9-2xN than native Cas9, in particular with two or three *boxB* HTs (Fig 1f). This Cas9-2xN activity was specific for *boxB* HTs, as no increase was observed with two upstream $MS2^U$ HTs. In addition, 2xN fusions partially increased editing with Wt PIDs at suboptimal N<u>AG</u> or N<u>GA</u> PAMs with a single *boxB* HT upstream, though activity at N<u>GG</u> PAMs was independent of 2xN (Extended Data Fig. 3c).

To rule out the possibility that nascent RNA binding non-specifically increases the activity of interacting Cas9 fusion proteins, e.g., by generally capturing higher concentrations of Cas9 fusions in the nucleus rather than specifically increasing editing near HTs, we tested whether HTs stimulate editing in *trans* using SSA dual-reporter yeast strains. These strains contain a second identical protospacer and PAM insertion, accompanied by zero or two upstream $MS2^U$ HTs, that disrupts the constitutively transcribed *CAN1* ORF on a separate chromosome from *LYS2* (Fig. 2a). Once reconstituted through SSA, the *CAN1* locus sensitizes cells to canavanine[61] (Can) and therefore serves as a counterselection marker: only edited cells suffer reduced colony formation efficiency in the presence of Can. Editing at *LYS2* by the NG variant was increased with scMCP fusions only when HTs were present at the *LYS2* locus (Fig. 2b–c). Conversely, whereas scMCP fusions increased editing by NG at the *CAN1* locus when HTs were present there, they did not increase editing at the *CAN1* locus when HTs were present at *LYS2* (Fig. 2c; Extended Data Fig. 4a). Wt control activities at each locus were not affected by fusion to scMCP or the presence of HTs (Extended Data

Figs. 4b–c). Taken together, these results indicate that auxiliary RNA binding domains (RBDs) fused to Cas9 partially compensate for subPAM interactions when their RNA substrates are transcribed from one or more templates near the protospacer in *cis*. Given their co-transcriptional binding capabilities[47-49,54-57], we infer that these RBDs stimulate editing through a TraCT mechanism.

TraCT can drive allele-specific editing

To further challenge the hypothesis that Cas9-RBD fusions function with *cis*-acting HTs via a TraCT mechanism we examined whether they would stimulate transcription-dependent *allelic* editing. This was investigated with a diploid SSA reporter yeast strain that contains identical target and $MS2^U$ insertions disrupting each of its *LYS2* alleles, but with different promoters controlling their transcription (Fig. 3a). One allele contains a β-Estradiol (β-E) inducible promoter (*pZEVi*) that can achieve high levels of transcription[62,63], while the other contains the native *pLYS2* promoter that is constitutive but weak[64,65]. We first tested haploid yeast strains containing either the constitutive *pLYS2* promoter or the inducible *pZEVi* promoter. In the *pLYS2* strain, editing with $Cas9^{MT3}$ or $Cas9^{MT3}$-scMCP was unaffected by β-E as expected (Fig. 3b). In the *pZEVi* strain, however, editing was significantly increased by β-E at 1000nM relative to 0 and 15nM, and $Cas9^{MT3}$-scMCP activity at 1000nM approached that of Wt (Fig. 3b; Extended Data Fig. 5a). Similar trends were obtained when plating diploid transformants after 24h outgrowth in liquid medium that selects for Cas9(-scMCP) and sgRNA maintenance but lacks β-E (Extended Data Fig. 5b). To definitively determine whether editing by the MT3 variant after outgrowth is preferentially stimulated at the *pZEVi* allele

upon induction, we subcultured the diploid transformants in liquid media containing 0, 15, or 1000nM β-E and then harvested genomic DNA for targeted deep sequencing (see Methods). This confirmed that the ratio of unedited *pLYS2*:*pZEVi* alleles increased significantly in MT3 populations with 1000nM β-E relative to 15nM, and to a much greater extent for the scMCP fusion (Fig. 3c). A similarly significant increase in this ratio, though smaller, was also detectable with $Cas9^{MT3}$-scMCP at 15nM β-E relative to 0nM, yet it was undetectable with native $Cas9^{MT3}$. Note that in the absence of β-E we measured a mean ratio slightly *less* than 1:1 (~0.95) with the scMCP fusion, but not with native Cas9 (~1.02), implying that nascent RNA binding at the weakly transcribed *pLYS2* allele was sufficient to license preferential editing relative to the repressed *pZEVi* allele. Collectively, these results establish an engineering strategy that facilitates spatiotemporal coupling between in vivo transcription near particular protospacers and Cas9 targeting activity at those sites.

TraCT can stimulate editing with exogenous donor DNA

We investigated whether TraCT activity is detectable in non-SSA repair contexts to probe its generalizability. First we designed a haploid *pZEVi-LYS2* yeast reporter strain that lacks the downstream repeat used for SSA, such that perfect reconstitution of its *LYS2* ORF through homology-directed repair requires co-transformation with an exogenous DNA donor fragment (Extended Data Fig. 6a). This strain contains one *boxB* HT upstream of the protospacer for testing with Cas9-2xN fusions (Extended Data Fig. 6a). Constitutive Wt Cas9 cleavage suppresses yeast colony formation until target sequences are genetically disrupted[66,67], and colony formation with $Cas9^{Wt}$ or $Cas9^{Wt}$-

2xN was increased over 50-fold in our system by supplying a DNA donor fragment that eliminates the target upon repair (Extended Data Fig. 6b). However, unlike Wt neither $Cas9^{NG}$ nor $Cas9^{NG}$-2xN suppressed colony formation under any conditions tested; they instead produced only reduced-growth phenotypes with targeting sgRNAs at 1000nM β-E (Extended Data Fig. 6b), and their phenotypes were indistinguishable in the absence of donor DNA (Extended Data Fig. 6b; Extended Data Figs. 7a–d; Supplementary Note 3). In parallel, we quantified *LYS2*$^{+}$ fractions by plating immediately (Extended Data Fig. 6c) or by replating after 44 h outgrowth on full plates containing 0, 15, or 1000nM β-E (Extended Data Figs. 6d–e). This revealed significant increases in donor-dependent editing with $Cas9^{NG}$-2xN after outgrowth at 0nM and 15nM (Extended Data Fig. 6d; Supplementary Note 3).

Protospacer orientation influences TraCT outcomes

The preceding sections defined several editing conditions in which RBD-mediated interactions can detectably compensate for subPAM activity. Given that subPAM interactions are expected to limit Cas9 recruitment to DNA at the PAM-recognition step, RBDs may license TraCT through a facilitated recruitment mechanism. However, the experiments with *pZEVi* strains also showed that β-E induction can stimulate native $Cas9^{NG}$ (Extended Data Figs. 6b–e, Extended Data Fig. 7a) or native $Cas9^{MT3}$ (Figs. 3b–c, Extended Data Fig. 5b) in certain contexts; i.e., when RBDs are absent. This prompted us to further investigate two ways in which transcription might inherently affect Cas9 editing activity and potentially mask RBD-mediated TraCT.

First, high transcriptional activity has been shown to increase histone eviction[68], and occlusion of PAM sequences within nucleosomes was previously found to impair Cas9 cleavage activity[69]. Therefore, high-level transcription may increase access to naked DNA and thus subvert kinetic or energetic barriers for subPAM interactions. Using MNase-seq and a *pZEVi-LYS2*$^{+}$ haploid yeast strain (see Methods), we confirmed that induction with 1000nM β-E for 7 h reduced median relative nucleosome occupancy at the *LYS2* locus to significantly lower levels than 15nM (~ 58% and 84% of the 0nM median, respectively; Extended Data Figs. 7e–f). Under these conditions, *LYS2* RNA levels were induced more than 100-fold (Extended Data Fig. 7g).

Second, transcribing RNA polymerases (RNAPs) might dislodge native Cas9 from cleaved protospacers in the PAM-downstream orientation employed, thereby facilitating repair outcomes[70]. Relative to the direction of transcription, this corresponds to an sgRNA spacer that anneals to the RNAP's template strand. High transcription can thus ensure frequent Cas9 turnover and rapid processing of cleaved DNA, whereas weak transcription may allow Cas9 to remain bound for longer periods of time and thereby delay processing. When we tested a protospacer in the inverted (PAM-upstream) orientation with 1 or 3 upstream *boxB* HTs using our original *pLYS2* haploid background (Extended Data Fig. 8a), TraCT was virtually undetectable (Extended Data Fig. 8b). Control experiments with Cas9-scTetR confirmed that the inverted protospacer orientation does not limit MT3 activity if the *tetO2* element is also relocated upstream (Extended Data Figs. 8c–d). Importantly, baseline editing activity with native Cas9 was unaltered by these inversions. In other words, weak transcription from *pLYS2* does not detectably stimulate native Cas9^{MT3} or Cas9^{NG} editing outcomes with one protospacer

orientation relative to the other, though it does stimulate Cas9^{MT3}-2xN or Cas9^{NG}-2xN via boxB with the PAM-downstream protospacer orientation (Supplementary Note 4).

TraCT can function in human cells

We also tested whether TraCT can stimulate indel formation through donor-independent non-homologous end-joining (NHEJ) in EGFP reporter lines derived from human U2OS osteosarcoma cells. To this end, two *boxB* HTs were inserted into either the 5' UTR of the *EGFP* marker cassette or an intron within the *EGFP* ORF, and downstream protospacers with NGA or NAG subPAMs were each targeted with one of four sgRNAs (Fig. 4a). When the U2OS.EGFP control line lacking *boxB* HTs was co-transfected with transient expression plasmids for Cas9 and sgRNA, we measured lower indel formation with an alternative Wt Cas9-2xN fusion (Cas9-2xN$^{Wt\text{-}alt}$) relative to the isogenic native Cas9$^{Wt\text{-}alt}$ architecture (Fig. 4b; Extended Data Fig. 1c; Supplementary Note 1). Despite this apparent impairment in the absence of *boxB*, indel formation by Cas9-2xN$^{Wt\text{-}alt}$ was significantly higher (1.61- to 2.07-fold) relative to Cas9$^{Wt\text{-}alt}$ with three out of four sgRNAs in the *boxB*-containing lines (Fig. 4b), indicating that TraCT can function in human cells.

TraCT requires a direct Cas9-RBD linkage

An assumption of our proposed model for facilitated recruitment is that direct physical interactions between Cas9 and nascent RNA, as afforded by the covalent linker used in our Cas9-RBD fusions, may be essential for engineered TraCT. To investigate this possibility and further refine our engineering guidelines, we expressed RBDs in *trans* to

test whether they suffice to stimulate native Cas9^{MT3} when a compatible HT is provided. We tested three different constitutive promoter strengths driving ectopic expression of a nuclear-localized mCherry-2xN (mC-2xN) fusion protein in a *LYS2* SSA reporter strain containing one *boxB* HT upstream of its protospacer insertion (Fig. 5a; Extended Data Figs. 9a–b). Editing with native Cas9^{MT3} was not appreciably stimulated at any mC-2xN expression strength (Fig. 5b). Importantly, we also observed a complete abrogation of Cas9^{MT3}-2xN's TraCT activity when expressing moderate (++) and high (+++) levels of mC-2xN (Fig. 5b) even though expression of mC-2xN had no effect on Wt editing levels (Extended Data Fig. 9c). The inhibitory effect of mC-2xN on Cas9^{MT3}-2xN likely arises from competitive binding at nascently transcribed boxB hairpins. Related experiments were conducted wherein the same three promoters were used to express nuclear-localized TetR proteins in *trans* with MT3 mutants of native Cas9 or Cas9-scTetR (Extended Data Figs. 9a, d). In contrast to results with mC-2xN, native Cas9^{MT3} activity in the absence of Dox was partially stimulated by *trans* TetR expression when compared with maximal Cas9^{MT3}-scTetR activity measured without *trans*-expressed TetR (Extended Data Fig. 9e–f; Supplementary Note 5). Altogether, these results demonstrate that *trans*-expressed RBDs are insufficient to compensate for native Cas9's subPAM interactions near *cis*-acting HTs, though *trans*-expressed TetR can provide partial stimulation through a proximal *tetO2* binding site. They furthermore indicate that physical linkages between RBDs and Cas9, or scTetR and Cas9, were essential for engineered TraCT activity or maximal *tetO2*-dependent activity, respectively.

Cas9-RBDs can exploit anchored hairpins

Postulating that engineered TraCT functions at least partially through an RBD-facilitated recruitment mechanism, we next examined whether Cas9-RBD fusions can compensate for subPAM interactions by binding RNA hairpins anchored in *cis*, rather than strictly relying on HTs transcribed in *cis*. To test this, we constitutively localized *trans*-expressed boxB hairpins near a protospacer that contains a subPAM but lacks *cis*-acting HTs. This was achieved with a dual sgRNA system for Cas9^{Wt}-2xN and two closely paired protospacers, in which one contains a suboptimal NGA PAM and the other contains an NGG PAM that can be targeted with partially mismatched spacers which license constitutive binding but not cleavage (Fig. 6a). An array of three boxB hairpins was fused to the 3' end of one sgRNA scaffold (Supplementary Note 2) while the other was left unmodified, and spacers for each target site were then tested in either scaffold. In the first configuration, hairpins are fused to the NGG-targeting sgRNA with spacer 1 (Fig. 6a); in the second, hairpins are fused to the NGA-targeting sgRNA with spacer 2 (Extended Data Fig. 10a). Either configuration could in principle stimulate Cas9^{Wt}-2xN recruitment to the NGA PAM via constitutive interactions at the NGG-adjacent protospacer. Indeed, the 2xN fusion robustly stimulated editing with each sgRNA configuration when 5 mismatches (Mm), but not 20 Mm that completely abrogate binding, were present in spacer 1 (Fig. 6b; Extended Data Fig. 10b; Supplementary Note 6). We also tested modified sgRNA scaffolds with only one or two boxB hairpins and this revealed smaller but still significant increases in editing with Cas9-2xN when spacer 1 contained 5 Mm (Extended Data Figs. 10c–d). These results establish that nascently transcribed RNA substrates are not strictly required to facilitate

Cas9-RBD targeting at subPAMs, as even *trans*-expressed RNA substrates can facilitate targeting if they are recruited near the protospacer via Cas9-mediated R-loops. However, anchored boxB hairpins are not strictly sufficient to stimulate targeting either, as inversion of the NGA-linked protospacer eliminated the 2xN fusion's advantage over native $Cas9^{Wt}$ (Extended Data Figs. 10e–f; Supplementary Note 6). It should moreover be emphasized that the strong NGG PAM we exploited for R-loop formation in this assay is expected to constitutively stimulate targeting at the subPAM, rather than license transcription-associated targeting.

Discussion

We report here that spatiotemporal control over Cas9 targeting in eukaryotes can be achieved using engineered Cas9 fusion proteins that conditionally interact with specific DNA or RNA substrates in proximity to protospacers. Unlike other engineered strategies for inducible Cas9 targeting that modulate the expression or activity of Cas9's core domains and sgRNA components[21-29], our approach offers conditional targeting activity only at protospacers with subPAM interactions and suitably positioned *cis* elements. Hence, our Wt PID fusions were fully active and targeted strong PAMs constitutively under all conditions tested. This implies that conditional *and* constitutive targeting could be simultaneously achieved at two unlinked protospacer sites if one has a subPAM and the other a strong PAM.

Central to this work, fusion of Cas9 to our scMCP or 2xN RBDs compensated for subPAM interactions through a TraCT mechanism that exploits *cis*-acting HTs. This activity is inherently conditional because the RNA hairpin substrates for these RBDs are

only transiently tethered near their HT DNA during active transcription in *cis.* Transcription of HTs in eukaryotes can thus serve as an endogenous physiological signal to control Cas9 targeting spatiotemporally, analogous to the role of protospacer transcription in DNA targeting by type III systems in prokaryotes[30-36]. TraCT activity scaled dynamically with the level of HT transcription and thereby generated allelic editing preferences; i.e., even identical targets in allelic loci could be partially distinguished based on their transcriptional activity. We anticipate that TraCT could be further applied as a genetic surveillance or tracking system in animal cell populations with heterogeneously transcribed DNA targets[71-75]. For example, stable expression of RBD fusions and sgRNAs in differentiating cells could selectively enrich for editing in subpopulations that upregulate transcription at HT-linked protospacers. The utility of these Cas9-RBD fusions for TraCT may prove limited outside of genetically engineered systems though, owing to their reliance on heterologous *cis*-acting HTs. Notwithstanding, natural eukaryotic systems also contain proteins capable of binding nascent RNA, such as splicing factors that are co-transcriptionally localized near the DNA templates of introns, splice junctions, or their downstream exons[76-78]. We therefore speculate that fusion of Cas9 to splicing factors could facilitate TraCT at such native sites. Relatedly, our yeast reporter system offers a generalizable two-hybrid platform for testing whether other putative or known protein-RNA interactions can function co-transcriptionally. It might also be possible to program TraCT virtually anywhere in the genome by targeting pre-mRNA with antisense RNAs fused to boxB hairpins[79,80] for Cas9-2xN recruitment, or, fused directly to Cas9's sgRNA scaffold. Alternatively, fusion of Cas9 to an RNA-targeting (d)Cas13 effector protein[81-83] from type VI CRISPR-Cas

systems[3] might prove effective. Whether (d)Cas13 complexes are capable of binding target RNAs co-transcriptionally remains to be elucidated, however.

