Insights into Bidirectional Gene Expression Control Using the Canonical GAL1/GAL10 Promoter

Gregory L. Elison\textsuperscript{1,2}, Yuan Xue\textsuperscript{1,2}, Ruijie Song\textsuperscript{2,3}, and Murat Acar\textsuperscript{1,2,3,4,5,*}

\textsuperscript{1}Department of Molecular Cellular and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06511, USA
\textsuperscript{2}Systems Biology Institute, Yale University, 850 West Campus Drive, West Haven, CT 06516, USA
\textsuperscript{3}Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, 300 George Street, Suite 501, New Haven, CT 06511, USA
\textsuperscript{4}Department of Physics, Yale University, 217 Prospect Street, New Haven, CT 06511, USA
\textsuperscript{5}Lead Contact

SUMMARY

Despite advances made in understanding the effects of promoter structure on transcriptional activity, limited knowledge exists regarding the role played by chromatin architecture in transcription. Previous work hypothesized that transcription from the bidirectional GAL1/GAL10 promoter is controlled through looping of its UAS region around a nonstandard nucleosome. Here, by editing the GAL1/GAL10 promoter at high resolution, we provide insights into bidirectional expression control. We demonstrate that the first and fourth Gal4 binding sites within the UAS do not functionally contribute to promoter activation. Instead, these sites, along with nearby regulatory regions, contribute to the directional regulation of gene expression. Furthermore, Gal4 binding to the third binding site is critical for gene expression, while binding to the other three sites is not sufficient for transcriptional activation. Because the GAL1/GAL10 UAS can activate gene expression in many eukaryotes, the regulatory mechanism presented is expected to operate broadly across the eukaryotic clade.

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

AUTHOR CONTRIBUTIONS

G.L.E. and M.A. designed the experiments and analyses, interpreted the data, and wrote the manuscript. G.L.E. constructed the strains, performed the experiments, and collected and analyzed the data. Y.X. performed the qPCR and ChIP-qPCR experiments, analyzed their data, wrote their methods, and contributed to their interpretation for the model. R.S. analyzed the flow cytometry data. G.L.E. and R.S. prepared the figures. M.A. guided the study. All of the authors read and approved the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures, three tables, and one supplemental methods file and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.09.050.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND SOFTWARE AVAILABILITY

All DNA sequences used in or relevant to this study are presented in the document Methods S1. The DNA sequences have also been deposited in the GenBank database under ID codes MH818048 – MH818098, and MH888036 – MH888039. In Methods S1, each sequence is also identified by its specific ID code.
In Brief

Elison et al. demonstrate how transcription from the canonical GAL1/GAL10 promoter is controlled bidirectionally. Editing the promoter content using CRISPR, they uncover distinct regulatory sites that are essential for directional gene expression control. They elucidate the differential nature of the four Gal4 binding sites in affecting transcriptional activation and directionality.

INTRODUCTION

For several decades, one of the major goals in genetics research has been to associate base pair sequences with phenotypic outputs. Much of that research has focused on protein coding regions and the ability to associate each base pair with a specific amino acid within a protein. However, an equally important part of the question, regulation of the expression of those proteins, has met with much less progress in linking genotype to phenotype. Regulation of gene expression is controlled on several levels, but a primary control mechanism (Weinhandl et al., 2014) involves the promoter region that is specific to each gene. A greater understanding of how those promoter regions contribute to transcriptional activity will be of great importance to the broader understanding of genome structure and function.

The GAL1/GAL10 bidirectional promoter of the yeast Saccharomyces cerevisiae is arguably the best studied promoter in all eukaryotic organisms (Johnston, 1987; Weinhandl et al., 2014). It is activated in the presence of galactose and the absence of glucose (Johnston and Davis, 1984; Weinhandl et al., 2014). The promoter is believed to contain four activating sites on which Gal4 has been thought to bind (Johnston and Davis, 1984; Kellis et al., 2003), in addition to two TATA boxes (Johnston, 1987; Kellis et al., 2003) that facilitate transcription in each direction (Figure 1A). When Gal4 binds to its binding sites, it is
thought to first interact with the Spt-Ada-Gcn5-acetyltransferase complex (SAGA) (Bhaumik and Green, 2001; Lin et al., 2012) and then recruit a variety of other protein complexes, to ultimately result in gene expression via RNA polymerase II (Bryant et al., 2008; Bryant and Ptashne, 2003; Larschan and Winston, 2001; Lin et al., 2012). Previous studies have offered contradictory explanations for the interaction among the various proteins and the SAGA complex (Bryant et al., 2008; Bryant and Ptashne, 2003; Larschan and Winston, 2001; Melcher and Johnston, 1995); however all of them agree that Gal4 initially binds to the promoter and activates the entire chain of interactions among the involved components (Bhaumik and Green, 2001; Bryant et al., 2008; Bryant and Ptashne, 2003; Larschan and Winston, 2001; Lin et al., 2012; Melcher and Johnston, 1995).

Beyond the knowledge pertaining to the binding of Gal4 on the promoter to initiate gene expression, limited information exists regarding the structure of the GAL1/GAL10 promoter sequence and how that structure mechanistically leads to the initiation of gene expression. The promoter contains 688 base pairs (Kellis et al., 2003), many of which are conserved within the Saccharomyces clade (Kellis et al., 2003) and currently have no known function other than to act as placeholders to keep known binding sites at the proper distance relative to one another (Elison et al., 2017). However, this lack of information should not be interpreted as a lack of importance. The main reason behind the limited knowledge about the promoter structure and activation mechanism has been the difficulty in making specific genetic edits to the promoter elements to reliably test their effects. For a faithful interpretation of the resulting phenotypic changes, these edits would need to be introduced on one copy of the promoter integrated in the yeast genome, instead of using plasmids, and they would need to be introduced in a marker-free fashion so as not to affect chromatin dynamics. Such editing capability has been made available to the research community only recently through a two-step CRISPR editing approach (Elison et al., 2017) that we developed to facilitate scarless genome editing at long promoter regions.

While understanding of the base pair architecture of the GAL1/GAL10 promoter has been limited, a previous study (Floer et al., 2010) provided insights into the general chromatin structure of the promoter in both its active and inactive states. Rather than being completely free of nucleosomes when activated as was previously thought to be the case, the study showed that the up-stream activation sequence (UAS) is wrapped around a nucleosome containing a nonstandard H2A.Z histone (Floer et al., 2010). The RSC protein complex was shown to be necessary for proper GAL1 expression and was thought to load the nucleosome onto the UAS of the GAL1/GAL10 promoter. The authors proposed a model in which the first and fourth Gal4 binding sites in the UAS region are not accessible to nuclear proteins and are not bound by Gal4 proteins.

Using the two-step CRISPR approach, we investigated genotype-phenotype relations in the GAL1/GAL10 bidirectional promoter and elucidated a number of unexpected findings. A set of mutations on regions outside the known Gal4 binding sites on the promoter caused the elimination of GAL1 expression, but not that of GAL10. Subsequently, additional regulatory regions were found, the mutations on which eliminated GAL10 expression, but not that of GAL1. This was surprising, because several recent papers (Floer et al., 2010; Zhou and O’Shea, 2011) had suggested that the base pairs outside the four Gal4 binding sites were
irrelevant for the control of gene expression. Two of the recently identified regulatory regions partially overlapped with the first and fourth Gal4 binding sites, which made us question the function of all four binding sites. We thus examined whether all four binding sites are involved in gene expression regulation via the conventional mechanism or a different one. We found that the first and fourth Gal4 binding sites helped facilitate a different regulatory mechanism and could be mutated without negative impact on the expression level of GAL1 or GAL10; however, mutating the recently identified regulatory regions before and after these sites eliminated gene expression. The second or third Gal4 binding sites do operate through the traditional gene expression control mechanism; mutating them lowered or eliminated gene expression in all of the cases.

