Analysis of Transcription Activation Distance as a Polygenic Trait in Saccharomyces cerevisiae

Analysis of Transcription Activation Distance as a Polygenic Trait in *Saccharomyces cerevisiae*

A dissertation presented

by

Caitlin Teresa Reavey

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University

Cambridge, Massachusetts

June 2013

Dissertation Advisor: Professor Fred Winston Caitlin Teresa Reavey

Analysis of Transcription Activation Distance as a Polygenic Trait in *Saccharomyces cerevisiae*

Abstract

Much of the eukaryotic transcriptional machinery is conserved from yeast to human. However, the distance over which transcriptional activation can occur differs between *Saccharomyces cerevisiae* and metazoans. In *S. cerevisiae,* the upstream activating sequence (UAS) is generally found within 300 base pairs of the transcription start site; when the UAS is moved too far away, activation no longer occurs. In contrast, metazoan enhancers can activate from as far as 100 kilobases from the start site. In past work, our lab identified five genes that, when mutant, allow transcription activation to occur at a greater-than-normal distance from the *GAL1* UAS. As this long-distance activation phenotype was weak, we have now studied long-distance activation as a polygenic trait, isolating strains with multiple mutations that together confer a strong phenotype. To do this, we constructed strains containing two reporters, *HIS3* and *URA3*. For each reporter, the *GAL1* UAS was placed approximately 800 base pairs upstream of the transcription start sites. By iterative selection for stronger and stronger expression of *HIS3*, followed by screening for stronger expression of *URA3*, we isolated three strains, each containing multiple mutations that contribute to the strength of the long distance activation phenotype. Causative mutations were identified in *MOT3*, *GRR1*, *MIT1*, *PTR3*, *YOR019W*, and *MSN2* that contribute to the long distance activation phenotype. Strains containing multiple mutations were found to activate the reporter construct at distances up to 2 kilobases. Microarray analysis revealed genome wide transcriptional

changes in the mutant strains. Statistical analysis of the microarray results suggests other potential sites of long distance activation throughout out the genome. These results have extended our understanding of mutations that allow long distance activation and have demonstrated the value of studying a phenotype as a polygenic trait.

Table of Contents

List of Figures

List of Tables

Chapter 2

Chapter 3

Appendix 1

Appendix 2

Appendix 3

Chapter 1

Introduction

The activation of transcription initiation is one of the most important steps in the regulation of gene expression. Work over the past few decades has provided vast insight into the proteins and mechanisms governing this step. In eukaryotes, most fundamental aspects of transcription initiation are conserved from yeast to humans. This conservation, as well as the ability to perform high-resolution genetic and genomic studies, has made yeast a valuable organism for studies on transcription activation. However, one striking difference between transcription activation in yeast and metazoans is the distance between the core promoter and the sites where transcription activators bind. In yeast, activator binding sites (upstream activation sequences or UASs) are generally found within 600 base pairs (bp) of the transcription start site (Xue et al., 2004). In contrast, in metazoans, the activator binding sites (enhancers) are sometimes located as far as one megabase away (Lettice et al., 2003). This raises the intriguing issues of whether activation distance is regulated in yeast and, if so, by what factors.

The focus of this dissertation was to take a genetic approach to study the factors that regulate transcription activation distance in yeast. In this introduction, I will briefly review transcription initiation by discussing the *trans*-acting factors and the cis-regulatory elements involved in initiation. I will introduce the models for long-distance transcription activation and summarize previous studies on regulation of activation distance in yeast. Finally, I will review the study of complex traits in yeast, as our genetic approach involved the isolation of polygenic mutants.

TRANS-ACTING FACTORS INVOLVED IN TRANSCRIPTION INITIATION

General Transcription Factors

The set of factors involved in initiation of RNA polymerase II are referred to as general initiation factors (Hahn and Young, 2011; Sikorski and Buratowski, 2009; Thomas and Chiang, 2006). These factors were originally biochemically purified from human cells and shown to be necessary for transcription initiation from the adenovirus major late promoter (Matsui et al., 1980; Samuels et al., 1982). The general initiation factors, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, are named for the biochemical fraction in which they were isolated (Thomas and Chiang, 2006). Together these factors are required for recognition of the promoter, start site selection, clearing of the template, and opening promoter DNA (Hahn and Young, 2011; Sikorski and Buratowski, 2009; Smale and Kadonaga, 2003; Thomas and Chiang, 2006).

Transcription Activators

Transcription activation is regulated by the binding of activators to DNA in a site-specific manor. Once an activator binds a regulatory site, it must communicate this activating signal to RNA polymerase II to promote transcription. Activators often have a role in the regulation of genes in response to an environmental or developmental change. Therefore, it is important that the activators themselves are regulated, either at the level of binding to DNA or at the level of activation once bound. Promoters often contain

multiple activator binding sites and multiple activators can function together to promote transcription (Hahn and Young, 2011).

There are several structural features shared by most activators. Activators typically contain a DNA binding domain and an activation domain that contacts other proteins. These domains usually function independently of each other and can in some cases be swapped between activators. This was first demonstrated when the LexA DNA binding domain from *E. coli* was fused to the *S. cerevisiae* Gal4 activation domain. The fusion protein was able to activate transcription in yeast, dependent on the presence of a LexA operator site (Brent and Ptashne, 1985). DNA binding motifs contain several common conserved structural domains such as helix turn helix motifs, leucine zippers, and zinc fingers (Gill and Tjian, 1992). Activation domains are less well structured than DNA-binding domains; however, there are several common motifs of activation domains (Hahn and Young, 2011). One such motif is an acidic domain containing critical hydrophobic residues, seen in activators such as Gal4, Gcn4, and VP16 (Jackson et al., 1996; Jonker et al., 2005). Other common motifs are a glutamine rich domain seen in Sp1 and Oct2 and a proline rich domain seen in p53, CTF1, and AP2 (Mitchell and Tjian, 1989; Tanaka and Herr, 1990; Williams and Tjian, 1991; Zarrinpar et al., 2003).

Co-activators

Co-activators are an additional class of component required for activation-dependent transcription (Sikorski and Buratowski, 2009; Smale and Kadonaga, 2003;

Thomas and Chiang, 2006). Co-activators are generally recruited to the promoter through an interaction with the activation domain of activators. Co-activators can function by bridging the signal from activators to RNA polymerase II or other components of the general initiation machinery, or by chromatin modification (Hahn and Young, 2011). These activities are important for the formation of the pre-initiation complex. The order of recruitment of factors to the promoter is dependent on the gene. For example, in yeast at *GAL1*, the SAGA co-activator complex helps recruit Mediator, although several studies found lower levels of Mediator are recruited in the absence of SAGA. ((Bhaumik et al., 2004; Bryant and Ptashne, 2003; Larschan and Winston, 2005; Lemieux and Gaudreau, 2004). In this example, co-activators are functioning cooperatively to promote transcription. In contrast, at Gcn4 dependent promoters, SAGA and the Mediator co-activator complex are recruited independently (Govind et al., 2005; Qiu et al., 2005).

Mediator

The Mediator co-activator complex is of particular interest to this study as two components of Mediator, Sin4 and Rgr1, were identified as factors involved in regulation of activation distance (Dobi and Winston, 2007; Dobi and Leeman, unpublished). Mediator is a co-activator that was originally identified in yeast using two approaches. The first was a genetic approach looking for suppressors of CTD-truncations of the Rpb1 subunit of RNA polymerase II (Koleske et al., 1992; Koleske and Young, 1994; Thompson et al., 1993). Mediator was also identified by biochemical purification based

on its ability to promote transcription activation in vitro (Kim et al., 1994). Yeast Mediator can be purified as part of a holoenzyme complex with RNA polymerase II or on its own (Kim et al., 1994).

Mediator is a multi-subunit complex comprised of a head, middle, tail, and kinase module (Dotson et al., 2000). The head module interacts with RNA polymerase II (Takagi 2006). The tail module is the site of activator binding; in yeast, the Med15 tail subunit has been shown to bind Gcn4 (Herbig et al., 2010). The kinase module is comprised of the Srb8, Srb9, Srb10, and Srb11 subunits, where Srb10 is a cyclin dependent kinase and Srb11 is the cyclin. This module is not always found associated with the complex (Borggrefe et al., 2002).

Mediator is an evolutionarily conserved co-activator. However, mammalian Mediator is more complex ((Dotson et al., 2000). For example, human Mediator can exist in a variety of forms, with the major forms known as PC2 and TRAP (Malik et al., 2005). Electron microscopy of yeast Mediator, murine Mediator, and human TRAP complex revealed that the structure of the head module is the most conserved of the modules, while greater differences are seen between yeast and mammalian Mediator in the middle and tail modules. (Asturias et al., 1999; Dotson et al., 2000). In fact, Sin4, which is present in the yeast tail domain, is absent in the tail domain of murine Mediator and human TRAP complex (Asturias et al., 1999; Dotson et al., 2000).

Sin4 and Rgr1 are components of the tail module of yeast Mediator. Sin4 is non-essential while Rgr1 is essential for viability. Truncation of the Rgr1 C-terminus results in a decrease in Rgr1 function and has been used to study the role of Rgr1 in gene regulation (Sakai et al., 1988). Sin4 and Rgr1 have been identified as negative regulators of a number of genes including *HO*, *IME1*, and *SUC2* (Covitz et al., 1994; Jiang et al., 1995; Jiang and Stillman, 1995; Sakai et al., 1988). Sin4 is also involved in the positive regulation of several genes including *HIS4*, *CTS1*, *Ty1*, and *MATα2*. (Jiang and Stillman, 1992, 1995)*sin4* and *rgr1* mutants display a similar spectrum of phenotypes including altered transcription and altered chromatin structure (Jiang and Stillman, 1992).

CIS-REGULATORY ELEMENTS OF TRANSCRIPTION

In this section, I will introduce five classes of *cis*-regulatory elements that have been studied in eukaryotic promoters. The first two classes, the TATA box and initiator element are common to yeast and metazoans. The next class is the UAS, the site of activator binding in yeast. Lastly, I will introduce the upstream promoter element, a component of the metazoan core promoter, and enhancers, the site of activator binding in metazoans.

The TATA Box

The first eukaryotic core promoter element identified was the TATA box, which is the binding site of TATA binding protein. Original studies of the TATA box suggested

that it was a conserved core promoter element (Sikorski and Buratowski, 2009; Smale and Kadonaga, 2003). A genome-wide analysis indicates that only ~24% of human core promoters contain a TATA box (Carninci et al., 2006; Yang et al., 2007), and in yeast only ~20% of yeast core promoters contain a TATA box (Basehoar et al., 2004; Yang et al., 2007). The same study of yeast TATA boxes also yielded a consensus sequence of TATA(A/T)A(A/T)(A/G). Interestingly, 99% of the pre-initiation complexes at TATA-less promoters are bound to a sequence that has two or less mismatches to the consensus TATA sequence (Rhee and Pugh, 2012). These TATA-like elements seem to function similarly to the TATA box, suggesting very few yeast genes are truly TATA-less.

The position of the TATA box is well defined in metazoans and in *Schizosaccharomyces pombe*, occurring 25-30 bp upstream of the transcription start site; however, in *S. cerevisiae*, it is much more variable, occurring 40-120 bp upstream from the start site (Sikorski and Buratowski, 2009; Smale and Kadonaga, 2003). This difference in distance is determined by TFIIB and RNA polymerase II. When TFIIB and RNA polymerase II are purified from *S. pombe* and combined with *S. cerevisiae* general initiation factors, initiation occurs *in vitro* 25-30 bp downstream of the TATA box. Conversely, when *S. cerevisiae* TFIIB and RNA polymerase II are combined with *S. pombe* basal transcription factors, initiation occurs 40-120 bp downstream of the TATA box (Li et al., 1994).

Initiator Element

Another core promoter sequence is the initiator element, the sequence surrounding the site at which RNA polymerase initiates transcription. 5'-SAGE mapping or transcription start sites and subsequent sequence alignment defined the *S. cerevisiae* consensus initiator to be $A(A_{rich})_5NPyA(A/T)NN(A_{rich})_6$ (Zhang and Dietrich, 2005). In metazoans, the initiator element is sufficient for determining the transcription start sites in TATA-less promoters and is able to enhance the strength of promoters containing a TATA box (Javahery et al., 1994).

