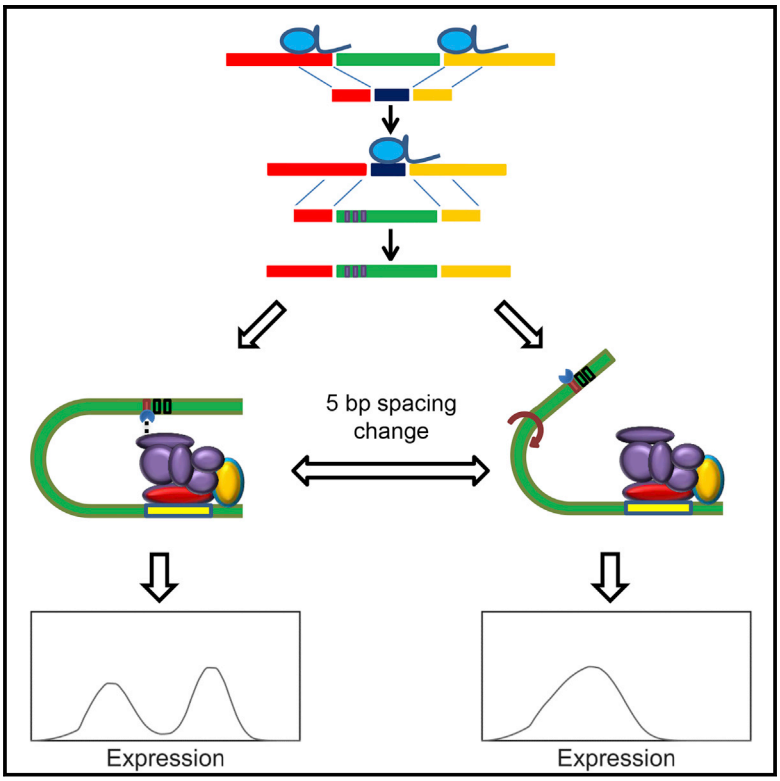
Cell Reports

Resource

A Precise Genome Editing Method Reveals Insights into the Activity of Eukaryotic Promoters

Graphical Abstract



Authors

Gregory L. Elison, Ruijie Song, Murat Acar

Correspondence

murat.acar@yale.edu

In Brief

Current methods to introduce specific genetic edits result in undesired changes at targeted genomic locations. Elison et al. describe a CRISPR-based technique facilitating scarless editing of endogenous genomic locations in yeast. Applying the technique to galactose network promoters demonstrates that spacing between key promoter elements is critical for activity.

Highlights

- CRISPR-based method allows scarless editing in yeast cells
- The method facilitates in vivo assessment of genotype-phenotype relationships
- Endogenous promoters are edited and the effect on promoter activity is measured
- Spacing between key promoter elements is found to be crucial for promoter activity

Elison et al., 2017, Cell Reports *18*, 275–286
January 3, 2017 © 2017 The Author(s).
http://dx.doi.org/10.1016/j.celrep.2016.12.014

CellPress

Cell Reports
Resource

OPEN ACCESS
CellPress

A Precise Genome Editing Method Reveals Insights into the Activity of Eukaryotic Promoters

Gregory L. Elison,[1,2] Ruijie Song,[2,3] and Murat Acar[1,2,3,4,5,*]
[1]Department of Molecular Cellular and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06511, USA
[2]Systems Biology Institute, Yale University, 850 West Campus Drive, West Haven, CT 06516, USA
[3]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, 300 George Street, Suite 501, New Haven, CT 06511, USA
[4]Department of Physics, Yale University, 217 Prospect Street, New Haven, CT 06511, USA
[5]Lead Contact
*Correspondence: murat.acar@yale.edu
http://dx.doi.org/10.1016/j.celrep.2016.12.014

SUMMARY

Despite the availability of whole-genome sequences for almost all model organisms, making faithful predictions of gene expression levels based solely on the corresponding promoter sequences remains a challenge. Plasmid-based approaches and methods involving selection markers are not ideal due to copy-number fluctuations and their disruptive nature. Here, we present a genome editing method using the CRISPR/Cas9 complex and elucidate insights into the activity of canonical promoters in live yeast cells. The method involves the introduction of a novel cut site into a specific genomic location, followed by the integration of an edited sequence into the same location in a scarless manner. Using this method to edit the *GAL1* and *GAL80* promoter sequences, we found that the relative positioning of promoter elements was critically important for setting promoter activity levels in single cells. The method can be extended to other organisms to decode genotype-phenotype relationships in various gene networks.

INTRODUCTION

In the few years since its development, the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing technique has been extensively used for the genetic modification of a variety organisms (Cong et al., 2013; DiCarlo et al., 2013; Jiang et al., 2013; Jao et al., 2013; Mali et al., 2013; Wang et al., 2014), including the budding yeast *S. cerevisiae* (Bao et al., 2015; DiCarlo et al., 2013; Horwitz et al., 2015; Mans et al., 2015; Ryan et al., 2014). However, while the technique has been successful, it fails to edit the genome in a scarless manner. For successful editing to take place, the final sequence must have either the protospacer adjacent motif (PAM) sequence or the targeting sequence altered (Horwitz et al., 2015; Mans et al., 2015). Papers and protocols acknowledge that they can only achieve scarless genome editing if the desired edit happens to disrupt the targeting or PAM sequence of the guide RNA (gRNA) being used (Mans et al., 2015; Ryan et al., 2016). This has not been a problem for most applications involving proteins, because these edits can be made in such a way that they do not change the resulting polypeptide (Mans et al., 2015). However, when investigating less well-defined genomic regions, such as promoters, in which the effects of minor base pair changes are unknown, there is a great need for a truly scarless version of CRISPR, which introduces only desired edits with no unwanted changes across a fairly large region of DNA (Ryan et al., 2014). This problem has led to a dearth of in vivo promoter studies aiming to dissect genotype-phenotype relationships, as introducing a variety of scarless changes has previously been too time consuming.

Despite the existence of other editing methods, even in *S. cerevisiae*, a model system known for its amenability to genomic editing, there are numerous problems when attempting to make scarless edits. Homologous recombination, the well-known method to introduce novel genetic material into yeast, is dependent on the presence of a selectable marker to increase editing efficiency to a usable range. This marker is a large scar in the genome, which may have a variety of unknown effects on sensitive genomic functions. In addition, the scarless method known as Uracil-5-Fluoroorotic acid (URA-FOA) counter-selection (Boeke et al., 1987) has such low efficiency that finding a correctly edited yeast strain takes a significant amount of time.

To address this challenge, we have adapted the CRISPR genome editing technique to modify genomic regions with no undesired base pair changes. We use CRISPR/Cas9 to create double-strand breaks flanking a region of interest to remove it from the genome and introduce a novel targeting sequence and PAM site in its place. This new site is then cut in a second round of editing, in which the original region is reintroduced, along with any desired modifications. This two-step editing method allows any desired edits to be introduced across the genomic region in question while maintaining all other base pairs as they are in the parent strain. In addition, once the intermediate strain is created, it may be reused indefinitely; thus, every subsequent new edit to the same genomic region only requires one transformation instead of two.

We used this method for the in vivo editing of the canonical *GAL1* promoter in yeast. We systematically removed or recoded the Gal4 binding sites from the promoter to study the phenotypic

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

consequences of these changes at the single-cell level. The activity from the edited promoter architectures was compared to the bimodal activity profile of the wild-type *GAL1* promoter, and deviations from the wild-type behavior were analyzed in terms of the fraction of "on" cells and the expression level of the "on" state. We found that the fourth binding site does not have any effect on transcriptional activity. Removing or recoding the first site prevented any activity from the promoter despite the presence of the second and third binding sites. However, removing the third Gal4 binding site (together with the first one) partially restored the wild-type activity in the *GAL1* promoter. Using our method to edit the *GAL80* promoter, analyzing the activity of the edited *GAL80* promoter architectures, and interpreting the results in the context of the results from the edited *GAL1* promoters supported the conclusion that the relative positioning of promoter elements is of critical importance for determining in vivo promoter activity levels at endogenous chromosomal locations. This was confirmed by edits; these changed only the spacing between the Gal4 binding sites and the TATA box, which resulted in expression level changes only attributable to spacing changes within the promoter.

RESULTS

Scarless Genome Editing in Live Cells

To demonstrate the viability of this technique, we chose to use a strain of *S. cerevisiae* in which one copy of the canonical *GAL1* promoter driving the yellow fluorescent protein (YFP) has been integrated into the *ho* locus (Acar et al., 2005, 2010) (Figure 1A) (Table S1). By targeting this promoter, we could see the influence of any edits on the output of YFP fluorescence in environments containing various concentrations of galactose.