In light of prior work and findings herein, several lines of evidence offer insight into the molecular details underlying engineered TraCT in eukaryotes. Nucleosomal DNA is thought to principally impede Cas9 targeting at the PAM-recognition step[69,84], and DNA binding sites for proteins that displace nucleosomes in vivo have been shown to locally stimulate targeting[85]. Active transcription was also confirmed to reduce nucleosome occupancy in our system (Extended Data Figs. 7e–f), though it need not evict histones from DNA entirely[86]. Hence, we speculate that recruitment of Cas9-RBD fusions to transcribing RNAPs, via tethered nascent RNA, provides temporal windows during which Cas9 can rapidly interrogate naked PAM DNA in *cis*. Such effects would transiently overcome the kinetic and/or energetic limitations of subPAM interactions in eukaryotes. In other words, TraCT might function through a dynamic hit-and-run mechanism that relies on concomitant Cas9 recruitment and exposure of target DNA by RNAPs. This two-pronged model could explain why upstream HTs offered a pronounced advantage over downstream HTs; downstream HTs will only yield RNA hairpin substrates after an RNAP has already transcribed beyond the target, at which point target DNA may once again be occluded within nucleosomes[86]. In contrast, hairpins constitutively anchored downstream using our paired-protospacer strategy robustly stimulated editing by Cas9-RBDs (Fig. 6), beyond the proxy-CRISPR effect[87] proposed to result from DNA exposure alone. This qualitatively mirrored our results comparing Cas9^{MT3}-scTetR to native Cas9^{MT3} with TetR (Extended Data Fig. 9e) and ostensibly supports our two-pronged TraCT model. Interactions with static *tetO2*

elements or NGG-adjacent mismatched protospacers licensed an advantage over native Cas9 only when positioned on the PAM side of the cleavable protospacer, however (Extended Data Figs. 8c–d; Fig. 6; Extended Data Figs. 10e–f). This might reflect Cas9's reported requirement for a sequence-independent 'post-PAM' interaction at PAM-distal DNA[88], which, should be nucleosome-depleted in these configurations. If we furthermore surmise that transcription could dynamically expose DNA for the post-PAM interaction, their relative directionalities should also be considered when interpreting our result that TraCT was influenced by protospacer orientation at a weakly transcribed locus (Extended Data Figs. 8a–b). Additional experiments will be required to dissect these mechanistic points and investigate whether transcription-associated DNA targeting can be achieved with other CRISPR-Cas effectors in eukaryotes.

References

1 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

2 Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911 (2018).

3 Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *CRISPR J* **1**, 325-336 (2018).

4 Kantor, A., McClements, M. E. & MacLaren, R. E. CRISPR-Cas9 DNA Base-Editing and Prime-Editing. *Int J Mol Sci* **21** (2020).

5 Scholefield, J. & Harrison, P. T. Prime editing - an update on the field. *Gene Ther.* **28**, 396-401 (2021).

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

7 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579-2586 (2012).

8 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

9 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).

10 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390-1400 (2008).

11 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).

12 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233-239 (2013).

13 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113 (2015).

14 Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027 (2017).

15 Boyle, E. A. *et al.* Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. *Sci Adv* **7** (2021).

16 Pacesa, M. *et al.* Structural basis for Cas9 off-target activity. *Cell* **185**, 4067-4081 e4021 (2022).

17 Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191-196 (2022).

18 Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).

19 Goldberg, G. W. *et al.* Engineered dual selection for directed evolution of SpCas9 PAM specificity. *Nat Commun* **12**, 349 (2021).

20 Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390-394 (2015).
21 Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139-142 (2015).
22 Liu, K. I. *et al.* A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12**, 980-987 (2016).
23 Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646-651 (2016).
24 Kundert, K. *et al.* Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat Commun* **10**, 2127 (2019).
25 Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **137**, 5642-5645 (2015).
26 Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755-760 (2015).
27 Jain, P. K. *et al.* Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors. *Angew Chem Int Ed Engl* **55**, 12440-12444 (2016).
28 Liu, Y. *et al.* Very fast CRISPR on demand. *Science* **368**, 1265-1269 (2020).
29 Oakes, B. L. *et al.* CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification. *Cell* **176**, 254-267 e216 (2019).
30 Samai, P. *et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**, 1164-1174 (2015).
31 Estrella, M. A., Kuo, F. T. & Bailey, S. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes Dev.* **30**, 460-470 (2016).
32 Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C. & Siksnys, V. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol. Cell* **62**, 295-306 (2016).
33 Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633-637 (2014).
34 Jiang, W., Samai, P. & Marraffini, L. A. Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**, 710-721 (2016).
35 Goldberg, G. W. *et al.* Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts. *Nature Communications* **9**, 61 (2018).
36 Rostol, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nat Microbiol* **4**, 656-662 (2019).
37 Sugawara, N., Ira, G. & Haber, J. E. DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Mol. Cell. Biol.* **20**, 5300-5309 (2000).
38 Bhargava, R., Onyango, D. O. & Stark, J. M. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet.* **32**, 566-575 (2016).

39 Bolukbasi, M. F. *et al.* DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* **12**, 1150-1156 (2015).
40 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).
41 Ichikawa, D. M. *et al.* A universal deep-learning model for zinc finger design enables transcription factor reprogramming. *Nat. Biotechnol.* (2023).
42 Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr. Gene Ther.* **16**, 156-167 (2016).
43 Krueger, C., Berens, C., Schmidt, A., Schnappinger, D. & Hillen, W. Single-chain Tet transregulators. *Nucleic Acids Res.* **31**, 3050-3056 (2003).
44 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262 (2018).
45 Lowary, P. T. & Uhlenbeck, O. C. An RNA mutation that increases the affinity of an RNA-protein interaction. *Nucleic Acids Res.* **15**, 10483-10493 (1987).
46 Valegard, K. *et al.* The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* **270**, 724-738 (1997).
47 Tutucci, E. *et al.* An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **15**, 81-89 (2018).
48 Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796-806 (2007).
49 Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat. Methods* **10**, 119-121 (2013).
50 Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623-626 (1994).
51 Ni, C. Z. *et al.* Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* **3**, 255-263 (1995).
52 Weisberg, R. A. & Gottesman, M. E. Processive antitermination. *J. Bacteriol.* **181**, 359-367 (1999).
53 Lange, S. *et al.* Simultaneous transport of different localized mRNA species revealed by live-cell imaging. *Traffic* **9**, 1256-1267 (2008).
54 Wilson, H. R., Yu, D., Peters, H. K., 3rd, Zhou, J. G. & Court, D. L. The global regulator RNase III modulates translation repression by the transcription elongation factor N. *EMBO J.* **21**, 4154-4161 (2002).
55 Wilson, H. R., Zhou, J. G., Yu, D. & Court, D. L. Translation repression by an RNA polymerase elongation complex. *Mol. Microbiol.* **53**, 821-828 (2004).
56 Conant, C. R., Goodarzi, J. P., Weitzel, S. E. & von Hippel, P. H. The antitermination activity of bacteriophage lambda N protein is controlled by the kinetics of an RNA-looping-facilitated interaction with the transcription complex. *J. Mol. Biol.* **384**, 87-108 (2008).
57 Said, N. *et al.* Structural basis for lambdaN-dependent processive transcription antitermination. *Nat Microbiol* **2**, 17062 (2017).

58 Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289-299 (1998).
59 Mogridge, J. *et al.* Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol. Cell* **1**, 265-275 (1998).
60 Austin, R. J., Xia, T., Ren, J., Takahashi, T. T. & Roberts, R. W. Designed arginine-rich RNA-binding peptides with picomolar affinity. *J. Am. Chem. Soc.* **124**, 10966-10967 (2002).
61 Whelan, W. L., Gocke, E. & Manney, T. R. The CAN1 locus of Saccharomyces cerevisiae: fine-structure analysis and forward mutation rates. *Genetics* **91**, 35-51 (1979).
62 McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57 (2013).
63 McIsaac, R. S., Gibney, P. A., Chandran, S. S., Benjamin, K. R. & Botstein, D. Synthetic biology tools for programming gene expression without nutritional perturbations in Saccharomyces cerevisiae. *Nucleic Acids Res.* **42**, e48 (2014).
64 Datta, A. & Jinks-Robertson, S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616-1619 (1995).
65 Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127-131 (2013).
66 DiCarlo, J. E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336-4343 (2013).
67 Satomura, A. *et al.* Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci Rep* **7**, 2095 (2017).
68 Schwabish, M. A. & Struhl, K. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **24**, 10111-10117 (2004).
69 Hinz, J. M., Laughery, M. F. & Wyrick, J. J. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* **54**, 7063-7066 (2015).
70 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 e48 (2018).
71 Johnston, R. J., Jr. *et al.* Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**, 956-968 (2011).
72 Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A. & Dougherty, J. D. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J. Neurosci.* **34**, 1420-1431 (2014).
73 Monahan, K. & Lomvardas, S. Monoallelic expression of olfactory receptors. *Annu. Rev. Cell. Dev. Biol.* **31**, 721-740 (2015).
74 Symmons, O. *et al.* Allele-specific RNA imaging shows that allelic imbalances can arise in tissues through transcriptional bursting. *PLoS Genet.* **15**, e1007874 (2019).
75 Xu, P. *et al.* The landscape of human tissue and cell type specific expression and co-regulation of senescence genes. *Mol Neurodegener* **17**, 5 (2022).

76 Kotovic, K. M., Lockshon, D., Boric, L. & Neugebauer, K. M. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol. Cell. Biol.* **23**, 5768-5779 (2003).
77 Gornemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* **19**, 53-63 (2005).
78 Lacadie, S. A. & Rosbash, M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol. Cell* **19**, 65-75 (2005).
79 Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci U S A* **110**, 18285-18290 (2013).
80 Montiel-Gonzalez, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res.* **44**, e157 (2016).
81 Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284 (2017).
82 Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027 (2017).
83 Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676 e614 (2018).
84 Hinz, J. M., Laughery, M. F. & Wyrick, J. J. Nucleosomes Selectively Inhibit Cas9 Off-target Activity at a Site Located at the Nucleosome Edge. *J. Biol. Chem.* **291**, 24851-24856 (2016).
85 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 U S A* **115**, 9351-9358 (2018).
86 Filipovski, M., Soffers, J. H. M., Vos, S. M. & Farnung, L. Structural basis of nucleosome retention during transcription elongation. *Science* **376**, 1313-1316 (2022).
87 Chen, F. *et al.* Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* **8**, 14958 (2017).
88 Zhang, Q. *et al.* The post-PAM interaction of RNA-guided spCas9 with DNA dictates its target binding and dissociation. *Sci Adv* **5**, eaaw9807 (2019).
89 Gu, X. *et al.* SAMTOR is an S-adenosylmethionine sensor for the mTORC1 pathway. *Science* **358**, 813-818 (2017).
90 Faedo, A. *et al.* Differentiation of human telencephalic progenitor cells into MSNs by inducible expression of Gsx2 and Ebf1. *Proc Natl Acad Sci U S A* **114**, E1234-E1242 (2017).
91 Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115-132 (1998).
92 Gietz, R. D. Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol Biol* **1205**, 1-12 (2014).

93 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).
94 Xie, Z. X. *et al.* Rapid and Efficient CRISPR/Cas9-Based Mating-Type Switching of Saccharomyces cerevisiae. *G3 (Bethesda)* **8**, 173-183 (2018).
95 Agmon, N. *et al.* Yeast Golden Gate (yGG) for the Efficient Assembly of S. cerevisiae Transcription Units. *ACS Synth Biol* **4**, 853-859 (2015).
96 Ikushima, S., Zhao, Y. & Boeke, J. D. Development of a Tightly Controlled Off Switch for Saccharomyces cerevisiae Regulated by Camphor, a Low-Cost Natural Product. *G3 (Bethesda)* **5**, 1983-1990 (2015).
97 Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).
98 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95 (2007).
99 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261-272 (2020).
100 Kubik, S. *et al.* Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. *Mol. Cell* **60**, 422-434 (2015).
101 Truong, D. M. & Boeke, J. D. Resetting the Yeast Epigenome with Human Nucleosomes. *Cell* **171**, 1508-1519 e1513 (2017).
102 Chen, K. *et al.* DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341-351 (2013).
103 Kogenaru, M. & Isalan, M. Drug-Inducible Control of Lethality Genes: A Low Background Destabilizing Domain Architecture Applied to the Gal4-UAS System in Drosophila. *ACS Synth Biol* **7**, 1496-1506 (2018).

Supplementary Information
Supplementary Notes 1-6.
Supplementary References
Supplementary Data 1-6 (separate files).

Author Contributions
G.W.G. conceived the study. G.W.G. designed experiments with help from J.D.B, M.A.B.H., L.K., and M.K. G.W.G., M.A.B.H., L.K., M.K., and K.O. generated biological reagents and/or conducted experiments. S.K. analyzed high-resolution growth curve data and colony size data. M.A.B.H. analyzed MNase-seq data. M.A.A. ran Seqtk processing for single-read NextSeq data files. S.A., J.D.B., D.F., M.A.A., and M.B.N. contributed other resources and/or technical advice. J.D.B. and M.B.N. provided funding. G.W.G. wrote the paper with help from J.D.B., M.A.B.H., L.K., S.K., and M.K.

Acknowledgements
We thank former Noyes lab members David Ichikawa and Jeff Spencer, as well as Ran Brosh and other current or former members of the Boeke and Noyes labs, for sharing helpful discussions and resources. Paolo Mita provided the pPM193 plasmid template

from the Boeke lab for amplifying the Wt TetR PCR fragment of scTetR. Mike Shen (Boeke lab alumnus) designed the pMS101 plasmid in collaboration with the Noyes lab. Makiha Fukuda of the Boeke lab provided *TAF10* primers for qRT-PCR. We also thank Addgene for providing plasmid templates to amplify a recoded TetR PCR fragment and an isogenic RevTetR PCR fragment: the pCW57.1-MAT2A plasmid[89] (#100521; RRID:Addgene_100521) and the pCAG-TetON-3G plasmid[90] (#96963; RRID:Addgene_96963), respectively, gifts from David Sabatini's lab and Elena Cattaneo's lab. The BPK1520 plasmid[18] was likewise acquired through Addgene (#65777; RRID:Addgene_65777) and is a gift from Keith Joung's lab. We gratefully acknowledge NYU Medical Center's Research Support Service team for preparation of various reagents and yeast media used in this work, the Genome Technology Center (RRID: SCR_017929) for low-complexity amplicon sequencing on a MiSeq as well as library preparation and sequencing for MNase-seq on a NextSeq 2000, and the IDT rhAmpSeq CRISPR Analysis Tool team for their services and tech support. We are also grateful for thoughtful discussions with Timothee Lionnet and other members of his lab, as well as Rachel Leicher from Shixin Liu's lab. G.W.G. was supported by an NRSA Postdoctoral Fellowship (F32GM137482) and an NIH K99 Pathway to Independence Award (K99GM147604), both from the NIGMS, as well as a Centers of Excellence in Genomic Science award (RM1HG009491) to J.D.B. from the NHGRI. Additional funding support was provided from NIGMS awards R01GM118851 and R01GM133936 to M.B.N.

Competing Interests

Jef Boeke is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to ReOpen Diagnostics, LLC and serves or served on the Scientific Advisory Board of the following: Logomix, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Sangamo, Inc., Tessera Therapeutics, Inc. and the Wyss Institute. Marcus Noyes is a founder of TBG Therapeutics, Inc. and consultant to Tessera Therapeutics, Inc. New York University is pursuing patent protection for findings described in this work, with Gregory Goldberg, Marcus Noyes, and Jef Boeke listed as inventors. The other authors declare no competing interests. Correspondence and requests for materials should be addressed to Jef.Boeke@nyulangone.org, Gregory.Goldberg@nyulangone.org, or Marcus.Noyes@nyulangone.org.

Main Figures

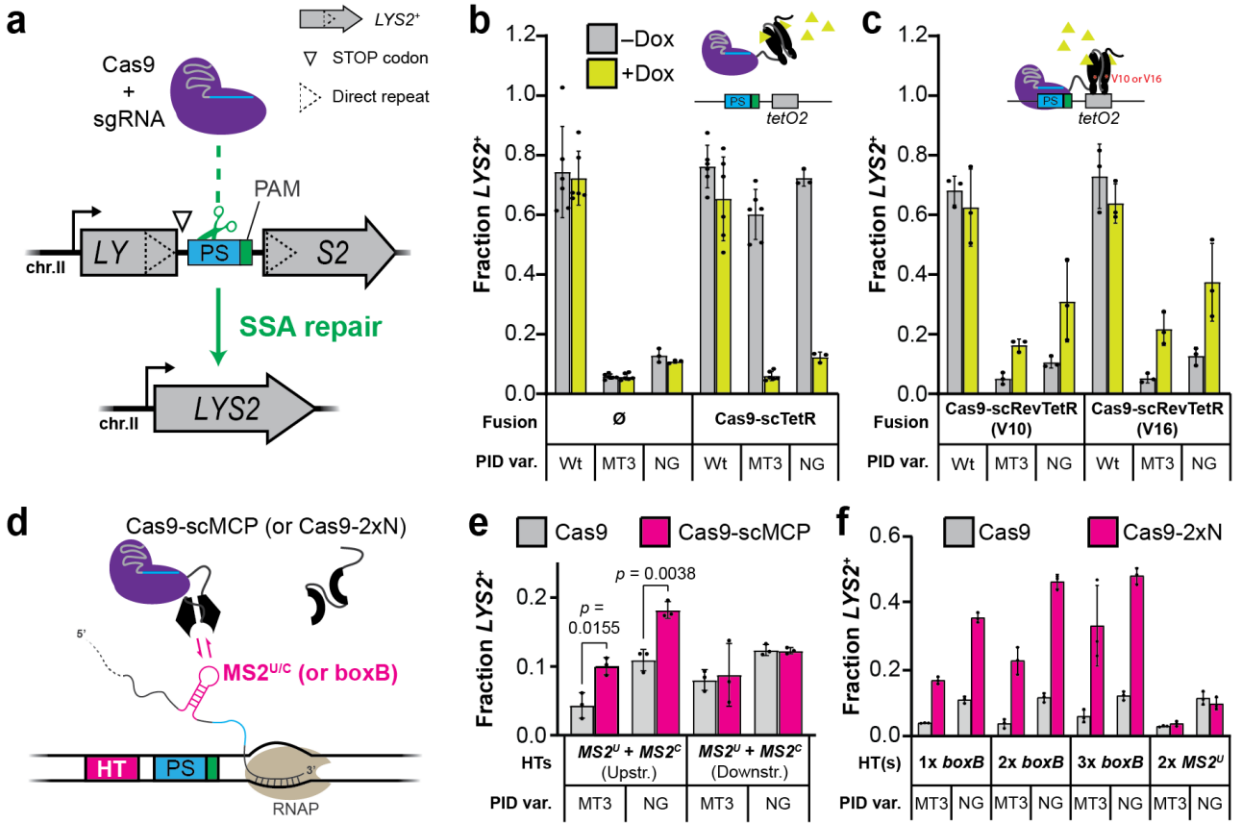


Fig. 1 | Cas9 activity can be controlled by conditional interactions between auxiliary fusion domains and protospacer-proximal *cis* elements. a, General schematic of the *LYS2* genetic reporter system, engineered with a heterologous protospacer (PS) insertion between 100 bp direct repeats. Co-expression of Cas9 with a targeting sgRNA licenses cleavage and deletion of the repeat-intervening sequences through SSA-based recombination that restores *LYS2*$^+$. The native *pLYS2* promoter (bent arrow) and an NGG PAM are used throughout the work, except where stated otherwise. **b**, Transformation-associated editing with the Cas9-scTetR fusion or native Cas9 control (Ø) in yeast containing a *tetO2* element 13 bp downstream of the PS. The edited fraction of *LYS2*$^+$ transformants (out of total transformants) was measured for Wt, MT3, and NG PID variants using synthetic dropout plates (see Methods) that lack (–) or contain (+) Dox. Inset schematic illustrates Dox-induced dissociation of the scTetR domain. Error bars, mean ± s.d. (n = 6 or n = 3, biological replicates). **c**, Same as in **b** but with Cas9-scRevTetR fusions as indicated (V10 or V16). Inset schematic illustrates Dox-induced binding of the scRevTetR domain. **d**, Simplified schematic illustration of the Cas9-scMCP fusion interacting with a nascently transcribed MS2$^{U/C}$ RNA hairpin, generated from a single Hairpin Template (HT) upstream of the PS. The scMCP domain can be replaced with a 2xN domain that in principle allows binding of two boxB hairpins simultaneously. The nascent transcript is elongated at its 3' end by an RNA polymerase (RNAP) transcribing left to right. **e**, Transformation-associated editing fractions measured for MT3 and NG variants of Cas9-scMCP or native Cas9 in strains containing

the indicated HTs. Error bars, mean ± s.d. (n = 3, biological replicates). **f**, Same experimental setup as in **e** but performed with the 2xN domain in place of scMCP and different HT strains as indicated.