Upstream Activation Sequences

Upstream activation sequences (UASs) are the promoter elements in yeast that serve as the binding sites for activators. UASs are required for activated transcription, function in either orientation, and are unable to function when placed downstream of the TATA box (Guarente and Hoar, 1984; Struhl, 1984). Several genome-wide datasets, including global transcription factor binding site ChIP data, identification of conserved regulatory motifs, and nucleosome location data, have increased the ease of identifying UASs (Borneman et al., 2007; Cliften et al., 2003; Hu et al., 2010; Narlikar et al., 2007). Most UASs are present within 600 bp upstream of the transcription start site (Xue et al., 2004) and in some cases, multiple UASs are present in the promoter of a gene (McBride et al., 1997; Simon et al., 2001). Many UASs are bound by activators in response to a change in environment; for example, the *CYC1* UAS is bound by Hap1 in the presence of

heme and the *CUP1* UAS is bound by Cup2/Ace1 in response to increased copper concentrations (Guarente and Hoar, 1984; Thiele, 1988).

The first UAS identified, which has become one of the most well studied, is the *GAL1* UAS (Guarente et al., 1982). This UAS is a 108 bp region containing four 17 bp sequences that serve as binding sites for Gal4. The binding sites have a consensus sequence of 5'-cggrnnrcynyncnccg-3'; the four binding sites display different degrees of variation from the consensus and are not functionally equivalent (Lohr et al., 1995). At least two of these Gal4 binding sites are required for the promoter to function bi-directionally. The *GAL1* UAS is able to function on heterologous genes, causing them to be induced by galactose (Guarente et al., 1982). This trait has made the *GAL1* UAS a useful tool for many genetic studies. For example, the *GAL1* UAS is able to function as a promoter element in metazoans; this ability allowed for the development of the Gal4 UAS system in Drosophila, a system that has had a major impact in Drosophila studies (Brand and Perrimon, 1993). Interestingly, the *GAL1* UAS is able to function in metazoans when integrated downstream of the TATA box (Webster et al., 1988), suggesting yeast also has a mechanism for preventing activation from downstream UASs.

Upstream Promoter Elements

Upstream promoter elements are found in metazoans and share several features with yeast UASs. They occur 100 to 200 bp upstream of the core promoter. These elements are typically recognition sites for a group of sequence specific transcription

factors such as Sp1, CTF, and CBF (Blackwood and Kadonaga, 1998). They serve to increase the rate of transcription by promoting the formation of the pre-initiation complex. Upstream promoter elements may be necessary for some enhancers to act at a distance (Marsman and Horsfield, 2012). The natural human IFN-β enhancer is located immediately upstream of the core promoter. Placement of the IFN-β enhancer several kilobases upstream of a reporter gene does not allow activation unless an upstream promoter element is present upstream of the core promoter (Nolis et al., 2009). This activation is dependent on looping of the DNA between the enhancer and the upstream promoter element. Although this was an artificial system, it raises the question of what upstream promoter elements may tell us about long-distance activation.

Enhancers

Enhancers are promoter elements similar to UASs in that they are binding sites for activators and function to promote transcription efficiency. They were first identified as sequences that, in either orientation, increase transcription of reporter constructs(Banerji et al., 1981; Benoist and Chambon, 1981; Gillies et al., 1983; Lohr et al., 1995). Like UASs, enhancers can be activated in response to environmental stimuli, but they are also activated during developmental changes and are involved in mediating tissue specific expression (Dickmeis et al., 2004; Kim et al., 2010; Maniatis et al., 1987). Enhancers differ from UASs in that they are able to activate transcription at greater distances, discussed in further detail below. Additionally, enhancers are able to function

when positioned downstream or within a gene, in addition to functioning when upstream of the gene (Khoury and Gruss, 1983).

Extensive efforts have been made to identify and map enhancers genome-wide. Enhancers are typically in regions depleted of nucleosomes, and thus are hypersensitive to nucleases. DNaseI and formaldehyde-assisted isolation of regulatory elements have used this trait to identify a number of enhancer sequences (Zentner and Scacheri, 2012). However, when applied on a whole-genome scale, these techniques will also identify other regulatory elements such as the core promoter and insulators. The ENCODE project has helped to identify and map 400,00 enhancers to human cell lines using ChIP-seq data on chromatin modification and transcription factor binding data (Calo and Wysocka, 2013; Zentner and Scacheri, 2012). H3K4me1 is associated with active enhancers, while H3K27ac is associated with poised enhancers (Creyghton et al., 2010 ; Heintzman et al., 2009; Heintzman et al., 2007).

Another striking difference between yeast UASs and metazoan enhancers is the ability of enhancers to activate over long-distance. For example, the five DNase I hypersensitive sites of the human β globin locus are spaced over a 15 kb region and are able to activate over approximately 50kb (Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999). In some cases, enhancers act on genes as far as a megabase away (Lettice et al., 2003). The large distances over which enhancers can act raises the complication of determining the gene(s) on which each enhancer acts. Detection of long-range interactions may give some insight into this complication (Chepelev et al., 2012;

Sanyal et al., 2012). The ability to act over such distances also requires regulation to prevent aberrant activation from enhancers. Insulators serve as barriers to activation, thus helping to maintain the fidelity of activation (Krivega and Dean, 2012).

MECHANISMS OF LONG-DISTANCE ACTIVATION

While many types of experiments have established that enhancers can activate transcription over distances as far as 1 Mb, the intriguing question is how this activation occurs. There are three prevailing models (Fig. 1-1) for how long-distance activation is mediated: linking, scanning (also called tracking), and looping (Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999; Marsman and Horsfield, 2012). In recent years, looping has become the favored model and has been shown to occur at many enhancers (Lenhard et al., 2012; Marsman and Horsfield, 2012). However the majority of enhancers have not been examined. It remains unknown if the same mechanism of activation occurs for every enhancer or if long-distance activation is enhancer specific and context dependent.

Figure 1-1. Models for long-distance transcription activation. **A.** The linking model. A chain of proteins is formed along the DNA from the enhancer to the core promoter upon activation. These proteins transduce the activation signal from the enhancer to the promoter. Purple circles represent activators, green circles represent the proteins that bind and transduce signal, red oval represents RNA polymerase II, and teal ovals represent general transcription factors. **B.** The scanning model. RNA polymerase is recruited to the enhancer, from which point it scans along the DNA to the core promoter. Purple circles represent activators, red oval represents RNA polymerase II, and teal ovals represent general transcription factors. **C.** The looping model. The enhancer and core promoter are brought into close proximity by looping out the intervening DNA. Purple circles represent activators, red oval represents RNA polymerase II, and teal ovals represent general transcription factors.

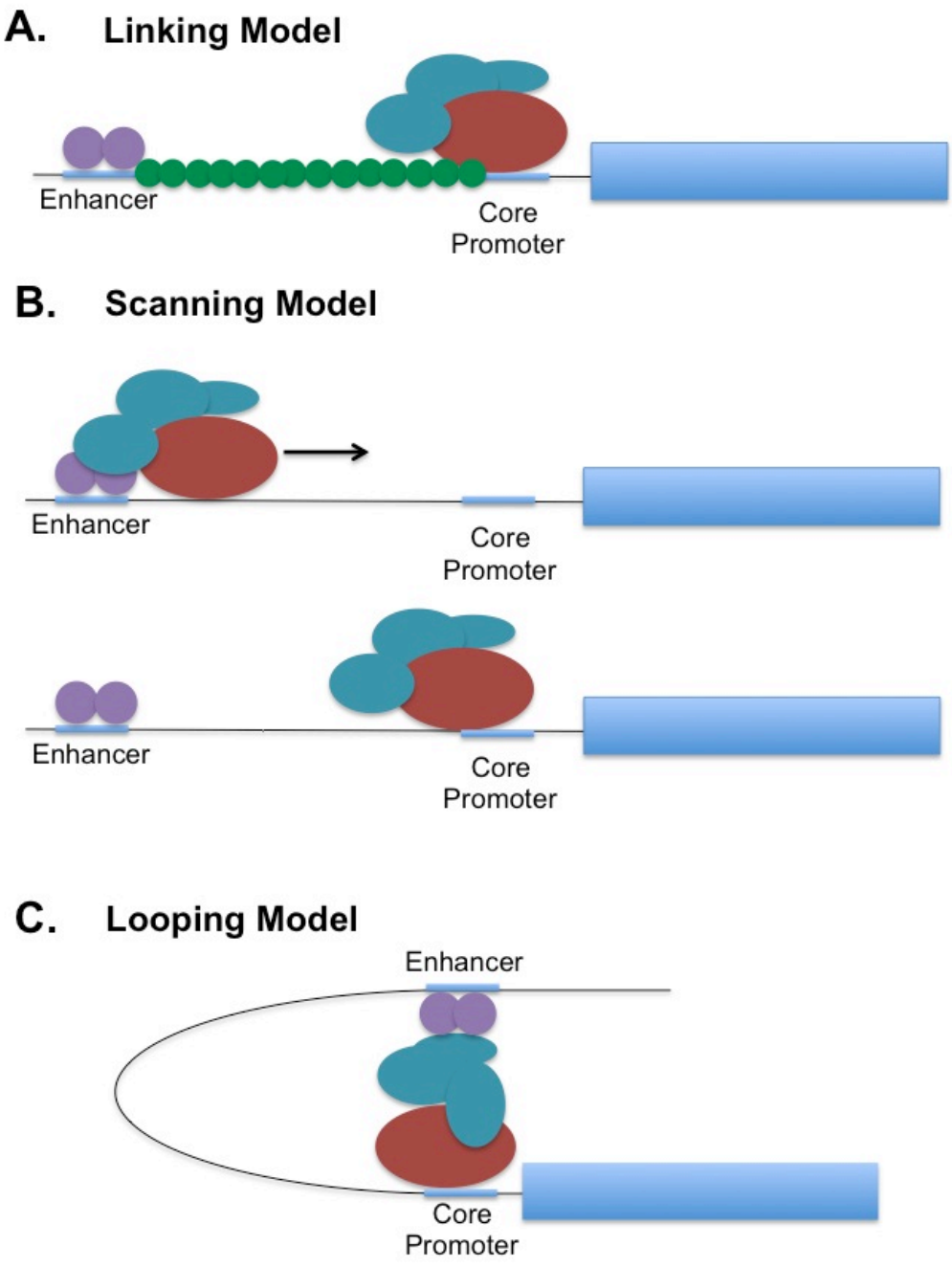


Figure 1-1. (Continued)

Linking

In the linking model of transcription activation, activation is mediated by the binding of proteins along the region between the enhancer and promoter. Upon activation the activator binds the enhancer. By this model, binding of the activator nucleates the formation of a chain of proteins along the chromatin to the transcription start site. These bound proteins serve to propagate the signal from the activators bound to the general transcription machinery at the core promoter. The core promoter would then act as a boundary to prevent further spreading of the signal (Bulger and Groudine, 1999). The linking model was proposed as a mechanism to explain activation at the human β-globin LCR; however there is no experimental evidence to support the linking model.

Scanning

The scanning model of activation involves linear tracking of the transcription machinery along the DNA from enhancer to core promoter. The strongest evidence for the scanning model come from experiments with the bacteriophage T4 late genes (Herendeen et al., 1992). Activation of the 40 T4 late genes requires three DNA polymerase accessory proteins. These proteins recognize an enhancer that can function to activate transcription thousands of bp away. This study used an *in vitro* transcription system to distinguish between activation by looping or scanning. In this system, binding sites for a mutant form of EcoRI, Eco RI Gln-111, were introduced on a plasmid between the enhancer and the promoter. This protein binds without cleaving the DNA. If tracking

is occurring, the binding of Eco RI Gln-111 will prevent activation by preventing the progression of the transcription machinery along the template. However, if looping is occurring, Eco RI Gln-111 binding should not impede the transcription machinery. The investigators found that binding of Eco RI Gln-111 does indeed prevent activation, thus supporting the scanning model.

Additional support for the scanning model comes from studies of UAS function in *S. cerevisiae*. Using a similar experimental approach to the Herendeen study, LexA binding sites were introduced between the *GAL1* UAS and promoter. Expression of LexA caused a six-fold loss of Gal4-dependent activation, suggesting that LexA acts to block progression of the transcription machinery (Brent and Ptashne, 1984). Insertion of a *CYC1* or *ADH1* terminator also caused reduced levels of activation, the *ADH1* terminator displayed the strongest effect with a 100-fold loss of activation, while the *CYC1* terminator displayed a 6-fold loss, similar to LexA binding.