RESULTS

Identification of Additional Regulatory Sites in the GAL1 Promoter

The GAL1/GAL10 bidirectional promoter is a canonical promoter used for expression studies in yeast specifically (Liu et al., 2017; Peng et al., 2015) and eukaryotes more broadly (Johnston, 1987). After several decades of study, it is generally agreed to contain eight important binding sites. In the 5’ to 3’ direction, these sites are a TATA box for GAL10 expression, three canonical Gal4 activator binding sites (Johnston and Davis, 1984), one noncanonical Gal4 activator binding site (Johnston and Davis, 1984), two Mig1 inhibitor binding sites (Nehlin et al., 1991), and a TATA box for GAL1 expression. The position of each of these except for the Mig1 sites is shown in Figure 1A. Because both genes are driven by the same promoter, we refer to the binding site order from 5’ to 3’ as given above for clarity, even when referencing the GAL10 promoter. When grown in a medium lacking glucose but containing both galactose and a neutral sugar such as mannose, cells will begin to express GAL1 and GAL10 at low galactose levels, and upon the addition of sufficient levels of galactose, nearly all of the cells will show expression from this promoter (Acar et al., 2005; Peng et al., 2016). As the promoter displays bimodal activity (Acar et al., 2005, 2010), the neutral sugar is needed together with galactose to avoid fitness-based enrichment of the population in the ON expression state, because only the ON cells may metabolize galactose. The addition of mannose leads to equal growth rates for both OFF and ON cells, and therefore any enrichment in the ON state is solely attributable to the induction dynamics of the galactose network, instead of fitness (Acar et al., 2005, 2010).

To measure the GAL1 expression level, we integrated one copy of the promoter driving yellow fluorescent protein (YFP) into the ho locus in haploid yeast. We used flow cytometry to examine 10,000 cells under activating conditions to analyze the effects of any genetic changes on YFP expression (Figure 1A). Expressing the YFP from the ho locus did not lead to gene expression differences compared to expressing it from the endogenous GAL1 locus (Figures S1 and S2A). Due to the positive feedback loop mediated by the proteins functioning as inducers in the GAL network, we observed a bimodal (Acar et al., 2005, 2010) gene expression distribution from the wild-type GAL1 promoter, with approximately 30%–40% of cells in the ON state when grown in media containing 0.1% mannose and 0.175% galactose. We used these media conditions to grow and analyze all of the strains in
terms of both the percentage of cells in the ON state and the mean expression level of the ON state (Figures 1A, 1B, and S2B).

Using the two-step CRISPR method we previously developed (Elison et al., 2017) to make scarless edits in the yeast genome, we chose to investigate how the regions outside the known binding sites on the GAL1 promoter affected gene expression. We recoded the 10- to 15-bp region immediately upstream of the first Gal4 binding site and the same length immediately downstream of the fourth Gal4 binding site, thus ensuring different base pair content while maintaining the original DNA length. We were surprised to find that both of these edits resulted in a total lack of expression from the GAL1 promoter. Conventional wisdom in the field was that only the Gal4 binding sites themselves were relevant for expression of the promoter, but clearly additional bases outside the Gal4 binding sites were necessary for transcription. Most changes to these bases did not show any change in expression after editing (Figure 1C), but after completing a series of experiments to identify the most crucial bases in these regions, we were able to narrow the scope of these regions to the three base pairs (GTA) immediately before the first Gal4 binding site (Figure 1D), and the two base pairs (AT) 11 bases downstream of the fourth Gal4 binding site (Figure 1E). These bases were absolutely necessary for any GAL1 expression. Figure 1F shows these locations in relation to the previously known regulatory regions on the GAL1 promoter; complete sequence information about all of the edited content and the editing locations is provided in Methods S1.

Identification of Additional Regulatory Sites in the \textit{GAL10} Promoter

The activities of the \textit{GAL1} and \textit{GAL10} promoters are highly similar to one another (Figure 2A). The “symmetric” expression from this bidirectional promoter is hypothesized to be important for maintaining fitness because producing Gal1 but not Gal10 enzymes leads to the buildup of a toxic metabolic product (Mumma et al., 2008). There is a fitness deficit for cells that are able to produce Gal1 but not Gal10 (Douglas and Hawthorne, 1964; Mumma et al., 2008) (Figure S3). To determine whether the regulatory regions we identified would break this expression symmetry, we measured single-cell YFP expression in the \textit{GAL10} direction from cells carrying the edited promoters. We unexpectedly found that the two recently identified regions that eliminated \textit{GAL1} expression lacked any influence on \textit{GAL10} expression (Figure S4).

We noted that the recently identified regions are strongly conserved among most species of the \textit{Saccharomyces} clade (Table S2). Treating the region between the first and fourth Gal4 binding sites as the UAS, we also noted that there are several conserved regions that are roughly symmetrical in their locations flanking the UAS region (Table S2). It therefore made intuitive sense that the \textit{GAL10} promoter regions that are symmetrical to the position of the recently identified \textit{GAL1} regions may provide additional regulation in the \textit{GAL10} direction. As such, we edited these regions in an attempt to identify any changes in \textit{GAL10} expression. Consistent with our hypothesis, we found two regions that, when mutated, caused complete \textit{GAL10} expression loss. The first was an 8-bp region positioned 7–14 bases upstream of the first Gal4 binding site of the UAS (Figure 2B). Of this 8-bp region, the 3-bp region (ATT) positioned 7–9 bases upstream is expected to be the critical element, because extending this
3-bp region by including more bases around it did not have a major impact on GAL10 expression (Figure S5). The second was a 4-bp region (AAAG) positioned 7–10 bp downstream of the fourth Gal4 binding site of the UAS (Figure 2C). While these recently discovered regions do regulate expression in the GAL10 direction, their disruption has no effect on expression in the GAL1 direction (Figure S6). This behavior mirrors that of the identified GAL1 regulatory regions that do not affect GAL10 expression. We have identified four additional regulatory sites within the GAL1/GAL10 bidirectional promoter. Two of these sites control only GAL1 expression, while the other two sites control only GAL10 expression. Figure 2D shows the regions that affect GAL10 but not GAL1, while Figure 2E shows all four additional regions.

Characterization of External Gal4 Binding Sites

Because the discovery of additional regulatory sites within the GAL1/GAL10 promoter calls into question the traditional model of the initiation of transcription from this canonical promoter, we decided to further our characterization by also investigating the effect of the Gal4 binding sites on GAL1/GAL10 expression. Our previous work indicated that some Gal4 binding sites were more critical than others when removed from the promoter (Elison et al., 2017). The first and fourth binding sites especially seemed unnecessary in most cases, while the third binding site seemed to be the most critical. The proximity of our additional regions in the present study to the first and fourth Gal4 binding sites caused us to hypothesize that these two Gal4 binding sites may actually be part of a larger regulatory mechanism rather than functioning as Gal4 binding sites.

We began with the characterization of the first Gal4 binding site because it is immediately adjacent to the critical GTA bases needed for GAL1 expression. We considered the possibility that some of the 17 bases that made up the first Gal4 binding site could be part of a larger regulatory region, including GTA. We therefore recoded the following specific bases of the first Gal4 binding site: the first three bases (CGG), the next three bases (ATT), and the last three bases (CCG). If the first Gal4 binding site were not actually a site for the binding of Gal4 but a part of the larger regulatory mechanism involving the GTA bases, then we hypothesized that the first three bases (CGG) should be critical for the operation of the larger mechanism, the next three (ATT) may or may not be critical, and the last three bases (CCG) would not be essential for the mechanism. Compared to the wild-type expression profile (Figure 3A), we found that the first three bases of the first Gal4 binding site were critical for YFP expression (Figure 3B), but that neither the next three nor the last three bases were necessary (Figures 3C and 3D). We note that the binding of Gal4 proteins on a Gal4 binding site requires both the first three (CGG) and the last three bases (CCG). These findings support the presence of a larger regulatory mechanism, with the first Gal4 site contributing to this mechanism instead of being directly involved in promoter activation.

Our previous work demonstrated that the fourth Gal4 binding site had no role in the expression of GAL1. Together with our observation described above, we saw a possibility that the fourth Gal4 binding site may also be part of the larger regulatory mechanism instead of binding Gal4 itself. This was further supported by the fact that the fourth Gal4 binding site is non-canonical in its sequence, with no explanation in the field as to how Gal4 may still
bind to it. To characterize the potential roles of the first and fourth Gal4 sites in the larger regulatory mechanism, we edited four specific promoter regions and measured their phenotypic consequences for both GAL1 and GAL10 expression (Figure 3E). We found that GAL1 expression was only critically affected by a loss of the first three base pairs in the first Gal4 binding site (Figure 3E). In addition, we found that the fourth Gal4 binding site does not affect GAL1 expression in any way. However, the opposite was true for GAL10 expression; sequence alterations within the first Gal4 binding site did not affect GAL10 expression. In contrast, the three base pairs on the 3′ edge of the fourth Gal4 binding site were critical for GAL10 expression, with the three base pairs on the 5′ edge being essentially unimportant (Figure 3E). These results further support the idea that the first and fourth Gal4 binding sites are not actually activating sites, but rather have been adapted into other regulatory sites controlling the direction of expression from the promoter. All measured mean YFP values were somewhat lower than the wild-type values. This may indicate that there are more subtle processes being influenced by these regions.