The first step of the genome editing process was to replace the existing P_{GAL1} with a short (30 bp) region containing a novel CRISPR cut site (Figure 1A). To do this, gRNAs targeting the nearest available cut sites immediately before and after the *GAL1* promoter were added to a pRS315 plasmid containing a yeast centromeric element and a *LEU2* marker. In addition, a donor oligonucleotide was created containing the novel cut site flanked by 50 bp of homology to each of the regions immediately adjacent to the cut sites described earlier. The plasmid and donor oligonucleotide were co-transformed into a yeast strain containing *CAS9* driven by a constitutive promoter, as well as the P_{GAL1}-YFP construct described earlier. Because of the relative proximity of the cut sites, the cellular repair machinery treats the two as a single break, which is repaired through the use of the added donor oligonucleotide while the intervening region is lost. The result of this step is the replacement of the original P_{GAL1} and small adjacent regions, including the original PAM sites, with a 30 bp novel DNA sequence containing a new CRISPR cut site.

The second step of the editing process begins upon creation of the new strain described earlier. This involves inserting a gRNA targeting the introduced PAM site into a pRS315 plasmid and co-transforming it into the new yeast strain, along with a new piece of donor DNA, which consists of the original promoter containing any desired edits and 50 bp of homology to each end of the new break (Figure 1B). The repair replaces the adjacent regions, which were lost in the previous step and adds the newly edited P_{GAL1} region. Ultimately, the only changes to the original strain are those desired. The targeting and the PAM sites targeted in the first step were reintroduced such that no trace of the editing process remains in the final genome (Figure 1B).

Characterization of Editing Efficiency

We measured the editing efficiency for both steps of our technique (Figure 1C), because efficiency is a key factor for any genome editing technique. The initial step had an efficiency of 63%. The first step must only be completed once for any editing to be conducted at the relevant locus. The efficiency of the second editing step is more important, and this step was found to have an efficiency of 77% (Figure 1C). This editing efficiency is comparable to the efficiencies reported in literature (Horwitz et al., 2015; Mans et al., 2015).

Editing *GAL1* Promoter Architecture and Measuring Promoter Activity in Single Cells

To functionally validate our technique, we decided to explore the phenotypic consequences of introducing rational edits to a promoter. For this, we investigated the effects of removing Gal4 binding sites from the canonical *GAL1* promoter, which is a faithful reporter of the yeast galactose (GAL) network (Figure S1) activity (Acar et al., 2005). The constitutively expressed Gal4 protein serves as the main transcriptional activator of the GAL network (Mizutani and Tanaka, 2003). It activates transcription from all network promoters, including that of the *GAL1* promoter, by binding to a 17 bp site (CGG-N(11)-CCG) on the promoters (Johnston, 1987). However, Gal80 repressor proteins negatively regulate the network activity by binding to Gal4 on promoter sites (Mizutani and Tanaka, 2003) (Figure 2A). Galactose relieves Gal80 repression via Gal3 proteins. Activated Gal3 binds to Gal80, leading to the dissociation of Gal80 from Gal4 proteins (Peng and Hopper, 2002). There is no known crosstalk between the GAL network and any other gene except for an inhibitor that is dependent on the presence of glucose, making it ideal for study in conditions without the presence of glucose.

The GAL network is induced by galactose (St John and Davis, 1981) in a bimodal fashion (Acar et al., 2005). To use as a reference, we measured the activity of the unedited wild-type *GAL1* promoter driving YFP. For this, we grew yeast cells for 22 hr in the presence of various galactose concentrations (0%, 0.025%, 0.05%, 0.1%, 0.2%, 0.4%, and 0.5%), in addition to 0.1% mannose. Mannose was used as a non-inducing sugar that is metabolized by both "off" and "on" cells of the bimodal distribution. At the end of the 22 hr period, single-cell YFP expression levels of ~3,000 cells were measured using a flow cytometer (fluorescence-activated cell sorting [FACS]), and two important phenotypes associated with a bimodal distribution were quantified for each condition: fraction of "on" cells and mean expression level of "on" cells (Figure 2B).

Altering the architecture of the *GAL1* promoter can potentially change the fraction of "on" cells, the mean expression level of the "on" state, or both. The native *GAL1* promoter (Johnston and Davis, 1984; Kellis et al., 2003) contains three Gal4 binding sites immediately adjacent to one another, followed by a fourth binding site with a different sequence 45 base pairs downstream

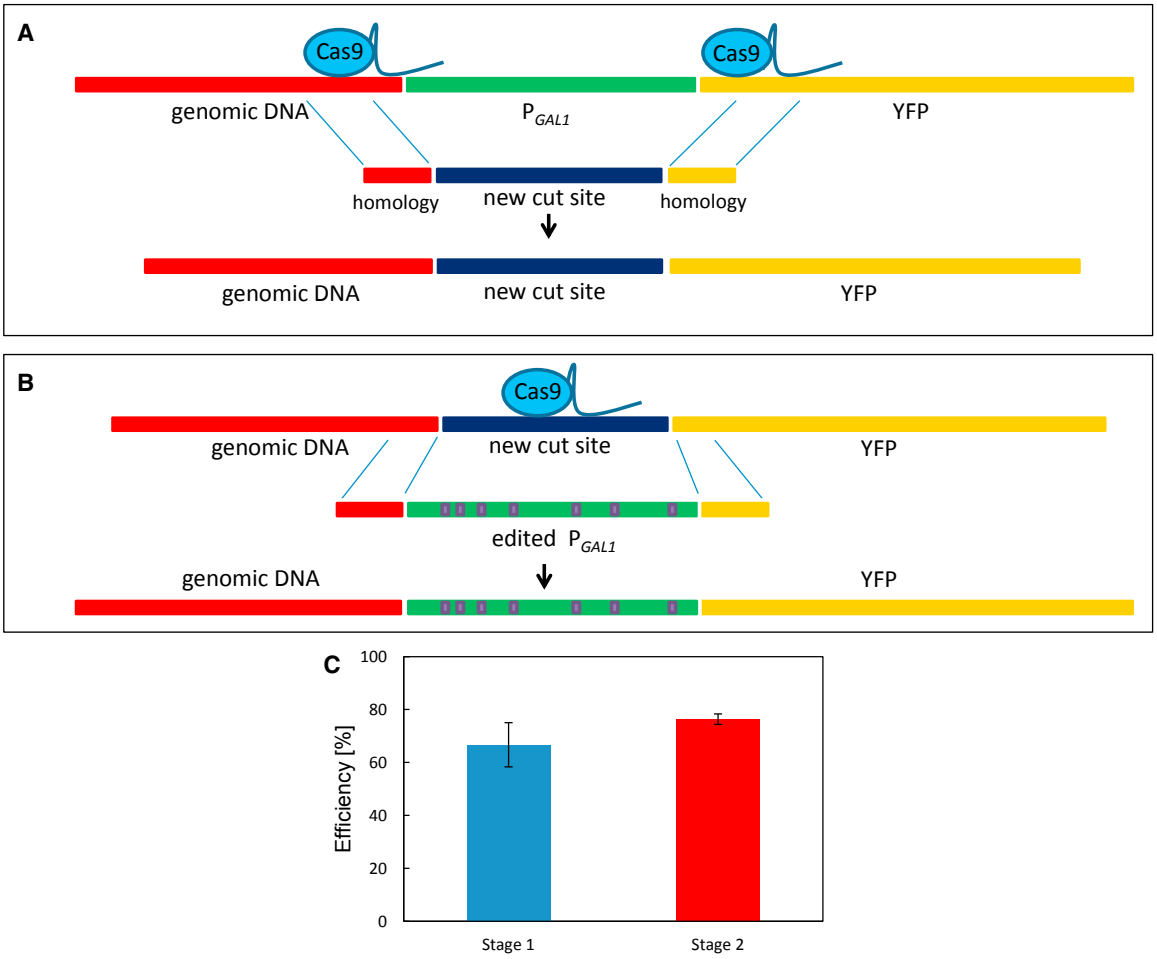


Figure 1. The Two-Step CRISPR Editing Method

(A) Introducing a novel CRISPR cut site to the *GAL1* promoter of the P_{GAL1}-YFP construct integrated in the *ho* locus of the yeast genome. P_{GAL1} is flanked by the YFP gene downstream and genomic DNA upstream. Two gRNA/Cas9 complexes target cut sites immediately adjacent to P_{GAL1}, resulting in its loss from the genome. The addition of a donor oligonucleotide with a novel CRISPR cut site flanked by regions of homology to the upstream genome and YFP allows repair. The repair results in a new region in which the old cut sites have been removed and a novel one is inserted.

(B) Introducing an edited P_{GAL1} oligonucleotide back to the original location. The novel CRISPR cut site is cut by a gRNA/Cas9 complex. Repair is facilitated by the addition of a donor oligonucleotide consisting of an edited P_{GAL1} and homology to the upstream genome and YFP. The repair gives rise to an edited P_{GAL1} upstream of YFP.

(C) Efficiency of CRISPR editing. Of the colonies tested for CRISPR editing, 63% contained the correct sequence after the first stage of editing for P_{GAL80} and 77% contained the correct sequence after the second stage of editing for P_{GAL1}.

Error bars indicate SEM (n = 2 for stage 1; n = 3 for stage 2).

of the previous three (Figure 2B). The 185-190 base pairs downstream of the fourth site form a TATA box, and the 5′-ATG-3′ start codon is 145 base pairs downstream of the TATA box (Figure 2B). We altered the four Gal4 binding sites in two ways: by simply removing the 17 bp binding site from the promoter outright and by recoding the conserved 5′-CGG-3′ and 5′-CCG-3′ sequences of the site into randomized A/T sequences of the same length, abolishing the binding of Gal4 on that position while keeping the number of base pairs intact.