Looping

The favored model for long-distance activation in metazoans is the looping mechanism, in which the enhancer is brought close to the core promoter via the formation of protein-mediated chromatin loops. These loops are detectable using chromosome conformation capture (3C) (Dekker et al., 2002) and have been observed for multiple loci in organisms such as Drosophila, mouse, and human cells. The first detection of long-distance interactions by 3C was of the murine β-globin LCR, where the DNase I

hypersensitive sites are located 40-60 kb from the genes activated (Tolhuis et al., 2002). These chromatin loops only form when the gene is active. Adaptations to 3C technology have allowed for increased identification of long-range enhancer promoter-interactions. One such modification identified 6,520 long-distance interactions between 2,067 putative enhancers and 1,619 target promoters in CDK^+ T cells (Chepelev et al., 2012). These results also suggested that 25% of promoters interact with two or more enhancers.

A limited number of studies have focused on the role of DNA looping mediating long-distance interactions in yeast, using artificial constructs. Formation of a DNA loop via LexA binding sites positioned 5' and 3' of *SNR6* allows activation of the gene by a downstream *GAL1* UAS (Petrascheck et al., 2005). Similarly, when linked to the telomere, *URA3* can be activated by a *GAL1* UAS 1.4 kb downstream, via the formation of telomere loops (de Bruin et al., 2001). Although these systems both show that in yeast long-distance activation can occur by looping, looping has not been shown to mediate transcription activation at any wild-type yeast promoters.

Several factors have been identified that are involved in the formation of chromatin loops in metazoans. Some of these factors are cell-type specific, such as the erythroid specific EKLF, required for loop formation at the β-globin locus (Drissen et al., 2004). Two general factors that are involved in loop formation are CCCTC-binding factor (CTCF) and the Cohesin complex. CTCF brings specific enhancers into contact with promoters, for example at the β-globin locus (Splinter et al., 2006). Cohesin, which has another known role in sister chromatid-cohesion during mitosis, also functions to

promote loop formation (Hadjur et al., 2009; Mishiro et al., 2009). CTCF and Cohesin cooperate to form loops at some loci, but also act independently of one another (Ross-Innes et al., 2011; Rubio et al., 2008). In embryonic stem cells, Mediator and Cohesin interact to promote formation of chromatin loops (Kagey et al., 2010). Transcription factors can also act to prevent loop formation, providing an additional layer of gene regulation. One example of this is OCT4, which prevents Cohesin binding at the HoxA locus (Kim et al., 2011).

The formation of chromatin loops to mediate long-distance interactions is well established, yet it is unknown how these loops form between the correct enhancer/gene pairings. By one model, the facilitated tracking model, activators bind to the enhancer and recruit the transcription machinery (Blackwood and Kadonaga, 1998). The transcription machinery then scans along the DNA until the promoter is found, at which point a stable DNA loop is formed. At the human β-globin LCR, the RNA polymerase II complex assembles at the HS2 enhancer and transcribes a series of short poly-adenylated RNAs across the 10kb between the enhancer and ε-globin gene (Zhu et al., 2007). 3C analysis demonstrates a loop forms between the enhancer and ε-globin, suggesting the loop may be formed after scanning has occurred. This study suggests some overlap between the scanning model and looping model of activation; the extent of this overlap remains to be seen. Wang et al. also observe looping and tracking at the human androgen receptor and propose a combined looping and tracking mechanism, in which polymerase continues to track along the DNA after the loop is formed (Wang et al., 2005).

ACTIVATION DISTANCE IN YEAST

Early studies of UASs revealed that the ability to activate is dependent on the distance between the UAS and the transcription start site (Guarente and Hoar, 1984; Struhl, 1984). Based on these results, a former graduate student in the Winston lab, Krista Dobi, constructed a series of reporters within a nonessential, long open reading frame, *BPH1*, that placed the *GAL1* UAS varying distances 5' of the TATA box of the *HIS3* gene. When the UAS was placed as far as 574 bp from the TATA box, activation occurred in inducing conditions (the presence of galactose), indicated by a His^+ phenotype. However, when the UAS was 799 bp from the TATA box, activation did not occur in galactose, as seen by a His^- phenotype (Dobi and Winston, 2007). These reporters were used to perform genetic studies on how activation distance is controlled in yeast.

Using several screening and selection methods, mutants were identified that activation of the reporter at the normally non-permissive distance of 799 bp. Mutants that allow transcriptional activation at a distance are referred to as having a long-distance activation (Lda^-) phenotype. Mutations in *SIN4*, *RGR1*, *SPT2*, *SPT10*, and *HTA1-HTB1* were demonstrated to cause an Lda^- phenotype. All of the factors identified have known roles in transcription regulation or chromatin structure. Of these mutations, *sin4Δ* caused the strongest Lda^- phenotype at 799 bp. Furthermore, weak activation was observed in a *sin4Δ* mutant at a distance as great as 1995 bp.

The Lda$^-$ phenotype of *sin4Δ* and *spt2Δ* mutants was tested by Northern analysis. Interestingly, these strains were shown to have two *HIS3* transcripts: a transcript that initiates adjacent to the UAS (long transcript), regardless of the position of the UAS, and a transcript the size of wild-type *HIS3* (short transcript). The 799 bp reporter strain in a wild-type produces only the long transcript, suggesting that it is nonfunctional for *HIS3* expression. The level of short transcript in *sin4Δ* and *spt2Δ* correlated with strength of Lda$^-$ phenotype. Using 5'-RACE, the transcripts in both the wild-type and *sin4Δ* strains were mapped. Both strains contained transcripts initiating in the spacer region between the UAS and *HIS3* open reading frame. Only the *sin4Δ* strain shows a *HIS3* transcript that initiates downstream of the TATA box.

Attempts were made to determine the mechanism by which long-distance activation occurs in a *sin4Δ* mutant. The two proposed mechanisms of long-distance activation are scanning and looping. To test the scanning model of activation, an *ADH1* terminator was inserted into the reporter region, between the UAS and the *HIS3* ORF. If activation is occurring in the *sin4Δ* mutant through scanning of the region between the UAS and *HIS3*, the *sin4Δ ADH1* terminator strain should not display activation of the reporter. The *sin4Δ ADH1* terminator strain displays the Lda$^-$ phenotype, which suggests long-distance activation is not occurring by scanning. To test the looping model of activation, 3C experiments were performed on the reporter for wild-type and *sin4Δ* strains. The results of the 3C experiments were inconclusive. Together, these experiments suggest long-distance activation of the reporter does not occur by the scanning mechanism, but there is no conclusive evidence for looping.

In an effort to identify additional regulators of activation distance, a mutation was identified that enhanced the long-distance activation phenotype of the *sin4Δ* mutant (Leeman and Winston, unpublished). The mutation was identified as resistant to 1 mM 3-aminotriazole (3-AT). 3-AT is a competitive inhibitor of the *HIS3* gene product; increases in *HIS3* expression correspond to increased 3-AT resistance. The enhancer mutation and *sin4Δ* together result in higher *HIS3* levels and therefore the double mutant has a stronger Lda$^-$ phenotype than a *sin4Δ* mutant alone. Interestingly, the enhancer single mutant does not have an Lda$^-$ phenotype. The isolation of the enhancer mutation demonstrated the existence of other factors that work with Sin4 to regulate activation distance.

These experiments resulted in the identification of mutations in *SIN4*, *RGR1*, *HTA1-HTB1*, *SPT2*, and *SPT10* that allowed long-distance activation of a reporter. The mechanism by which these changes to these factors allow long-distance remains unknown. Isolation of the enhancer suggests the existence of other factors that are involved in regulating activation distance. Additionally, the stronger Lda$^-$ phenotype of the enhancer *sin4Δ* double mutant suggests that only studying single mutants limits the strength of phenotype of the mutants isolated. To isolate additional mutants with stronger long-distance activation phenotypes, we decided to study transcription activation distance by isolating polygenic mutants.

COMPLEX TRAITS

Classical genetic studies in model organisms usually focus on single mutations in order to facilitate gene identification and subsequent studies. However many phenotypes in nature are the result of the combined effects of mutations in many genes (Glazier et al., 2002; McCarthy et al., 2008; Womack et al., 2012). A phenotype caused by multiple mutations is referred to as a complex trait. Many human diseases that have a genetic basis, including Type 2 diabetes, schizophrenia, and hypertension, are complex. Genome-wide association studies (GWAS) are focused on determining the causes of complex diseases (Glazier et al., 2002; McCarthy et al., 2008). There are two major challenges faced in performing GWAS analysis. The first is generating large enough data sets to identify all loci involved. The second challenge is, once a region has been identified as being associated with a trait, determining the causative single nucleotide polymorphisms (SNPs) within the region (McCarthy et al., 2008). These challenges are far less daunting in yeast. First, the short generation time of yeast allows for the generation of large numbers of progeny, and therefore large data sets. Second, the genetic tools available in yeast allow alleles to be tested for causality. These characteristics have made yeast an ideal organism for understanding the mechanisms behind natural genetic variation.

S. cerevisiae strains found in nature display a broad range of phenotypic variance, much of which is complex (Liti and Louis, 2012; Swinnen et al., 2012). These phenotypes that vary in degree are quantitative traits; regions of DNA that are linked to

the genes causing the phenotype are quantitative trait loci (QTLs). Studies of QTLs in yeast have been influential in development of strategies to study quantitative traits. They have established techniques for identification of QTLs, for mapping of the SNPs within the QTL, and for establishing SNPs as causative (Liti and Louis, 2012; Swinnen et al., 2012).

A significant amount of knowledge of quantitative traits has come from the generation of hybrid strains and the analysis of progeny; two landmark studies in yeast quantitative trait genetics were performed using this strategy. In 2002, the Kruglyak lab performed microarrays on segregants derived from a cross of the distantly related BY and RM *S. cerevisiae* strains. Measuring global changes in expression between segregants allowed for analysis of quantitative traits throughout the genome (Brem et al., 2002). Analysis of the microarrays revealed a subset of transcripts that show a much wider range of expression changes than observed for either parent. Computational analysis of the expression changes suggests that at least 50% of transcripts are affected by at least five QTLs (Brem and Kruglyak, 2005). A concurrent study analyzed the segregants derived from a cross of the YJM145 and S288c strains. This study looked specifically at QTLs associated with high temperature growth. Sequence analysis of the QTLs led to the identification of three SNPs associated with high temperature growth (Steinmetz et al., 2002).

Analysis of hybrid strains has since led to the identification of QTLs, and in some cases causative SNPs, for a number of yeast quantitative traits. Several studies have

identified SNPs associated with sporulation efficiency by the analysis of hybrids of high sporulating and low sporulating strains (Ben-Ari et al., 2006; Deutschbauer and Davis, 2005; Gerke et al., 2006). Similar analyses have been performed to study QTLs between more diverged *S. cerevisiae* strains (Cubillos et al., 2011) as well as between *S. cerevisiae* and *S. paradoxus*, a closely related *Saccharomyces* species, (Liti et al., 2009).

The study of quantitative traits has also been useful in understanding aspects of yeast biology. In some studies, in-lab evolution experiments have been performed to select for a desired trait. The evolved strains can then be used to identify the causative mutations conferring the selected trait. Romano *et al.* used in-lab evolution to select for mutants able to grow on media with varying pH. These strains were used to identify mutations that conferred ability to grow at high pH (Romano et al., 2010). Mutations were established as causative through allele replacement or reciprocal hemizygosity, a method that tests the affect of each allele in the hybrid diploid background. Recently, a yeast in-lab evolution experiment was performed to study the evolution of multi-cellularity in unicellular organisms (Koschwanez et al., 2013). Interestingly, the evolved strains had accumulated different sets of mutations.