Interior Bases of Gal4 Binding Sites Are Critical for GAL1 and GAL10 Expression

After elucidating that the first and fourth Gal4 binding sites in the GAL1/GAL10 bidirectional promoter are involved in directional regulation instead of direct activation of the promoter, we next wondered whether there were any distinguishing characteristics between the four GAL4 binding sites that may explain the differences found in their impact on expression. Specifically, we asked whether the 11 base pairs between the “consensus” CGG and CCG base pairs were the cause of these expression differences.

The interior 11 base pairs differ considerably among the four Gal4 binding sites (Figure 4A; Table S3). We investigated the effects of recoding some or all of the interior 11 base pairs within each of the first three Gal4 binding sites to different bases and examined their impact on both GAL1 and GAL10 expression in comparison to the wild-type expression profile (Figure 4B). Changes to the interior of the first Gal4 binding site had little to no effect on the expression of the promoter in either direction (Figure 4C), supporting the idea that this binding site has evolved away from Gal4 binding and is instead serving a role in the larger regulatory mechanism. In contrast, changes to the interior of the second Gal4 binding site resulted in a loss of approximately half of the expression when compared to the wild-type (Figure 4D). Changes to the interior of the third Gal4 binding site resulted in a complete loss of expression in both directions (Figure 4E).

Our experimental observations contradict the conventional wisdom in the field, which is that only the three base pairs on each side of a 17-bp Gal4 binding site are relevant for Gal4-activated gene expression and that all four binding sites with the 17-bp consensus binding motif are actually involved in Gal4-mediated promoter activation in vivo. Instead, our results indicate that the template for a Gal4 binding site needs to change from the conventional [CGG … N_{11} … CCG] motif to more specific alternate motifs that take into account differences in the ability of each motif to lead to full expression. At the very least, the ability of each motif to drive expression in vivo should be fully tested, because previous experiments in vitro are unlikely to apply to this situation. These results strongly support the idea that the first and fourth Gal4 binding sites have evolved to take part in a broader...
directional expression control mechanism and are no longer functionally involved in promoter activation, despite their carrying the outer motif elements. The observation that only the third Gal4 binding site is critical for expression in both directions strongly indicates that this site plays a key role in functional promoter activation in vivo, with the impact of the second Gal4 binding site being relatively less important in comparison.

In-Depth Analysis of the Effect of the Third Gal4 Binding Site on Gene Expression

We further investigated our hypothesis that the third Gal4 binding site is critical for activation of the GAL1/GAL10 promoter. To this end, we first constructed four strains in which we altered the interior 11 base pairs of the third Gal4 binding site by recoding either two or three base pairs at a time for GAL1 (Figures S7A–S7D) and for GAL10 (Figures S8A–S8D). Compared to the wild-type, the results from the edited strains indicate that the middle region of the third Gal4 binding site is critical for expression in both directions, because disrupting the middle six base pairs results in a total loss of GAL1 and GAL10 expression.

As a result of these findings, we investigated how recoding the seventh, eighth, and ninth base pairs individually within the full 17-bp-long sequence of the third Gal4 binding site would affect gene expression (Figures S7E–S7G), because recoding all three of these bases had resulted in a total lack of GAL1 promoter activity (Figures S7B and S7C). Both the seventh and ninth base pairs turned out to be critical for expression from the GAL1 promoter, while the eighth base pair could be changed without a consequence for promoter activity. The pattern for the GAL10 promoter was similar, although changing the ninth base pair did not change GAL10 expression (Figures S8E–S8G).

We then chose to compare Gal4 binding between the wild-type strain and the strain that contains a 3-bp recoding within the third Gal4 binding site (Figure S7B), causing a total lack of expression from the GAL1 promoter. By doing so, we hoped to be able to shed light on two distinct mechanisms on which we hypothesized earlier in this work. First, we hypothesize that the first and fourth Gal4 binding sites are not occupied by Gal4 in vivo due to the results presented in Figures 3 and 4; instead, these sites act as part of the recently identified regulatory regions facilitating directional control. Second, we hypothesize that Gal4 binding to the third Gal4 binding site is critical for expression from the GAL1 promoter, while the other three binding sites are to a large degree unnecessary for expression (Figure 4). By performing chromatin immunoprecipitation (ChIP)-qPCR experiments, we aimed to answer both questions simultaneously.

Starting with the second of these questions, Gal4 is known to bind to the wild-type GAL1/GAL10 promoter (Mizutani and Tanaka, 2003). Using the same galactose induction level as used above (0.175%), we performed a ChIP-qPCR experiment on the strain carrying the wild-type promoter driving YFP and measured a 6.4% Gal4 binding (Figures 5A, 5B, and 5D). In this strain, the endogenous GAL1/GAL10 promoter was removed so that the ChIP results would solely indicate the Gal4 binding level at one copy of the reporter-driving promoter; measuring Gal4 binding on the promoter driving YFP was important to associate ChIP results with the YFP results obtained from the same strain. The necessity to remove the endogenous GAL1/GAL10 promoter meant that this strain lacked endogenous GAL1 gene
expression. Because Gal1 proteins carry a Gal3-like inducer function (Bhat and Hopper, 1992), not having Gal1 proteins in this strain affected the expression level distribution (Figure 5B).

Next, using the same strain background missing the endogenous GAL1/GAL10 promoter, we performed ChIP-qPCR experiments and measured Gal4 binding on the GAL1/GAL10 promoter (driving YFP), which was recoded in the third Gal4 binding site to eliminate this site. We saw a 43.5% reduction in Gal4 binding on this recoded promoter compared to the wild-type promoter driving YFP (Figures 5C and 5D). We interpret this result to be due to the disruption of the third Gal4 binding site in this strain, as this disruption was the only genetic change introduced to the wild-type promoter driving YFP. The fact that Gal4 is still present on the promoter while being unable to produce YFP implies that Gal4 binding to some or all of the other three binding sites is not sufficient to activate the promoter. Thus, we conclude that the third binding site is the critical site within the promoter; Gal4 binding on the third site is strictly necessary for promoter activation.

Regarding the answer to the first question raised above, while the resolution limitation of the ChIP-qPCR technique makes a direct interpretation of our results somewhat difficult, we still gained additional insights. The length of the DNA covering the first three Gal4 binding sites together is 54 bp. When the fourth Gal4 binding site is included, the total length to probe for Gal4 binding becomes 118 bp (Figure 5A). It is unfortunate that the resolution of the ChIP technique is in the range of 200–500 bp (Skene and Henikoff, 2015), preventing us from directly confirming that the first and fourth Gal4 binding sites do not bind Gal4 in an active promoter. However, we still were able to make observations supporting our hypothesis. As noted previously, the disruption of solely the third Gal4 binding site reduces total Gal4 binding to the promoter by 43.5% (Figure 5C). Considering that each binding site would attract the same number of Gal4 proteins, a loss of 43.5% binding indicates that only one additional binding site binds to Gal4, as otherwise we would expect a binding loss of only 25% or 33% (in the case that four or three are bound, respectively).

**Interpretation of the Findings in Regard to Promoter Structure**

Having determined that a regulatory scheme exists in the GAL1/GAL10 promoter architecture, we next aimed at providing a model for our experimental findings. Any such model would need to take into account the fact that loss of the additional regulatory regions could completely eliminate expression in one direction while leaving expression in the other direction intact. In addition, it would need to account for the fact that the main role of the first and fourth Gal4 binding sites is not to bind Gal4, but instead to form a part of the recently identified regulatory regions. Finally, the model would need to explain why the second Gal4 binding site is needed to maintain full expression in both directions and why the third Gal4 binding site is absolutely necessary for expression in both directions.

Results from a previous study (Floer et al., 2010) suggested that the region between the first and fourth Gal4 binding sites on the GAL1 promoter may be wrapped around a nucleosome containing a non-standard H2A.Z histone. The authors proposed that such a topology would leave the second and third Gal4 binding sites accessible to Gal4 proteins, while the first and fourth Gal4 binding sites would be near the base of the DNA loop; this structure would
maintain integrity even during promoter activation. We note that the regulatory scheme proposed in the previous study has not been directly tested by introducing precise edits on the promoter and by examining their phenotypic consequences at the single-cell level.

Our results are consistent with the model proposed by the previous study (Floer et al., 2010) but also lead to considerable expansion of the model. The concept of having the DNA looped around a nucleosome with only the second and third Gal4 binding sites accessible to Gal4 proteins fits with our data showing that only those two “inner” binding sites are critical for expression in both directions. We further expand this idea by showing that, of these Gal4 binding sites, the third one is critical for expression, while the second one is of lesser importance. In addition, our work ascribes alternative regulatory roles to the first and fourth Gal4 binding sites by demonstrating that they contribute to the directional regulation of gene expression.