To investigate the phenotypic consequences of these binding site alterations on the activity of the *GAL1* promoter, we performed galactose induction experiments and measured the resulting promoter activity profiles in single cells. Figure 2C shows results from the wild-type strain, while Figure 2D shows, as an example, results from one of the promoter-edited strains. In the edited strain, the first and second Gal4 binding sites were recoded to neutral sites on the *GAL1* promoter. Compared to the wild-type promoter activity, the recoding of these two binding sites resulted in a reduction of the "on" state mean expression level but a nearly identical fraction of "on" cells. These results verify that edits made using this method can produce phenotypic changes in vivo.

Editing the *GAL80* Promoter Sequence

To move the phenotypic characterization of our editing approach to the gene network level and to test the approach on the editing

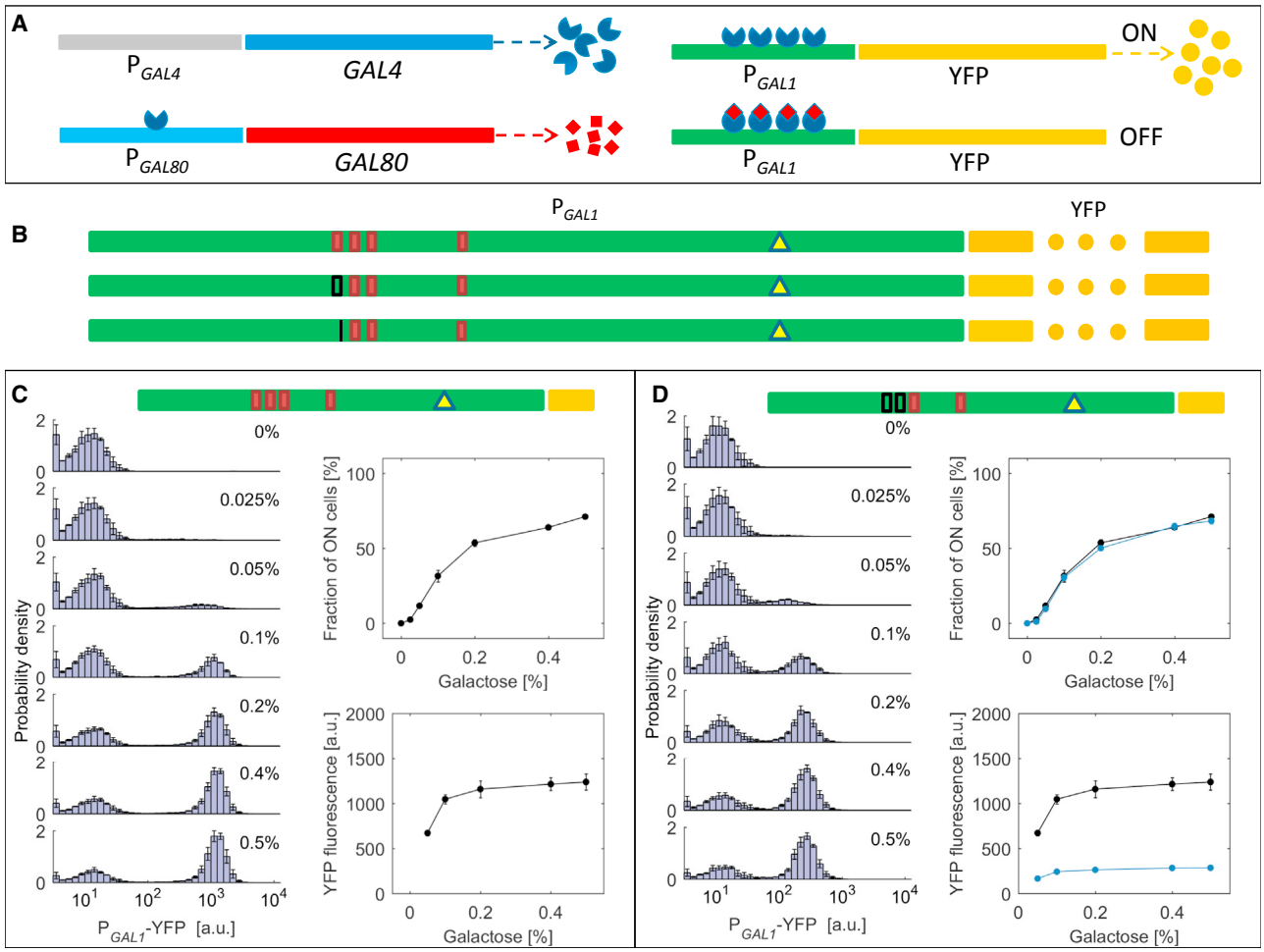


Figure 2. Components of the GAL Network and Characterization of *GAL1* Promoter Activity

(A) Relevant components of the yeast GAL network. *GAL4* is constitutively expressed and activates P_{GAL80} and P_{GAL1}. Gal80 inhibits the activity of Gal4, thus inhibiting both its own and P_{GAL1} expression.

(B) Architecture of the *GAL1* promoter. The first line represents the wild-type *GAL1* promoter. The red boxes are Gal4 binding sites and the yellow-filled triangle is the TATA box. The second line represents a *GAL1* promoter in which the first Gal4 binding site has been recoded (empty black box). The third line represents a *GAL1* promoter in which the first Gal4 binding site has been removed entirely (black line), shortening the length of the promoter by 17 bp.

(C) Phenotypic characterization of wild-type P_{GAL1}-YFP activity. The bar at the top indicates that the data refer to the wild-type architecture of the *GAL1* promoter. The seven histograms show the flow cytometry data at seven concentrations of galactose as indicated, from 0% to 0.5%. The top right panel shows the fraction of "on" cells as galactose concentration increases, and the bottom right panel shows the mean expression level of the "on" state for concentrations with at least 5% "on" cells.

(D) Phenotypic characterization of an edited *GAL1* promoter used in P_{GAL1}-YFP. The bar at the top indicates that the data refer to an edited *GAL1* promoter with the first and second Gal4 binding sites recoded. The top and bottom right panels show the fraction of "on" cells and mean "on" expression levels, respectively, as galactose concentration increases (blue). The data from the wild-type *GAL1* promoter (black) are included for comparison purposes. The mean "on" expression level plots only include galactose concentrations at which both strains plotted have at least 5% "on" cells.

Error bars indicate SEM (n = 3). See also Figure S1 and Table S1.

of a promoter functioning in its natural location, we next introduced rational edits on the endogenous *GAL80* promoter, which contains a single Gal4 binding site (Figure 3A, red box). For this purpose, we constructed four strains carrying different edits on the *GAL80* promoter and measured the resulting activity of the GAL network using YFP driven by the wild-type *GAL1* promoter.

In the first strain, we inserted a new Gal4 binding site immediately before the existing one (Figure 3B) on the *GAL80* promoter, while in the second strain we recoded the existing base pairs before the existing Gal4 binding site into a second Gal4 binding site (Figure 3C), a change doubling the number of binding sites but not increasing the length of the promoter. Both strains showed a massive increase in the fraction of "on" cells—in the recoded strain, virtually all cells were in the "on" state even in the absence of galactose. Because Gal80 is a repressor protein, these findings indicate that the alterations we made on the *GAL80* promoter greatly reduced or eliminated *GAL80*