The study of a complex trait in yeast requires a method for determining the mutations that cause the phenotype of interest. The mutation rate of yeast is approximately $5x10^{-10}$ mutations per base pair per generation (Lang and Murray, 2008) which means culturing of yeast strains results in the accumulation of mutations. Simply sequencing a single yeast strain would not distinguish between mutations accumulated

through passaging of the strain and mutations causing the phenotype of interest. A technique that has been used successfully to determine causative mutations is bulk segregant analysis (Brauer et al., 2006; Ehrenreich et al., 2009). By this technique, a large numbers of mutant and wild-type isolates, generated through a back-cross, are analyzed. This analysis allows for distinguishing between causative and non-causative mutations. Candidate causative mutations can then be verified by allele replacement or reciprocal hemizygosity analysis.

OVERVIEW OF DISSERTATION

Previous studies have shown that transcriptional activation distance is regulated in yeast (Dobi and Winston, 2007; Guarente and Hoar, 1984; Struhl, 1984). The most systematic of these studies(Dobi and Winston, 2007)identified several factors that constrain activation distance. However, additional results (Leeman and Winston, unpublished) suggested that additional factors contribute to this regulation and would only be identified by the isolation of complex, or polygenic mutants. That is the focus of this dissertation.

In Chapter 2, I describe the isolation of polygenic mutants that allow strong transcriptional activation over a distance that is normally too great for any activation. To isolate the mutants, we constructed a reporter system adapted from Dobi and Winston, (2007). In the revised system, we constructed two reporters on different chromosomes, both with the *GAL1* UAS approximately 800 bp upstream of the TATA box. At one, the

reporter gene is *HIS3* and at the other the reporter gene is *URA3*. Beginning with strains with both reporters and a *sin4Δ* mutation, new mutations were selected using 3-aminotriazole (3-AT), a competitive inhibitor of the *HIS3* gene product. We then screened all mutants for increased *HIS3* and *URA3* expression by Northern analysis. By iteratively selecting for resistance to increased levels of 3-AT, we successfully isolated three lineages of strains, all showing a greatly increased Lda$^-$ phenotype compared to the *sin4Δ* parent. Two of the final strains are genetically related and the third strain is unrelated. Then by bulk segregant analysis followed by allele replacement tests, several mutations were demonstrated to be causative. Interestingly, all three strains differ in their causative mutations and the unrelated lineages contain completely different sets of mutations, with the exception of the starting mutation *sin4Δ*. Reconstruction experiments for each strain can account for some (in two cases) or all (in one case) of the Lda$^-$ phenotypes.

In Chapter 3, I describe additional analysis of the mutations isolated and identified in Chapter 2. We first constructed combinations of mutations and found mutations isolated in different lineages in combination could give increased Lda$^-$ phenotypes to varying degrees. Then, we constructed polygenic mutants strains with reporter distances of 1397 bp and 2027 bp. We found that the polygenic mutants are able to activate transcription of the longer distance reporters more strongly than the *sin4Δ* single mutant. Next, we performed microarray analysis on a wild-type strain, our three original mutant isolates, and two reconstructed strains to investigate the overall transcriptional changes in the mutants. We saw a significant overlap in genes with a two-

fold expression change between all three original mutant isolates, as well as a significant overlap between the genes changed in the original mutant and the reconstructed strain. We then looked for potential regions of long-distance activation by analyzing tandem adjacent gene pairs showing an increased expression at both genes, all strains analyzed displayed a greater number of upregulated pairs than expected by chance. Finally, we identified a site of long-distance activation outside of the reporter in the mutant strains.

Taken together, the results presented in this study provide an increased understanding of activation distance regulation in yeast. By taking an alternate genetic approach, we were able to identify additional factors involved in the regulation of activation distance. These mutants show an altered transcriptional profile and may allow long-distance activation at non-reporter regions of the genome. Although the mechanism by which long-distance activation occurs remains elusive, we are still pursuing the idea that chromatin loops are formed between the UAS and transcription start site in the mutants.

References

Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., and Kornberg, R.D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. Science *283*, 985-987.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699-709.

Ben-Ari, G., Zenvirth, D., Sherman, A., David, L., Klutstein, M., Lavi, U., Hillel, J., and Simchen, G. (2006). Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genet *2*, e195.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Bhaumik, S.R., Raha, T., Aiello, D.P., and Green, M.R. (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev *18*, 333-343.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Borggrefe, T., Davis, R., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (2002). A complex of the Srb8, -9, -10, and -11 transcriptional regulatory proteins from yeast. J Biol Chem *277*, 44202-44207.

Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. Science *317*, 815-819.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Brauer, M.J., Christianson, C.M., Pai, D.A., and Dunham, M.J. (2006). Mapping novel traits by array-assisted bulk segregant analysis in Saccharomyces cerevisiae. Genetics *173*, 1813-1816.

Brem, R.B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A *102*, 1572-1577.

Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. Science *296*, 752-755.

Brent, R., and Ptashne, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. Nature *312*, 612-615.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729-736.

Bryant, G.O., and Ptashne, M. (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. Mol Cell *11*, 1301-1309.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev *13*, 2465-2477.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res *22*, 490-503.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Covitz, P.A., Song, W., and Mitchell, A.P. (1994). Requirement for RGR1 and SIN4 in RME1-dependent repression in Saccharomyces cerevisiae. Genetics *138*, 577-586.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Cubillos, F.A., Billi, E., Zorgo, E., Parts, L., Fargier, P., Omholt, S., Blomberg, A., Warringer, J., Louis, E.J., and Liti, G. (2011). Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol Ecol *20*, 1401-1413.

de Bruin, D., Zaman, Z., Liberatore, R.A., and Ptashne, M. (2001). Telomere looping permits gene activation by a downstream UAS in yeast. Nature *409*, 109-113.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Deutschbauer, A.M., and Davis, R.W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet *37*, 1333-1340.

Dickmeis, T., Plessy, C., Rastegar, S., Aanstad, P., Herwig, R., Chalmel, F., Fischer, N., and Strahle, U. (2004). Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo. Genome Res *14*, 228-238.

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol Cell Biol *27*, 5575-5586.

Dotson, M.R., Yuan, C.X., Roeder, R.G., Myers, L.C., Gustafsson, C.M., Jiang, Y.W., Li, Y., Kornberg, R.D., and Asturias, F.J. (2000). Structural organization of yeast and mammalian mediator complexes. Proc Natl Acad Sci U S A *97*, 14307-14310.

Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev *18*, 2485-2490.

Ehrenreich, I.M., Gerke, J.P., and Kruglyak, L. (2009). Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BYxRM cross. Cold Spring Harb Symp Quant Biol *74*, 145-153.

Gerke, J.P., Chen, C.T., and Cohen, B.A. (2006). Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. Genetics *174*, 985-997.

Gill, G., and Tjian, R. (1992). Eukaryotic coactivators associated with the TATA box binding protein. Curr Opin Genet Dev *2*, 236-242.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell *33*, 717-728.

Glazier, A.M., Nadeau, J.H., and Aitman, T.J. (2002). Finding genes that underlie complex traits. Science *298*, 2345-2349.

Govind, C.K., Yoon, S., Qiu, H., Govind, S., and Hinnebusch, A.G. (2005). Simultaneous recruitment of coactivators by Gcn4p stimulates multiple steps of transcription in vivo. Mol Cell Biol *25*, 5626-5638.

Guarente, L., and Hoar, E. (1984). Upstream activation sites of the CYC1 gene of Saccharomyces cerevisiae are active when inverted but not when placed downstream of the "TATA box". Proc Natl Acad Sci U S A *81*, 7860-7864.

Guarente, L., Yocum, R.R., and Gifford, P. (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc Natl Acad Sci U S A *79*, 7410-7414.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al*. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al*. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., and Hahn, S. (2010). Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Mol Cell Biol *30*, 2376-2390.

Herendeen, D.R., Kassavetis, G.A., and Geiduschek, E.P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. Science *256*, 1298-1303.

Hu, M., Yu, J., Taylor, J.M., Chinnaiyan, A.M., and Qin, Z.S. (2010). On the detection and refinement of transcription factor binding sites using ChIP-Seq data. Nucleic Acids Res *38*, 2154-2167.

Jackson, B.M., Drysdale, C.M., Natarajan, K., and Hinnebusch, A.G. (1996). Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. Mol Cell Biol *16*, 5557-5571.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol *14*, 116-127.

Jiang, Y.W., Dohrmann, P.R., and Stillman, D.J. (1995). Genetic and physical interactions between yeast RGR1 and SIN4 in chromatin organization and transcriptional regulation. Genetics *140*, 47-54.

Jiang, Y.W., and Stillman, D.J. (1992). Involvement of the SIN4 global transcriptional regulator in the chromatin structure of Saccharomyces cerevisiae. Mol Cell Biol *12*, 4503-4514.

Jiang, Y.W., and Stillman, D.J. (1995). Regulation of HIS4 expression by the Saccharomyces cerevisiae SIN4 transcriptional regulator. Genetics *140*, 103-114.

Jonker, H.R., Wechselberger, R.W., Boelens, R., Folkers, G.E., and Kaptein, R. (2005). Structural properties of the promiscuous VP16 activation domain. Biochemistry *44*, 827-839.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al*. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Khoury, G., and Gruss, P. (1983). Enhancer elements. Cell *33*, 313-314.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al*. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599-608.

Kim, Y.J., Cecchini, K.R., and Kim, T.H. (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. Proc Natl Acad Sci U S A *108*, 7391-7396.

Koleske, A.J., Buratowski, S., Nonet, M., and Young, R.A. (1992). A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. Cell *69*, 883-894.

Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. Nature *368*, 466-469.

Koschwanez, J.H., Foster, K.R., and Murray, A.W. (2013). Improved use of a public good selects for the evolution of undifferentiated multicellularity. Elife *2*, e00367.

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions-long-distance calls. Curr Opin Genet Dev *22*, 79-85.

Lang, G.I., and Murray, A.W. (2008). Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics *178*, 67-82.

Larschan, E., and Winston, F. (2005). The Saccharomyces cerevisiae Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription. Mol Cell Biol *25*, 114-123.

Lemieux, K., and Gaudreau, L. (2004). Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II. EMBO J *23*, 4040-4050.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13*, 233-245.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet *12*, 1725-1735.

Li, Y., Flanagan, P.M., Tschochner, H., and Kornberg, R.D. (1994). RNA polymerase II initiation factor interactions and transcription start site selection. Science *263*, 805-807.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V.*, et al*. (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337-341.

Liti, G., and Louis, E.J. (2012). Advances in quantitative trait analysis in yeast. PLoS Genet *8*, e1002912.

Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J *9*, 777-787.

Malik, S., Baek, H.J., Wu, W., and Roeder, R.G. (2005). Structural and functional characterization of PC2 and RNA polymerase II-associated subpopulations of metazoan Mediator. Mol Cell Biol *25*, 2117-2129.

Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science *236*, 1237-1245.

Marsman, J., and Horsfield, J.A. (2012). Long-distance relationships: enhancer-promoter communication and dynamic gene transcription. Biochim Biophys Acta *1819*, 1217-1227.

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem *255*, 11992-11996.

McBride, H.J., Brazas, R.M., Yu, Y., Nasmyth, K., and Stillman, D.J. (1997). Long-range interactions at the HO promoter. Mol Cell Biol *17*, 2669-2678.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet *9*, 356-369.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J *28*, 1234-1245.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Narlikar, L., Gordan, R., and Hartemink, A.J. (2007). A nucleosome-guided map of transcription factor binding sites in yeast. PLoS Comput Biol *3*, e215.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. Proc Natl Acad Sci U S A *106*, 20222-20227.

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W., and Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res *33*, 3743-3750.

Qiu, H., Hu, C., Zhang, F., Hwang, G.J., Swanson, M.J., Boonchird, C., and Hinnebusch, A.G. (2005). Interdependent recruitment of SAGA and Srb mediator by transcriptional activator Gcn4p. Mol Cell Biol *25*, 3461-3474.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Romano, G.H., Gurvich, Y., Lavi, O., Ulitsky, I., Shamir, R., and Kupiec, M. (2010). Different sets of QTLs influence fitness variation in yeast. Mol Syst Biol *6*, 346.
Ross-Innes, C.S., Brown, G.D., and Carroll, J.S. (2011). A co-ordinated interaction between CTCF and ER in breast cancer cells. BMC Genomics *12*, 593.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A *105*, 8309-8314.

Sakai, A., Shimizu, Y., and Hishinuma, F. (1988). Isolation and characterization of mutants which show an oversecretion phenotype in Saccharomyces cerevisiae. Genetics *119*, 499-506.