In an attempt to confirm certain features of this model, we reasoned that if the promoter is wrapped around a nucleosome with the second and third Gal4 binding sites exposed, then the overall looped structure should be extremely sensitive to the length of the UAS region, which is composed of all four Gal4 binding sites and the base pairs between them. We therefore shortened the length of the UAS region by deleting two DNA segments (covering different locations in each of two separate strains) between the third and fourth Gal4 binding sites. The wild-type distance between the third and fourth Gal4 binding sites is 45 bp (Figure 5E). In one of the two strains we constructed, we shortened the length of the loop by 11 bp (Figure 5F), while in the second strain, we shortened it by 21 bp (Figure 5G). We chose these two specific length reductions in an attempt to maintain the position of all sites with respect to the turn of the DNA helix, which is roughly 10.5 bp long. Consistent with the looping hypothesis, we found that both deletions caused a complete lack of expression in both the GAL1 and GAL10 directions (Figures 5F and 5G).

To address the possibility that these deletions caused the lack of expression due to the potential deletion of critical base pairs (e.g., uncharacterized binding sites) rather than due to a change in UAS length, we recoded the 11-bp region instead of deleting it, such that every base within the region is changed to a different base while retaining the same UAS length. The activity of the promoter carrying this recoded region was indistinguishable from the activity of the wild-type promoter (Figure 5H), suggesting that the change in length was the cause of the loss of expression in the strain carrying the 11-bp deletion.

Because both of the above length reductions and the 11-bp recoded region resulted in the removal of one of the three putative RSC binding sites (each being 4 bp long) postulated by previous work (Floer et al., 2010), we wanted to see how the removal of all three RSC binding sites would affect promoter activity. Yet another reason for performing this experiment was that the 11-bp region recoded or deleted, as described above, included additional base pairs beyond the intended recoding or deletion of one RSC binding site, potentially confounding the interpretation of the results. Therefore, we recoded only the three RSC binding sites in the GAL1 promoter while leaving all other base pairs of the promoter and the length of the promoter undisturbed. We found that the removal of all three RSC binding sites from the UAS region was sufficient to completely eliminate GAL1
expression (Figure 5I), leading to the conclusion that RSC binding is necessary for expression from the promoter. Because RSC binding helps associate the UAS with a nucleosome containing a nonstandard H2A.Z histone (Floer et al., 2010), this result supports our hypothesis that loop formation is necessary for expression from the \textit{GAL1} promoter in the yeast genome.

**Measuring mRNA Levels to Probe Promoter Activity at the Transcriptional Level**

To provide insights into promoter activity at the transcriptional level, we measured YFP mRNA levels by performing RT-qPCR experiments on a set of strains (Figure 6A) containing the wild-type \textit{GAL1} (Figures 6B and 6C) or \textit{GAL10} (Figures 6B and 6D) promoters driving YFP and their edited versions. The mRNA data agreed with the data obtained at the protein (YFP) level in each strain; the promoter edits that led to nearly identical YFP profiles compared to the wild-type promoter also had similar mRNA profiles compared to the mRNA level of their wild-type counterpart (\textit{GAL1} or \textit{GAL10}). In addition, the edits that led to a total lack of YFP induction (Figure 6B) over the basal level compared to wild-type YFP also had no or very low mRNA levels compared to the mRNA levels of their wild-type counterparts (Figures 6C and 6D). While the strain in which an 11-bp region was deleted from the \textit{GAL1} UAS (edit 3) had no detectable YFP mRNA, two strains that had no detectable YFP induction showed a small level of YFP mRNA.

**Assessing the Effect of Galactose Induction Level on the Behavior of Edited Promoters**

In addition to providing additional insights into the bidirectional \textit{GAL1}/\textit{GAL10} promoter activity control at the transcriptional level, we also performed control experiments to see whether growing cells at higher galactose concentrations would lead to changes in our main results. For this, we tested a set of strains by growing them in 0.5% galactose or 2% galactose together with 0.1% mannose and measured the resulting YFP expression. For the strains expressing YFP at the 0.175% galactose induction, increasing the galactose concentration increased the fraction of ON cells, as expected, but for the strains not expressing YFP at 0.175% galactose induction, adding more galactose to the growth media did not change the results (Figure S9).

**Proposed Model of Directional Gene Expression Control**

Under our expanded model (Figure 7A), the first and fourth Gal4 binding sites near the base of the loop have evolved away from being sites for mediating Gal4-activated gene expression. These two outer Gal4 binding sites and the recently identified sites neighboring them (ATT and GTA before the first Gal4 binding site, and AAAG and AT after the fourth Gal4 binding site) primarily act as elements in a larger mechanism controlling directional gene expression. These two Gal4 binding sites are not responsible for general promoter activation. Their acting as part of a directional activation mechanism would explain why each set (ATT, GTA or AAAG, AT) is crucial for expression in only one direction. When the regulatory elements controlling only one direction are abolished, the promoter may only express in the opposite direction (Figure 7B). Finally, when the third Gal4 binding site is not bound by Gal4 due to introducing specific sequence alterations eliminating the binding motif, then Gal4 may still bind to its second binding site on the promoter, but the promoter is incapable of being activated in either direction (Figure 7C).
DISCUSSION

In this study, we experimentally demonstrated the existence of a directional regulatory mechanism controlling the activity of the *S. cerevisiae* GAL1/GAL10 bidirectional promoter. We also identified additional regulatory sites outside the UAS region and showed that the contribution of the Gal4 binding sites on transcriptional activity is not as straightforward as previously thought. For several decades, the binding site motif for Gal4 has been [CGG ... N11 ... CCG], with little thought given to the potential effect of the interior 11 base pairs on transcriptional activation. This was in part due to the fact that only recently has it been possible to perform high-fidelity (free from plasmid copy number variations and marker effects) precise genome editing on yeast and other cell types. Using a recently developed application of the CRISPR method, we showed how the interior sites can have specific effects on promoter activation and they cannot be categorized as a simple N11.

These results will have long-lasting consequences for bioinformatics studies hunting binding sites across genomes and genetic studies aiming to link genotype to phenotype. Here, we further showed that changing only the interior base pairs within the third Gal4 binding site completely eliminated gene expression by also affecting Gal4 binding on the promoter. In addition, our finding that the loss of the first or fourth Gal4 binding sites has no effect on either GAL1 or GAL10 expression so long as their outer edges (CGG for the first site and CCG for the fourth site) are kept intact indicates that not all sites having a consensus Gal4 binding motif are functionally identical for promoter activation.

To mechanistically explain our observations, we expanded upon a previously proposed model (Floer et al., 2010) in which the GAL1/GAL10 UAS region loops around a nucleosome in such a way as to expose the second and third Gal4 binding sites to mediate Gal4-facilitated activation, while the first and fourth Gal4 binding sites are not involved in promoter activation. We uncovered additional regulatory sites outside the UAS region. Our results strongly suggest that the first and fourth Gal4 binding sites have evolved to take part in a larger regulatory mechanism that ensures equal expression of Gal1 and Gal10 proteins.

The presence of evolutionary pressure for establishing such a regulatory mechanism that facilitates equal or symmetric GAL1 and GAL10 expression is further supported by the fact that the product of the enzymatic reaction catalyzed by the Gal1 galacto-kinase is toxic to yeast cells unless it is processed by the Gal7 and Gal10 enzymes (Mumma et al., 2008). In the absence of Gal10, yeast cells grown on galactose will exhibit growth defects as the toxic byproduct builds up. This presents the yeast cells with an evolutionary pressure to keep the expression of Gal1 and Gal10 at similar levels, because it may be worse for the cells to express one enzyme than to express neither of them.

We note that a recent paper (Jin et al., 2017) explored the evolution of promoter directionality in yeast. The authors concluded that newly created promoters are inherently bidirectional and are evolutionarily tuned to eventually support expression in only one direction. The directional control mechanism we present here may be one of those used to generate unidirectionality during the course of evolution. More specifically, we hypothesize that for the GAL1/GAL10 bidirectional promoter, the directional control regions we have discovered are products of an evolutionary process to ensure bidirectionality. These regions
would not be likely to exist in a random sequence of DNA representing a newly created promoter. However, the evolution of these DNA regions immediately flanking new UAS regions would be expected to be prime targets for evolution to increase expression in a preferred direction or to reduce expression in an unpreferred direction.