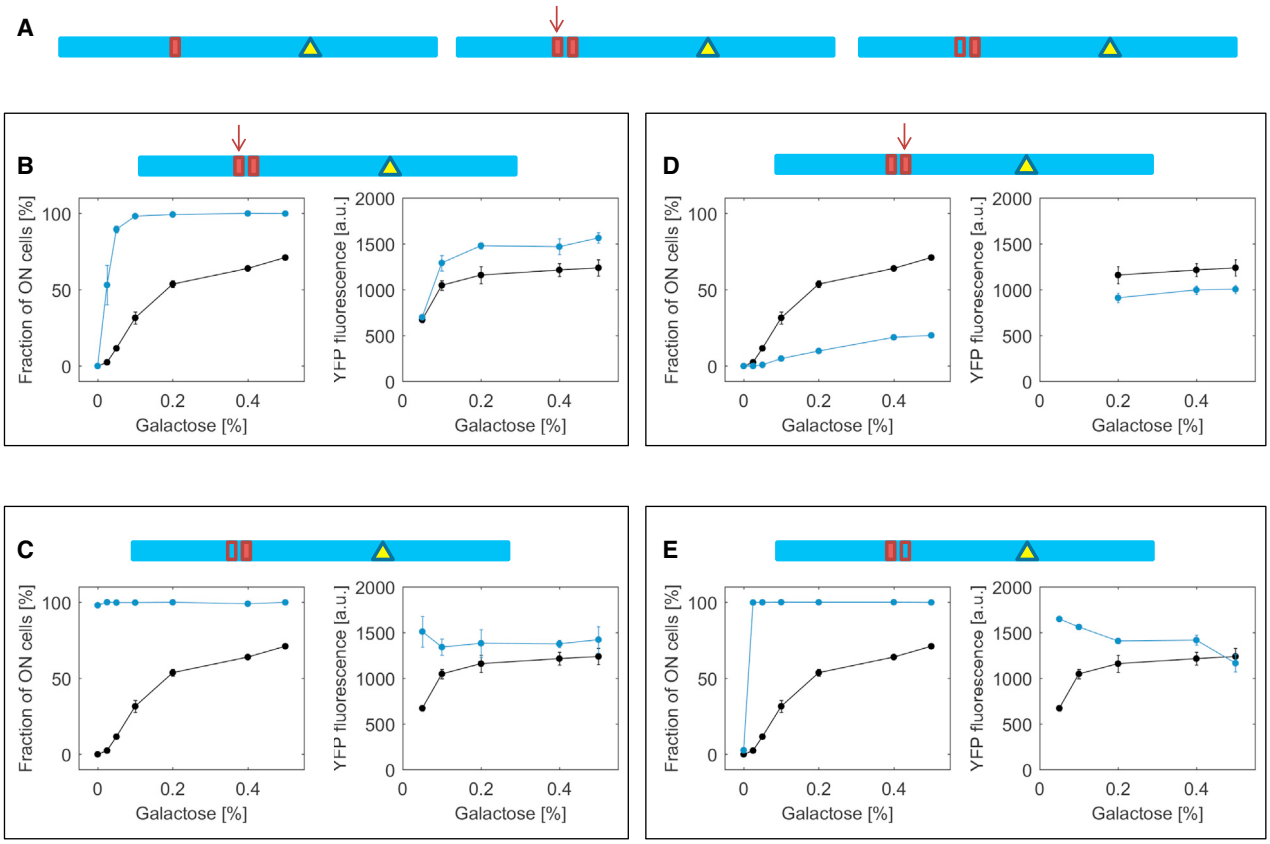


Figure 3. Editing the *GAL80* Promoter and Phenotypic Characterizations

(A) Cartoons illustrating promoter elements and edits on the *GAL80* promoter. The first depicts the wild-type *GAL80* promoter with a single Gal4 binding site (red box) and a TATA box (yellow triangle). The second shows a second Gal4 binding site (red arrow) inserted before the original site. The third has a recoded Gal4 binding site (empty red box) before the original site.

(B–E) Each horizontal bar indicates one strain and illustrates the specific edits introduced in the *GAL80* promoter. The filled red boxes are Gal4 binding sites, while recoding an existing sequence to a Gal4 binding site is depicted by an empty red box. A pointing red arrow indicates the second Gal4 binding site inserted. Wild-type data are plotted in black as the reference, while the data of each bar-illustrated strain are shown in blue. Each panel depicts either the fraction of "on" cells as a percentage or the mean expression level of the "on" cells (a.u.), reported by the unedited P_{GAL1}-YFP. The mean "on" expression level plots only include galactose concentrations at which both strains plotted have at least 5% "on" cells.

Error bars indicate SEM (n = 3). See also Table S1.

expression, thus allowing the *GAL1* promoter to be activated in a larger fraction of cells.

In the third strain, we added a new Gal4 binding site immediately downstream of the existing one, while in the fourth strain, we recoded the existing base pairs downstream of the existing Gal4 binding site into a second Gal4 binding one. Compared to the unedited strain, the strain with one Gal4 binding site added downstream of the original site showed a reduction in the fraction of "on" cells while keeping the mean YFP expression levels of the "on" cells roughly the same (Figure 3D). This is in contrast to the results obtained from the recoded strain (Figure 3E), which resembled the initial two strains (Figures 3B and 3C) discussed earlier in that the expression of *GAL80* seemed to be eliminated at all galactose concentrations used. This result from the promoter containing an additional downstream binding site (Figure 3D) indicates higher *GAL80* expression levels compared to wild-type. These results suggest that while binding site number is clearly important for setting promoter expression level or strength, it is not the only factor. The location of binding sites within the promoter is also of utmost importance. To give a specific example, when two Gal4 binding sites are adjacent to each other (Figures 3B–3E), their relative positioning to the other promoter elements determines whether transcription will proceed (Figure 3D) or be blocked (Figures 3B, 3C, and 3E). Future studies using single-molecule imaging techniques may directly elucidate how such transcriptional inactivation may occur.

Introducing Simultaneous Edits on the *GAL1* and *GAL80* Promoters

To further investigate our findings and expand the applicability of our method, we set out to simultaneously introduce edits on two genomic locations: *GAL80* promoter at its endogenous location and *GAL1* promoter driving YFP at the *ho* locus. To do this, we first constructed a strain containing novel targeting and PAM sites replacing both *GAL1* and *GAL80* promoters. We then transformed this strain with a plasmid containing two gRNAs, one for

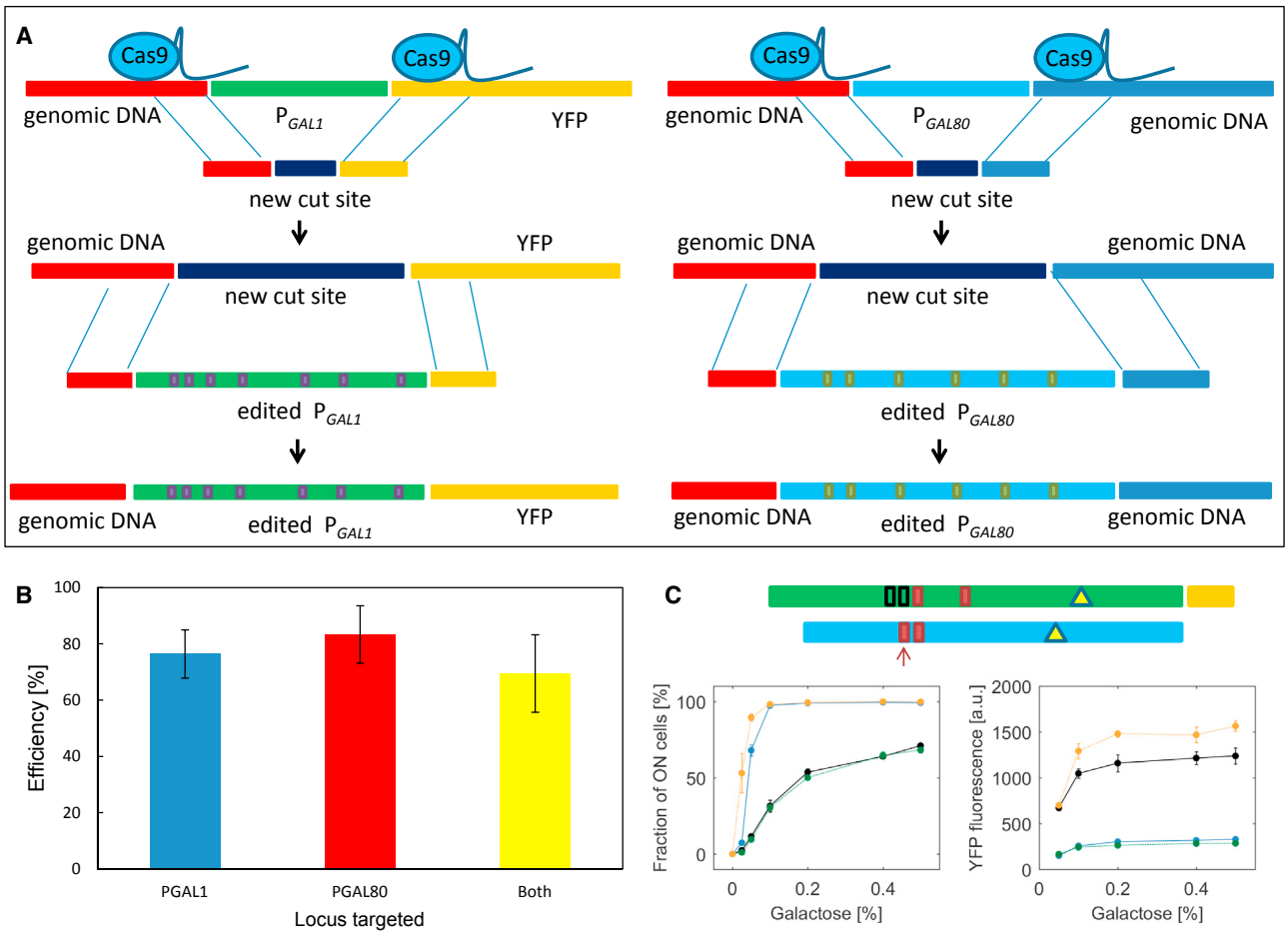


Figure 4. Introducing Simultaneous Edits into the *GAL1* and *GAL80* Promoters

(A) Cartoon diagram showing the two-step CRISPR editing technique. The second step of the technique introduces two edited sequences (*GAL1* and *GAL80* promoters) simultaneously.

(B) Editing efficiency for the second step of the technique, where 76% of the colonies had the correct *GAL1* promoter edits, 83% had the correct *GAL80* promoter edits, and 69% were correct for both *GAL1* and *GAL80* promoters.