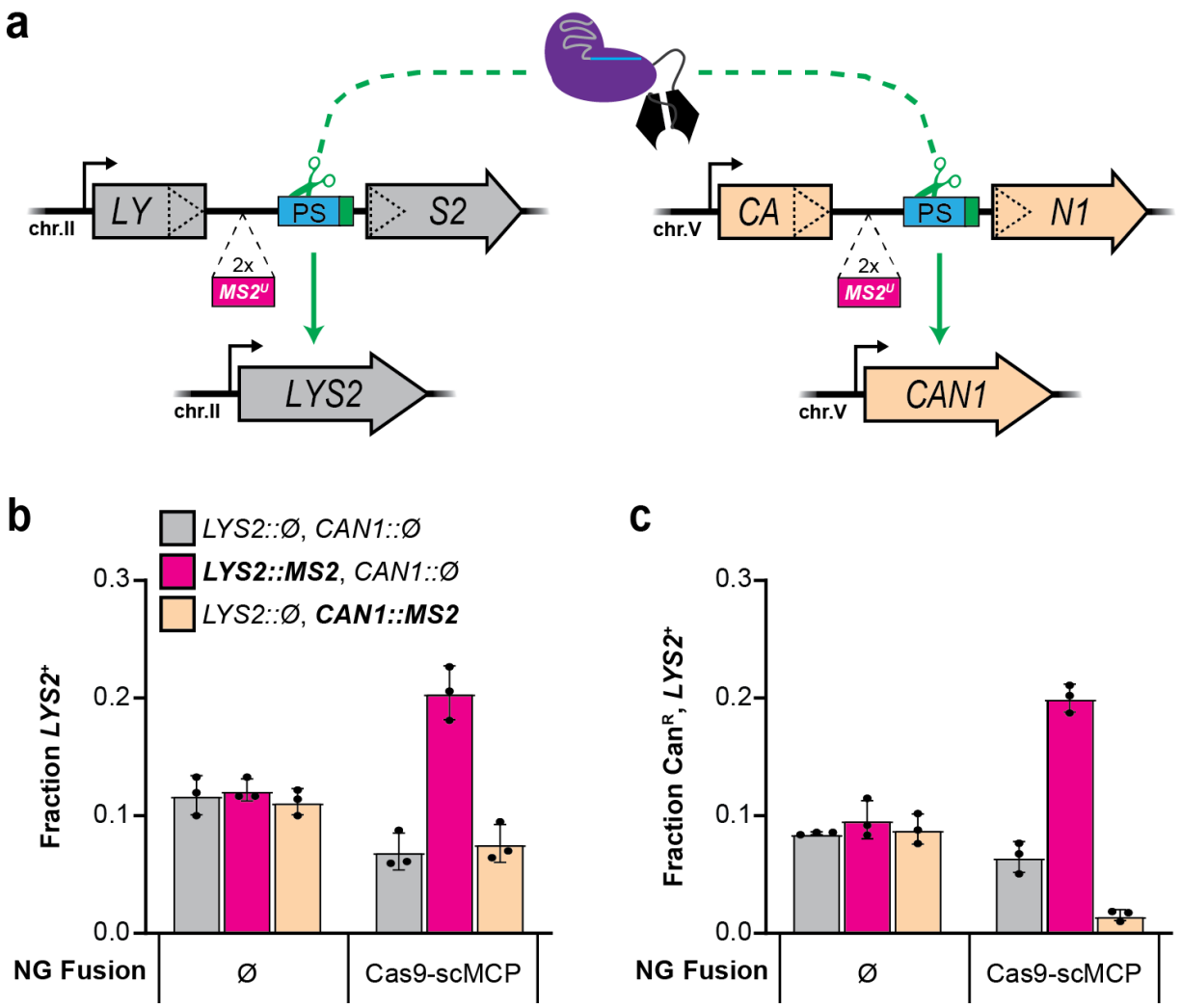


Fig. 2 | Nascent RNA binding is insufficient to compensate for subPAM interactions in *trans*. a, Schematic summary of the *LYS2* and *CAN1* dual-reporter system, engineered with identical protospacer and PAM insertions between direct repeats at each locus. SSA-based editing can occur via cleavage at either or both loci. Reconstitution of the *LYS2* gene confers lysine prototrophy as in Fig. 1a, whereas reconstitution of the *CAN1* gene confers canavanine sensitivity for counterselection; only cells that were unedited at their *CAN1* locus can form canavanine-resistant (Can^R) colonies. Strains contained either no HTs, or two upstream $MS2^U$ HTs inserted at one of their loci. The native constitutive promoters (bent arrows) were left unmodified at each locus. **b**, Transformation-associated editing fractions measured for $Cas9^{NG}$-scMCP or native $Cas9^{NG}$ in dual-reporter strains harboring (::MS2) or lacking (::Ø) the HT insertions at each locus. Measurements were performed as in Fig. 1e. Error bars, mean ± s.d. (n = 3, biological replicates). **c**, Editing measurements obtained with the same transformation samples tested in panel **b** except using canavanine-supplemented synthetic dropout plates that select only for Can^R, $LYS2^+$ colonies. Error bars, mean ± s.d. (n = 3, biological replicates).

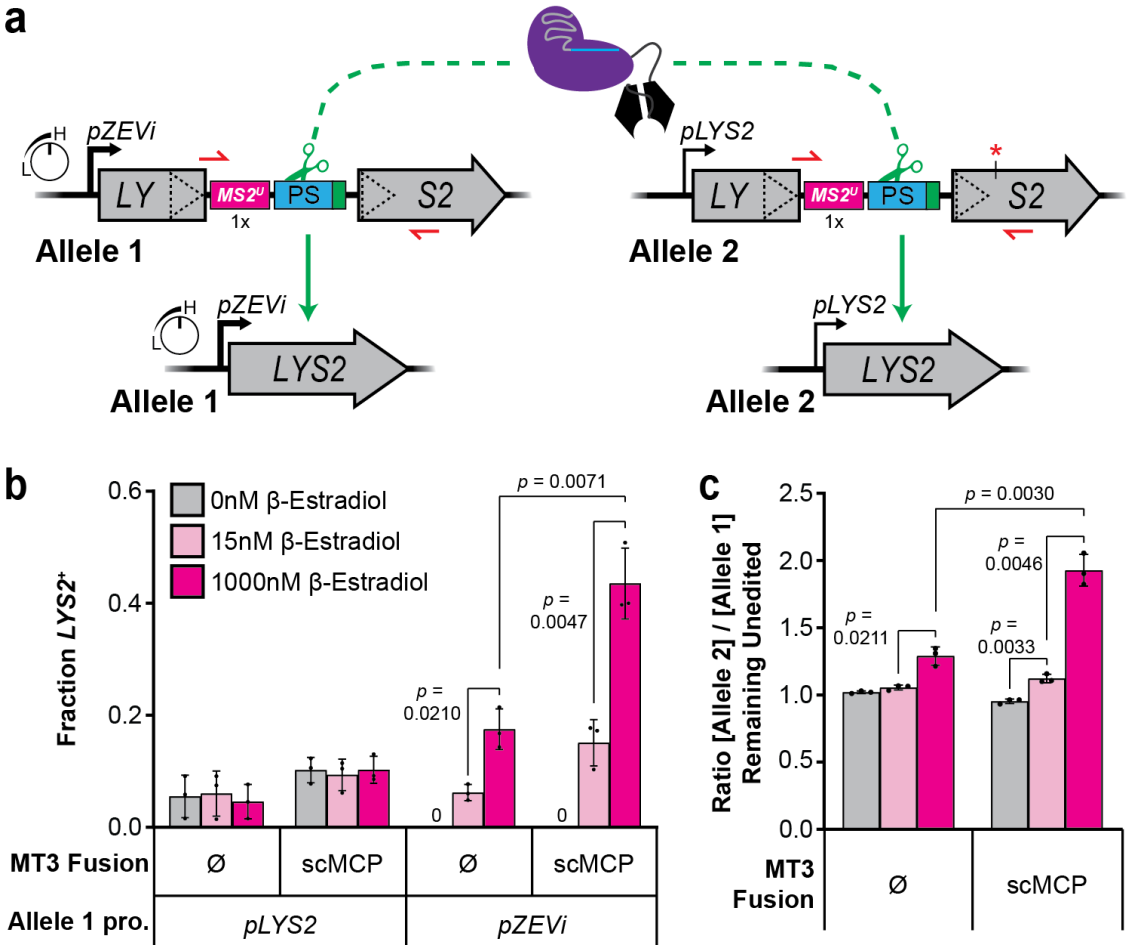


Fig. 3 | TraCT can be exploited for preferential editing at differentially transcribed alleles with identical target sequences. a, Schematic summary of the diploid SSA reporter system. Each allele differs only in its promoter sequences and a single synonymous point mutation (red asterisk) downstream of identical direct repeat and intervening sequences that disrupt *LYS2* as shown. Allele 1's *pZEVi* promoter can be highly induced with β-Estradiol (β-E) but is repressed in its absence; hence, Allele 1 cannot confer a *LYS2*$^{+}$ phenotype in the absence of β-E, even if reconstituted via SSA. **b**, Transformation-associated editing fractions measured for Cas9^{MT3}-scMCP or native Cas9^{MT3} in haploid strains containing only Allele 1 with the *pZEVi* or native *pLYS2* promoter. Measurements were performed as in Fig. 1e except that transformants were also plated in parallel on synthetic dropout plates supplemented with β-E at the indicated concentrations. Error bars, mean ± s.d. (n = 3, biological replicates). Zero (0) *LYS2*$^{+}$ colonies of the *pZEVi* strain were detected in the absence of β-E at the time of scoring. **c**, Allelic editing in diploid transformants with Cas9^{MT3}-scMCP or native Cas9^{MT3}, determined by deep sequencing PCR amplicons to distinguish between alleles via the point mutation shown in panel **a**. Amplicons were generated with a primer pair that specifically amplifies *unedited* alleles (depicted as small red half arrows in panel **a**). Data are plotted as the ratio of unedited Allele 2:Allele 1 genotypes measured at each β-E concentration (color coded as in panel **b**). Error bars, mean ± s.d. (n = 3, biological replicates).

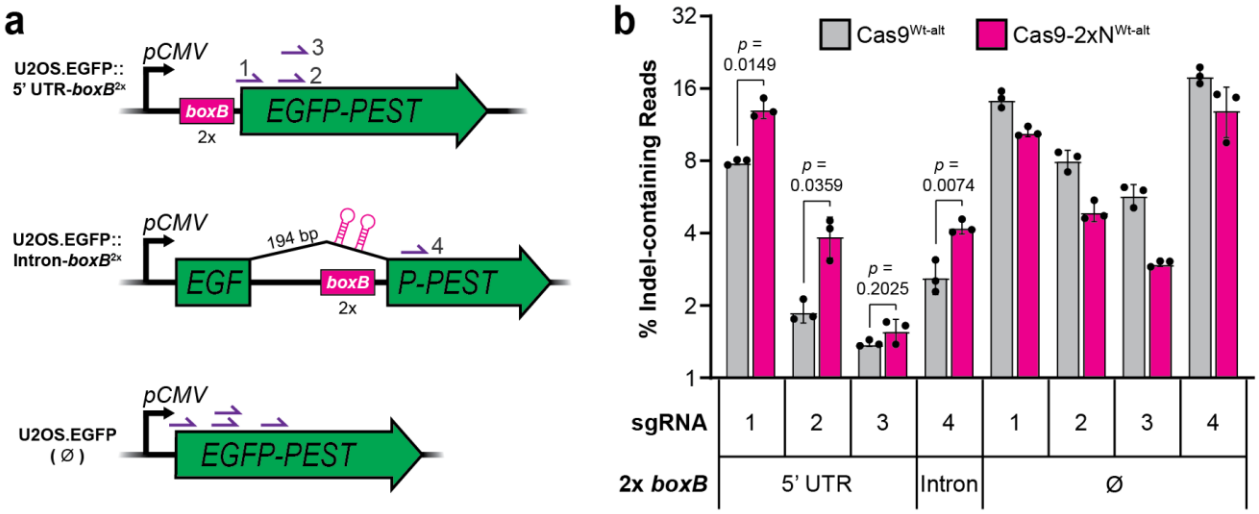


Fig. 4 | Engineered TraCT can stimulate indel formation in human cells. a, Schematic summary of the non-essential *EGFP-PEST* ORF targeted in human cell culture experiments with the U2OS.EGFP (Ø) parent line and its U2OS.EGFP::5' UTR-*boxB*^{2x} or U2OS.EGFP::Intron-*boxB*^{2x}(V1.0) derivatives. Relative positions of the four sgRNA targets (purple half arrows with numbering) and upstream 2x *boxB* HT insertions within the 5' UTR or 194 bp synthetic intron are indicated. Transcription is driven by the strong constitutive *pCMV* promoter in all cell lines. **b**, Deep sequencing of protospacer-spanning amplicons to quantify the percentage of reads that contain indels, inferred to result from donor-independent NHEJ. gDNA templates were extracted 48 h after co-delivering one of four targeting sgRNA plasmids (1 through 4) and the Cas9-2xN$^{Wt\text{-}alt}$ fusion or control Cas9$^{Wt\text{-}alt}$ plasmid into the indicated cell lines. The PAMs CGAGG, CGAGC, GAGCT, and TGACC license targeting for sgRNAs 1 through 4, respectively. Error bars, mean ± s.d. (n = 3, biological replicates).

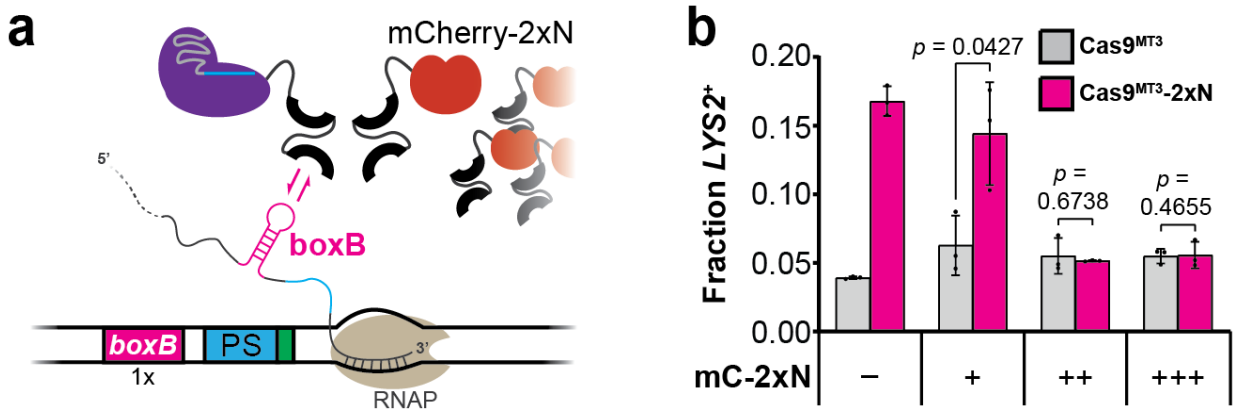


Fig. 5 | 2xN peptides expressed independently of Cas9 or Cas9-2xN fusions do not compensate for *boxB*-proximal subPAM interactions and can limit TraCT. a, Simplified schematic illustration similar to Fig. 1d, except that a Cas9-2xN fusion is shown co-expressed with an mCherry-2xN (mC-2xN) fusion in the same strain; either can interact with a nascently transcribed boxB RNA hairpin. **b**, Transformation-associated editing fractions measured for Cas9^{MT3}-2xN or native Cas9^{MT3} in SSA reporter strains with one *boxB* HT upstream of their protospacer and mC-2xN fusion proteins expressed in *trans* at the indicated levels (+/++/+++), or not at all (–). Measurements were performed as in Fig. 1f; data shown for the strain lacking mC-2xN (–) are replotted from Fig. 1f (1x *boxB*, MT3) to facilitate comparisons. Error bars, mean ± s.d. (n = 3, biological replicates).