The fact that the $GAL1/GAL10$ promoter is known to be able to drive gene expression in a wide variety of eukaryotes suggests that the regulatory regions we have identified are likely recognized by the transcriptional regulation machineries in other eukaryotic cell types. Conserved recognition after millions of years of evolution would strongly suggest that the regulatory mechanism we describe here is not specific to yeast. Our results suggest the need to reevaluate results from previous studies on promoter structure and function. The widespread application of the CRISPR-based precise promoter editing technology will bring the research community many steps closer to having more reliable genotype-phenotype relations and will likely elucidate several additional regulatory mechanisms, with some of them potentially operating in other “well-characterized” promoters.

**STAR METHODS**

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Murat Acar (murat.acar@yale.edu).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

*Saccharomyces cerevisiae with the W303 genetic background*—All yeast *Saccharomyces cerevisiae* strains constructed carry the haploid W303 strain background (mating type alpha). Complete genomic descriptions of all constructed strains are provided in Table S1 with specific promoter edits detailed in Methods S1. When grown for transformations, yeast cells were grown overnight in 5 mL of YPD (yeast peptone dextrose) in a shaking incubator. When grown for FACS analysis, yeast cells were first grown overnight in 5 ml of minimal media (complete synthetic media minus the appropriate amino acids) containing 0.1% mannose as a carbon source. Upon reaching OD$_{600}$ = 0.1 after 22h of overnight shaker growth, cells were diluted and transferred to 5ml of the same minimal media containing 0.1% mannose as a carbon source (during this phase, the appropriate concentration of galactose was also added). FACS analysis was conducted after the cells reached OD$_{600}$ = 0.1 after 22h of shaker growth. Throughout these experiments, the shaker speed was set to 225 rpm and the temperature was 30 C. Any deviations from these standard growth conditions are detailed in the maintext or in the relevant methods sections.

**METHOD DETAILS**

**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. In order to create plasmids containing a single gRNA cassette, we ordered (Integrated DNA Technologies) a 388bp gRNA cassette sequence modified from the template created by DiCarlo et al. (DiCarlo et al., 2013) (Methods S1). The cassette consists of a promoter, terminator, and gRNA sequence which 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. In order 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.

**Yeast strain background and introducing the CAS9 gene into yeast**—All strains were created from WP35, a W303 strain with one copy of the $P_{\text{GAL1}}$-YFP construct inserted into the ho locus. The rest of the yeast genome was unaltered. Complete genomic descriptions of all strains are provided in Table S1. To insert the CAS9 gene into the yeast genome, we obtained plasmid #43802 from Addgene. This plasmid (p414-TEF1p-Cas9-CYC1t), constructed by DiCarlo et al. (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 and then linearized the plasmid in the TRP1 gene using MfeI enzyme and 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 promoter integrated in the ho locus**—The first step of the strain construction process for measuring expression from the wild-type and edited promoters introduced at the ho locus consisted of transforming the strain GE1, which contains a $P_{\text{GAL1}}$-YFP construct inserted into the ho locus as well as the CAS9 gene stably integrated into the yeast genome, with a plasmid containing gRNAs targeting the two sides flanking $P_{\text{GAL1}}$, with the goal of cutting out the entirety of $P_{\text{GAL1}}$. A donor oligonucleotide carrying a 30bp CRISPR cut and PAM site in between 50bp regions homologous to each genomic region flanking the GAL1 promoter was co-transformed (using the lithium acetate method) with the plasmid containing two gRNAs. This allowed the transformed cells to use the donor as a repair template after the two gRNAs facilitated cutting on both sides of the promoter, followed by the removal of the promoter from the genome. The transformed cells were then grown on minimal media plates missing leucine for two days, after which colonies were PCR tested and sequenced to verify that editing had occurred. After this, replica plating between rich media (YPD) and CSM minimal media missing leucine was performed to locate colonies which had lost the centromeric plasmid used in the first step of the 2-step CRISPR editing (Elison et al., 2017). At this point, the strain was ready to be integrated in its ho locus with any promoter (wild-type or edited) to drive YFP.

The second step of the strain construction process consisted of transforming the strain (constructed as described in the previous paragraph) carrying the 30bp inserted CRISPR cut site in its ho locus with a plasmid containing a gRNA cassette targeting the inserted CRISPR cut site in the presence of a donor DNA sequence whose content was either the wild-type or an edited $\text{GAL1/GAL10}$ promoter. The donor DNA sequence also included 50bp sequences (on both sides) that were homologous to the genomic regions flanking the inserted CRISPR cut site. The transformed cells were then grown for two days on CSM minimal media plates.
missing leucine, after which the colonies formed were PCR tested and sequenced to verify that editing had occurred. The lithium acetate method was used for the yeast transformations.

**Introducing rational edits in the endogenous GAL1/GAL10 locus**—The first step of strain construction process for measuring expression from the wild-type and edited promoters introduced at the endogenous GAL1/GAL10 locus consisted of transforming the strain GE1, which contains a P\textsubscript{GAL1}-YFP construct inserted into the \textit{ho} locus as well as CAS9 gene stably integrated into the genome, with a plasmid containing gRNAs targeting two sites flanking the endogenous P\textsubscript{GAL1/GAL10} GAL1 region, with the goal of cutting out the entirety of the P\textsubscript{GAL1/GAL10} GAL1. A donor oligonucleotide carrying a 30bp CRISPR cut and PAM site followed by a full copy of YFP with a cyc terminator in between 50bp regions homologous to each genomic region flanking P\textsubscript{GAL1/GAL10} GAL1 was co-transformed using the lithium acetate method with the plasmid containing two gRNAs. This allowed the transformed cells to use the donor as a repair template after the gRNAs cut on either side of the promoter and gene and removed them from the genome. The transformed cells were then grown on –LEU plates for two days after which colonies were PCR tested and sequenced to verify that editing had occurred. After this, replica plating between rich media (YPD) and CSM minimal media missing leucine was performed to locate colonies which had lost the centromeric plasmid used in the first step of the 2-step CRISPR editing (Elison et al., 2017). At this point, the native GAL locus consisted of the GAL10 gene (which contained a small truncation removing bases including the ATG as a result of the CRISPR/Cas9 action), a CRISPR cut site we introduced, and the YFP.

The second step of the strain construction process consisted of genome-integrating the GAL10P\textsubscript{GAL1/GAL10} GAL1 construct into the \textit{ho} locus. The entire GAL10P\textsubscript{GAL1/GAL10} GAL1 region from a wild-type strain was PCR-amplified using long primers carrying 60bp flanking regions homologous to both sides of the \textit{ho} region. The resulting linear construct (PCR product) was transformed (using the lithium acetate method) into the yeast strain that had been constructed as described in the previous paragraph. The construct was integrated in the \textit{ho} locus, and the resulting strain was tested for its capacity to grow on galactose-only media. The strain not containing the GAL10P\textsubscript{GAL1/GAL10} GAL1 construct could not grow on media containing galactose as the sole carbon source, but the strain containing the GAL10P\textsubscript{GAL1/GAL10} GAL1 integration in its \textit{ho} locus grew equally as well as the wild-type (Figure S1).

The final step of the strain construction process consisted of transforming (using the lithium acetate method) the strain of the previous paragraph (carrying a 30bp inserted CRISPR cut site followed by YFP integrated in the place of the endogenous P\textsubscript{GAL1/GAL10} GAL1 region and the GAL10P\textsubscript{GAL1/GAL10} GAL1 construct integrated in the \textit{ho} locus) with a plasmid containing a gRNA cassette targeting the inserted CRISPR cut site in the presence of a donor DNA sequence whose content is either the wild-type or an edited GAL1/GAL10 promoter. The donor DNA sequence also included 50bp sequences (on both sides) that were homologous to the genomic regions flanking the inserted CRISPR cut site. The transformed cells were then grown for two days on CSM minimal media plates missing leucine, after
which the colonies formed were PCR tested and sequenced to verify that editing had occurred.

**Growth conditions, media, and flow cytometry data analysis**—All growths testing for \( \text{GAL1} \) or \( \text{GAL10} \) expression were conducted in duplicate at 30°C in a shaking incubator in 5 mL of the appropriate synthetic dropout media. Cells were first grown overnight for 22 hours in minimal media containing 0.1% mannose as the carbon source, reaching a cell density (\( \text{OD}_{600} \)) between 0.075 and 0.15 after the 22 hours growth period. Cells were subsequently diluted into the induction media (minimal media containing 0.1% mannose and 0.175% galactose (or another galactose concentration as described in the paper)), and grown for another 22 hours, reaching a cell density between 0.075 and 0.15. After the induction period, single cell fluorescence values were analyzed using flow cytometry (BD FACSVerse). Each FACS sample had on average 10,000 cells after gating. Log-amplified fluorescence measurements for the gated cells were converted to linear scale for analysis. A threshold for ON state (150 a.u.) was selected based on fluorescence measurements from both unedited and uninduced cells and applied uniformly to all samples. The fraction of ON cells and the mean expression level of such cells was then quantified for relevant samples.