(C) Phenotypic characterization of the strain carrying the dual promoter edits. Each horizontal bar indicates the *GAL1* (green) or *GAL80* (sky blue) promoter and illustrates the specific edits introduced in them. The filled red boxes are Gal4 binding sites, while recoding an existing sequence to a Gal4 binding site is depicted by an empty red box. An empty black box shows recoding of a previously existing Gal4 binding site to a null sequence with the same length. A yellow triangle indicates the TATA box. Wild-type data are plotted in black as the reference, while the data of the dual-bar-illustrated edited strain are shown in blue. The left and right panels depict, respectively, the fraction of "on" cells as a percentage or the mean expression level of the "on" cells (a.u.), reported by the edited P_{GAL1}-YFP. Green (orange) data points show the phenotypic levels measured from a strain carrying the same edits but only on the *GAL1 (GAL80)* promoter, not both. The mean "on" expression level plot only includes galactose concentrations at which all strains plotted have at least 5% "on" cells.

Error bars indicate SEM (n = 3). See also Table S1.

each promoter site, as well as donors carrying the edited promoters (Figure 4A). To find the combined efficiency of simultaneous editing at both genome locations, we examined editing efficiency for both locations individually and together. Using targeted sequencing, we found that 76% of the colonies were successfully edited at the *GAL1* promoter, 83% were successfully edited at the *GAL80* promoter, and 69% were successfully edited at both locations (Figure 4B). These efficiency levels are in line with efficiencies reported in yeast by CRISPR papers (Horwitz et al., 2015).

To measure the phenotypic consequences of the dual edits, we constructed a strain by combining two specific sets of edits previously introduced on *GAL1* (Figure 2D) and *GAL80* (Figure 3C) promoters. When introduced into the wild-type background, the *GAL1* promoter edits (the first and second binding site locations recoded) gave rise to "on" cell fractions similar to the wild-type strain, but the expression level of the "on" state was drastically reduced (Figure 2D, blue; Figure 4C, green). However, when introduced into the wild-type background, the *GAL80* promoter edit (recoding a second Gal4 binding site upstream of the original one) led to increases in the fraction of "on" cells and "on" state expression levels (Figure 3C, blue; Figure 4C, orange). Analyzing the fraction of "on" cells in the strain carrying the *GAL1* and *GAL80* dual edits (Figure 4C, left panel,

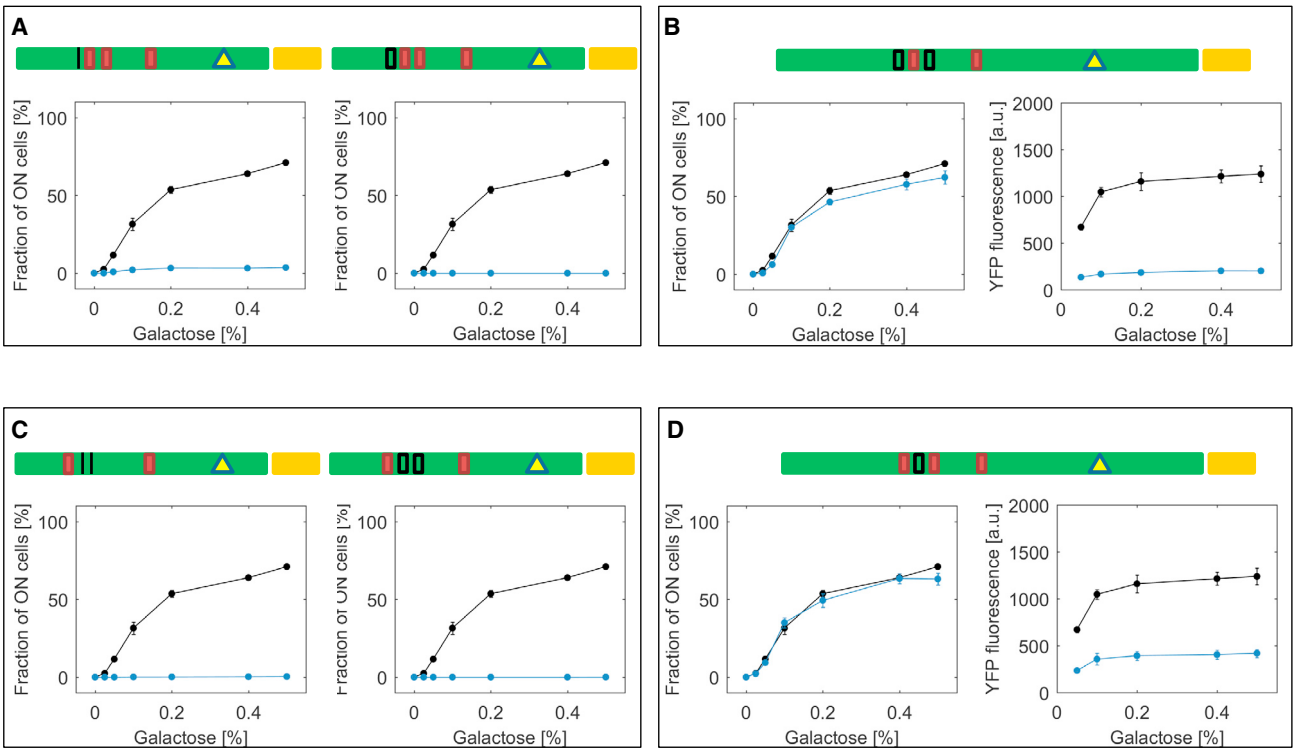


Figure 5. Systematically Editing the Canonical Gal4 Binding Sites on the *GAL1* Promoter

Each horizontal bar indicates one strain and illustrates the specific edits introduced in the *GAL1* promoter. The filled red boxes are Gal4 binding sites, while removing a Gal4 binding site by recoding is depicted by an empty black box. A yellow triangle indicates the TATA box, and the black line indicates the removal of a Gal4 binding site.

(A–D) Wild-type data are plotted in black as the reference, while the data of each bar-illustrated strain are shown in blue. Each panel depicts either the fraction of "on" cells as a percentage or the mean expression level of the "on" cells (a.u.). The mean "on" expression level plots only include galactose concentrations at which both strains plotted have at least 5% "on" cells.

Error bars indicate SEM (n = 3). See also Figure S3 and Table S1.

blue) showed us that the *GAL80* edit, which lowered *GAL80* protein expression, dictated the phenotype of the dual-edit strain. This was expected, because the phenotypic effect of the *GAL1* edits was neutral (Figure 4C, left panel, green). However, when analyzing the phenotype of mean "on" state expression levels, it was clear that the *GAL1* edit, which lowered the *GAL1* expression mean (Figure 2D, right panel, blue), dictated the phenotype of the dual-edit strain. This was also expected, because the phenotypic effect of the *GAL80* edits was considerably less intense than that of the *GAL1* edit (Figure 4C, right panel, orange).

Mechanistic Insights into the Differential Activities of the Edited *GAL1* Promoters

In an effort to further understand how the changes in promoter activity are caused by the addition or removal of Gal4 binding sites, we combinatorially removed or recoded each of the three main Gal4 binding sites on the P_{GAL1}-YFP construct and measured the resulting promoter activity profiles compared to the unedited promoter's activity. We first examined YFP profiles from the two strains in which all three of the immediately adjacent Gal4 binding sites were removed or recoded and found that these strains lacked any YFP expression (Figure S2A). In addition, when we removed or recoded the fourth binding site, we found that these changes did not have any effect on *GAL1* activity (Figures S2B and S2C). These results indicate that the fourth binding site is neither necessary nor sufficient for promoter activity. We thus concluded that this site was not relevant to the *GAL1* promoter's activity and treated it as such for the remainder of this study. The difference of this site from the canonical motif (Supplemental Information) supports our phenotypically validated conclusion.

Focusing our attention on the first three canonical Gal4 binding sites, we next examined the promoter-edited strains to determine whether removing or recoding the first Gal4 binding site (by itself or together with the second site) had an effect on transcriptional activity from the *GAL1* promoter. Similar to the results observed by editing the *GAL80* promoter (Figure 3), removing or recoding the first site prevented any activity from the promoter despite the presence of the remaining two binding sites (Figure 5A). More surprising was the observation that recoding both the first and third Gal4 binding sites restored the wild-type activity in the promoter in terms of the fraction of "on" cells but not in terms of the expression level of the "on" state (Figure 5B). Removing the first and third binding site locations had the same effect (Figure S3A). Similar phenotypes were seen for

other *GAL1* promoter edits (Figures S3B and S3C). Removing or recoding the third binding site prevented activity from the promoter despite the presence of the other two binding sites (Figures S3B and S3C). While the hypothesis that the number of transcription factor binding sites on a promoter is the key determinant of promoter activity was initially appealing, results from three promoter architectures contradicted this idea. Two of these represent promoter designs in which the second and third binding sites had both been either removed or recoded, leaving only the first Gal4 binding site intact (Figure 5C). According to the binding site number hypothesis, the activity from these two promoters should have been similar to the activity from those promoters carrying only the second Gal4 binding site intact (Figure 5B; Figure S3A). Instead, having only the first Gal4 binding site intact resulted in the abolishment of all activity from the promoter, while having only the second Gal4 binding site resulted in a promoter activity profile matching the wild-type in terms of the fraction of "on" cells. The other promoter architecture, which generated results contradictory to the binding site number hypothesis, was the one containing intact first and third Gal4 binding sites with the second site recoded (Figure 5D). Despite having two functional binding sites, the phenotypic characteristics of this promoter resembled those promoters (Figures 2D and 5B; Figure S3A) with only one functional binding site. These findings, together with the ones observed from the edited *GAL80* promoters (Figure 3), made us reject the hypothesis that the number of transcription factor binding sites was the sole key factor in setting promoter activity levels.