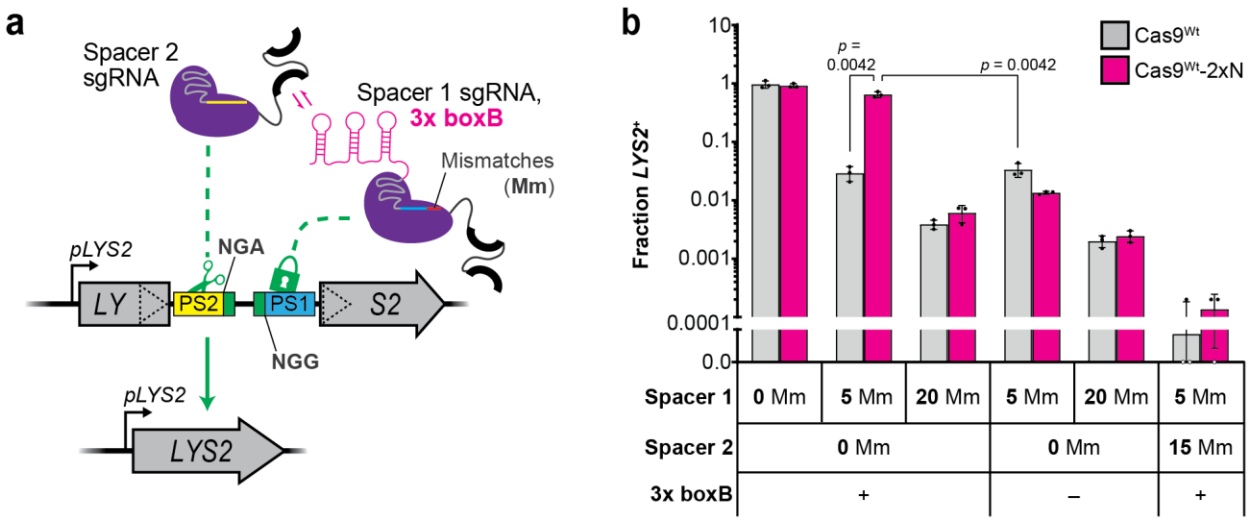


Fig. 6 | Constitutive R-loop formation with boxB-fused sgRNAs that bind DNA but do not license cleavage can stimulate Cas9^{Wt}-2xN targeting at nearby subPAMs. **a**, Schematic summary of the paired-protospacer SSA reporter system with dual sgRNA programming for Cas9^{Wt}-2xN. Three boxB hairpins are fused to the Spacer 1 sgRNA targeting Protospacer 1 (PS1) but not fused to the Spacer 2 sgRNA targeting Protospacer 2 (PS2). Mismatches (Mm) at the 5' end of Spacer 1 (depicted as the red segment) do not prevent binding but prevent cleavage at PS1. The PAM sequences at each target are denoted 5' to 3' irrespective of their orientation. **b**, Transformation-associated editing fractions measured for Cas9^{Wt}-2xN or native Cas9^{Wt} in the SSA reporter strain summarized in panel **a**. Various combinations of dual sgRNAs with (+) or without (–) the 3x boxB fusion were delivered via plasmid expression vectors. The number of Mm present in each spacer, relative to their intended protospacer targets, is indicated. Error bars, mean ± s.d. (n = 3, biological replicates).

Methods

Microbial strains and growth conditions. Cultivation of *Escherichia coli* strains XL1-Blue or TOP10 was carried out as previously described[19] for petri dish plates or 5mL liquid cultures. Media were supplemented with carbenicillin (100 µg/mL) and/or kanamycin (50 µg/mL) for plasmid cloning and maintenance. When chemically competent or electrocompetent XL1-Blue cells were prepared for downstream cloning, media was supplemented with tetracycline (15 µg/mL) alone.

The *S. cerevisiae* yeast strains used in this work were all derived from BY4741 or BY4731 haploids[91]; their genotypes are provided in Supplementary Data 2. Except where noted otherwise, all liquid yeast culture and competent cell preparation was performed sterilely with YEPD broth or synthetic complete (SC) dropout broths lacking one or more amino acid (SC–) in 18 x 150 (for 5mL volumes) or 25 x 150 mm (for 20mL volumes) disposable round-bottom glass tubes as described previously[19]. This included our adaption of the high-efficiency lithium acetate transformation protocol[92]. Briefly, 5mL YEPD or SC starter cultures were first inoculated from single colonies and grown for 24 h overnight. After subculturing overnight starter cultures by diluting to ~ 0.15 Attenuance ($D_{600\ nm}$) in 20mL of the same broth (1:100 for YEPD or 1:70 for SC) and allowing 4.5 h (YEPD) or 5.25 h (SC) growth on a roller drum at 30 °C (~ 0.6 $D_{600\ nm}$), cells were harvested by centrifugation at 2103 x g for 3 min and washed once with resuspension in an equal volume of sterile double-distilled water (ddH_2O). After washing a second time with resuspension in an equal volume of filter-sterilized lithium acetate (0.1M) and re-pelleting, cells were transferred to 1.5mL tubes in their residual supernatant, centrifuged again, and resuspended to a final volume of 200µl in 0.1M Lithium Acetate. Each 20mL subculture yielded four transformation reaction aliquots; when scaling up for more transformations, additional subcultures were grown in parallel and cells were pooled in 1.5mL tubes at this step. A modified transformation mix (TRAFO mix)[92] was prepared fresh in parallel, and 50µl of cells were combined with 300µl of TRAFO mix in 1.5mL tubes for each transformation reaction. DNA (≤ 20µl in ddH_2O or buffer EB from Qiagen) was then added directly to this mixture, and tubes were homogenized by vortexing before transferring to a roller drum at 30 °C. After rotating for 30 min, 36µl pure DMSO was added to each tube, and tubes were re-homogenized by vortexing and transferred to a 42 °C heat block. After incubating for 15 min, cells were pelleted by centrifugation and resuspended in filter-sterilized 5mM $CaCl_2$ (400µl, unless noted otherwise) and then incubated at room temperature (r.t.) for 10 min. After 10 min at r.t., resuspended cells were spread directly on solid media (150-200µl for 100 x 15 mm plates) using sterile 4mm glass beads (Fisher), or spotted as droplets (5µl on YEPD or 10µl on SC) from serial dilutions prepared with sterile ddH_2O. Our fresh TRAFO mixes were homogenized from 240µl PEG 3350 (44%), 36µl 1M Lithium Acetate, and 25µl Promega Herring Sperm DNA (10 mg/mL) per reaction by vortexing vigorously in a 50mL sterile falcon tube. Immediately before preparing a TRAFO mix, pre-aliquoted Herring Sperm DNA was thawed and resuspended by vortexing vigorously, boiled on a heat block at 95 degrees for 5 minutes, and then chilled on ice for 5 minutes. To select for transformation-associated integration of DNA fragments linked to a nourseothricin-resistance (Nat^R) resistance marker, cells were plated on solid media supplemented with clonNAT (100 ug/mL) after recovering half the transformant suspension for 16-24 h

in the same clonNAT-free liquid media that was used to prepare their parental competent cells. ClonNAT was generally maintained during Nat^R strain re-streaks but not included in the media during other assay procedures. For all experimental yeast assays in this work, biological replicates were initiated by launching liquid starter cultures from fresh single colonies grown on SC dropout plates at 30 °C for 48 h after streaking. Except where noted otherwise, SC media formulations (Sunrise Science) with leucine and/or uracil dropped out were used throughout the work to select for maintenance of the Cas9 and/or sgRNA episomes, respectively. Slight colony size reductions were observed when selecting for strains that express $Cas9^{Wt}$-scMCP with or without a sgRNA (Extended Data Fig. 5c), suggesting low-level toxicity from expression.

DNA preparation and cloning. All circular plasmid DNA used in this work was purified from saturated 5mL liquid cultures of *E. coli* (grown for 16 h at 37 °C as described previously[19]) using Qiagen miniprep reagents. Centrifugation at 4000 x *g* for 6 min was used to pellet the cultures before miniprepping. Plasmid identifiers and descriptions are provided in Supplementary Data 3. All plasmid-derived linear DNA fragments used for cloning, yeast strain construction, or transformation-associated editing experiments were isolated from miniprepped DNA via restriction digest and gel purification. PCR amplicons used for cloning, CRISPR-assisted knock-in, or Illumina sequencing were also gel-purified. For cloning and CRISPR-assisted knock-ins, amplicons were generally obtained using high-fidelity Q5 Hot Start (2X Master Mix) or Phusion polymerases from NEB; otherwise, GoTaq Green (Promega) was used for fragments up to ~ 2 kb that could not be obtained with Q5 or Phusion. GoTaq was also used for all PCR amplification from genomic DNA (gDNA) templates, including yeast colony PCRs as described previously[19]. Cycling conditions recommended by the manufacturer were usually followed, except that 90 s of annealing was used during each cycle when amplifying from gDNA with GoTaq. All Wt, MT3, and NG PID fragments used for assembly with gel-purified acceptor vector backbone fragments in transformation-associated editing experiments were amplified from plasmid templates using Q5 or Phusion with oligos oGG1019/oGG1052 (Supplementary Data 4) and then gel-purified. Gel purifications were typically carried out with disposable x-tracta tools (Sigma-Aldrich) and MinElute kits (Qiagen). In addition to PCR- or plasmid-derived fragments, commercially synthesized DNA fragments (e.g., gBlocks from IDT) were also used directly in cloning procedures.

The majority of plasmid cloning in this work was carried out with Gibson[93] or NEBuilder HiFi (NEB) master mixes for isothermal assembly of linear DNA fragments at 50 °C for 1 h. In some cases these assemblies included a single-stranded DNA bridging oligo for introducing small inserts, such as sgRNA spacers. Otherwise, cloning was achieved via one-pot restriction-ligation reactions with type IIS restriction enzymes (NEB or Thermo Scientific) and T4 DNA ligase (NEB). These 20µl reactions included circular plasmid DNA (destination vector) and annealed oligos with appropriate overhangs for ligation, 1µl type IIS restriction enzyme, 1µl T4 ligase, and buffers (CutSmart and T4 ligase buffers, each at 1X final concentrations); they were incubated at 37 °C for 4-24 h. In most cases XL1-Blue or TOP10 chemically competent cells were used for cloning and before plating were recovered into 300µl SOC broth for 1 h on a roller drum at 30 or 37 °C after the 45 s heat shock at 42 °C. If these did not yield clones initially,

electrocompetent XL1-Blue cells were occasionally used in accordance with dialysis and electroporation procedures described previously[19]. Following electroporation, transformants were recovered into 10mL SOC broth for 1 h on a roller drum at 30 or 37 °C before plating. At a minimum, all cloned sequences derived from PCR amplicons or commercially synthesized DNA and their adjacent backbone-derived sequences at recombinant junctions were verified by Sanger sequencing. In some cases, entire plasmids were sequenced with long-read sequencing by Plasmidsaurus. The expected sequences of all plasmids used in this work are provided in Supplementary Data 3.

All Cas9 acceptor vectors were cloned in *E. coli* and served as the source of backbone fragments for assembly with PID inserts in yeast. These were based off the pGG211, pGG431, and pGG442 acceptor vectors[19], which were themselves derived from a minimal Cas9^{Wt}-SV40 expression plasmid (pNA0306) described previously[94]. The kanamycin-resistance (Kan^{R}) cassette that replaces PIDs in all acceptor vectors is flanked by restriction sites for backbone isolation and ensures that any uncut or singly cut plasmid carryover does not contribute functional Cas9 activity upon transformation. All backbone fragments were isolated by restriction digest and gel purification; NheI and NotI were used for pGG574, pGG547, pGG548, or pGG618 plasmids (Dox induction experiments) while NheI and PspOMI were used for pGG532, pGG500, or pGG573 plasmids (TraCT experiments). When functional Cas9 plasmids were needed for strain construction they were pre-assembled with PID inserts and cloned in *E. coli* before delivery into yeast. Expression of sgRNAs in yeast was achieved with derivatives of the single sgRNA (pNA0304) or double sgRNA (pNA0308) plasmids described previously[94]. The pGG197 single sgRNA plasmid was used for the majority of targeting assays in this work and its spacer matches a heterologous *EGFP* protospacer (which served as Protospacer 1 in paired-protospacer experiments) that was employed previously[19]. The Protospacer 2 sequence used in paired-protospacer experiments is also a heterologous *EGFP* protospacer which was targeted by the 'NAG 3' spacer employed previously[18], but its PAM was altered in this work (<u>CAG</u>CTTGC→<u>CGA</u>AGTAA). All derivatives of pNA0308 also included a *SUP4* terminator insertion between the second sgRNA scaffold and the *RPR1* terminator. Constitutive expression of Z_3EV in yeast, the β-E-inducible activator protein for our *pZEVi* promoter, was achieved with an *ACT1* promoter and *RPS5* terminator cassette stably integrated at the *YKL162C-A* locus. This cassette was cloned along with a Nat^{R} marker and homology arms in a NotI fragment of pGG613, as in pMS101, but without extraneous sequences between the Nat^{R} marker and downstream homology arm. The pMS101 plasmid includes previously described parts[62] but was redesigned with a pAV10.KN backbone for integration at the *YKL162C-A* locus as described previously[95]. The *pZEVi* promoter used in this work is based off the P4 promoter[63], a *GAL10-1* hybrid promoter with binding sites for Z_3EV. Constitutive expression cassettes for nuclear-localized TetR and mCherry-2xN proteins were similarly cloned in *E. coli* plasmids and then isolated on NotI fragments for stable integration at the *YKL162C-A* locus in yeast. These plasmids were built off the pSIB619 architecture previously used for constitutive expression of the cam-TA protein[96] with a *TDH1* promoter and *STR1* terminator (modified to lack BsmBI restriction sites). The *TDH1* promoter was replaced with *ADH1* or *TDH3* promoters to obtain higher constitutive expression strengths in a subset of the pSIB619 derivatives. The Cas9$^{Wt\text{-}alt}$ (pAG5) and Cas9-2xN$^{Wt\text{-}alt}$ (pGG768) plasmids used for transient expression in human

cells were cloned in *E. coli* and derived from the pGG439 plasmid[19], which was itself derived from the MSP469 plasmid[18]. Transient expression of sgRNAs in human cells was achieved with derivatives of the BPK1520 single sgRNA plasmid[18], also cloned in *E. coli*.

Construction of yeast strains. All haploid yeast strains constructed in this work are described within Supplementary Data 2 and were generated using one of the following three transformation schemes. (1) To introduce circular episomes carrying selection markers that confer prototrophy, plasmid DNA (~ 30-60ng) was delivered directly through transformation and cells were plated with appropriate selection. (2) To modify the native *LYS2* or *CAN1* loci (including *LYS2's* promoter region) we used a CRISPR-assisted knock-in method essentially as described previously[19]. Sequences of the sgRNA plasmids used for knock-in at the *LYS2* ORF (pGG202), *CAN1* ORF (pGG203), or *LYS2* promoter (pGG578) are provided in Supplementary Data 3. Sequences of the linear double-stranded DNA donor fragments used for knock-in, including their homology arms, are listed in Supplementary Data 5. When patching colonies for initial phenotypic screening in this work, SC–Arg–Leu–Lys–Ura plates were used in place of SC–Lys, and SC–Arg–Leu–Lys–Ura+Can plates were used in place of SC–Arg+Can plates. Furthermore, during *pZEVi* strain isolations, 1000nM β-E was included in the SC–Arg–Leu–Lys–Ura plates to turn on the *pZEVi* promoter and facilitate *LYS2*$^{+}$ colony screening. CRISPR-assisted edits were validated by Sanger sequencing of gDNA-templated PCR amplicons that covered, at a minimum, all recombinant sequences and the flanking homology arms used for knock-in. In some cases, intermediate strains with genotypes that differ from BY4741's were used as parent strains for a second round of CRISPR-assisted knock-in, as outlined in Supplementary Data 2. Before delivering different sgRNA and/or Cas9 plasmids into CRISPR-edited strains for downstream experiments or for secondary rounds of CRISPR-assisted editing, the original sgRNA and/or Cas9 episomes used for knock-in were cured via replica plate screening as described previously[19]. (3) To integrate DNA into the *YKL162C-A* locus[95], linear DNA fragments (~ 200-300ng) bearing a Nat^{R} marker and sufficient homology arm lengths to select for integrative recombination were directly transformed without CRISPR co-selection; cells were then recovered for 24h in liquid YEPD or SC dropout media to select for plasmids before plating on solid media of the same type but with clonNAT present. These linear DNA fragments were excised from circular *E. coli* plasmids listed in Supplementary Data 3 via restriction digest with NotI, followed by gel purification. Integration of the pGG613 fragment in yGWG307, yGWG311, and yGWG312 lineages or the pGG658 fragment in yGWG315 was further validated by colony PCR and Sanger sequencing of regions spanning the homology arm junctions in *YKL162C-A*.

The diploid SSA reporter strain yGWG422 was mated from yGWG312 and yGWG381 haploids (Supplementary Data 2) on solid media. In brief, single colonies of each strain were streaked from opposite sides of a YEPD plate into a central overlapping region where mating could occur. After 7 h growth at 30 °C, this plate was replica-plated onto an SC–Ura–His plate to select for segregants with prototrophies unique to each parent strain (*URA3*+ from pGG197 in yGWG312 and *HIS3*+ from yGWG381's genome). A diploid clone was isolated from the segregant pool by re-streaking onto an SC–Ura–His plate; its SSA target region was amplified by colony PCR

and Sanger sequencing confirmed the expected heterozygosity (~ 1:1 allelic ratio) at the position of the point mutation.

Quantitative transformation-associated editing assays in yeast. Editing of haploid SSA reporter strains with a single sgRNA was performed essentially as described previously for clonal selection experiments in yeast[19]. In brief, haploid reporter strains were co-transformed via the lithium acetate procedure described above using 150 ng linearized backbone DNA and ~ 41 ng PID insert DNA (~ 1:3 molar ratio). Transformant populations resuspended in $CaCl_2$ were then serially diluted and spotted in parallel on different SC dropout plates, including SC–Leu–Ura plates to select for all viable transformants maintaining both the Cas9 and sgRNA episomes, as well as SC–Arg–Leu–Lys–Ura plates that select for transformants with both episomes and a reconstituted *LYS2* marker. Where applicable, additional spotting on SC–Arg–Leu–Lys–Ura plates supplemented with Dox, Can, or β-E was performed in parallel using the same serial dilutions. Plates were incubated for 47 h at 30 °C (or 72 h for Can-supplemented plates) before immediately photographing and then quantifying colony forming units (CFUs) by eye. After initial scoring, plates were stored at benchtop for up to 16 h (or 24 h for Can-supplemented plates) and counted once more; the higher of two CFU counts was used in determining each *LYS2*$^+$ fraction, calculated as the ratio of *LEU2*$^+$, *LYS2*$^+$, *URA3*$^+$ (and Can^R, where applicable) CFUs over *LEU2*$^+$, *URA3*$^+$ CFUs obtained on SC–Leu–Ura plates. Typically, each Cas9 PID variant fusion protein replicate was tested in parallel with a corresponding native Cas9 PID variant control replicate performed on the same day. The same procedures were followed for editing of paired-protospacer reporter strains with dual sgRNAs except that recipient strains harbored pre-assembled native Cas9^{Wt} or Cas9^{Wt}-2xN episomes instead of the sgRNA episome; competent cells were therefore prepared in SC–Leu rather than SC–Ura media and were transformed with circular dual sgRNA plasmid DNA (2.5 µg) instead of linear fragments.