Samuels, M., Fire, A., and Sharp, P.A. (1982). Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J Biol Chem *257*, 14419-14427.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Curr Opin Cell Biol *21*, 344-351.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S., *et al*. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. Cell *106*, 697-708.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., and Davis, R.W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. Nature *416*, 326-330.

Struhl, K. (1984). Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. Proc Natl Acad Sci U S A *81*, 7865-7869.

Swinnen, S., Thevelein, J.M., and Nevoigt, E. (2012). Genetic mapping of quantitative phenotypic traits in Saccharomyces cerevisiae. FEMS Yeast Res *12*, 215-227.

Tanaka, M., and Herr, W. (1990). Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell *60*, 375-386.

Thiele, D.J. (1988). ACE1 regulates expression of the Saccharomyces cerevisiae metallothionein gene. Mol Cell Biol *8*, 2745-2752.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell *73*, 1361-1375.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell *19*, 631-642.

Webster, N., Jin, J.R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. Cell *52*, 169-178.

Williams, T., and Tjian, R. (1991). Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev *5*, 670-682.

Womack, J.E., Jang, H.J., and Lee, M.O. (2012). Genomics of complex traits. Ann N Y Acad Sci *1271*, 33-36.

Xue, W., Wang, J., Shen, Z., and Zhu, H. (2004). Enrichment of transcriptional regulatory sites in non-coding genomic region. Bioinformatics *20*, 569-575.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52-65.

Zarrinpar, A., Bhattacharyya, R.P., and Lim, W.A. (2003). The structure and function of proline recognition domains. Sci STKE *2003*, RE8.

Zentner, G.E., and Scacheri, P.C. (2012). The chromatin fingerprint of gene enhancer elements. J Biol Chem *287*, 30888-30896.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res *33*, 2838-2851.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res *35*, 5532-5544.

Chapter 2

Identification of mutations that contribute to long-distance activation in *Saccharomyces cerevisiae*

Attribution of experiments in Chapter 2

Mutant isolation, generation of segregants, and isolation of DNA were performed by Caitlin Reavey. Library preparation and sequencing were performed at the Lewis-Sigler Institute for Integrative Genomics at Princeton University. Mapping of sequence reads to the reference genome, SNP identification, and normalization of sequencing reads were performed by Dr. Mark Hickman. Identification of causative alleles, mutant reconstructions, deletion analysis, stability tests, and transformant tests of reporter duplication were performed by Caitlin Reavey

Abstract

Most fundamental aspects of transcription initiation are conserved from yeast to humans; however a striking difference between transcription in yeast and metazoans is the distance over which transcription activation occurs. Relocation of the UAS too far 5' of the transcription start site results in loss of transcription, indicating that in yeast activation distance is regulated. Mutations in *SIN4, RGR1, HTA1-HTB1, SPT2,* and *SPT10* have been identified as allowing long-distance activation in yeast, although this activation was weak. We have studied long-distance activation as a polygenic trait, isolating three strains with multiple mutations that together confer a strong phenotype. To do this, we constructed strains containing two reporters, *HIS3* and *URA3*. For each reporter, the *GAL1* UAS was placed approximately 800 bp upstream of the transcription start sites. We have identified mutations in *MOT3, GRR1, MIT1, PTR3, YOR019W*, and *MSN2* that contribute to the long-distance activation phenotype. Additionally, we isolated two strains disomic for chromosome III, the site of the reporter containing *HIS3*.

Introduction

Among eukaryotes, much of the transcriptional machinery is conserved; however, the distance over which transcriptional regulation occurs differs between yeasts and metazoans. In *S. cerevisiae*, where the genome is compact, the UAS is generally found within 600 bp 5' of the transcription start site (Xue et al., 2004). In contrast, metazoan enhancers can be spaced over a range from less than 1 kb to as far as 1 Mb from the start site and can be found upstream or downstream of the start site. Previous studies on transcription activation distance in yeast revealed that relocation of a UAS too far 5' of the transcription start reduces or abolishes initiation (Dobi and Winston, 2007; Guarente and Hoar, 1984; Struhl, 1984). This suggests yeast has some mechanism for regulating the distance over which a UAS can activate transcription. Without regulation of activation distance, UASs could aberrantly activate neighboring genes.

A previous study took a genetic approach to identify factors that control activation distance in yeast (Dobi and Winston, 2007). In this study, a series of reporters were constructed within a nonessential, long open reading frame, *BPH1*, that placed the *GAL1* UAS varying distances 5' of the TATA box of the *HIS3* gene. When the UAS was placed as far as 574 bp from the TATA box, activation occurred under inducing conditions (the presence of galactose), indicated by a His^+ phenotype. However, when the UAS was placed 799 bp from the TATA box, activation did not occur in galactose, as seen by a His^- phenotype. Using a strain with this reporter, mutations were identified in *HTA1-HTB1*, *SPT2*, *SPT10*, *RGR1,* and *SIN4* that allowed long-distance activation, albeit

only at a modest level (Dobi and Winston, 2007 and Dobi and Leeman, unpublished). Mutations that allow activation of the reporter are designated as having the long-distance activation (Lda$^-$) phenotype.

Attempts to isolate stronger mutants resulted in the isolation of strains containing more than one mutation contributing to the phenotype. First, a mutation was identified that enhances the Lda$^-$ phenotype of a *sin4Δ* mutation (Leeman and Winston, unpublished). Second, in an attempt to identify additional factors involved in the regulation of activation distance, mutations were selected that allow activation of the *CUP1* UAS over a normally non-permissive distance. This selection yielded multiple strains containing two mutations that both contribute to the Lda$^-$ phenotype. These results suggest that additional factors involved in regulating activation distance may be identified by studying strains containing multiple mutations that contribute to the Lda$^-$ phenotype.

S. cerevisiae has been established as an ideal organism to study complex traits. The short generation time allows for the isolation of large numbers of progeny for analysis. This, combined with the genetic tractability of yeast, makes it feasible to identify the causative alleles for a given phenotype. Complex traits that occur as natural variants have been studied in *S. cerevisiae,* including meiosis, oxidative stress response, and high temperature growth (Ben-Ari et al., 2006; Deutschbauer and Davis, 2005; Gerke et al., 2006; Sinha et al., 2006; Witten et al., 2007). Experimental evolution experiments have successfully produced polygenic mutants by selecting for a specific trait. In one

such study, they selected for strains resistant to varying pH levels and were able to identify genes involved in alkali stress (Romano et al., 2010). In a more recent study, ten strains were selected that were able to grow at low levels of sucrose (Koschwanez et al., 2013). Interestingly, some of the identified causative mutations appeared in more than one of the ten evolved strains.

The research presented in this chapter focuses on the isolation of three polygenic mutants that each show an increased Lda$^-$ phenotype compared to the strongest single mutants. We identified candidate causative mutations using bulk segregant analysis and confirmed mutations as causative by allele replacement. Our results show that the causative mutations differ between unrelated lineages. For one of our strains, we have been able to reconstruct the phenotype from the identified causative mutations. For the other two, the missing heritability not found by bulk segregant analysis is likely explained by a duplication of the reporter. While the majority of the mutations identified contribute to the Lda$^-$ phenotype through loss of function, two are caused by altered function. Taken together, this study has identified additional factors that are involved in the regulation of transcriptional activation distance and demonstrated the feasibility of polygenic mutant analysis to understand a fundamental aspect of gene expression.

Materials and Methods

S. cerevisiae strains

The *S. cerevisiae* strains used in this study (Table 2-1, 2-2) are isogenic with a *GAL2*$^{+}$ derivative of S288C (Winston et al., 1995). In the course of the study, we isolated a number of strains by mutagenesis. The two strain tables distinguish between strains that did not go through the selection process (Table 2-1) and the mutant strains and their derivatives (Table 2-2). Rich (YPD) and synthetic complete (SC) dropout media were prepared as previously described (Rose et al., 1990). SC Gal and SC-His Gal media contained 2% galactose as the carbon source. YP Raffinose contained 2% raffinose as the carbon source. Specified concentrations of 3-aminotriazole were added to SC-His Gal medium. Strains were constructed by standard methods, either through crosses or transformation (Ausubel et al., 1991).

The strains with dual reporters that were used to select mutants allowing long-distance activation were derived from previously described strains (Dobi and Winston, 2007) by standard procedures. The dual reporter strains contained the *bph1::kanMX-UAS*$_{GAL1}$*799-HIS3* and *ybr281c::TRP1-UAS*$_{GAL1}$*806-URA3* reporters and *sin4Δ0::LEU2*. To construct the *URA3* reporter, the *HIS3* open reading frame (ORF) was replaced with the *URA3* ORF using the oligos in Table 2-3. Both reporters contain the *HIS3* TATA element, transcription start sites, and transcription termination sequence. These reporter strains were used in the selection of mutants.

Allele replacements, either to correct a mutant allele to wild-type or the reverse, were done by a two-step transformations method, using strains in which the *URA3* ORF was deleted from the *ybr281c* reporter. Allele replacements to correct mutant alleles to wild-type were made in strains CR101, CR109, and CR110. Mutant reconstructions to replace wild-type alleles with mutations were done in *sin4 Δ 0::LEU2* strains. The first step in each case was the integration of *URA3* at the relevant site, using the oligos listed in Table 2-3. DNA containing *URA3* were made by PCR using pRS406 as template DNA. For replacement of *URA3* by wild-type alleles, PCR fragments approximately 500 bp in length were synthesized using DNA from strain CR70, the wild-type parental strain, as template. PCR fragments were used to transform the *URA3* transformants to 5-FOA resistance. For replacement of *URA3* by mutant alleles, PCR fragments approximately 500 bp in length were amplified off CR98 (*mot3* and *sgm1*), CR91 (*mit1, ptr3,* and *yor019w*), or CR92 (*msn2*) genomic DNA as template. CR194, the *grr1 sin4Δ0* reconstruction strain was made by transformation of a 100 bp PCR fragment made with overlapping oligos. Correct alleles were all verified by Sanger sequencing.

Construction of complete deletions of the *MOT3*, *GRR1*, *MIT1*, *PTR3*, and *MSN2* ORFs were constructed by replacement with *URA3*, using pRS406 as template DNA to generate PCR fragments with *URA3*. CR101 was transformed with *mot3Δ100::URA3* and *grr1Δ0::URA3*. CR109 was transformed with *mit1Δ0::URA3* and *ptr3Δ0::URA3*. CR110 was transformed with *msn2Δ0::URA3*. All strains were verified by PCR.

Table 2-1. Strains used in this study not subjected to selection

Strain	Lineage Alias	Genotype
FY76		*MAT***a** *lys2-128 δ*
CR70	2.0	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mkc7-1307*
CR71	1.0	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806-URA3-natMX*
CR74		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX*
CR111		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mkc7-1307*
CR113		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806*
CR169		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 mot3- 1162*
CR194		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 grr1-531*
CR204		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531*
CR214		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806*
CR217		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 mit1-560 ptr3-1088 mkc7-1307 yor019w-1659*
CR218		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 msn2-1956 ptr3-1088 mkc7-1307 yor019w-1659*
CR219		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531*

Table 2-2. Selected strains and their derivatives used in this study

Strain	Lineage Alias	Genotype
CR76	2.1	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mkc7-1307 yor019w-1659*
CR79	1.1	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806-URA3-natMX grr1-531*
CR82	2.2	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX ptr3-1088 mkc7-1307 yor019w-1659*
CR91	2.3a	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mit1-560 ptr3-1088 mkc7-1307 yor019w-1659*
CR92	2.3b	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mkc7-1307*
CR98	1.2	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806-URA3-natMX mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251*
CR101	1.2	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251*
CR109	2.3a	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 mit1-560 ptr3-1088 mkc7-1307 yor019w-1659*
CR110	2.3b	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 msn2-1956 ptr3-1088 mkc7-1307 yor019w-1659*
CR171		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mit1-560 PTR3 mkc7-1307 yor019w-1659*
CR198		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806-URA3-natMX mit1-560 ptr3-1088 mkc7-1307 YOR019W*

Table 2-2. (Continued)