As an alternative hypothesis, we propose that the relative position of the Gal4 binding site or sites and other promoter elements, such as the TATA box, is an important parameter influencing the promoter activity levels. The fact (Dion and Coulombe, 2003) that the Gal4 binding site or sites are involved in the looping of DNA for a stable transcriptional initiation provides a mechanistic basis for this hypothesis. Accepting distance between important promoter elements as a key factor does not fully discard potential contributions from the number of binding sites on promoter activity. Removing a binding site automatically alters the relative distances between promoter elements, meaning that the parameters of binding site number and distance cannot be fully orthogonal. In addition, it is more than plausible to think that when crucial spacing needs between promoter elements are not met, the effects of transcription factor binding site number could be muted due to disruption of the spacing necessary for proper folding of DNA and transcriptional initiation.

Going back to our experimental observations, the hypothesis based on the relative positioning of promoter elements would explain why the two strains with only the first Gal4 binding site intact (Figure 5C) failed to be inducible. Having only one binding site was not the culprit, because having only the second binding site had produced a wild-type-like fraction of "on" cells (Figure 5B; Figure S3A). This indicates that compared to the phenotypic effect of the second binding site, having only the first site increased the distance between the site and the downstream promoter elements, leading to the deactivation of the promoter. The relative positioning hypothesis also explains the reactivation of the *GAL1* promoter when two Gal4 binding sites are separated by 20 bp (Figure 5D), as the promoter was inactive when two binding sites were separated by 2 bp (Figure 5A).

Systematic Alterations of Distance between Specific Promoter Elements

To confirm the hypothesis that the spacing of key promoter elements can have a drastic impact on promoter activity, we constructed three additional strains in which the distance between the third Gal4 binding site and the TATA box was shortened by 5, 10, or 17 base pairs. This was done by moving the 5, 10, or 17 bp long *GAL1* promoter region immediately downstream of the third Gal4 binding site to a location immediately upstream of the first Gal4 binding site (Supplemental Information). These strains were built off of the strain seen in Figure 2D, because we reasoned that having only one canonical binding site would make results easier to interpret. The resulting three strains therefore had the single intact Gal4 binding site moved 5, 10, or 17 bp closer to other promoter elements, such as the transcription start site (TSS) and the TATA box, and moved the same distance from the 5′ end of the promoter. The parent strain (Figures 2D and 6A) displayed a fraction of "on" cells virtually identical to that of the wild-type. In contrast, the strain in which the Gal4 binding site was moved 5 bp closer to the TSS and TATA box (Figure 6B) displayed an almost complete lack of promoter activity. However, the strain in which the Gal4 binding site was moved 10 bp closer to the TSS and TATA (Figure 6C) closely resembled its parent (Figure 6A). Even more interestingly, the strain in which the Gal4 binding site was moved 17 bp closer to the TSS and TATA box (Figure 6D) had a fraction of "on" cells similar to that of both its parent and the wild-type strain but had a higher mean expression level than its parent, even higher than the wild-type.

These results definitively show that the spacing between various promoter elements is crucial for promoter activity. The stark difference among strains with their Gal4 binding sites located only a few base pairs closer to or farther from the TSS and TATA box illustrates that a change in the relative position of various promoter elements is capable of almost eliminating activity from a promoter (Figure 6B), keeping activity almost the same (Figure 6C), or massively increasing expression compared to the parent strain (Figure 6D). The changes made to the number of Gal4 binding sites also resulted in small spacing changes, which we originally ignored, such as those in Figure 5C (right panel), but these are likely to be the cause of the loss of expression from the promoter. We hypothesize that these phenotypic changes can be explained by the positioning of the binding sites on the DNA helix. A helical turn of the DNA is known to be ~10.5 bp, so a change of 5 bp should place the Gal4 binding sites on the opposite side of the DNA compared to the original, potentially resulting in a loss of interaction between the DNA-bound Gal4 and the transcriptional machinery. However, a change of 10 bp would bring the Gal4-bound site back to its original position in three-dimensional space, potentially explaining why expression is restored. The increased expression level resulting from the 17 bp shortening is harder to fully explain, because the 17 bp change is expected to lead to both a partial helical turn and a change in position along the DNA strand. Still, this result reinforces the main hypothesis that the spacing of key

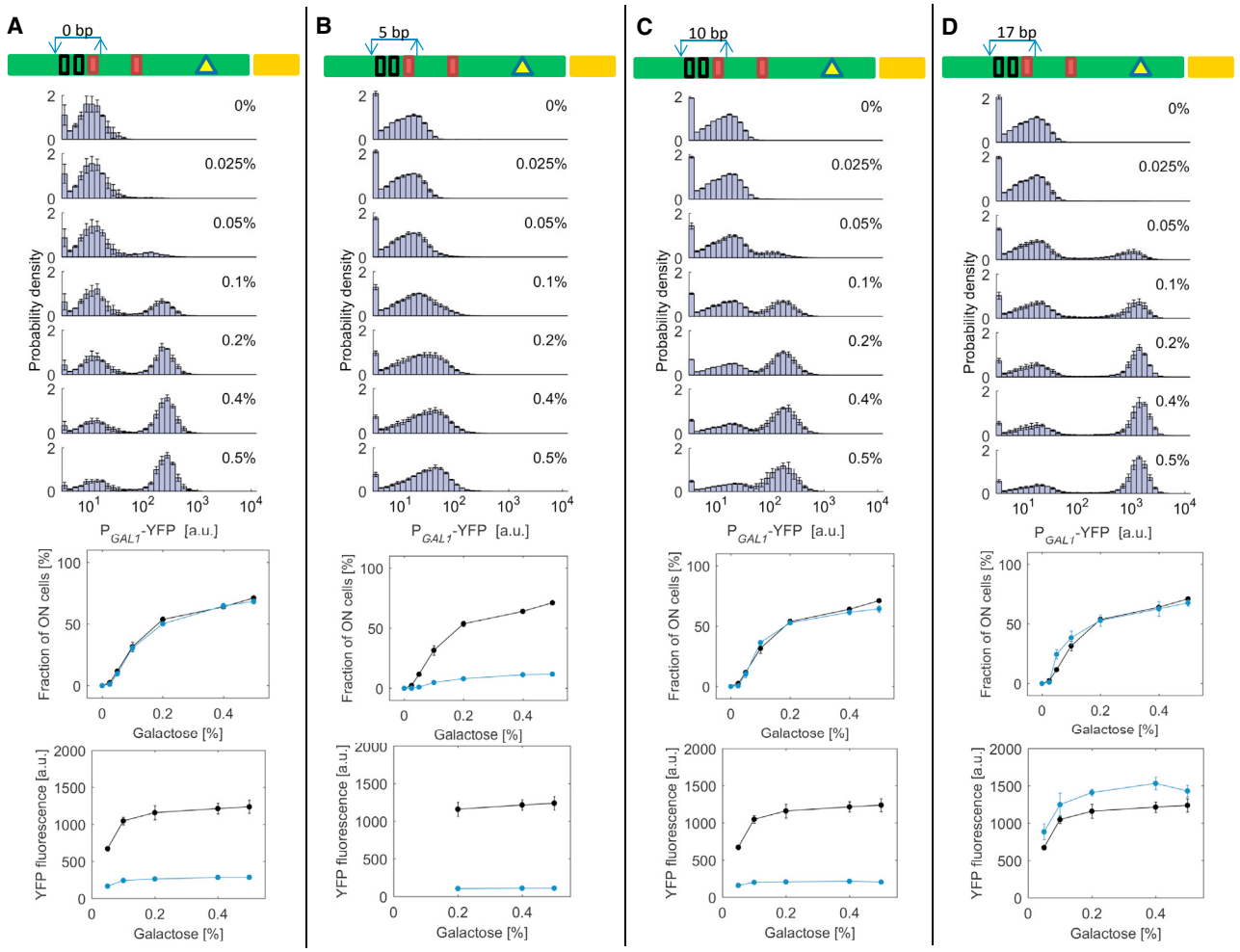


Figure 6. Phenotypic Characterization of Systematic Spacing Changes in the *GAL1* Promoter

Each horizontal bar indicates one strain and illustrates the specific edits introduced in the *GAL1* promoter. The filled red boxes are Gal4 binding sites, while removing a Gal4 binding site by recoding is depicted by an empty black box. A yellow triangle indicates the TATA box, and the black line indicates the removal of a Gal4 binding site. Arrows above the promoter indicate the number and position of base pairs moved from one location within the promoter to another. Wild-type data are plotted in black as the reference, while the data of each bar-illustrated strain are shown in blue. Histograms show the flow cytometry data at seven concentrations of galactose as indicated, from 0% to 0.5%. Each panel depicts either the fraction of "on" cells as a percentage or the mean expression level of the "on" cells (a.u.), reported by the edited P_{GAL1}-YFP. The mean "on" expression level plots only include galactose concentrations at which both strains plotted have at least 5% "on" cells.