For donor-dependent editing of haploid yeast, all recipient strains harbored pre-assembled native Cas9$^{Wt/NG}$ or Cas9$^{Wt/NG}$-2xN plasmids. These strains were transformed with 75 ng of the targeting sgRNA plasmid (pGG197) alone or co-transformed with the targeting or non-targeting (pGG202) sgRNA plasmids (75 ng) and 120 bp commercially synthesized double-stranded DNA donor fragment (500 ng). Transformants were resuspended in 700µl $CaCl_2$ prior to initial platings, performed by spreading aliquots (154µl per full plate) from the $CaCl_2$ suspension on SC–Leu–Ura media with or without β-E at 15nM or 1000nM. These plates were grown for 44 h at 30 °C before scrape-homogenizing for serial dilution and replating to determine their *LYS2*$^+$ fractions with SC–Arg–Leu–Lys–Ura plates that contained β-E (1000nM). Replates were incubated 38.5 h at 30 °C before immediately photographing and quantifying CFUs to determine *LYS2*$^+$ fractions as described above, except only 12 h were allowed at benchtop before recounting. In parallel with initial platings, aliquots were withdrawn from the remaining $CaCl_2$ suspensions for serial dilution and spotting on the same three solid media conditions. For strains with Cas9^{NG} or Cas9^{NG}-2xN, the same serial dilutions were also spotted on SC–Arg–Leu–Lys–Ura plates supplemented with β-E at 15nM or 1000nM to quantify *LYS2*$^+$ fractions obtainable immediately after transformation.

Plate reader growth assays. To obtain 1mL SC–Leu–Ura liquid starter cultures for each biological replicate, single colonies isolated by streaking on a single plate were vigorously inoculated and cultured in parallel on the same day. These were grown with continuous shaking (800 RPM) for 22 h at 30 °C in 96-well V-bottom assay blocks (Corning, 2mL capacity) that were pre-autoclaved (dry) with a 4mm glass bead in each well. Starter cultures were then subcultured 1:100 into 96-well SensoPlate microplates with a clear, flat bottom (Greiner Bio-One) containing 200µl of SC–Leu–Ura media in each well, and grown for 32 h at 30 °C with their lids on and continuous shaking in a Cytation 5 microplate reader (BioTek). High-resolution growth curves were generated from Attenuance ($D_{600\ nm}$) measurements taken every 10 minutes. For Dox induction experiments, microplates were briefly withdrawn after 5 h of growth and cultures were supplemented with Dox (10µM final concentration) or solvent alone by pipetting directly into the wells, then returned to the Cytation 5 for an additional 27 h of growth. Lids were kept warm in a 30 °C incubator during the withdrawal step. For β-E induction experiments, media containing β-E was supplemented at 15nM or 1000nM before aliquoting into microplate wells and was thus present throughout the 32 h.

Analysis of high-resolution growth curves from plate reader assays. All growth curves were analyzed using custom python scripts. For each curve, $D_{600\ nm}$ measurements were corrected for background by subtracting their minimum measurements from each value, and then doubling times (t_d) were calculated by fitting a logistic equation to each curve[97]:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}}$$

$$N_t = \text{ Attenuance } (D_{600nm})\ at\ time\ t$$

$$t_d = {}^{\ln 2}/_{r}$$

To calculate time-resolved differences in growth, $D_{600\ nm}$ measurements from +Dox cultures were subtracted from –Dox cultures at each time point. Significance was calculated at each time point by comparing between pooled control samples (n = 18) and *tetO2*+, Cas9^{NG}-scRevTetR (V16) samples (n = 6). Before calculating significance, the median differences calculated for each pair of cultures over their first 5 h (during which neither culture contains Dox) were assumed due to background variation and subtracted from the differences at each timepoint. Difference plots were generated using the python package matplotlib[98]. Curve fitting and statistical tests were performed in python with SciPy[99].

Quantitative editing assays with diploid reporter yeast. Editing of our diploid SSA reporter strain was initiated by co-delivering native Cas9 or Cas9-scMCP backbones with Wt or MT3 PID inserts as described above for transformation-associated editing of haploid SSA reporter strains with a single sgRNA. However, transformants were not plated immediately after the 10 min incubation in $CaCl_2$; rather, half the resuspended

mixture (200µl) was inoculated into 20mL SC–Leu–Ura broth to provide selection for cells which maintain both Cas9 (assembled in vivo) and sgRNA episomes, as well as time for steady-state segregation of the newly assembled Cas9 episomes. These 20mL cultures were grown for 24 h at 30 °C with rotation on a roller drum and then subcultured 1:20 in a fresh 20mL volume of the same media, either lacking β-E or supplemented with β-E at 15nM or 1000nM. Subcultures were grown at 30 °C with rotation on a roller drum until $D_{600\ nm}$ ~ 1.0 (12-17 h), pelleted by centrifugation at 3000 x *g* for 3 min, washed once with resuspension in an equal volume of sterile ddH_2O, and then re-pelleted and frozen at -20 °C for downstream gDNA isolation by bead beating and phenol-chloroform extraction (25:24:1 phenol:chloroform:isoamyl alcohol). Bead beating was carried out at 4 °C on a FastPrep-24 5G machine from MP Biomedicals using their 2mL tubes pre-filled with Lysing Matrix Y; eight cycles of 15-second bead beating was performed with 30 s pauses between each cycle. For the phenol-chloroform extraction, treatment with RNAse A (Thermo Scientific) at a final concentration of 30 mg/mL was performed for 15 min at 37 °C before the final 70% ethanol precipitation step. In parallel, aliquots were withdrawn from the post-24 h transformant cultures to determine their *LYS2*$^+$ fractions immediately prior to the 12-17 h subculture step. These aliquots were first pelleted by centrifugation at 2103 x g for 4 minutes, washed twice in an equal volume of sterile ddH_2O with resuspension and re-pelleting with the same centrifugation conditions, and then resuspended in a final volume of 1mL ddH_2O before serially diluting and spotting on plates. Plates were incubated 36 h at 30 °C before immediately photographing and quantifying CFUs as described for haploid editing assays, except only 12 h were allowed at benchtop before recounting.

To analyze allele frequencies at *LYS2* loci within the above gDNA preparations, targeted deep sequencing using an Illumina MiSeq (v2 reagent kit) was performed on amplicons that spanned a 230 bp region with the PS and allele-specific point mutation. These were amplified from unedited alleles in a single round of PCR (25 cycles) with GoTaq Green and unique combinations of primers oGG1246-oGG1254 (Supplementary Data 4) for multiplexed sequencing with dual indexing in a single run. Demultiplexed paired-end (230 x 230) reads were analyzed with IDT's rhAmpSeq CRISPR Analysis Tool (CRISPAltRations 1.1.0) using a modified protocol to quantify allele frequencies rather than Cas9 editing per se. Briefly, the 230 bp target region at Allele 1 was provided to the tool as 'Amplicon' sequence, while the same region but with the Allele 2 point mutation was provided as 'Donor' sequence. In addition, the 'Guide RNA' sequence provided to the tool was a mock spacer that overlapped the point mutation position, which was thus the fourth bp upstream of the PAM. Using these parameters, the 'Percent Unedited' output corresponds to the percentage of reads containing the Allele 1 sequence, while the 'Percent Perfect HDR' output corresponds to the percentage of reads containing the Allele 2 sequence. These output values summed to ≥ 98.5% for each sample analyzed, confirming that the majority of data were considered in our allelic ratio calculations. Approximately 0.204-0.473 million merged reads were analyzed per sample.

Plating assay to measure colony sizes. To obtain liquid cultures for plating, 18 x 150 glass tubes containing 5mL SC–Leu or SC–Leu-Ura broth were inoculated with single

colonies and grown for 22 h on a roller drum at 30 °C. SC–Leu–Ura was used for strains with the sgRNA plasmid while SC–Leu was used for strains without. Once grown, cultures were diluted in sterile ddH_2O and then plated on solid media of the same type to obtain ~ 150-400 colonies per plate. Plates were photographed directly from above after 47 h growth at 30 °C and colonies were then analyzed as 2-dimensional objects using custom python scripts to quantify colony areas. This was achieved using the python package scikit-image: colonies were segmented by the Otsu thresholding algorithm and touching colonies were separated by the watershed method. Objects detected near plate edges were excluded to avoid measuring artifacts. Swarm plots were generated using the python package seaborn and *t* tests were performed with SciPy[99].

MNase-seq procedures and analyses. Experiments employed a *pZEVi-LYS2*$^+$ haploid strain (yGWG538 in Supplementary Data 2) harboring an intact *LYS2*$^+$ ORF, the native Cas9^{Wt} plasmid (pGG541), and the single *EGFP* sgRNA plasmid (pGG197) which is non-targeting in this background. These features were chosen (1) to ensure standard mRNA quantifications performed in parallel (see qRT-PCR section below) by avoiding nonsense-mediated decay expected in targeting strains and (2) to emulate the plasmid conditions of our other experiments without interfering with growth. The following procedures were adapted from MNase-seq methods described previously[100,101]. Liquid starter cultures grown for 24 h at 30 °C were obtained by inoculating 5mL SC–Leu–Ura media from single colonies and then subcultured (1:70) into 500mL flasks with 100mL of the same media, either lacking β-E or supplemented at 15nM and 1000nM. These were grown at 30 °C with shaking (220 RPM) until $D_{600\ nm}$ ~ 1.0 (~ 7 h) and then crosslinked by adding formalin (1% final concentration) and incubating at 25 °C with gentle shaking (110 RPM) for 15 min. After adding 5mL glycine (125mM final concentration) to quench formalin for 5 min at 25 °C with gentle shaking (110 RPM), cultures were pelleted by centrifugation for 5 min at 3000 x *g*, washed twice with resuspension in equal volumes of ice-cold sterile ddH_2O, and then pelleted again in 1.5mL tubes for freezing at -80 °C. Pellets were thawed on ice and resuspended in spheroplasting buffer (1.2M sorbitol, 100mM potassium phosphate, 1mM $CaCl_2$, 0.5mM β-mercaptoethanol, pH 7.5) with 1 mg/mL of 100T Zymolyase. Cells were spheroplasted at 37 °C for 30 minutes and spheroplasts were collected by centrifugation and washed twice with 1mL of spheroplasting buffer. Spheroplasts were divided into four 1.5mL tubes, pelleted by centrifugation, and resuspended in MNase digestion buffer (1M Sorbitol, 50mM NaCl, 10mM Tris-HCL (pH 7.4), 5mM $MgCl_2$, 1mM $CaCl_2$, 0.5mM spermidine, 0.075% NP-40, add 1mM β-mercaptoethanol) with 20 U of MNase (Thermo Scientific Cat. EN0181). Chromatin was digested with MNase at 37 °C for 15, 20, 45, and 60 minutes and reactions were stopped by addition of EDTA to a final concentration of 30mM. Cross-links were reversed by addition of SDS (to 0.5% final) and proteinase K (0.5 mg/ml final) and incubated at 37 °C for 1 hour followed by 65 °C for 2 hours. Digested DNA was extracted using 25:24:1 phenol:chloroform:isoamyl alcohol and precipitated with 70% isopropanol. To remove RNA, the mixture was then resuspended in TE buffer with RNAse A (6 Kunitz units) and incubated at 37 °C for 1 hour. Finally, DNA was purified with the Zymo Clean and Concentrator kit. MNase digestions were assessed by running on an agarose gel with ethidium bromide staining to confirm mono-nucleosome

fragments between 100-200bp. A multiplexed Illumina sequencing library was then prepared from samples digested for 15 and 60 min through direct adapter ligation using the automated KAPA general DNA-seq library preparation kit. Paired-end sequencing of this library on an Illumina NextSeq 2000 with a P2 reagent kit (100 cycles) yielded approximately 40 million reads per sample after demultiplexing. Demultiplexed reads were first analyzed with Trimmomatic (v0.39) processing to remove sequencing adaptors and then with FastQC (v0.11.4) to assess read quality. Processed reads were then aligned to a custom reference of chromosome II (chr.II) containing *pZEVi-LYS2*$^{+}$, using the Burrows Wheeler aligner (BWA) mem algorithm (v0.7.7). Aligned reads were then used as input for nucleosome analysis using the DANPOS (v2) pipeline[102]. Analyses of samples digested for 15 and 60 min revealed similar trends but suggested incomplete digestion at 15 min; hence, only 60 min data are presented. We observed a total of 4,670 nucleosomes across chr.II with ~ 1% of these nucleosomes showing statistically significant changes in occupancy levels when comparing between 0nM β-E and 15nM or 1000nM, of which about half were from the *LYS2*$^{+}$ locus (Supplementary Data 1*; p* value < 0.01; FDR-corrected Poisson tests performed by DANPOS). Tukey's tests for significant Log2FC differences across the *LYS2*$^{+}$ ORF (Extended Data Fig. 7f) were performed using GraphPad Prism software.

qRT-PCR and Fold Change calculations for *pZEVi-LYS2* induction. To purify total RNA for reverse transcription, 1.3mL aliquots were withdrawn from MNase-seq cultures immediately prior to formalin addition, pelleted at 4 °C, washed once with resuspension in 1.3mL ice-cold autoclaved ddH_2O, then re-pelleted at 4 °C for -80 °C storage until further processing. RNA was purified using the RNeasy Mini kit (Qiagen) and an in-solution DNase I (Qiagen) treatment prior to final cleanup. cDNA was synthesized using SuperScript IV Reverse Transcriptase (Invitrogen) according to the manufacturer's recommendations for a 20µl reaction with 1µg total RNA and random hexamers (IDT). qPCRs were performed with gene-specific primers (each at 500nM) from IDT and 2.5µL of the diluted cDNA (1:5 dilution) in a 10µL reaction with Fast SYBR Green Master Mix (Life Technologies) and a LightCycler® 480 Real-time PCR System (Roche) according to the manufacturers' protocols. Primers used for the *TAF10* housekeeping gene or *LYS2* target gene are listed in Supplementary Data 4. Raw Ct values were calculated using the second derivative method. ΔCt values were calculated relative to the *TAF10* levels in each sample; ΔΔCt values were calculated relative to the *LYS2* levels in replicate 3 of the 0nM samples. Ultimately, the plotted Fold Changes in *LYS2* RNA levels were calculated as Fold Change = $2^{-\Delta\Delta Ct}$.

Generation of the clonal U2OS.EGFP::5' UTR-*boxB*^{2x} and U2OS.EGFP::Intron-*boxB*^{2x}(V1.0) cell lines. Using CRISPR-assisted knock-in, the U2OS.EGFP cell line was modified to contain two *boxB* HTs embedded within the 5' UTR or an intron of its *EGFP-PEST* marker. Briefly, to generate the 5' UTR insertion in U2OS.EGFP::5' UTR-*boxB*^{2x}, 2 µg of Cas9 plasmid (pAG5), 666 ng of sgRNA plasmid (pGG771) and 20 pmol of a linear ssDNA repair template (oGG1238) with two phosphothiorate linkages on each end were mixed with 82 µL Nucleofector Solution V and 18 µL of Supplement Solution 1 according to manufacturer's recommendations for the Cell Line Nucleofector Kit V (Lonza). $5x10^5$ U2OS.EGFP parent cells were resuspended in the above mixture

and nucleofected with a Lonza Nucleofector 2b Device (Kit V, Program X-001). The cells were immediately transferred to a 6-well plate containing 2 mL of medium and incubated at standard conditions (37 °C and 5% CO_2). Five days post-nucleofection, single cells were sorted based on forward scatter area and width parameters into separate wells of 96-well plates using a BD FACSAria IIu flow cytometer. Wells that showed outgrowth within 7 days post-sorting were genotyped with primers P40 and P41 (Supplementary Data 4) for PCR-based screening of clones carrying the desired knock-in. Out of 26 clones screened, 7 contained an amplicon detectably longer than the parental genotype. These 7 clones were further verified by Sanger sequencing, and a single clone was chosen for downstream assays. To generate the intron insertion in U2OS.EGFP::Intron-*boxB*^{2x}(V1.0), transfections were performed using a modified Effectene reagent (Qiagen) protocol as previously described[103] except that the DMEM used for cell culture was also supplemented at ~ 1% each of sodium pyruvate (100mM stock), GLUTAMAX (100X), and non-essential amino acids (100X) from Gibco. Per transfection, 0.25 µg of Cas9 plasmid (pAG5), 0.1 µg of sgRNA plasmid (pGG443) and 5.5 pmol of a linear ssDNA repair template (oGG1239) was used in 100µL of buffer EC with a DNA:Enhancer ratio of 1:8 (3.2µL) and DNA:Effectene ratio of 1:15 (6µL). The resulting transfection complexes were added to each well in a 6-well plate that was freshly seeded with $2x10^5$ U2OS.EGFP parent cells in 2mL of media. Seven days post-transfection, single cells were sorted based on forward scatter area and width parameters into separate wells of 96-well plates using a BD FACSAria IIu flow cytometer. Wells that grew out within 7 days post-sorting were genotyped with primers P31 and P32 (Supplementary Data 4) for PCR-based screening of clones carrying the desired knock-in. Out of 91 clones screened, 18 contained an amplicon detectably longer than the parental genotype. Sanger sequencing of the 18 amplicons revealed only one sequence that matched the repair template sequence, except for two unexpected point mutations within the intron's polypyrimidine tract and splice acceptor elements, and the corresponding cell line was isolated for downstream assays conducted in the same medium.

Human cell culture assays to measure Cas9-mediated indels. To generate indels in the U2OS.EGFP cell line or its *boxB*^{2x}-containing derivatives, Cas9$^{Wt\text{-}alt}$ (pAG5) or Cas9-2xN$^{Wt\text{-}alt}$ (pGG768) plasmids were co-delivered with sgRNA plasmids (Supplementary Data 3) via nucleofection as described previously for EGFP knockdown assays[19] except that ptdTomato was excluded, media lacked Geneticin (G418 Sulfate), cells were harvested 48 h post-nucleofection for genomic DNA (gDNA) purification rather than flow cytometry, and all three biological replicates were performed on the same day. gDNA was purified from cells in 6-well plates using the DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol. For indel detection at the targets of sgRNAs 1 through 3, deep sequencing with an Illumina NextSeq 2000 (300-cycle Low Output P1 reagent kit) was performed on PCR amplicons spanning 297 bp of the protospacer region, and demultiplexed single-read (297 x 1) sequences were analyzed with IDT's rhAmpSeq CRISPR Analysis Tool (CRISPAltRations 1.1.0) using the default parameters for amplicon analyses. To employ this tool for our single-read data, it was necessary to first generate mock R2 read pair '.fastq' files from the R1 files—with identical but reverse-complemented sequences; then, generate sub-sampled files for

downstream analyses—each with 0.100 million random pairs of reads. Both file processing steps were performed with the Seqtk script for python. Ultimately, > 0.098 million merged reads were analyzed per sample. For indel detection at the sgRNA 4 target, deep sequencing with an Illumina MiSeq (600-cycle v3 reagent kit) was performed on PCR amplicons spanning 297 bp of the protospacer region, and demultiplexed paired-end (297 x 2) reads were analyzed with IDT's rhAmpSeq CRISPR Analysis Tool (CRISPAltRations 1.1.0) using the default parameters for amplicon analyses. Approximately 0.690-1.159 million merged reads were analyzed per sample without Seqtk pre-processing. Sequences of the amplified regions provided to the tool as "Amplicons" are included in Supplementary Data 6 along with raw results from the analyses. PCR amplicons were generated through a single round of GoTaq Green PCR (30 cycles) using unique combinations of primers listed in Supplementary Data 4 for dual-indexed sequencing runs. To facilitate this PCR reaction, the first 10 cycles annealed at 55 °C and the last 20 cycles annealed at 65 °C. Amplified DNA was size-selected (~ 300-500 bp) by gel purification prior to sequencing, accounting for the total amplicon length expected with Illumina adapters (433 bp for unedited genotypes).