CR200	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 GRR1 sgm1-1220 rim8-155 tma108-1374*
CR201	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 MOT3 grr1-531 sgm1-1220 rim8-155 tma108-1374*
CR202	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 SGM1 rim8-155 tma108-1374*
CR221	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 MIT1 ptr3-1088 mkc7-1307yor019w-1659*
CR224	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 MSN2 ptr3-1088 mkc7-1307 yor019w-1659*
CR225	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3Δ::URA3 grr1-531 SGM1 rim8-155 tma108-1374*
CR227	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1Δ::URA3 SGM1 rim8-155 tma108-1374*
CR230	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mit1Δ::URA3 ptr3-1088 mkc7-1307yor019w-1659*
CR231	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mit1-560 ptr3Δ::URA3 mkc7-1307 yor019w-1659*
CR233	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 msn2Δ::URA3 ptr3-1088 mkc7-1307yor019w-1659*

Table 2-3. Oligonucleotides used in this study

Oligo	Purpose	Sequence (5'-3')
CRO168	Replace *HIS3* ORF with *URA3* ORF in *YBR281c* reporter	TATACTAAAAAATGAGCAGGCAAGATAAAC GAAGGCAAAGATGTCGAAAGCTACATATAA
CRO169	Replace *HIS3* ORF with *URA3* ORF at *YBR281c* reporter	TATATATATCGTATGCTGCAGCTTTAAATAA TCGGTGTCATTAGTTTTGCTGGCCGCATC
FO201	Northern probe of *HIS3*	TGAGCAGGCAAGATAAAC
FO609	Northern probe of *HIS3*	GCCTCATCCAAAGGCGC
FO481	Northern probe of *URA3*	CCATGGAGGGCACAGTTAAGCCGC
F0483	Northern probe of *URA3*	CCCTTCCCTTTGCAAATAGTCCTC
FO1324	Northern probe of *SNR190*	GGCCCTGATGATAATG
FO1325	Northern probe of *SNR190*	GGCTCAGATCTGCATG
FO3662	Delete *URA3* from *ybr281c* reporter	GTTGGTTGGGTGACCCAACAAATCAT
FO3663	Delete *URA3* from *ybr281c* reporter	CTATTGAATACTTTAGACAAAATCTCA
CRO388	Integrate *URA3* at site of *mot3* mutation	TCGCAACAAAGACATTTTCTATGCCCTTGGT GCTTAAGCAAGATTGTACTGAGAGTGCAC
CRO389	Integrate *URA3* at site of *mot3* mutation	AATTGTTGTAGTTAAAGATGATGTTGTTTTTC TTGAGTTCCTGTGCGGTATTTCACACCG
CRO232	Replace *URA3* with *mot3* allele	ACGACAGCACCTAACCATCC
CRO233	Replace *URA3* with *mot3* allele	TCTTCATTTTCGGGAGCTGT

Table 2-3. (Continued)

CRO226	Integrate *URA3* at site of *grr1* mutation	AAAAAAATCCAAGAGTTTCTGGTTGTTATAG AGAAACGTAAGATTGTACTGAGAGTGCAC
CRO227	Integrate *URA3* at site of *grr1* mutation	TTTAAGGTTGTCTAGCTCAATTTCGTTCAGTA TTTTTTTTCTGTGCGGTATTTCACACCG
CRO228	Replace *URA3* with WT *GRR1* allele	CGCTGAACGGGATTGACATA
CRO229	Replace *URA3* with WT *GRR1* allele	GCTTGGAATGGCAGTATGCA
CRO300	Replace *URA3* with *grr1* mutation	AATAGTGGGCAAAAAAATCCAAGAGTTTCT GGTTGTTATAGAGAAACGTAAAAAAAAATA
CRO301	Replace URA3 with *grr1* mutation	CTCCTTTAGTTTAAGGTTGTCTAGCTCAATTT CGTTCAGTATTTTTTTTTACGTTTCTCT
CRO220	Replace *URA3* with *mit1* allele	GTACTGATTCCGCCGTCATT
CRO221	Replace *URA3* with *mit1* allele	TCAGGGGAGTGGAAGAGTTG
CRO386	Integrate *URA3* at site of *ptr3* mutation	AACTTCAACCAACTAACAGAGCAATCGTCAT CTTCACTCTAGATTGTACTGAGAGTGCAC
CRO387	Integrate *URA3* at site of *ptr3* mutation	TATCAAGAAATCATTGGAAAGTTTGCAAAA ACGTTGGCTCCTGTGCGGTATTTCACACCG
CRO224	Replace *URA3* with *ptr3* allele	GCACATGATCTGGACGAAGA
CRO225	Replace *URA3* with *ptr3* allele	ATGGGGAATCTCGACACGTA
CRO352	Integrate *URA3* at site of *yor019w* mutation	CCAGCTTTAAGAATGCTTTGATAGGCAATGG GTCGAAAAAAGATTGTACTGAGAGTGCAC
CRO353	Integrate *URA3* at site of *yor019w* mutation	CTGAGGAAGAATATGGTATTAAAGATTTTCT AAACTTTGTCTGTGCGGTATTTCACACCG

Table 2-3 (Continued)

CRO318	Replace *URA3* with *yor019w* allele	CCCAGCATTCAAGAAGGAAG
CRO319	Replace *URA3* with *yor019w* allele	GCACCGGCACTTTTAACTTT
CRO408	Integrate *URA3* at site of *msn2* mutation	AAGTGTCGTAATAGAATCAACAAAGGAACT CGAGGAGAAAAGATTGTACTGAGAGTGCAC
CRO409	Integrate *URA3* at site of *msn2* mutation	GGTCGTTCGTTAGAGTGAACAGATCTCACAT GCCTTTTCACTGTGCGGTATTTCACACCG
CRO406	Replace *URA3* with *msn2* allele	TATCACCATTTCCCACAGCA
CRO407	Replace *URA3* with *msn2* allele	TGACAAGCAAATGGTCGTTC
CRO476	Delete *PTR3* with *URA3*	ACACATACATAGGTACGAAATACACAACTG ATAGGCGTTCAGATTGTACTGAGAGTGCAC
CRO477	Delete *PTR3* with *URA3*	GTATACCAGAACCTTAAACATACGTATATAT TTAGATGCACTGTGCGGTATTTCACACCG
FO7014	Delete *MSN2* with *URA3*	TTTTTCAACTTTTATTGCTCATAGAAGAACTA GATCTAAAAGATTGTACTGAGAGTGCAC
FO7015	Delete *MSN2* with *URA3*	TTATGAAGAAAGATCTATCGAATTAAAAAA ATGGGGTCTACTGTGCGGTATTTCACACCG
CRO416	Replace *kanMX-GAL1* with *URA3*	ATTCACAACTTTGGTCAAACGCCTTTACAAA TATTTCAGGAGATTGTACTGAGAGTGCAC
CRO417	Replace *kanMX-GAL1* with *URA3*	AAGATTGTCTTCTCAAATATTGGCTTCATTG GAACCTTACCTGTGCGGTATTTCACACCG
FO6826	Check transformation of *bph1::kanMX-* UAS_{GAL1}	TTACCCAGGCGCTGTAAATC
FO6829	Check transformation of *bph1::kanMX-* UAS_{GAL1}	GGTTACCTGAAACCGAATGC

Table 2-3 (Continued)

FO1311	Check transformation of *bph1::kanMX-UAS$_{GAL1}$*	TGATTTTGATGACGAGCGTAAT

Selection of polygenic mutants

Starting with strains CR70 and CR71, each of which contain *sin4Δ*, we selected for mutants with stronger Lda$^-$ phenotypes using the selection scheme outlined in Fig. 2-1. For each round of selection, 10 independent cultures of each strain were grown in YPD overnight at 30°C. From each culture, 200 μL of cells were plated onto two SC-His Gal 3-AT plates, one of which was UV irradiated for two minutes at 5000μJ/cm^3. Plates were incubated at 30° and colonies that grew were purified on SC-His Gal 3-AT. The first round of mutants were selected using1 mM 3-AT, the second round using 5 mM 3-AT, and the third round using 10 mM 3-AT. CR76, CR79, CR82, and CR91 were identified as spontaneous mutants. CR92 and CR98 were isolated after UV mutagenesis. The 3-AT phenotypes of mutants were verified after purification by dilution spot tests. Spot tests were also used to test whether the His$^+$ phenotype was dependent on galactose. Mutants showing stronger 3-AT resistance than the parent strains were then tested for *HIS3* and *URA3* mRNA levels by Northern analysis. Only those mutants that showed increased mRNA levels for both reporters relative to the parent strain were used for further selection.

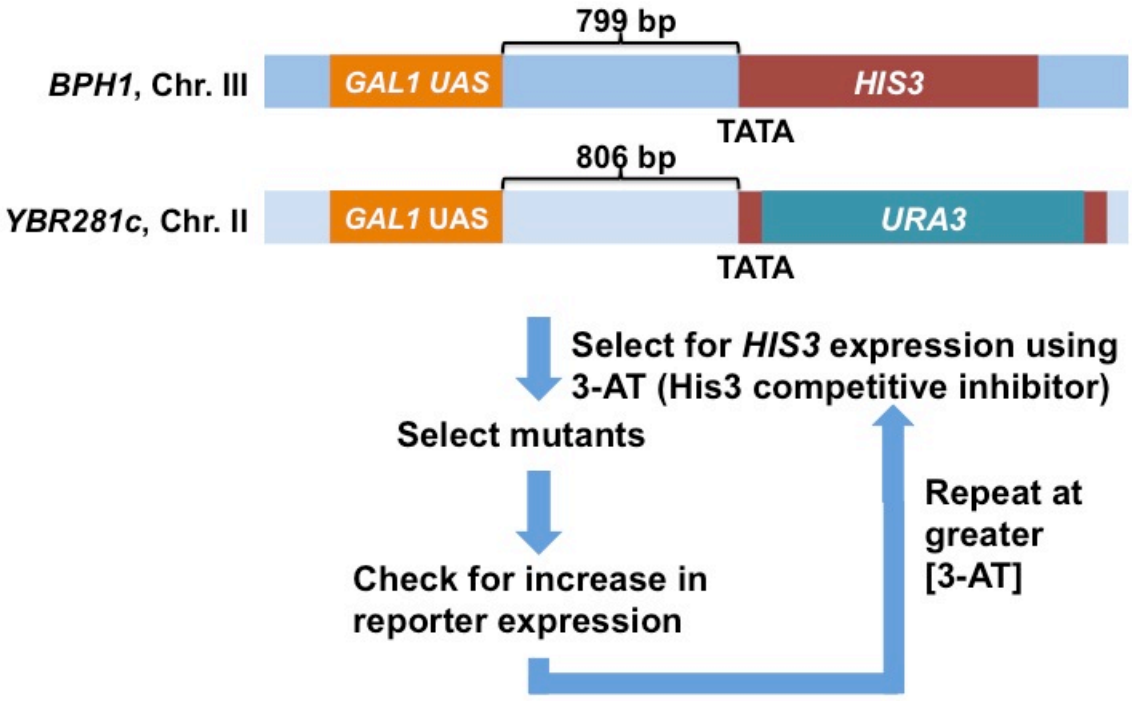


Figure 2-1. Selection of polygenic mutants with an increased Lda$^-$ phenotype. Two *sin4Δ* strains, CR70 and CR71, each containing the two diagrammed reporters, were used for the first round of selection. Mutants were selected for increased resistance to 3-AT. Northern analysis was used to check for increased reporter expression. Mutants displaying increased expression of both reporters were used for the next round of selection. The process was repeated using greater concentrations of 3-AT.

Bulk segregant analysis

Pools of strains were generated for bulk segregant analysis by backcrossing two of the polygenic mutants, CR91 and CR92, to CR71 and backcrossing the third one, CR98, to CR70. For each cross, approximately 300 tetrads were dissected. The Lda- phenotypes of the progeny were tested by replica plating to SC-His Gal media with 0 mM, 1 mM, 5 mM, and 10 mM 3-AT. The mutant pools for each cross were composed of segregants showing the same 10 mM 3-AT resistance seen in the polygenic mutant parent. The wild-type pools for each cross were composed of segregants that showed the phenotype of the *sin4Δ* single mutant, which is sensitivity to 1 mM 3-AT. Each pool was required to contain at least 40 segregants, based on prior bulk segregant analysis studies (Brauer et al., 2006; Wenger et al., 2010). CR98 segregants were all derived from complete tetrads. CR91 and CR92 segregants were from a mixture of complete and incomplete tetrads. The mutant and wild-type pools from each backcross contained the same number of segregants: 48 segregants (CR91), 42 segregants(CR92), or 45 segregants (CR98).