(A) Strain with first and second Gal4 binding sites recoded. This is the parent of the three strains (carrying specific spacing changes), whose phenotypes are displayed in the remaining panels of this figure.

(B) Strain with the 5 bp sequence immediately downstream of the third Gal4 binding site moved immediately upstream of the first Gal4 binding site.

(C) Strain with the 10 bp sequence immediately downstream of the third Gal4 binding site moved immediately upstream of the first Gal4 binding site.

(D) Strain with the 17 bp sequence immediately downstream of the third Gal4 binding site moved immediately upstream of the first Gal4 binding site.

Error bars indicate SEM (n = 3). See also Table S1.

promoter elements is critical for setting specific promoter activity levels.

In summary, while there may still be a number of unknown factors influencing the overall promoter activity from each configuration of the *GAL1* promoter, we have elucidated an initial model (Figure 7) based on strains in which the third Gal4 binding site is systematically moved toward the TSS and TATA box. The wild-type promoter architecture carries three canonical binding sites on which Gal4 proteins bind. The presence of transcriptional mediators induces the bending of the Gal4-bound DNA to facilitate stable assembly of the transcriptional machinery, including RNA polymerase (Figure 7A). On a promoter with a single active binding site, Gal4 binds to that site and interacts with the transcriptional machinery, causing promoter activation but with a lower expression level (Figure 7B). When the single Gal4 binding site is moved 5 bp closer to the TATA box, the site-bound Gal4 protein is rotated almost a full half turn around the DNA helix, causing all expression to be lost, likely due to the loss of

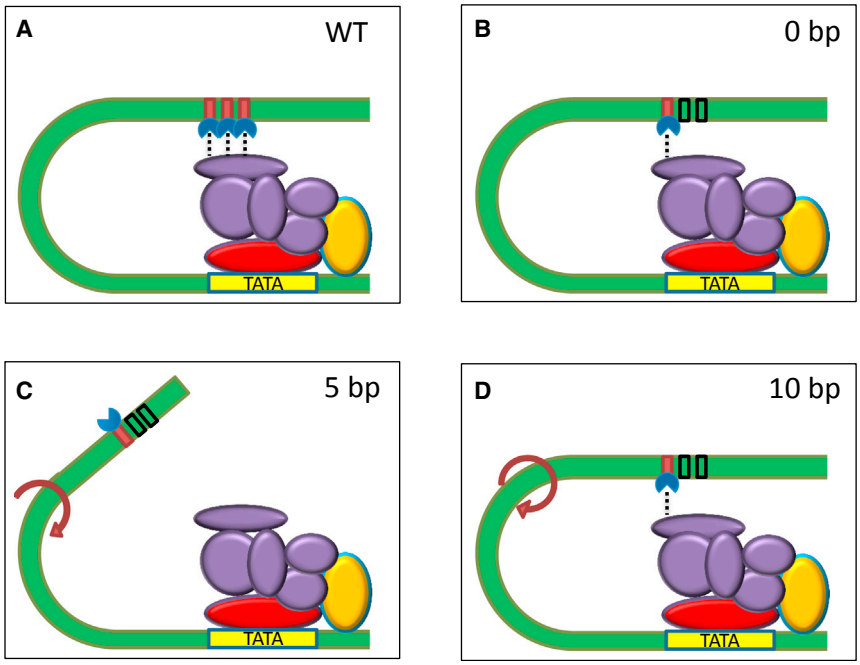


Figure 7. Model Incorporating Distance-Based Factors Influencing *GAL1* Promoter Activity

DNA is shown in green. The filled red boxes are Gal4 binding sites, while removing a Gal4 binding site by recoding is depicted by an empty black box. A bright yellow box indicates the TATA box. Gal4 proteins are depicted by blue Pac-Man symbols. TATA binding protein is shown in cherry. DNA polymerase is shown in dark yellow. Mediator proteins are shown in purple. Black dotted lines indicate interactions between Gal4 proteins and the mediator complex. Red arrows indicate the turn of the DNA helix.

(A) Schematic of the wild-type *GAL1* promoter.

(B) Schematic of the promoter architecture present in the parent strain from which the three strains carrying three spacing changes were constructed. The first and second Gal4 binding sites have been recoded.

(C) Schematic of the promoter with the third Gal4 binding site moving 5 bp toward the TSS and TATA box (rotated nearly one half circle around the DNA helix).

(D) Schematic of the promoter with the third Gal4 binding site moving 10 bp toward the TSS and TATA box (rotated almost one full circle around the DNA helix).

interactions between Gal4 and transcriptional machinery on the promoter (Figure 7C). However, when the single binding site is moved 10 bp closer to the TATA box, the site-bound Gal4 protein is rotated almost full circle around the DNA helix, leading to promoter activity levels similar to the levels seen without the 10 bp change. This result indicates that the 10 bp move restores the interaction of the Gal4 protein with the transcriptional machinery (Figure 7D). This model explains why small changes in spacing caused loss of activity from the *GAL1* promoter in many of the edited promoters examined in this study.

DISCUSSION

In this study, we have applied the CRISPR/Cas9 genome editing technology to introduce a technique that allows truly scarless editing of large genomic regions in live yeast cells by preserving endogenous PAM sites. We have shown that by removing the region of interest and then replacing it with an edited version, any number of edits can be made to a genomic region. We recognize that the necessity to first construct an intermediate strain by introducing a novel PAM site may be less than ideal; however, it is indispensable when scarless editing is needed. This is due to the inability to introduce donor oligonucleotides retaining the original PAM sites, which are present outside of the region to edit. If we were to use donors with the original PAM sites, the CRISPR machinery would not be able to differentiate between the editing donor and the to-be-edited genome and would cut both, a problem recognized in several previous papers (Horwitz et al., 2015; Mans et al., 2015). The use of this method to overcome this significant obstacle is a major advantage over any other in vivo editing method for sensitive genomic regions.

The method presented here can be extended to many other organisms, and we anticipate that it will be widely used in a range of fields and applications, including directed evolution (Arnold, 1993; Kuchner and Arnold, 1997; Shao and Arnold, 1996) performed in vivo. There is no easy method to create synthetically modified organisms containing libraries of mutations at specific regions in their genome. Such an ability would greatly enrich the technique of directed evolution, because edited genes could be expressed in their native loci and without copy number variations, both of which have been inevitable disadvantages of the traditional plasmid-based directed evolution technique.

As a potential limitation of our method, the necessity of having more than one CRISPR step may make the method relatively difficult to perform in animals. In addition, we anticipate that essential genes would not be amenable to editing with our method, because organisms could not survive the lack of such genes during the intermediate step of editing. Temporarily introducing essential genes during the intermediate step via plasmids should alleviate this limitation.

Using this method, we have elucidated previously unknown characteristics of the canonical *GAL1* and *GAL80* promoters that are commonly used not only in yeast studies but also in many other eukaryotic organisms. We have constructed yeast strains with edited promoters and examined the effect of transcription factor binding site number and location in vivo. We have found several surprising results that provide important insights into a more comprehensive understanding of genotype-phenotype relationships in eukaryotes. For example, having two Gal4 binding sites immediately adjacent to each other did not result in promoter activation even in high concentrations of galactose, while promoters with either one or three Gal4 binding sites immediately adjacent to each other had normal promoter

activation. When the two Gal4 binding sites were separated by only 20 bp, the promoter was still functional in an activity pattern similar to the wild-type.

Even more surprisingly, we found that small changes to the spacing between transcription factor binding sites and other promoter elements, such as the TATA box, can have dramatic effects on promoter activity. We acknowledge the existence of previous work (Sharon et al., 2012) in which Gal4 binding sites were added to synthetic promoters, including a modified *GAL1/10* promoter that had its native regulatory elements inactivated by mutation and that reported increasing the number of binding sites (up to five at predefined locations) generally increased the level of transcription (with one notable exception for two binding sites placed 1 bp apart), in seeming contradiction to our observations. However, the positions of the added Gal4 binding sites in that study were artificially determined and not related to their native position. As we have shown here, the distance between the binding site and other promoter elements (such as the TSS and TATA box) is critically important to the binding site's ability to activate transcription, so the fact that we reach different conclusions with respect to the number of promoter binding sites should be unsurprising. Our results, obtained using a marker-free precise genome editing technique on native promoters, reinforce the idea that the relative positioning of promoter elements is of critical importance to promoter activity in live single cells and demonstrate that simply adding transcription factor binding sites to a promoter can have nonlinear consequences for transcriptional activation.