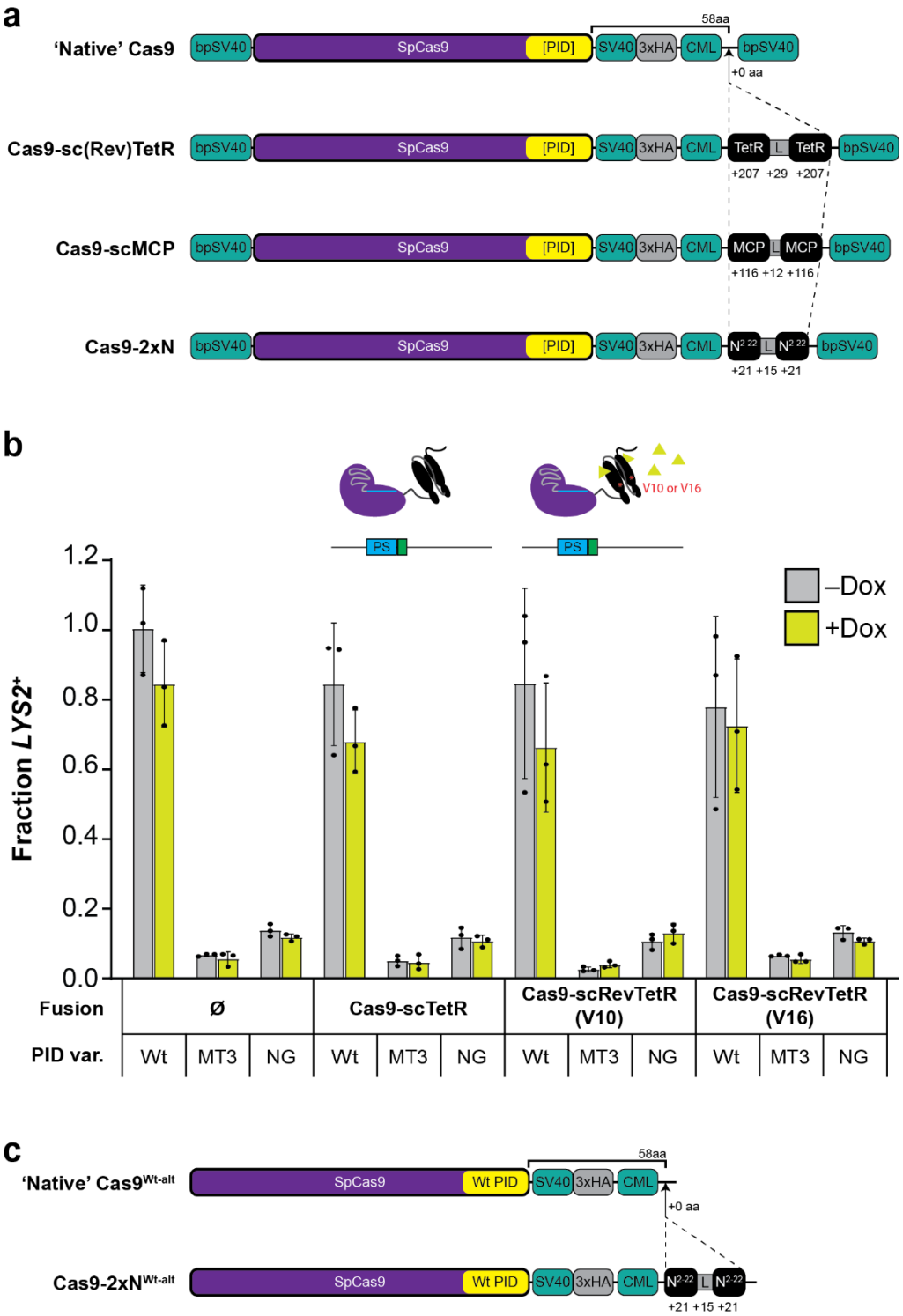
Analysis of yeast cell fluorescence by flow cytometry. To analyze the mCherry-2xN strains for fluorescence, 1mL starter cultures were first obtained for each biological replicate as described above in plate reader growth assays, except SC–Ura broth was used. Once grown, 33.3µl aliquots were withdrawn from starter cultures and analyzed on a SONY SA3800 spectral analyzer after dilution into 500µl aliquots of PBS+0.5% FBS. Mean mCherry values were measured in arbitrary fluorescence units from 30,000 events per culture according to the manufacturer's suggested settings for mCherry fluorescence, after gating on a majority population in the expected size range.

Statistical testing. GraphPad Prism, matplotlib (python), DANPOS v2 (python), IGV, or Microsoft Excel were used to generate all quantitative plots in this work before manually compiling and annotating them in Adobe Illustrator. Except where noted otherwise, *p* values were calculated with GraphPad Prism or python (SciPy) software using unpaired, two-tailed *t* tests with Welch's correction and Gaussian distributions assumed. Biological replicates were performed on separate days unless noted otherwise.

Data availability

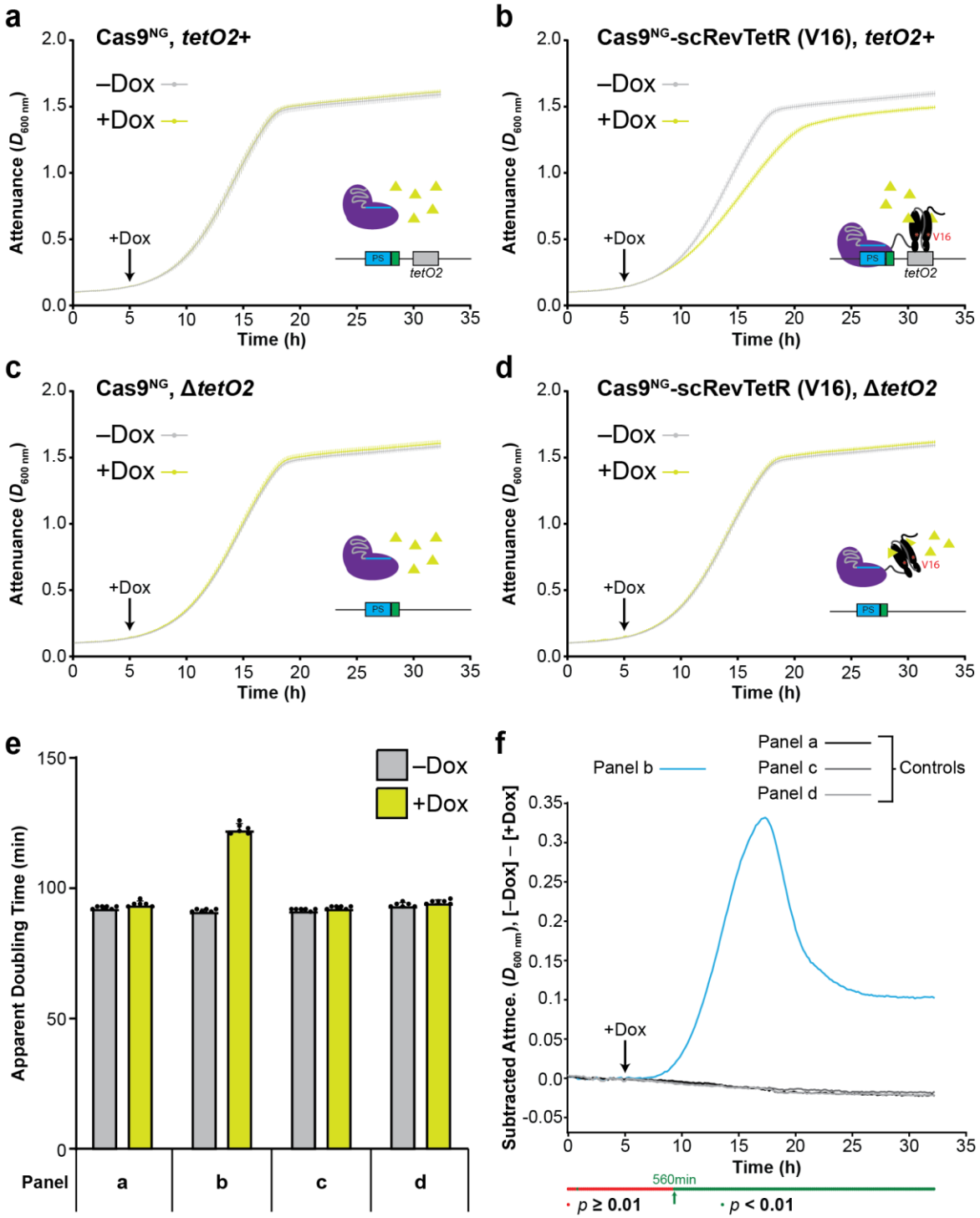
Relevant data supporting the findings of this study will be made available in the published article and its supplementary files, as well as the corresponding authors upon reasonable request.

Extended Data Figures



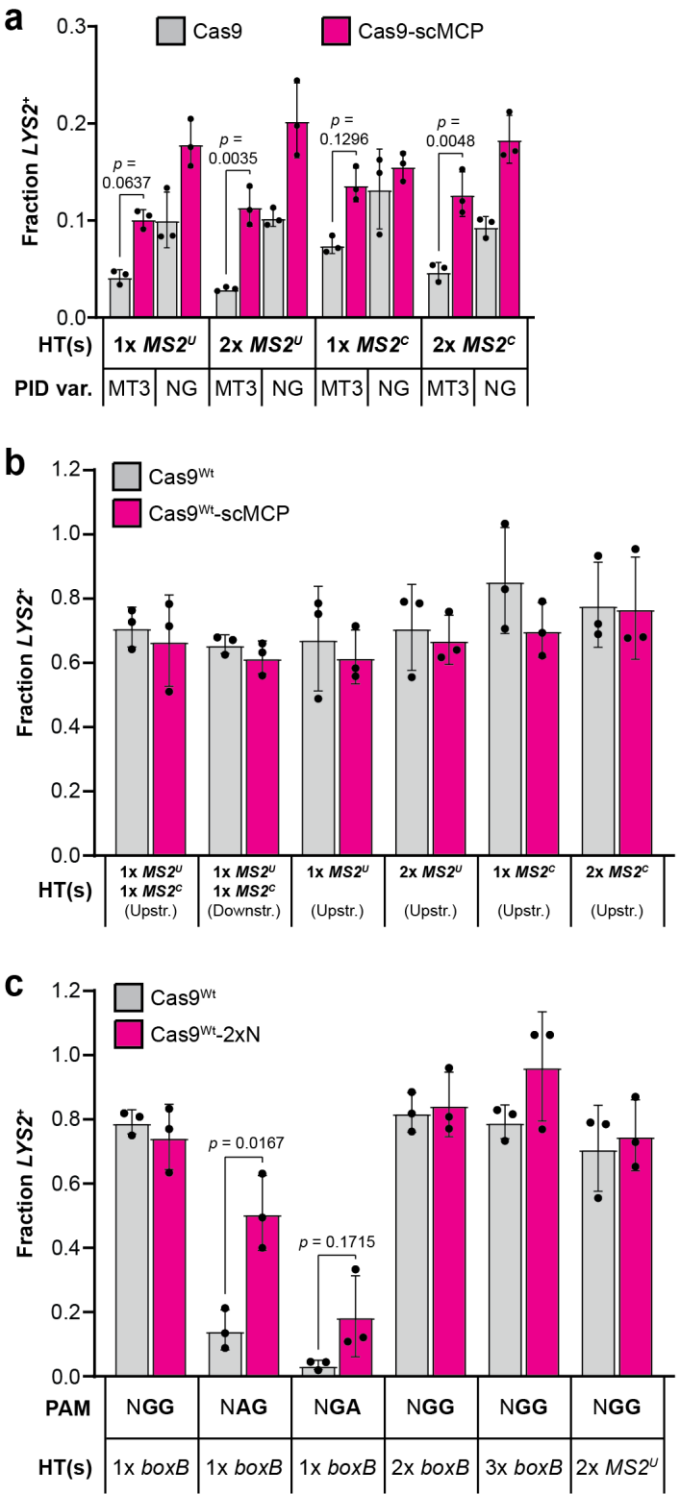
Extended Data Fig. 1 | Summary of Cas9 fusion proteins used in this work and control experiments for sc(Rev)TetR performed in a strain lacking *tetO2*. a, One-dimensional protein schematics diagramming the native Cas9 scaffold and isogenic

insertion site for Cas9 fusions tested in yeast. A previously described[39] 58 amino acid (aa) linker that includes SV40 and C-myc-like (CML) nuclear localization signals (NLSs) is fused to *Streptococcus pyogenes* Cas9 (SpCas9) immediately downstream of its C-terminal PAM-interacting domain (PID). The additional aa lengths of each fusion domain segment are indicated below; e.g., +29 for the linker (L) connecting tandem TetR monomers. Additional bipartite SV40 (bpSV40) NLSs are present at the extreme N- and C-termini of all Cas9 proteins. 3xHA; three tandem repeats of the influenza hemagglutinin tag. **b**, Transformation-associated editing with the Cas9-sc(Rev)TetR fusions or native Cas9 control (Ø) in yeast lacking a *tetO2* element. Measurements were performed as in Fig. 1b. Inset schematics depict binding-proficient TetR (–Dox) and RevTetR (+Dox) states that remain unbound in the absence of a *tetO2* site downstream of the protospacer (PS). Error bars, mean ± s.d. (n = 3, biological replicates). **c**, One-dimensional protein schematics with labeling as in panel **a** but for the alternative Cas9 architectures expressed in human cells without bpSV40 NLSs; only proteins with Wt PIDs were constructed.



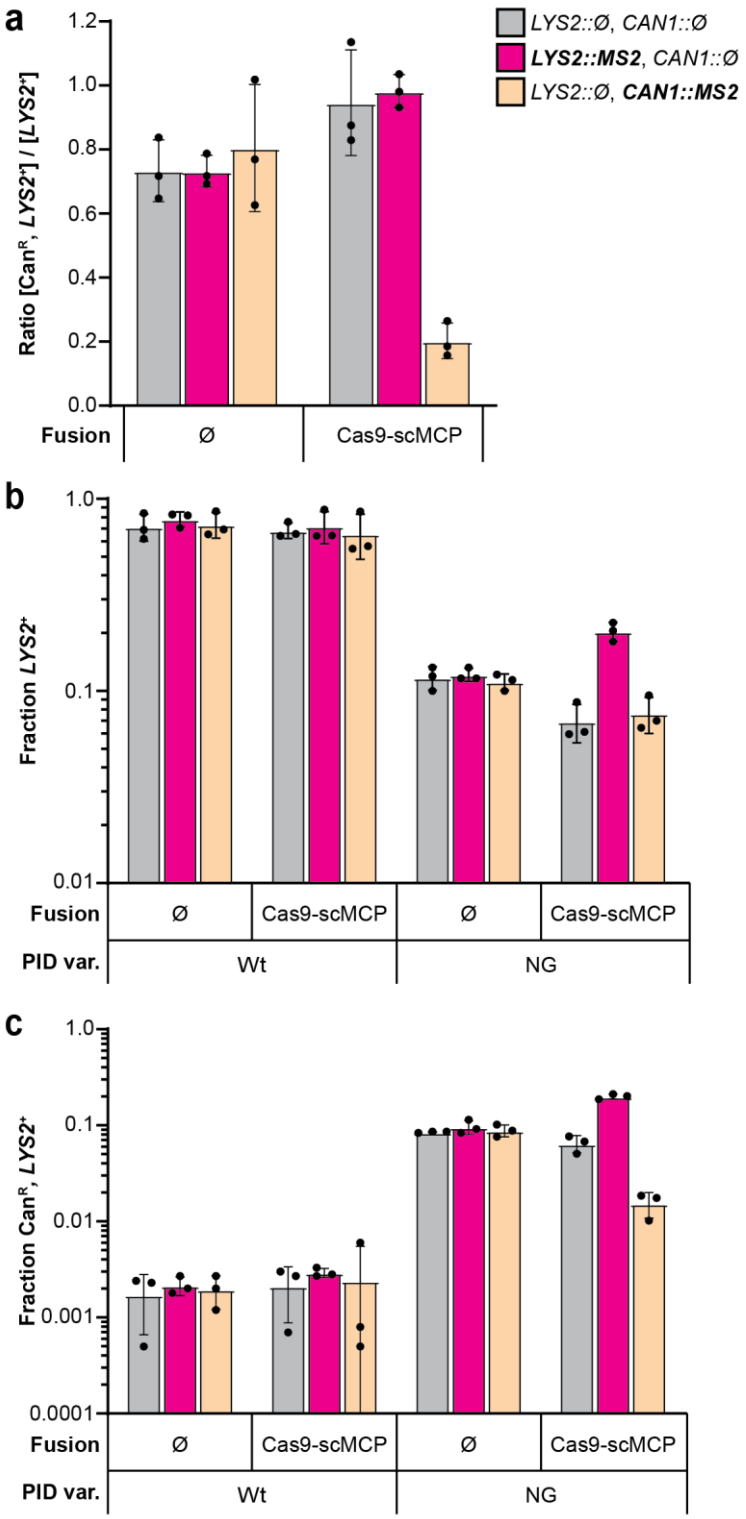
Extended Data Fig. 2 | Time-resolved detection of a Dox-induced targeting phenotype for Cas9^{NG}-scRevTetR (V16) in liquid culture assays. a, Growth of the specified strain in liquid cultures with (+Dox) or without (–Dox) induction at the indicated 5 h timepoint (black arrow) was monitored at high resolution using a plate reader. Inset illustration depicts a native Cas9^{NG} protein that is unable to bind the *tetO2* element in the presence of Dox because it lacks the scRevTetR fusion. Attenuance ($D_{600\ nm}$)

readings were taken every 10 min. Error bars, mean ± s.d. (n = 6, biological replicates). **b**, Same experimental setup as in **a** except the strain tested contains a scRevTetR (V16) fusion that is bound in the presence of Dox, as shown in the inset. **c**, Same experimental setup as in **a** except the strain tested lacks the *tetO2* element. **d**, Same experimental setup as in **b** except the strain tested lacks the *tetO2* element and thus its scRevTetR domain remains unbound even in the presence of Dox. **e**, Apparent doubling times calculated from all growth curves plotted above (panels **a**-**d**). **f**, Time-resolved detection of Dox-induced growth phenotypes. Differences in growth were quantified at each time point for panels **a**-**d** above by subtracting the $D_{600\ \mathrm{nm}}$ readings of +Dox cultures from those of –Dox cultures. Significant differences calculated for panel **b** ($p < 0.01$) at each time point are denoted with green dots below the plot and were first detected at 560 min growth. To improve readability, only the mean differences calculated for each panel are plotted.