**Fitness measurements**—The growth conditions used for the doubling-time quantification experiments presented in Figure S4 were as follows. Cells were grown on appropriate synthetic amino acid dropout plates (minimal media plates containing 2% glucose). From these, a single colony was picked and suspended in the appropriate synthetic amino acid dropout media containing 0.1% mannose and 0.2% galactose as the carbon sources. All experiments were performed in duplicates for each strain. Then, the cell density (\( \text{OD}_{600} \)) in the suspended cultures was measured and the cultures were diluted, using the fresh media of the same content, in such a way that the cell density (\( \text{OD}_{600} \)) would reach a value between 0.1 and 0.5, after 22 hours of growth at 30°C shaker incubator (225 rpm). The actual cell densities right before the 22 hours growth period were calculated by using the measured densities right before the dilution and the dilution factor. After the 22 hours growth period, the cell densities (\( \text{OD}_{600} \)) were measured and recorded again. The doubling times of the cultures during the 22 hours growth period were then calculated using the measured cell densities right before and after the 22 hours growth period. For the calculations, for each replicate of each strain, the following equation was used to determine doubling time: \( \ln(2)/\ln(\text{OD}_{22h}/\text{OD}_{0h}/22\text{hours}) \) with the \( \text{OD}_{600} \) values measured immediately before and after the 22 hour growth period. Error bars indicate SEM (n = 2).

**rt-qPCR experiments**—Cells were first grown in the appropriate synthetic dropout media (minimal media) overnight for 22 hours, reaching an optical density (\( \text{OD}_{600} \)) between 0.075 and 0.15. The media contained 0.1% mannose as a non-inducing carbon source. Cells were subsequently diluted into the induction media containing 0.1% mannose and 0.175% galactose, and grown for another 22 hours, reaching a cell density between 0.5 and 0.6. RNA was then extracted from these log-phase cells. For analysis of mRNA expression under \( \text{P}_{\text{GAL1}} \) control, qPCR primers were designed to bind immediately downstream of the transcription start site and inside the YFP reporter (GT TAATACCTC TACTA CT ACGTCAAGGA, forward; GGTAGTTTCCGT

*Cell Rep. Author manuscript; available in PMC 2018 November 29.*
ATGTTGCAT, reverse). For analysis of mRNA expression under PGAL10 control, qPCR primers were designed to bind on the YFP reporter sequence downstream of the PGAL10 (TGGCCAACACTTGTCACTACT, forward; ACGTGTCTTGTAGTTCCCGTC, reverse). Relative gene expression was analyzed using the –DDCT method: results from the measurements were first normalized to the mRNA expression level of ALG9, then results from the measurements of each edited strain were normalized to either the measurement of the wild-type (WT) GAL1 or WT GAL10 expression, as appropriate. Primers used for the analysis of ALG9 gene expression were CACGGATAGTGGCTTTGGTGA (forward) and TATGATTATCTGCGACAGGAAAAGA (reverse).

**ChIP-qPCR Experiments**—Cells were first grown in the appropriate synthetic dropout media (minimal media) overnight for 22 hours, reaching an optical density (OD600) between 0.075 and 0.15. The media contained 0.1% mannose as a non-inducing carbon source. Cells were subsequently diluted into the induction media containing 0.1% mannose and 0.175% galactose, and grown for another 22 hours, reaching a cell density (OD600) between 0.5 and 0.6. The cells were grown to a higher density than in other procedures due to harvesting requirements for this assay. ChIP experiments were performed as previously described (Xue and Acar 2018). Briefly, 4 × 10^8 cells were collected for each ChIP reaction, and cell densities were pre-adjusted to a density of 1× 10^7 cells/ml in 40 mL culture. Formaldehyde was added directly to the culture at a final concentration of 1%. To stop crosslinking, 6 mL of 2.5 M glycine was added, and samples were incubated at room temperature for 5 min before harvesting. ChIP enrichment values were calculated as: (IP-secondary anti-body)/ Input * 100%. Anti-Gal4 (ab1396, Abcam) and anti-rabbit (D2114, Santa Cruz) antibodies were used. Primers used for ChIP-qPCR were: TGTGATGTGAGAACTGTATCCTAGC (upstream of the UAS region, forward) and TAATCATTTGCATCCATAC ATTTTG (upstream of the UAS region, reverse); TAATACGCTTAACTGCTCATTGCT (within the UAS region, forward) and CCAA TTTTTCCTTCTTCCATAACCAT (within the UAS region, reverse); TGGCCAACACTTGTCACTACT (downstream of the UAS region, forward) and ACGTGTCTTGTAGTTCCCGTC (downstream of the UAS region, reverse).

**QUANTIFICATION AND STATISTICAL DATA ANALYSIS**—Each sample of flow cytometry data were analyzed in R using the Bioconductor flowCore software package (Ellis et al., 2018). Raw flow cytometry data were gated using a gate selected based on the forward scattering (FSC) and side scattering (SSC) measurements. The gate chosen is an ellipse whose bounding rectangle whose upper right corner is at FSC = 57000 a.u., SSC = 53000 a.u., whose lower left corner is at FSC = 33000 a.u., SSC = 25000 a.u., and whose edges are parallel to the axes. Each sample has about 10,000 cells on average after gating. Log-amplified fluorescence measurements for the gated cells were converted to linear scale for analysis. The resulting single-cell fluorescence measurements of all samples/replicates obtained from each strain were combined together by binning the data and plotting it as a histogram. The error bar on each bin represents the standard error of the mean for each bin. Then, a threshold for ON state (150 a.u.) was selected based on fluorescence measurements from both unedited and uninduced cells and applied uniformly to all samples. The fraction of ON cells and the mean expression level of such cells was then quantified for relevant samples.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Acar Lab members for helpful discussions and W. Peng for providing the parent strain WP35. M.A. acknowledges funding from the NIH (1DP2AG050461-01 and 1U54CA209992-01). G.L.E. was in part supported by a training grant from the NIH (T32 GM007499). p414-TEF1p-Cas9-CYC1t was a gift from George Church (Addgene plasmid no. 43802).

REFERENCES

Acar M, Becskei A, and van Oudenaarden A (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature 435, 228–232. [PubMed: 15889097]
Acar M, Pando BF, Arnold FH, Elowitz MB, and van Oudenaarden A (2010). A general mechanism for network-dosage compensation in gene circuits. Science 329, 1656–1660. [PubMed: 20929850]
Bhat PJ, and Hopper JE (1992). Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast GAL/MEL regulon. Mol. Cell. Biol 12, 2701–2707. [PubMed: 1317007]
Bhaumik SR, and Green MR (2001). SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes Dev 15, 1935–1945. [PubMed: 11485988]
Bryant GO, and Ptashne M (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. Mol. Cell 11, 1301–1309. [PubMed: 12769853]
Bryant GO, Prabhu V, Floer M, Wang X, Spagna D, Schreiber D, and Ptashne M (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol 6, 2928–2939. [PubMed: 19108605]
DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, and Church GM (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41, 4336–4343. [PubMed: 23460208]
Douglas HC, and Hawthorne DC (1964). Enzymatic expression and genetic linkage of genes controlling galactose utilization in Saccharomyces. Genetics 49, 837–844. [PubMed: 14158615]
Elison GL, Song R, and Acar M (2017). A precise genome editing method reveals insights into the activity of eukaryotic promoters. Cell Rep 18, 275–286. [PubMed: 28052256]
Ellis B, Haaland P, Hahne F, Meur NL, Gopalakrishnan N, Spidlen J and Jiang M (2018). flowCore: basic structures for flow cytometry data. Version 1.46.1 (R Foundation for Statistical Computing).
Floer M, Wang X, Prabhu V, Berrozpe G, Narayan S, Spagna D, Alvarez D, Kendall J, Krasnitz A, Stepansky A, et al. (2010). A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. Cell 141, 407–418. [PubMed: 20434983]
Jin Y, Eser U, Struhl K, and Churchman LS (2017). The ground state and evolution of promoter region directionality. Cell 170, 889–898. e10. [PubMed: 28803729]
Johnston M (1987). A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol. Rev 51, 458–476. [PubMed: 2830478]
Johnston M, and Davis RW (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. Mol. Cell. Biol 4, 1440–1448. [PubMed: 6092912]
Kellis M, Patterson N, Endrizzi M, Birren B, and Lander ES (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423, 241–254. [PubMed: 12748633]
Larschan E, and Winston F (2001). The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. Genes Dev 15, 1946–1956. [PubMed: 11485989]
Lin L, Chamberlain L, Zhu LJ, and Green MR (2012). Analysis of Gal4-directed transcription activation using Tra1 mutants selectively defective for interaction with Gal4. Proc. Natl. Acad. Sci. USA 109, 1997–2002. [PubMed: 22308403]