EXPERIMENTAL PROCEDURES

Construction of Plasmids

The base plasmid used for all plasmid construction was pRS315, a yeast centromeric plasmid containing a *LEU2* marker, which was a gift of Mark Hochstrasser. To create plasmids containing a single gRNA cassette, we ordered (Integrated DNA Technologies) a 388 bp gRNA cassette sequence modified from the template created by DiCarlo et al. (2013) (Supplemental Information). The cassette consists of a promoter, terminator, and gRNA sequence that function well in yeast. These sequences were constructed with a BamHI cut site on the 3′ end and a HindIII cut site on the 5′ end of the cassette. Upon receipt, these sequences and pRS315 were digested with the appropriate enzymes and ligated to form plasmids containing a gRNA cassette. To create plasmids containing two gRNA cassettes, the two cassettes were individually ordered, with the first containing a NotI site and a BamHI site and the second containing an XmaI site and a HindIII site. The two were then ligated into pRS315 in two separate transformations to create a plasmid containing two gRNAs separated by 12 base pairs.

Introducing the *CAS9* Gene into Yeast

All strains were created from WP35, a W303 strain with one copy of the P_{GAL1}-YFP construct inserted into the *ho* locus. The rest of the yeast genome was unaltered. To insert *CAS9* into this strain, we obtained plasmid #43802 from Addgene. This plasmid, p414-TEF1p-Cas9-CYC1t, described by DiCarlo et al. (2013), contains the *CAS9* gene under a constitutive *TEF1* promoter, along with a *TRP1* marker. To insert it into WP35, we removed the centromeric region from the plasmid, linearized the plasmid in the *TRP1* gene using MfeI enzyme, and then transformed the linear product into WP35 using the standard lithium acetate (LiOAc) transformation technique. This resulted in the strain GE1.

Introducing Rational Edits on the *GAL1* and *GAL80* Promoters

The first step of editing at the *GAL1* promoter (P_{GAL1}) consisted of transforming GE1 with a plasmid containing gRNAs targeting two sites flanking P_{GAL1}, with the goal of cutting out the entirety of P_{GAL1}. A donor oligonucleotide carrying a novel 30 bp CRISPR cut and PAM site between 50 bp regions homologous to each genomic region flanking the *GAL1* promoter was co-transformed (LiOAc technique), with the plasmid containing two gRNAs so that cells could use the donor as a repair template. The transformed cells were then grown on −LEU (leucine) plates for 2 days, after which colonies were PCR tested and sequenced to verify that editing had taken place. After this, replica plating between rich media (yeast extract peptone dextrose [YPD]) and −LEU media was used to locate colonies, which had lost the centromeric plasmid used in the first step. This was necessary because the plasmid used in step 1 would have been able to cut the desired product after the second round of editing if it remained in the cells.

The second step of the editing process consisted of transforming (using the LiOAc technique) the intermediate strain carrying the 30 bp novel PAM site with a plasmid containing a gRNA cassette targeting the new CRISPR cut site in the presence of a new donor oligonucleotide. The donor consisted of an edited version of the P_{GAL1} between 50 bp sequences homologous to the genomic regions flanking the cut site. The transformed cells were then grown on −LEU plates for 2 days, after which colonies were PCR tested and sequenced to verify that editing had taken place.

Editing of P_{GAL80} involved the same two-step process but used gRNAs and donors that were relevant to P_{GAL80}. Finally, to construct the multiplexed strains, the transformations for the first step of the editing process were performed successively to create an intermediate yeast strain with both P_{GAL1} and P_{GAL80} replaced by two novel PAM sites. The second step of the editing process was completed simultaneously for the P_{GAL1} and P_{GAL80} loci by using a single plasmid containing two gRNAs (each targeting one locus) and two separate donor oligonucleotides carrying the edited P_{GAL1} and P_{GAL80} sequences.

Quantifying the Efficiency of the CRISPR Editing

The editing efficiency was calculated by dividing the number of correctly edited colonies (as identified by PCR confirmation and sequencing) by the total number of colonies examined. The use of a selectable marker on the gRNA-containing plasmid ensured that only colonies that received a plasmid could grow and be used for analysis. To quantify the efficiency of the first step of the editing process, we used the results obtained from targeting the *GAL80* promoter site. Two independent transformations were conducted, 8 and 24 colonies were collected from the two transformations, and the editing efficiency values were calculated separately. As a result, 6 of 8 colonies and 14 of 24 colonies were counted as correct. The results were combined to provide the overall efficiency in terms of mean and SEM (n = 2) as presented in the manuscript. In quantifying the efficiency of the second step of the editing process, three independent transformations were conducted by simultaneously editing the *GAL80* and *GAL1* promoters, 24 colonies were collected from each transformation, and the editing efficiency values were calculated separately. The results were combined to provide the overall efficiency in terms of mean and SEM (n = 3) as presented in the manuscript.

Growth Conditions, Media, and Flow Cytometry Data Analysis

Cells were grown in the appropriate synthetic amino acid dropout media. All growths were conducted in duplicate at 30°C in a shaking incubator in 5 ml of media. Cells were first grown overnight for 22 hr in minimal media containing 0.1% mannose as the carbon source, reaching an optical density 600 (OD_{600}) between 0.075 and 0.15. They were subsequently diluted into the induction media containing 0.1% mannose and the appropriate concentration of galactose and grown for another 22 hr, reaching a cell density between 0.075 and 0.15. After the induction period, single-cell fluorescence values were analyzed using flow cytometry (Stratedigm-8 with high throughput auto sampler [HTAS]). Each FACS sample had on average 3,000 cells after gating. Log-amplified fluorescence measurements for the gated cells were converted to the linear scale for analysis. A threshold for the "on" state (75.7 a.u.) was selected based on fluorescence measurements from uninduced, unedited cells and uniformly applied to all samples. The fraction of "on" cells was then quantified for each sample. For each galactose concentration that resulted in at least 5% "on" cells, the mean expression level of such cells was also quantified.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.12.014.

AUTHOR CONTRIBUTIONS

G.L.E. and M.A. designed the experiments and analyses. G.L.E. constructed the strains, performed the experiments, and analyzed the data quantifying editing efficiency. R.S. analyzed the flow cytometry data. G.L.E. and R.S. prepared the figures. M.A. guided the study. G.L.E. and M.A. wrote the manuscript. All authors approved the manuscript.

ACKNOWLEDGMENTS

The authors thank M.A. lab members for useful discussions and W. Peng for providing the base strain WP35. M.A. acknowledges funding through the NIH Director's New Innovator Award (1DP2AG050461-01) and a junior faculty fellowship from Yale University. G.L.E. was partly supported by the NIH T32 GM007499. p414-TEF1p-Cas9-CYC1t was a gift from George Church (Addgene plasmid #43802). M.A. and G.L.E. have filed a provisional application with the US Patent and Trademark Office on this work.

Received: July 27, 2016
Revised: November 2, 2016
Accepted: December 5, 2016
Published: January 3, 2017

REFERENCES

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature *435*, 228–232.

Acar, M., Pando, B.F., Arnold, F.H., Elowitz, M.B., and van Oudenaarden, A. (2010). A general mechanism for network-dosage compensation in gene circuits. Science *329*, 1656–1660.

Arnold, F.H. (1993). Engineering proteins for nonnatural environments. FASEB J. *7*, 744–749.

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., and Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. ACS Synth. Biol. *4*, 585–594.

Boeke, J.D., Trueheart, J., Natsoulis, G., and Fink, G.R. (1987). 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. *154*, 164–175.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336–4343.

Dion, V., and Coulombe, B. (2003). Interactions of a DNA-bound transcriptional activator with the TBP-TFIIA-TFIIB-promoter quaternary complex. J. Biol. Chem. *278*, 11495–11501.

Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S., and Newman, J.D. (2015). Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Syst. *1*, 88–96.

Jao, L.E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc. Natl. Acad. Sci. USA *110*, 13904–13909.

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

Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol. Rev. *51*, 458–476.

Johnston, M., and Davis, R.W. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *4*, 1440–1448.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.

Kuchner, O., and Arnold, F.H. (1997). Directed evolution of enzyme catalysts. Trends Biotechnol. *15*, 523–530.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J., and Daran, J.M. (2015). CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. *15*, 1–15.

Mizutani, A., and Tanaka, M. (2003). Regions of GAL4 critical for binding to a promoter in vivo revealed by a visual DNA-binding analysis. EMBO J. *22*, 2178–2187.

Peng, G., and Hopper, J.E. (2002). Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proc. Natl. Acad. Sci. USA *99*, 8548–8553.

Ryan, O.W., Skerker, J.M., Maurer, M.J., Li, X., Tsai, J.C., Poddar, S., Lee, M.E., DeLoache, W., Dueber, J.E., Arkin, A.P., and Cate, J.H. (2014). Selection of chromosomal DNA libraries using a multiplex CRISPR system. eLife *3*, e03703.

Ryan, O.W., Poddar, S., and Cate, J.H. (2016). CRISPR-Cas9 genome engineering in *Saccharomyces cerevisiae* cells. Cold Spring Harb. Protoc. *2016*, t086827.

Shao, Z., and Arnold, F.H. (1996). Engineering new functions and altering existing functions. Curr. Opin. Struct. Biol. *6*, 513–518.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. *30*, 521–530.

St John, T.P., and Davis, R.W. (1981). The organization and transcription of the galactose gene cluster of *Saccharomyces*. J. Mol. Biol. *152*, 285–315.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.