Genomic DNA was isolated for each segregant pool to be used for library construction. To do this, a saturated culture of each segregant was grown in YPD to saturation at 30°C. For each pool, 1 ml of each culture was combined, the pooled culture was split into 6 fractions, and DNA was extracted from each fraction as previously described (Rose et al., 1990). The 6 fractions of extracted DNA were pooled and submitted for library construction.

High-throughput sequencing of yeast segregant pools to identify SNP

Genomic DNA was multiplexed using the Illumina TruSeq DNA PCR-Free Sample Preparation Kits. Using PCR-free sample preparation prevents additional mutations from being introduced during library amplification. The resulting libraries were run on a single lane of an Illumina HighSeq 2000. The sequencing resulted in 108-nucleotide single end reads. We had greater than 80 fold coverage (Table 2-4) for all of our pooled libraries, which is greater than the 50-fold coverage we hoped to get (Wenger et al., 2010). The sequence reads, compiled in a FASTQ file, were mapped to the *S. cerevisiae* S288C genome using BWA (Li and Durbin, 2009) producing a SAM file. Using SAMTools (Li et al., 2009), this SAM file was converted to a BAM file, which was searched for SNPs using Freebayes (Garrison and Marth, 2012). Freebayes generated a VCF file containing the SNPs. We then calculated, using a PERL script that is available upon request (M. Hickman, unpublished), the frequency of all SNPs in two matched pools: 1. segregants with the wild-type phenotype and 2. segregants with the Lda^- phenotype. SNPs with a higher frequency in pool 2 were further considered as candidate SNPs that might be causative for the Lda^- phenotype. There was no strict threshold cutoff, however SNPs were considered candidates if they were present at a frequency greater than 50% in pool 2. The SNPs were subsequently verified by observing their frequency and location in the BAM file containing all of the aligned reads, using IGV (Integrated Genome Viewer; James, 2011). Once the SNPs were confirmed by these methods, they were further confirmed to be present in the Lda^- parent and absent in the wild-type parent.

Table 2-4. Sequencing of segregant pools

Sequencing Pool	Pool Size	Number of Mapped Reads	Number of Unmapped Reads	Coverage
1.2 mutant	45	12,364,877	294,006	85x
1.2 wild-type	45	12,115,678	263,008	89x
2.3a mutant	48	14,174,495	258,696	97x
2.3a wild-type	48	13,097,764	139,359	101x
2.3b mutant	42	38,280,078	216,407	260x
2.3b wild-type	42	11,391,104	225,478	80x

Spot tests

Saturated cultures of indicated strains were grown in YPD at 30°C. Cultures were adjusted to the same concentration by OD_{600} and serially diluted by 10-fold steps. Dilutions were spotted on the indicated media and incubated at 30°C. Subsets of the media tested by spot tests are presented in figures within the chapter. The 3-AT concentrations presented were chosen to best represent each experiment.

Northern analysis

RNA isolation and Northern hybridization experiments were performed as previously described (Ausubel et al., 1991). As indicated, strains were grown to mid-log in YP Raffinose and then shifted to 2% glucose or 2% galactose for one hour. Northern hybridization analysis was conducted with probes to the coding regions of *HIS3* (-27 to +376, where +1 is the ATG), *URA3* (+206 to +727), and *SNR190* (+1 to +190), which was used as a loading control.

Stability tests

Three independent overnight cultures of CR91, CR92, and CR98 were grown in YPD at 30°C. Cultures were serially diluted and, for each culture, 100 μL of a 10^{-5} dilution was plated onto a YPD plate in duplicate, yielding approximately 100-150 colonies per plate. Each YPD plate was replica plated to SC Gal, SC-His Gal, SC-His

Gal 1 mM 3-AT, SC-His Gal 5 mM 3-AT, and SC-His Gal 10 mM 3-AT. The percent stability was calculated for each strain on each medium relative to growth on SC Gal.

Strain aliases

Throughout Chapters 2 and 3, strains isolated from the mutant selections are referred to by strain aliases that indicate their lineage and the round of selection in which they arose. For example, CR91 and CR92, which were isolated independently in the third round of selection of lineage 2 are designated strains 2.3a and 2.3b, respectively. Multiple strains can have the same alias, as we use this designation for any progeny that have the same phenotype. Alias designations are also used when referring to strains that were generated from that background. For example, a *mot3* allele replacement experiment, where the *mot3* mutant allele is replaced with the *MOT3* wild-type allele in strain 1.2, the resulting strain is designated 1.2 $MOT3^{+}$. Similarly, when *MOT3* is deleted in strain 1.2, the resulting strain is designated 1.2 *mot3Δ*.

Results

Selection of polygenic mutants that allow long-distance activation

To select for polygenic mutants, we started with two strains, each containing two reporters (Fig. 2-1) and *sin4Δ*. We began the selection with *sin4Δ* because it displays the strongest Lda^{-} phenotype of the previously identified mutants and *sin4* mutants were the

most frequent class of single mutant isolated (Dobi and Winston, 2007). The use of two reporters should reduce the number of *cis*-acting mutations isolated.

We selected for mutations that strengthened the Lda$^-$ phenotype of a *sin4Δ* mutant using multiple rounds of selection for 3-AT resistance, as described in Materials and Methods. This selection scheme resulted in the isolation of three strains showing an increased Lda$^-$ phenotype (Fig. 2-2). Two of the three strains, 2.3a and 2.3b, were initially part of the same lineage and were only separated for the third and final round of selection. Strain 2.3a was isolated as a spontaneous mutant and strain 2.3b was isolated using UV mutagenesis, so we suspected these strains contained different mutations. The third strain, 1.2, was isolated independently of 2.3a and 2.3b. Additional rounds of mutagenesis were attempted for each, but did not yield any strains with increased expression at both reporters.

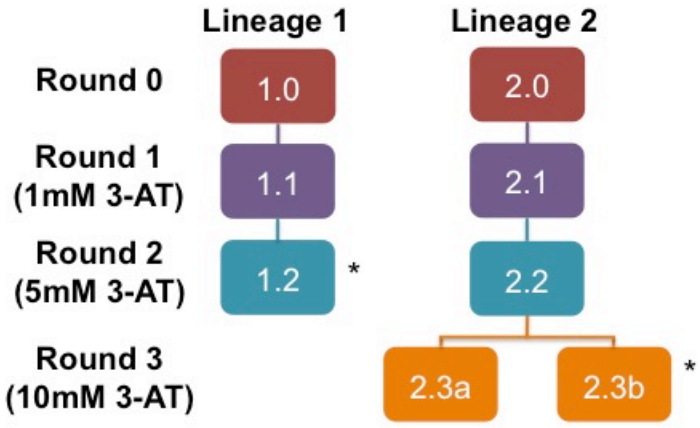


Figure 2-2. Isolation of polygenic mutants. Starting with *sin4Δ* strains, two independent lineages were isolated that display increased Lda$^-$ phenotypes. Two independent mutations were isolated at the last step of selection of lineage 2. This resulted in three final strains that were analyzed, 1.2, 2.3a, and 2.3b. Asterisks denote the use of UV mutagenesis.

We compared 3-AT resistance and reporter mRNA levels between the strains isolated at each round of selection. The strains show an increase in 3-AT resistance with each round of mutagenesis, with one exception (Fig. 2-3). In this one case, strain 1.1 shows a similar phenotype to its *sin4Δ* single mutant parent*;* we originally scored this strain as resistant to 1 mM 3-AT and proceeded to the next round of selection. However, a second test showed that it is actually 3-AT sensitive. Regardless, it still gave rise to stronger mutants. The three final strains are resistant to 10 mM 3-AT, but the growth of strain 1.2 is stronger than that of strains 2.3a and 2.3b. In addition, all the mutants require the presence of galactose to grow on media lacking histidine, showing that the activation of the reporter is galactose dependent. Finally, the 3-AT phenotype correlates with transcript levels at both reporters, as all three lineages show an increased level of *HIS3* and *URA3* mRNA as the number of mutations increases (Fig. 2-4). In addition to the *HIS3* and *URA3* transcripts, we also see a long transcript for each probe hybridization. Previous 5'-RACE experiments indicate the long transcript initiates proximal to the UAS (Dobi and Winston, 2007). This long transcript is at lower levels in 1.1 and 1.2 than in the *sin4Δ,* although remains fairly constant in lineage 2 mutants. *URA3* mRNA levels overall correlated well with *HIS3* mRNA levels; however, strain 2.3a shows higher a *HIS3* mRNA level than strain 2.3b, but the two strains show a similar level of *URA3* mRNA.

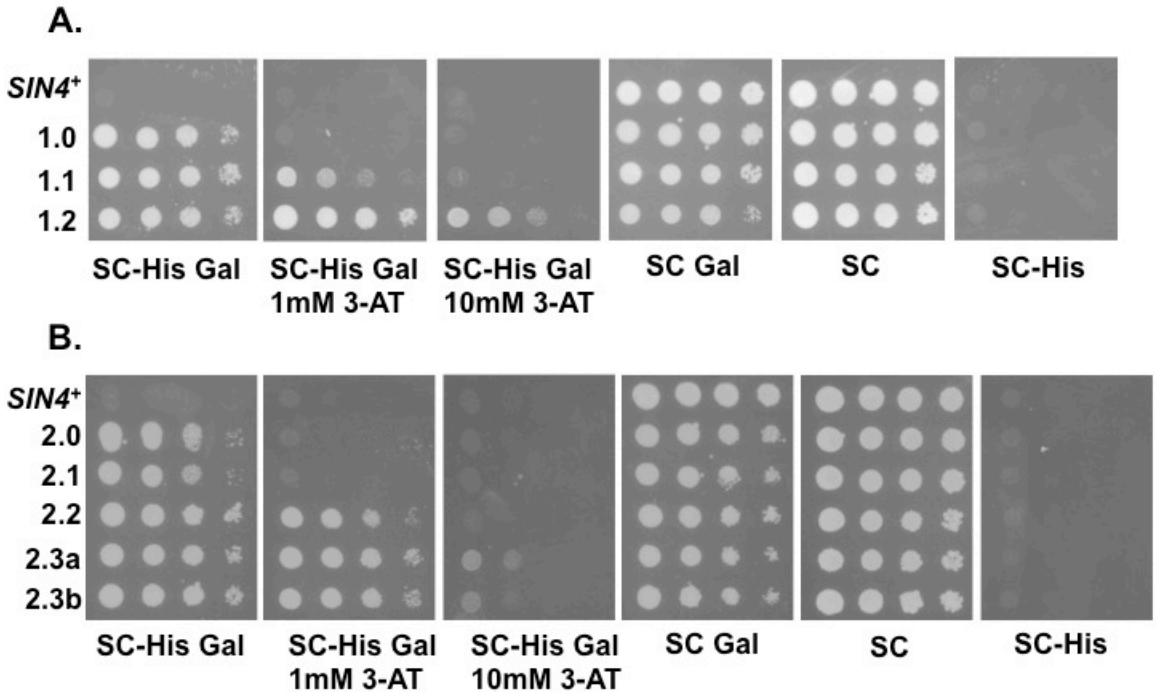


Figure 2-3. Lda$^-$ phenotypes of lineages. **A.** Growth of strains in lineage 1. The strains are *SIN4*$^+$ (CR74), 1.0, 1.1, and 1.2. Ten-fold serial dilutions were made of saturated YPD cultures. The dilutions were spotted to the specified media and incubated for four days at 30°. **B.** Growth of strains in lineage 2. The strains are *SIN4*$^+$ (CR74), 2.0, 2.1, 2.2, 2.3a, and 2.3b. Strains were grown and incubated as in panel A.