Cell Rep. Author manuscript; available in PMC 2018 November 29.
Liu P, Song R, Elison GL, Peng W, and Acar M (2017). Noise reduction as an emergent property of single-cell aging. Nat. Commun 8, 680. [PubMed: 28947742]

Melcher K, and Johnston SA (1995). GAL4 interacts with TATA-binding protein and coactivators. Mol. Cell. Biol 15, 2839–2848. [PubMed: 7739564]

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. [PubMed: 12727884]

Mumma JO, Chhay JS, Ross KL, Eaton JS, Newell-Litwa KA, and Fridovich-Keil JL (2008). Distinct roles of galactose-1P in galactose-mediated growth arrest of yeast deficient in galactose-1P uridylyltransferase (GALT) and UDP-galactose 4’-epimerase (GALE). Mol. Genet. Metab 93, 160–171. [PubMed: 17981065]

Nehlin JO, Carlberg M, and Ronne H (1991). Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J 10, 3373–3377. [PubMed: 1915298]

Peng W, Liu P, Xue Y, and Acar M (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat. Commun 6, 6226. [PubMed: 25670371]

Peng W, Song R, and Acar M (2016). Noise reduction facilitated by dosage compensation in gene networks. Nat. Commun 7, 12959. [PubMed: 27694830]

Skene PJ, and Henikoff S (2015). A simple method for generating high-resolution maps of genome-wide protein binding. eLife 4, e09225. [PubMed: 26079792]

Weinhandl K, Winkler M, Glieder A, and Camattari A (2014). Carbon source dependent promoters in yeasts. Microb. Cell Fact 13, 5. [PubMed: 24401081]

Xue Y, and Acar M (2018). Live-cell imaging of chromatin condensation dynamics by CRISPR. iScience 4, 216–235. [PubMed: 30027155]

Zhou X, and O’Shea EK (2011). Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. Mol. Cell 42, 826–836. [PubMed: 21700227]
Highlights

- Editing of the GAL1/GAL10 promoter reveals distinct regulatory sites
- The regulatory sites contribute to directional regulation of gene expression
- The first and fourth Gal4 binding sites do not contribute to promoter activation
- Gal4 binding to the third binding site is critical for promoter activation
Figure 1. Identification of Additional Regulatory Proteins within the Yeast GAL1 Promoter

(A) The layout of the GAL1 promoter. The Gal4 binding sites are colored such that red indicates the first Gal4 binding site, orange the second, green the third, and blue the fourth. The yellow rectangles are TATA boxes, and the white arrow indicates the direction in which expression is being measured. The histograms show the response of the wild-type strain (strain GE1) to a variety of galactose conditions from uninduced (0% galactose) to fully induced (2% galactose).

(B) Histogram of $P_{GAL1}$-YFP expression levels using the wild-type promoter (strain GE1).

(C) Expression histogram from an edited GAL1 promoter (strain GE117), with the edits not altering GAL1 expression compared to wild-type.

(D) Expression histogram from an edited GAL1 promoter (with edits immediately upstream of the first Gal4 binding site; strain GE210), indicating a loss of GAL1 expression.

(E) Expression histogram from an edited GAL1 promoter (edited 11 bp downstream of the fourth Gal4 binding site; strain GE163), indicating a loss of GAL1 expression.

(F) Schematic displaying all previously known and additional discovered regulatory sites within the GAL1 promoter. Previously known regions are depicted as in (A), and recently discovered regions are depicted by light blue hexagons. Black arrows indicate where the promoter has been edited. Upper bars show the wild-type genotype of the region being edited. Lower bars show the same region containing whatever edits were made, which are indicated by capital letters. The wild-type does not have bars because it is unedited. Bases written in a colored font are provided to indicate bases that are
part of the binding site depicted in that color in (A). Panels depicting a fluorescence-activated cell sorting (FACS) histogram were created using 10,000 cells of the strain represented by the lower bar. Strain data error bars indicate SEMs (n = 2). The percentage given in any histogram indicates the concentration of galactose used as a carbon source. In addition to galactose, all of the strains were provided with 0.1% mannose as a neutral sugar.
Figure 2. Identification of Additional Regulatory Regions within the Yeast GAL10 Promoter

(A) The layout of the GAL10 promoter. The Gal4 binding sites keep the coloring scheme used in Figure 1A, but the promoter is expressing in the upstream direction, as shown by the white arrow. The left histogram shows wild-type P_{GAL10}-YFP expression (strain GE105) and is compared to the right histogram, which shows wild-type P_{GAL1}-YFP expression (strain GE1). The percentages and means of ON cells are displayed for both.

(B) Expression histogram from an edited GAL10 promoter (edited 6 bp upstream of the first Gal4 binding site; strain GE135), indicating a loss of GAL10 expression.

(C) Expression histogram from an edited GAL10 promoter (edited 6 bp downstream of the fourth Gal4 binding site; strain GE168), indicating a loss of GAL10 expression.

(D) Schematic displaying all previously known and recently discovered regulatory sites within the GAL10 promoter. Previously unidentified regions are depicted by purple hexagons.

(E) Schematic displaying all previously known and recently discovered regulatory sites for both the GAL1 and GAL10 promoters. The additional GAL1 promoter-related regions are depicted by light blue hexagons, and recently identified GAL10 promoter-related regions are depicted by purple hexagons. Black arrows indicate where the promoter has been edited. Upper bars show the wild-type genotype of the region being edited. Lower bars show the same region containing whatever edits were made, which are indicated by capital letters. The wild-type does not have bars because it is unedited. Bases written in a colored font are provided to indicate bases that are part of the binding site depicted in that color in Figure 1A. Panels depicting an FACS histogram were created using 10,000 cells of the strain represented by the lower bar. Strain data error bars indicate SEMs (n = 2).
Figure 3. Characterization of the Phenotypic Impacts of the First and Fourth Gal4 Binding Sites

(A) Histogram of $P_{\text{GAL1}}$ YFP expression levels using the wild-type promoter (strain GE1).

(B) Expression histogram from an edited GAL1 promoter (the first three critical base pairs [CGG] of the first Gal4 binding site have been disrupted; strain GE164), indicating loss of GAL1 expression.

(C) Expression histogram from an edited GAL1 promoter (the three base pairs [ATT] immediately adjacent to those in [B] have been disrupted; strain GE149), indicating wild-type GAL1 expression.

(D) Expression histogram from an edited GAL1 promoter (the second three critical base pairs [CCG] of the first Gal4 binding site have been disrupted; strain GE165).

(E) Bar graph showing the mean ON expression for both GAL1 (blue) and GAL10 (red) expression of four different edits introduced on the shared promoter. The first edit disrupted the first three critical base pairs of the first Gal4 binding site (GAL1: strain GE164; GAL10: strain GE193). The second edit disrupted the second three critical base pairs of the first Gal4 binding site (GAL1: strain GE165; GAL10: strain GE194). The third edit disrupted the first three critical base pairs of the fourth Gal4 binding site (GAL1: strain GE182; GAL10: strain GE204). The fourth edit disrupted the second three critical base pairs of the fourth Gal4 binding site (GAL1: strain GE183; GAL10: strain GE196).

Black arrows indicate where the promoter has been edited. Upper bars show the wild-type genotype of the region being edited. Lower bars show the same region containing whatever
edits were made, which are indicated by capital letters. The wild-type does not have bars because it is unedited. Bases written in a colored font are provided to indicate bases that are part of the binding site depicted in that color in Figure 1A. Panels depicting an FACS histogram were created using 10,000 cells of the strain represented by the lower bar. Strain data error bars indicate SEMs (n = 2).
Figure 4. The Interior Bases of Gal4 Binding Sites Are Critical for GAL1 and GAL10 Expression

(A) Diagram of the GAL1/GAL10 promoter, with the base pairs making up the first three Gal4 binding sites indicated. Colors agree with the scheme used in Figure 1A.