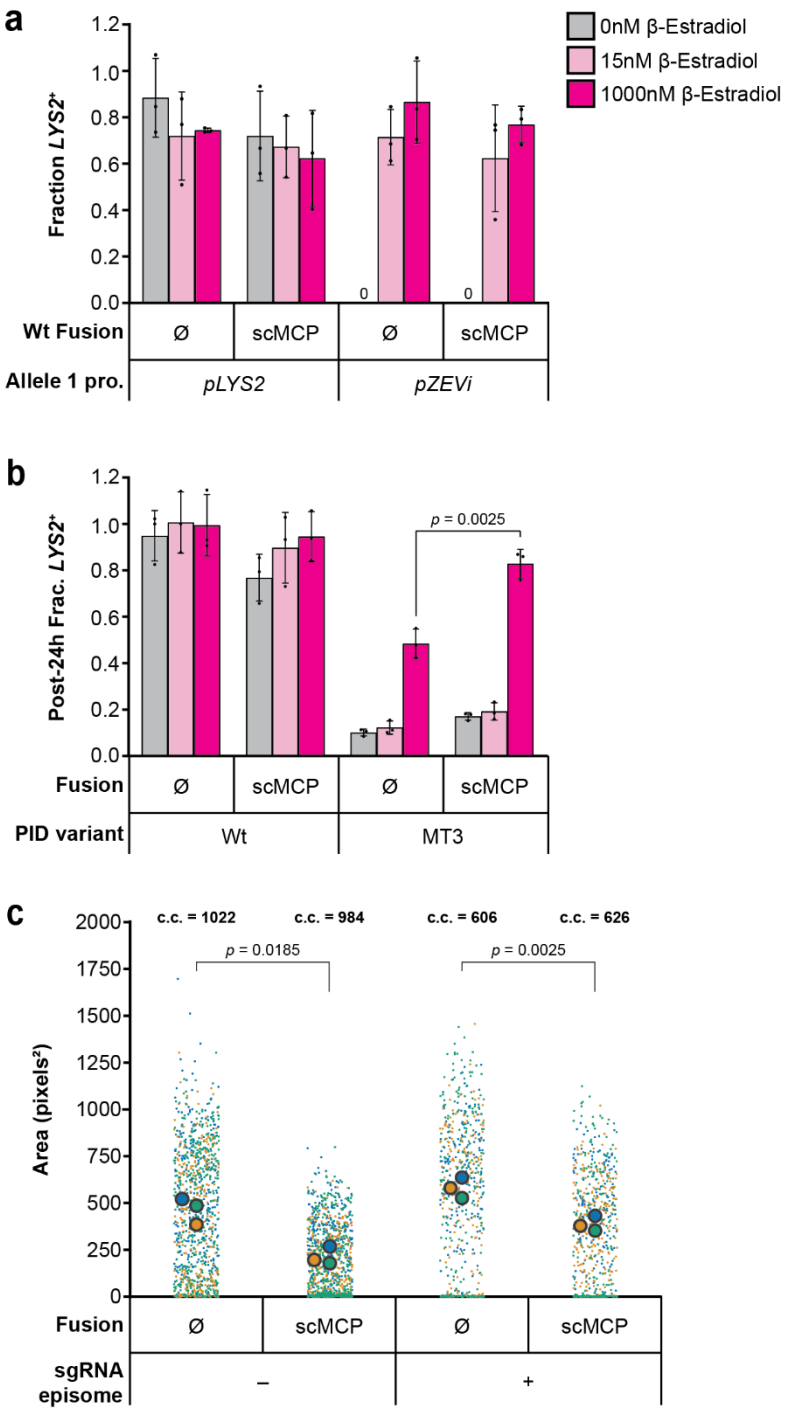
Extended Data Fig. 3 | Testing of alternative MS2 hairpin configurations, Wt PID controls, and Cas9^{Wt}-2xN with non-NGG subPAMs. a, Transformation-associated editing fractions measured for MT3 and NG variants of Cas9-scMCP or native Cas9 in strains containing alternative HT configurations as indicated. Error bars, mean ± s.d. (*n* = 3, biological replicates). **b**, Control experiments with Wt PIDs, using the same setup

and HT strains tested in panel **a** and Fig. 1e. Error bars, mean ± s.d. (*n* = 3, biological replicates). **c**, Experiments with Wt PID variants of Cas9-2xN or native Cas9 using the same setup and HT strains tested in Fig. 1f, as well as two additional strains containing one *boxB* HT and either NAG or NGA PAMs as indicated. Error bars, mean ± s.d. (*n* = 3, biological replicates).



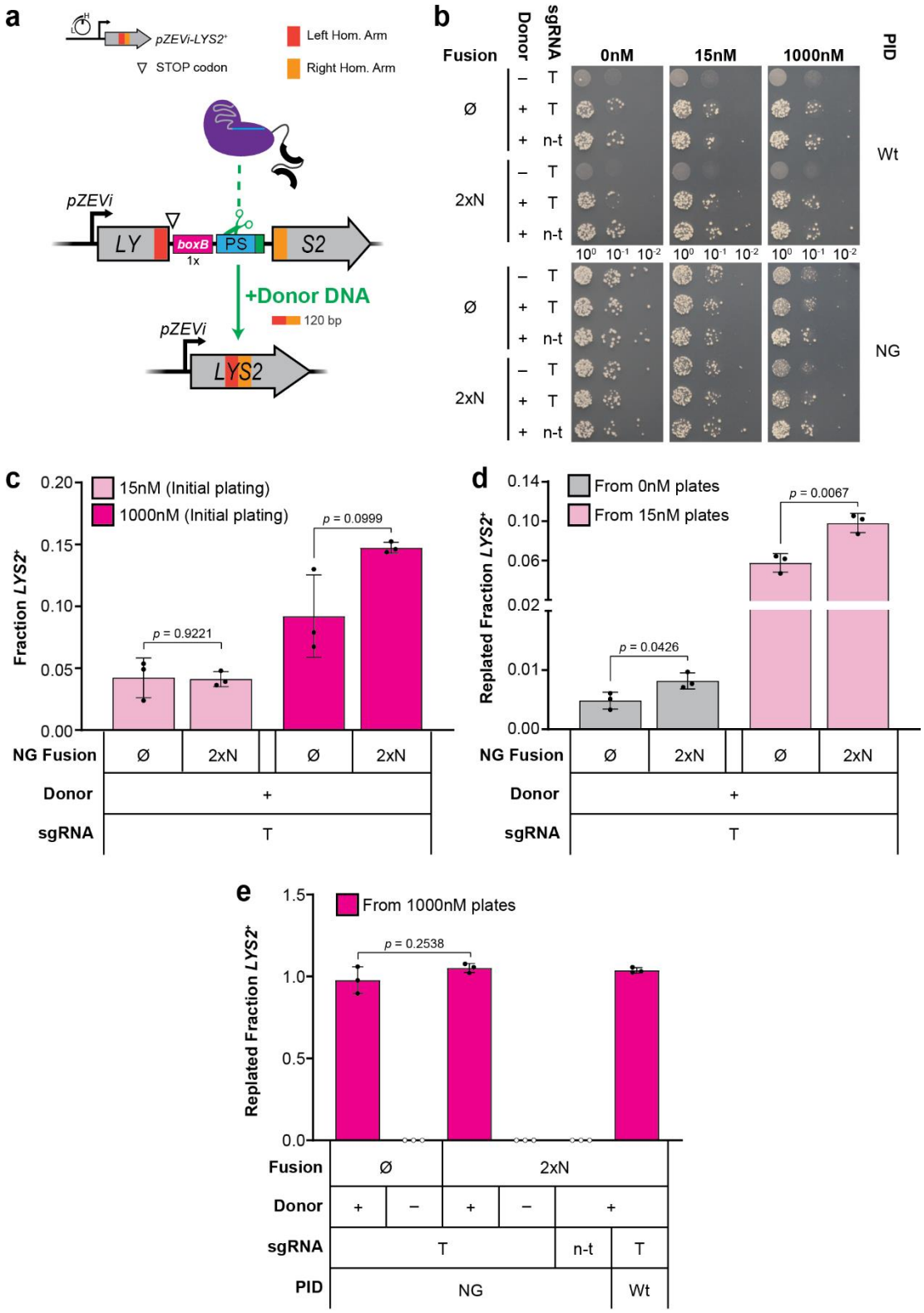
Extended Data Fig. 4 | Re-plotting of data from Fig. 2 and testing of Wt PID controls in dual-reporter strains. a, Raw counts of *LYS2*$^+$ and Can^R, *LYS2*$^+$ colonies from the experiments of Figs. 2b–c were used to calculate and plot ratios as indicated (Can^R, *LYS2*$^+$ counts over *LYS2*$^+$ counts). Error bars, mean ± s.d. (*n* = 3, biological

replicates). **b**, Re-plotting of data from Fig. 2b in logarithmic scale, alongside data from Wt PID control experiments. Error bars, mean ± s.d. (n = 3, biological replicates). **c**, Re-plotting of data from Fig. 2c in logarithmic scale, alongside data from Wt PID control experiments. Error bars, mean ± s.d. (n = 3, biological replicates).



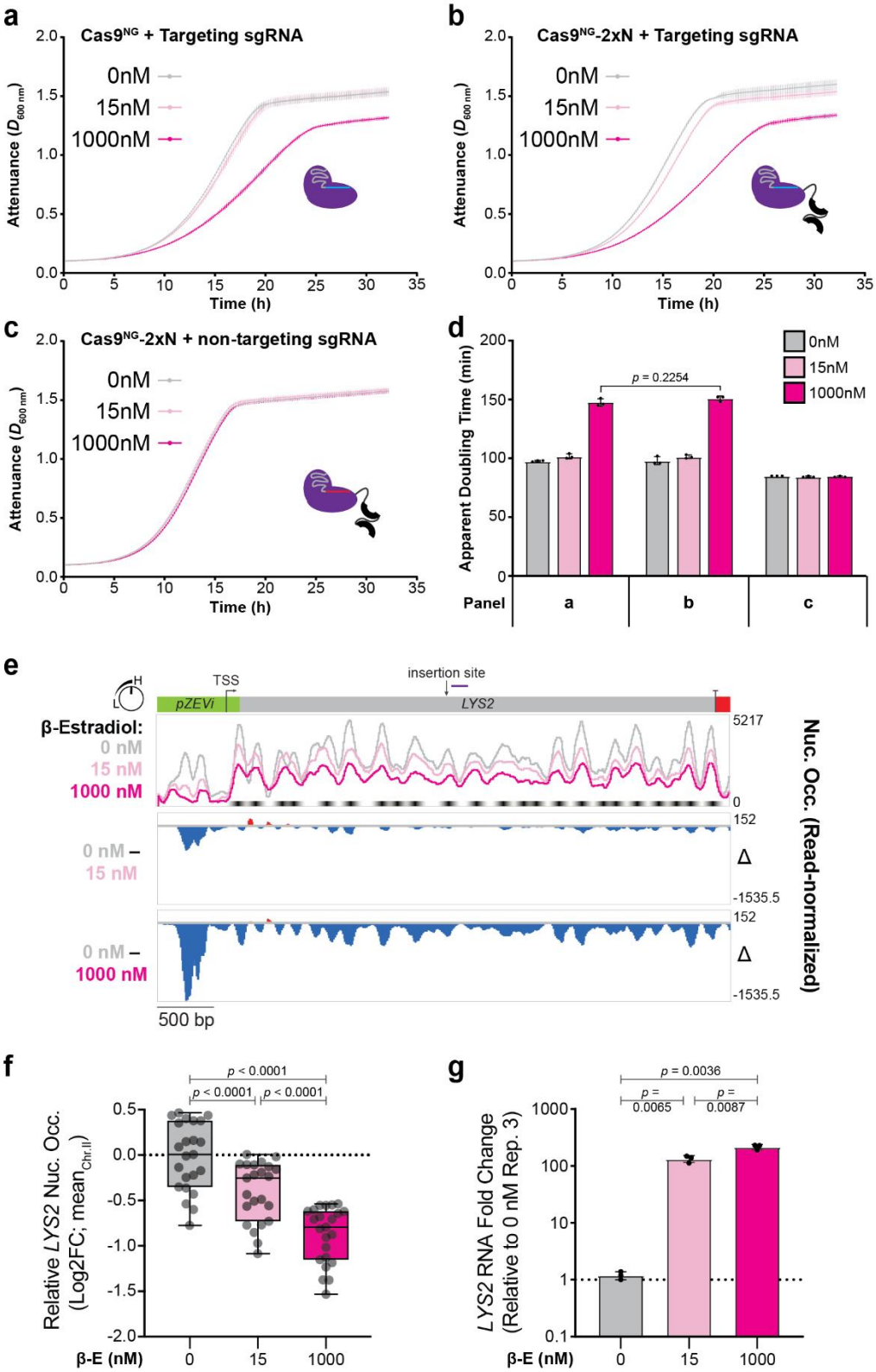
Extended Data Fig. 5 | Control experiments for Fig. 3 and colony size comparisons for strains expressing native Cas9^{Wt} or Cas9^{Wt}-scMCP under non-targeting conditions. a, Transformation-associated editing fractions measured under the same conditions as Fig. 3b experiments, but with Wt PID variants of Cas9-scMCP or native Cas9 in *pLYS2* or *pZEVi* haploid strains. Zero (0) *LYS2*$^+$ colonies of the *pZEVi* strain were detected in the absence of β-E at the time of scoring. Error bars, mean ± s.d. (n = 3, biological replicates). **b**, Diploid *LYS2*$^+$ fractions assayed with β-E at each

concentration (color coded as in panel **a**) by plating the Cas9^{MT3} and Cas9^{MT3}-scMCP transformant populations of Fig. 3c after the 24 h liquid outgrowth step and immediately before the subculture induction step. Wt PID control populations were also assayed here post-outgrowth, though not subcultured or studied further. All *LYS2*$^{+}$ colonies detected in the absence of β-E are inferred to result from repair of the constitutive *pLYS2* allele since repair of the *pZEVi* allele in non-induced haploids is insufficient to generate *LYS2*$^{+}$ colonies under the conditions tested. Error bars, mean ± s.d. (n = 3, biological replicates). **c**, Colony sizes measured for BY4741 strains carrying the indicated Cas9 plasmids with (+) or without (–) episomal expression of an sgRNA, which, is effectively non-targeting because this background lacks the protospacer insertion in *LYS2*. The total number of colonies counted (c.c.) for each strain is indicated in boldface above each column. Datapoints from each biological replicate (n = 3) are color-coded, and the corresponding mean values measured for each replicate are indicated with large outlined dots. Parametric t tests were performed on means (paired within replicates) to obtain the p values shown.



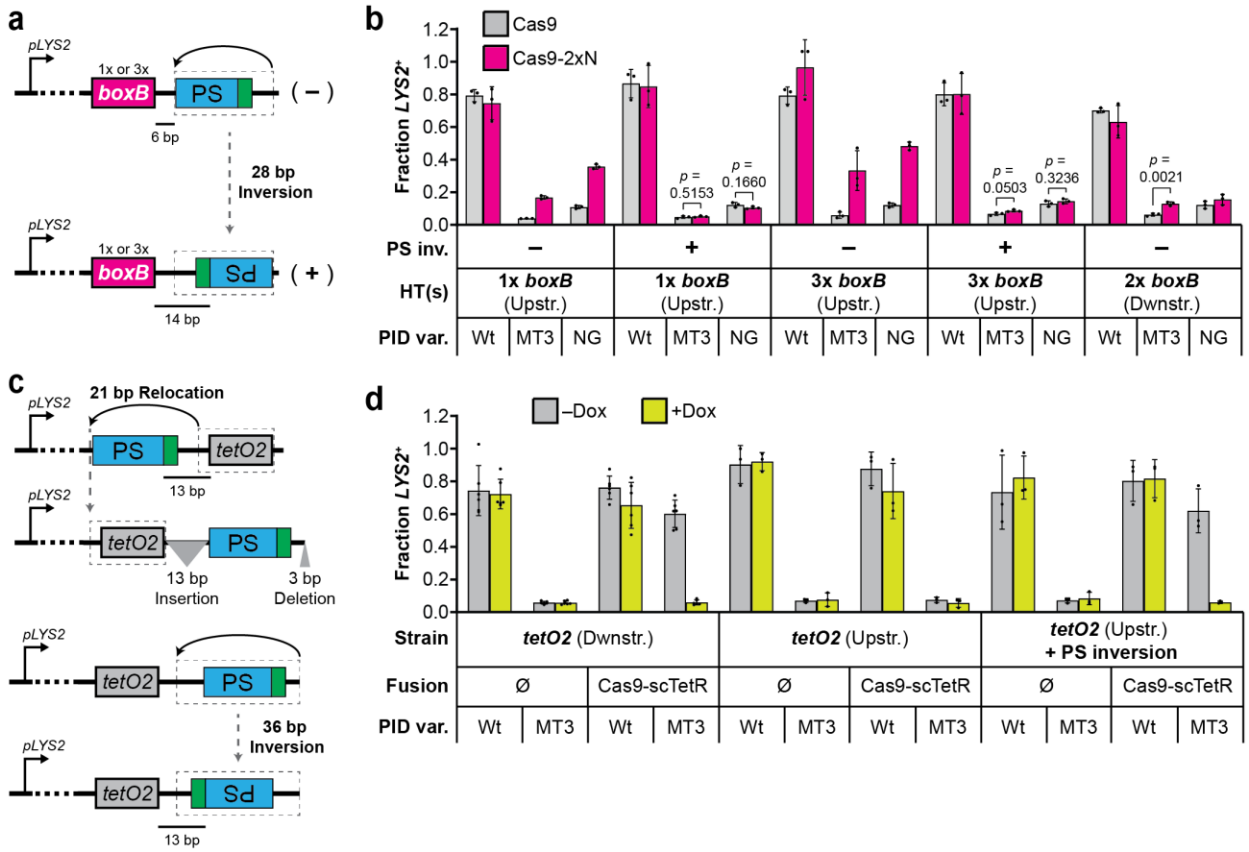
Extended Data Fig. 6 | TraCT can stimulate donor-dependent editing in *pZEVi* haploid yeast strains and donor-independent editing in human cells. a, Schematic summary of the donor-dependent *LYS2* editing system. Upon cleavage, the disrupted

pZEVi-LYS2 locus can be reconstituted through homology-directed repair with a 120 bp donor DNA fragment that scarlessly removes intervening sequences between the left (red) and right (orange) homology arms. **b**, Post-transformation colony phenotypes for strains harboring the Wt or NG PID variants of native Cas9 (Ø) or Cas9-2xN as indicated, plated on media with lysine to permit growth of all transformants. Transformations were performed with targeting (T) or non-targeting (n-t) sgRNA plasmids and co-delivered (+) or absent (–) donor DNA. The resulting populations were plated as serial dilutions on plates that contained β-E at the indicated concentrations (0nM, 15nM or 1000nM) in addition to lysine. Photos shown contain representative data from one of the three biological replicates (n = 3). **c**, Transformation-associated editing fractions measured for the indicated NG populations from panel **b**; plating was performed in parallel from the same serial dilutions but using lysine-free media with β-E at the indicated concentrations. Error bars, mean ± s.d. (n = 3, biological replicates). **d**, *LYS2*$^+$ fractions measured for the indicated NG populations from panel **b**, but after an outgrowth step with lysine present on plates containing β-E at the indicated concentrations. To activate *LYS2* expression in edited cells during the replating step, lysine-free plates contained 1000nM β-E. **e**, *LYS2*$^+$ fractions measured for the indicated NG populations from panel **b** as performed in panel **d**, except that 1000nM β-E was used during the outgrowth step.

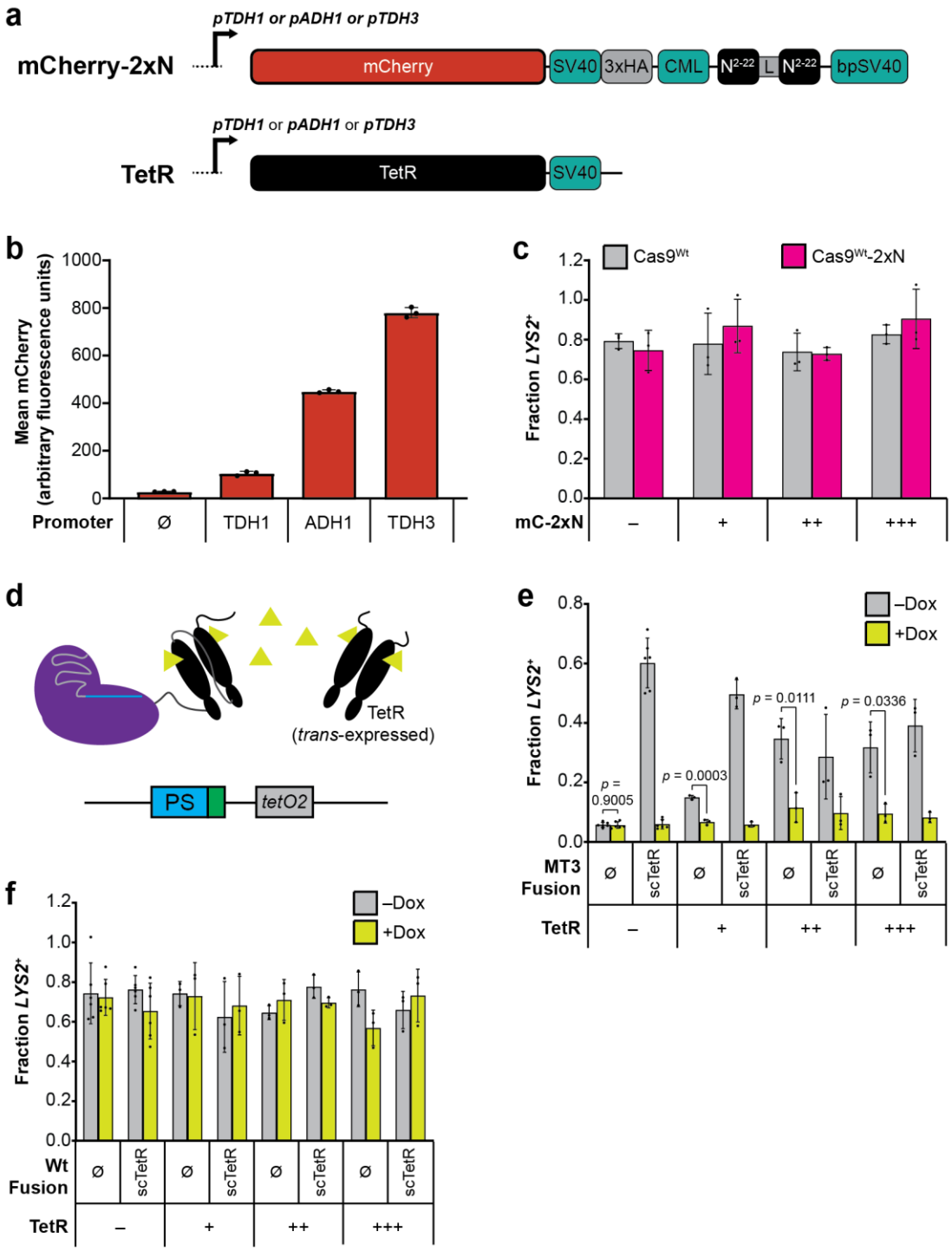


Extended Data Fig. 7 | Transcriptional induction at the *LYS2* locus can increase basal Cas9^{NG} targeting and reduce nucleosome occupancy. a, Plate reader experiments to monitor growth in liquid media with β-E (15nM or 1000nM) or lacking it

altogether (0nM). The *pZEVi* reporter strain tested here is an unedited isolate from donor-dependent editing experiments with native Cas9^{NG}; this protein lacks an auxiliary RBD fusion as shown in the inset illustration. Attenuance ($D_{600\ nm}$) readings were taken every 10 min. Error bars, mean ± s.d. (n = 3, biological replicates). **b**, Same experimental setup as in **a** except the strain tested contains a 2xN fusion, as shown in the inset, that can interact with boxB RNA transcribed upstream of the protospacer. **c**, Same experimental setup as in **b** except the strain tested has non-targeting spacer sequences (red line) in its sgRNA scaffold. **d**, Apparent doubling times calculated from all growth curves plotted above (panels **a**-**c**). **e**, Nucleosome occupancies determined from MNase-seq experiments with 0nM, 15nM, or 1000nM β-E induction were calculated after normalizing across samples for read depth and then plotted for the intact *LYS2* ORF (gray), *pZEVi* promoter (green), and native *LYS2* terminator (red) in scale with the schematic above. To emphasize changes in read-normalized nucleosome occupancies at each β-E concentration, the lower two panels plot differences (Δ) relative to 0nM as indicated. Bent arrow in *pZEVi* denotes position of the transcriptional start site (TSS), and straight arrow above *LYS2* denotes the site where insertions begin in target strains. The inferred positions of 23 nucleosomes within the *LYS2* ORF are indicated with black spots below the top panel coverage traces. Relative percentages of read-normalized nucleosome occupancies were also calculated (15nM / 0nM or 1000nM / 0nM) for each of the 23 nucleosomes (see Supplementary Data 1). All data plotted represent means calculated from n = 3 biological replicates. Purple line above *LYS2* and downstream of the target insertion site delineates the amplicon generated in qRT-PCR (see panel **g**). **f**, Box and whisker plot of read-normalized occupancies for *LYS2* ORF nucleosomes are presented for each β-E concentration in terms of a Log2 Fold Change (Log2FC). Log2FC values were calculated relative to the mean read-normalized nucleosome occupancy measured across chromosome II (chr.II) at each concentration (denoted with a black dotted line). Gray dots represent mean Log2FC values calculated for each of the 23 nucleosomes from n = 3 biological replicates. Significant differences in Log2FC across the entire *LYS2* ORF at each β-E concentration were determined with a Tukey's test. **g**, qRT-PCR experiments to determine the fold change in *LYS2* RNA levels for MNase-seq cultures at each β-E concentration. Fold changes were calculated relative to the lowest of three values measured at 0nM β-E, which was normalized to '1' (black dotted line). Considering that the maximal RNA levels previously obtained with a *pZEVi* promoter were ~ 100-fold increased after 2 h induction[62], *LYS2* transcript accumulation is likely already saturated at our timepoint. Error bars, mean ± s.d. (n = 3, biological replicates).

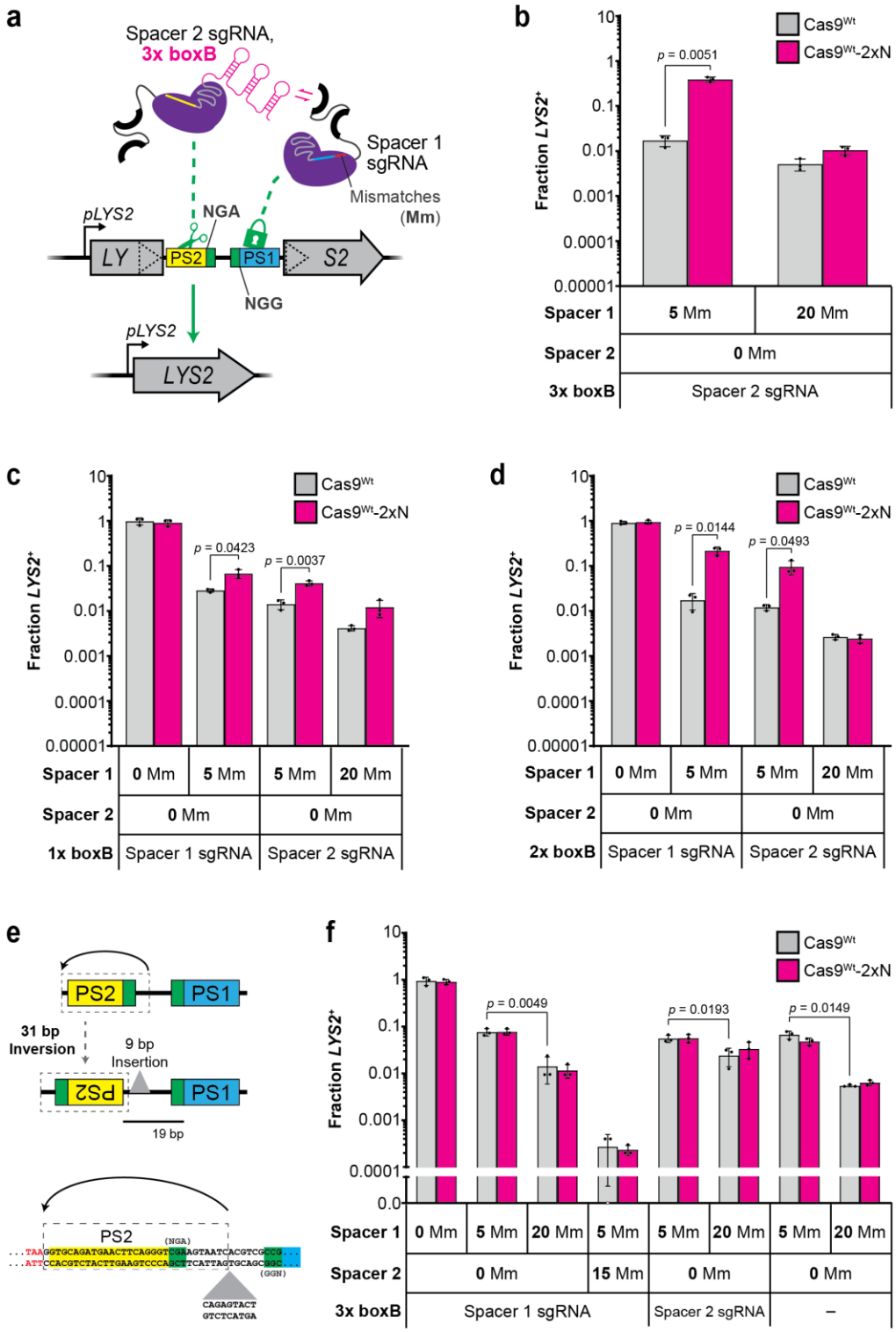


Extended Data Fig. 8 | Protospacer orientation can influence TraCT activity. a, Schematic summary of the 28 bp target inversion (dashed gray box) that includes the protospacer and 8 adjacent bp on its PAM side, including the NGG PAM, in SSA reporter strains with *pLYS2* and 1x or 3x *boxB* HTs upstream of the target. Distance in bp between the protospacer and nearest upstream *boxB* HT is also shown. **b**, Transformation-associated editing fractions measured for the indicated PID variants of native Cas9 or Cas9-2xN in strains with the inverted (+) or parental (–) protospacer orientation and various HT configurations. Error bars, mean ± s.d. (n = 3, biological replicates). **c**, Schematic summary of the *tetO2* relocation and protospacer inversion tested in SSA reporter strains with *pLYS2*. Top, 21 bp that included *tetO2* were relocated (dashed gray box) immediately upstream of the protospacer while 13 bp were inserted between *tetO2* and the protospacer to preserve spacing relative to the original *tetO2* (Dwnstr.) strain design; deletion of 3 bp at positions 9 through 11 downstream of the protospacer was intended to partially offset the SSA reporter insertion's overall increase in length. Bottom, 36 bp that included 8 bp on each side of the protospacer were inverted (dashed gray box) relative to the relocated *tetO2* (Upstr.) design, thus preserving 13 bp of spacing between *tetO2* and the protospacer. **d**, Transformation-associated editing fractions measured for Wt and MT3 variants of native Cas9 (Ø) or Cas9-scTetR in panel **c** strains as indicated, with the same experimental setup used for Fig. 1b; Wt and MT3 data collected with the original *tetO2* (Dwnstr.) strain is replotted from Fib. 1b for comparison. Error bars, mean ± s.d. (n = 6 or n = 3, biological replicates).



Extended Data Fig. 9 | Additional summary and supporting experiments for Fig. 5, including tests with ectopically expressed TetR proteins in SSA reporter strains. **a**, One-dimensional protein schematics diagramming features of the nuclear-localized mCherry-2xN (mC-2xN) and TetR proteins expressed from the ectopic *YKL162C-A* locus on chromosome XI, using one of three promoters as indicated above the bent arrow. C-terminal protein sequences beyond mCherry in mC-2xN are identical to sequences beyond Cas9-2xN's PID and include the 58 aa linker; the same

abbreviations from Extended Data Fig. 1 are used here. **b**, mCherry fluorescence measured for Fig. 5b yeast strains expressing mC-2xN from each of the indicated promoters; background fluorescence measured with the isogenic strain lacking an mC-2xN expression cassette is also plotted (Ø). Mean values were calculated from $n = 3$ (biological replicates) populations of single yeast cells cultured in parallel and analyzed by flow cytometry on the same day. Error bars, mean ± s.d. **c**, Transformation-associated editing fractions measured with the same strains and conditions tested in Fig. 5b but after delivery of $Cas9^{Wt}$ or $Cas9^{Wt}$-2xN. For comparison, data shown for the strain lacking mC-2xN (–) are replotted from Extended Data Fig. 3c (NGG, 1x *boxB*). Error bars, mean ± s.d. ($n = 3$, biological replicates). **d**, Schematic illustration summarizing the expected effect of Dox induction in strains containing $Cas9^{MT3}$-scTetR, *trans*-expressed TetR, and a protospacer-proximal *tetO2* element. **e**, Transformation-associated editing with native $Cas9^{MT3}$ or the $Cas9^{MT3}$-scTetR fusion in strains with a protospacer-proximal *tetO2* element and TetR absent (–) or expressed in *trans* at the indicated levels (+/++/+++). Measurements were performed as in Fig. 1b; MT3 data with strains lacking *trans*-expressed TetR are replotted from there for comparison. Error bars, mean ± s.d. ($n = 6$ or $n = 3$, biological replicates). **f**, Same experimental setup and strains tested in panel **e** but with $Cas9^{Wt}$ or $Cas9^{Wt}$-scTetR; the strain lacking *trans*-expressed TetR (–) is replotted from Wt data of Fig. 1b for comparison.



Extended Data Fig. 10 | Testing of alternative sgRNA configurations and target orientations with the paired-protospacer system for recruitment via anchored boxB hairpins. a, Schematic summary of the same SSA reporter strain used in Fig. 6 but tested with an alternative sgRNA configuration where Spacer 2 is fused to the 3x

boxB scaffold. **b**, Transformation-associated editing fractions measured for $Cas9^{Wt}$-2xN or native $Cas9^{Wt}$ with the sgRNA configuration of panel **a** and 5 or 20 Mm in Spacer 1 as indicated. Error bars, mean ± s.d. (*n* = 3, biological replicates). **c**, Transformation-associated editing fractions measured as in panel **b** but with a single boxB hairpin fused to the indicated sgRNA scaffold and spacer configurations. **d**, Same experimental setup as in panel **c** but with two boxB hairpins fused to sgRNA scaffolds. **e**, Schematic summary of the 31 bp target inversion (dashed gray box) that includes Protospacer 2 (PS2) and 10 adjacent bp on its PAM side, including the NGA PAM. A 9 bp insertion (gray triangle) was also introduced to preserve the 19 bp spacing between PS2 and Protospacer 1 (PS1). Sequence of the original strain is shown 5' to 3' at the bottom of the schematic, beginning with the upstream TAA stop codon (red text) and ending with the NGG PAM of PS1 (blue highlight). **f**, Same plasmids and experimental setup tested in panel **b** and Fig. 6b but with the PS2-inverted strain as summarized in panel **e**.