(B) Histogram of $P_{\text{GAL1}}$-YFP expression levels using the wild-type promoter (strain GE1). Mean expression of the ON peak is indicated in arbitrary units (a.u.).

(C) Expression histograms for GAL1 (left: strain GE140) and GAL10 (right: strain GE190) expression from a strain in which the interior of the first Gal4 binding site has been disrupted.

(D) Expression histograms for GAL1 (left: strain GE184) and GAL10 (right: strain GE197) expression from a strain in which the interior of the second Gal4 binding site has been disrupted. Mean expression of the ON peaks is indicated in arbitrary units.

(E) Expression histograms for GAL1 (left: strain GE153) and GAL10 (right: strain GE170) expression from a strain in which the interior of the third Gal4 binding site has been recoded.

Black arrows indicate where the promoter has been edited. Upper bars show the wild-type genotype of the region being edited. Lower bars show the same region containing whatever edits were made, which are indicated by capital letters. The wild-type does not have bars because it is unedited. Bases written in a colored font are provided to indicate bases that are part of the binding site depicted in that color in Figure 1A. Panels depicting an FACS
histogram were created using 10,000 cells of the strain represented by the lower bar. Strain data error bars indicate SEMs (n = 2).
Figure 5. Characterization of GAL1 Promoter Activation Mechanisms

(A) Models indicating where qPCR primer pairs bind to the wild-type (WT) GAL1 promoter (top) or the edited GAL1 promoter (bottom). The upstream pair is shown with black arrows, the UAS pair is shown with blue arrows, and the downstream pair is shown with red arrows. The empty green box on the bottom promoter indicates that the third Gal4 binding site has been recoded.

(B) Expression histogram of the WT GAL1 promoter (strain GE230), taken immediately before cells went through the ChIP-qPCR experiments. A total of 10,000 cells were analyzed and are shown in the histogram. Strain data error bars indicate SEMs (n = 2).

(C) Expression histogram of the edited GAL1 promoter (strain GE231), taken immediately before cells went through the ChIP-qPCR experiments. A total of 10,000 cells were analyzed and are shown in the histogram. Strain data error bars indicate SEMs (n = 2).

(D) ChIP-qPCR data for a strain containing the WT GAL1 promoter (strain GE230) and a strain containing an edited version of the GAL1 promoter (strain GE231), in which the third binding site had been recoded. Both strains had the endogenous GAL1 and GAL10 promoters and genes knocked out. The Gal4 binding on the promoter was quantified by having qPCR primers targeting three areas: upstream of the promoter, on top of the UAS region, and downstream of the promoter.
(E) Expression histogram from the WT *GAL1* promoter (strain GE1). Mean expression (in arbitrary units) and percentage of cells in the ON peak are indicated. The cartoon above the histogram indicates the four Gal4 binding sites and the distance in base pairs between the third and fourth Gal4 binding sites.

(F) Expression histogram from an edited *GAL1* promoter (edited such that an 11-bp region between the third and fourth Gal4 binding sites has been deleted; strain GE155), indicating loss of *GAL1* expression.

(G) Expression histogram from an edited *GAL1* promoter (edited such that a 21-bp region between the third and fourth Gal4 binding sites has been deleted; strain GE156), indicating loss of *GAL1* expression.

(H) Expression histogram from an edited *GAL1* promoter (edited such that the same 11-bp region between the third and fourth Gal4 binding sites deleted in [F] has been recoded while retaining the WT distance between the third and fourth Gal4 binding sites; strain GE206), indicating *GAL1* expression similar to WT. Mean expression (in arbitrary units) and percentage of cells in the ON peak are indicated.

(I) Expression histogram from an edited *GAL1* promoter (edited such that all three RSC binding sites have been recoded; strain GE241), indicating loss of *GAL1* expression.

In (E)–(I), bars indicate the four Gal4 binding sites, and numbers within the bars indicate the number of base pairs between the third and fourth binding sites. Empty black boxes indicate that a region has been recoded. Panels depicting an expression histogram were created using 10,000 cells of the strain represented by the lower bar. Strain data error bars indicate SEMs (n = 2).
Figure 6. Quantification of mRNA Levels between WT and Edited GAL1 and GAL10 Promoters
(A) Genotype identification of each of the eight strains analyzed for YFP mRNA levels. Upper bars show the WT genotype of the region being edited. Lower bars show the same region containing whatever edits were made, which are indicated by capital letters. The WT does not have bars because it is unedited. Bases written in a colored font are provided to indicate bases that are part of the binding site depicted in that color in Figure 1A.
(B) FACS profiles of the eight strains analyzed for YFP mRNA levels. All of the histograms were created using 10,000 cells of the strain, as indicated in the upper right corner of the individual histograms. Strain data error bars indicate SEMs (n = 2).
(C) YFP mRNA quantification from four edited promoters (strains GE210, GE155, GE151, and GE206) and the WT GAL1 promoter (strain GE1). Error bars indicate SEMs (n = 2–4).
(D) YFP mRNA quantification from two edited promoters (strains GE218 and GE168) and the WT GAL10 promoter (strain GE105). Error bars indicate SEMs (n = 2–4).
Figure 7. Proposed Model of Directional Gene Expression Control

(A) Model representing the interaction of Gal4 binding sites, Gal4, and the recently identified regulatory regions for GAL1/GAL10 promoter in the WT promoter. The DNA is looped such that only the second (orange) and third (green) Gal4 binding sites are involved in promoter activation. The recently identified regulatory regions (light blue hexagon and the first Gal4 site [red]; purple hexagon and the fourth Gal4 site [blue]) are necessary for directional activation of the promoter in the GAL1 or GAL10 direction. The yellow arrows demonstrate that both GAL1 and GAL10 expression is present.
(B) Model representing a strain in which the directional elements necessary for *GAL1* expression have been recoded at one of the three regions identified as necessary for promoter activation. The recoding is represented by the empty shapes. The yellow arrow demonstrates that only *GAL10* expression is present.

(C) Model representing a strain in which the interior of the third Gal4 binding site (green) has been recoded, indicated by the empty box. Although Gal4 may still bind to the second Gal4 binding site (orange) and the directional elements are still intact, the promoter is not able to express either *GAL1* or *GAL10*, as indicated by the lack of yellow arrows.
## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| Rabbit polyclonal anti-GAL4 | Abcam | Cat# Ab1396 |
| Goat anti-Rabbit IgG-HRP | Santa Cruz | Cat# D2114 |
| **Bacterial and Virus Strains** | | |
| Biological Samples | N/A | N/A |
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| Formaldehyde solution | Sigma-Aldrich | Cat# F8775 |
| Glycine | Americanbio | Cat# AB00730 |
| **Critical Commercial Assays** | | |
| RNeasy Mini Kit | QIAGEN | Cat# 74106 |
| Fast SYBR Green Master Mix | Thermo Fisher | Cat# 4385612 |
| SuperScript IV VILO Master Mix | Thermo Fisher | Cat# 11756050 |
| **Deposited Data** | | |
| *S. cerevisiae* GAL1/GAL10 promoter sequences: see Methods S1 | This Paper | N/A |
| CRISPR cut site replacing the endogenous GAL1/GAL10 promoter in yeast strain GE6: GenBank ID code: MH888038 | This Paper | N/A |
| gRNA cassette sequence: GenBank ID code: KT876200 | DiCarlo et al., 2013. | N/A |
| **Experimental Models: Organisms/Strains** | | |
| *S. cerevisiae* Strain background: W303 | Acar et al., 2005. | N/A |
| *S. cerevisiae* Strains created in this paper, see Table S1 | This Paper | N/A |
| **Oligonucleotides** | | |
| Primers for ChIP analysis, see Figure 5 | This paper | N/A |
| Primers for rt-qPCR analysis, see Figure 6 | This paper | N/A |
| Yeast gRNA template: see Methods S1 | DiCarlo et al., 2013. | N/A |
| **Recombinant DNA** | | |
| Plasmid: p414-TEF1p-Cas9-CYC1t | DiCarlo et al., 2013. Addgene. | Addgene #43802 |
| **Software and Algorithms** | | |
| flowCore | Ellis et al., 2018 | [https://doi.org/10.18129/B9.bioc.flowCore](https://doi.org/10.18129/B9.bioc.flowCore) |
| BD FACS Suite software | BD Biosciences | N/A |
| **Other** | | |
| BD FACSVerse | BD Biosciences | [https://wwwbdbiosciencescom/us/instruments/research/cell-analyzers/bd-facsverse/n/1298832/overview](https://wwwbdbiosciencescom/us/instruments/research/cell-analyzers/bd-facsverse/n/1298832/overview) |