Figure 2-4. Northern analysis of lineage 1 and lineage 2 strains. **A.** Northern analysis of *BPH1* and *YBR281c* reporters in lineage 1. A Northern blot was hybridized with a probe for *HIS3* (top panel), *URA3* (middle panel), and *SNR190* (bottom panel). Strains are as follows: lane 1 (FY76), lanes 2 and 6 (CR74), lanes 3 and 7 (1.0), lanes 4 and 8 (1.1), and lanes 5 and 9 (1.2). Strains in lanes 2-5 were shifted from growth in 2% raffinose to 2% glucose and strains in lanes 6-9 were shifted to 2% galactose. **B**. Northern analysis of *BPH1* and *YBR281C* reporters in lineage 2. A Northern blot was hybridized with a probe for *HIS3* (top panel) *URA3* (middle panel), and *SNR190* (bottom panel). Strains are as follows: lane 1 (FY76), lanes 2 and 8 (CR74), lanes 3 and 9 (2.0), lanes 4 and 10 (2.1), lanes 5 and 11 (2.2), lanes 6 and 12 (2.3a), and lanes 7 and 13 (2.3b). Strains in lanes 2-7 were shifted to 2% glucose and strains in lanes 8-15 were shifted to 2% galactose.

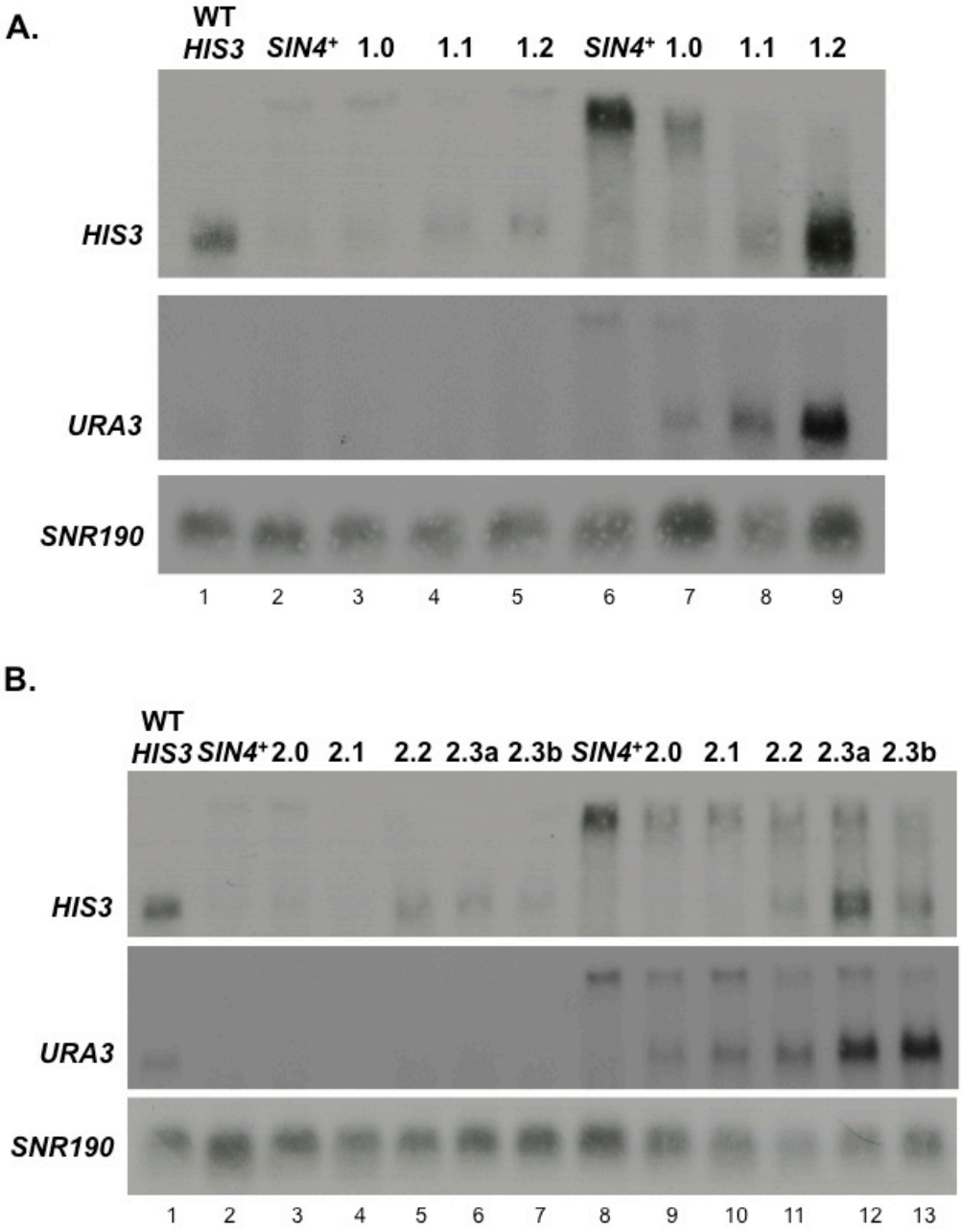


Figure 2-4. (Continued)

Identification of candidate causative mutations by bulk segregant analysis

We used bulk segregant analysis and whole-genome sequencing to identify candidate causative mutations (Fig. 2-5), as described in Materials and Methods. Briefly, the final three polygenic mutant strains, 1.2, 2.3a, 2.3b were crossed to *sin4Δ*, sporulated, and approximately 300 tetrads were dissected and analyzed. The progeny displayed 3-AT resistance phenotypes ranging from sensitivity to 1 mM 3-AT (seen in the *sin4Δ* single mutant parent) to resistance to 10 mM 3-AT (seen in the polygenic mutant parent) (Fig. 2-5). It is worth noting that mutants grow more quickly when replica plated to 3-AT (as in Fig. 2-5) than when they are spotted to 3-AT (as in Fig. 2-3). Resistance to 10 mM 3-AT was present in 6.8% (1.2), 6.4% (2.3a), 4.1% (2.3b), and of the segregants (Table 2-5), frequencies consistent with the polygenic mutants containing five causative mutations, including the *sin4Δ*. The crosses generated a high frequency of incomplete tetrads (Table 2-5), particularly the 2.3a and 2.3b backcrosses. The number of viable progeny in tetrads appeared random. Due to the lack of complete tetrads, it is difficult to access the accuracy of the 3-AT resistance frequencies.

We identified a set of candidate causative mutations for each of the mutant strains by two main criteria. First, mutations were candidates to be causal if the mutant allele frequency was greater than 50% of the sequence reads in the mutant pool. Frequencies of 50% are consistent with random segregation of mutations. We set the threshold below 100%, reasoning the final strains might contain distinct combinations of causative mutations that are sufficient to confer the strong mutant phenotype. By this reasoning,

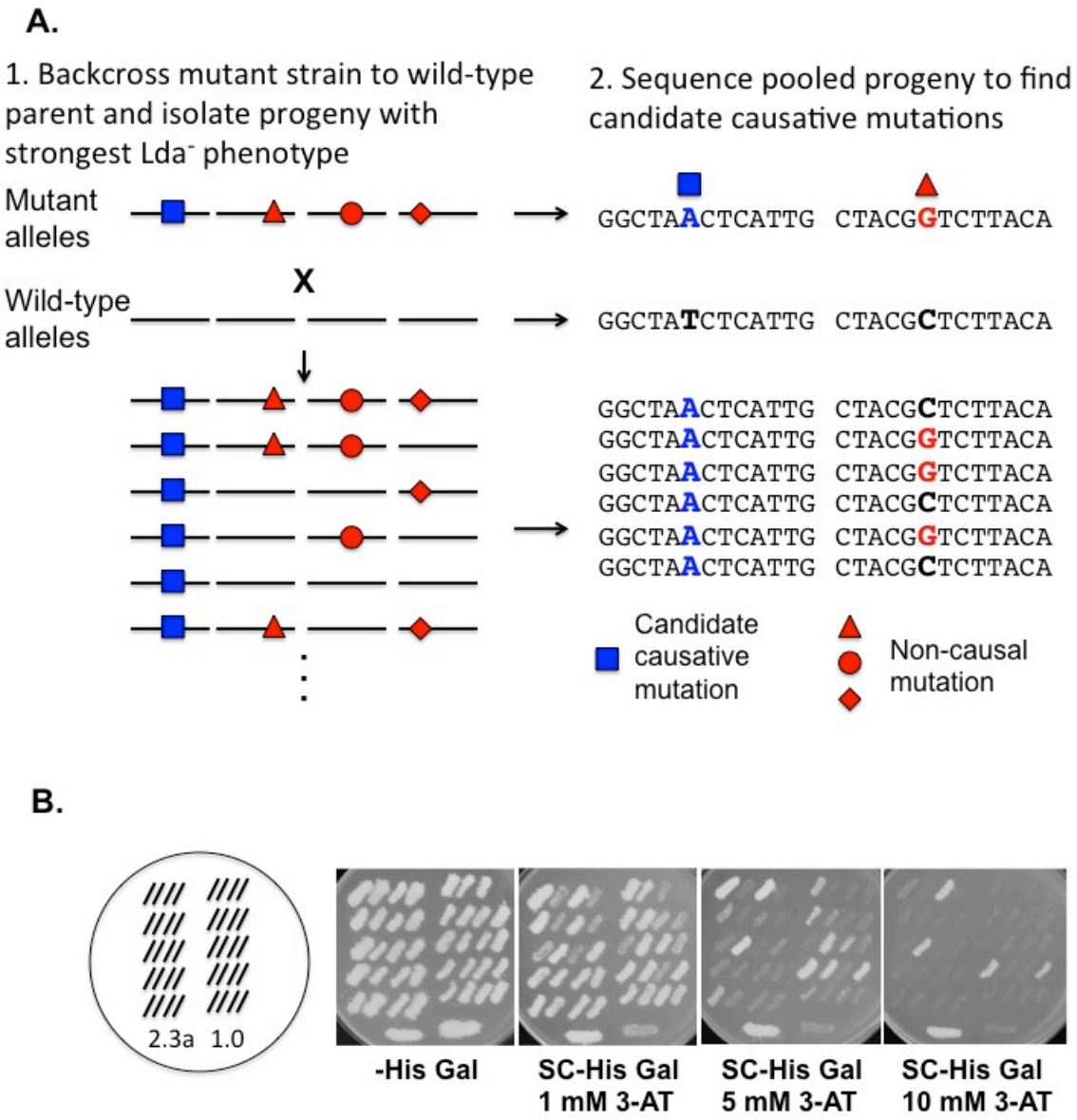


Figure 2-5. Bulk segregant analysis of mutants. **A.** Schematic of identification of causative mutations by bulk segregant analysis and whole genome sequencing. Adapted from Koschwanez et al. 2013. **B.** Growth of segregants isolated from the cross of 2.3a x 2.0. Each set of four strains is the progeny from a complete tetrad. The strains at the bottom of the plate, 2.3a and 1.0, are the parents for the cross. Strains were grown on YPD, incubated overnight at 30°, and replica plated to the indicated media. Replica plates were incubated for four days at 30°.

Table 2-5. Backcross of 1.2, 2.3a, and 2.3b

	Segregants with strong phenotype	Total number of tetrads	Frequency of strong phenotype	Percentage complete tetrads*
1.2	66	282	6.8%	61%
2.3a	59	297	6.4%	45%
2.3b	42	340	4.1%	38%

*Numbers of viable spores were: 969 (1.2), 920 (2.3a), and 1018 (2.3b)

every mutation might not be required in every segregant to produce the strong mutant phenotype. Setting the threshold lower than 100% should also account for any technical errors. Second, we required that the mutant allele be present in the mutant parent strain and wild-type allele to be present in the *sin4Δ* single mutant parent. Using these requirements, we established a list of genes to be tested for causality by allele replacement (Table 2-6).

Lineage 1contains mutations in *MOT3*, *GRR1*, and *SGM1*. Mot3 is a site-specific DNA-binding transcription factor that represses ergosterol biosynthetic genes during hypoxic and osmotic stress (Grishin et al., 1998; Hongay et al., 2002; Madison et al., 1998; Martinez-Montanes et al., 2013). Mot3 is also able to act as a transcription activator and contains a prion domain (Abramova et al., 2001; Grishin et al., 1998). Grr1 is one of two yeast F-box components of the SCF ubiquitin ligase complex (Li and Johnston, 1997). It is involved in the turnover of the G1 cyclins Cln1, Cln2, and Cln3 (Barral et al., 1995; Landry et al., 2012). Grr1 also has an established role in glucose transport (Bailey and Woodword, 1984; Conklin et al., 1993; Flick and Johnston, 1991). Sgm1 is a protein of unknown function required for wild-type growth on galactose and mannose (Entian et al., 1999).

The lineage 2 strains, 2.3a and 2.3b, share mutations in *PTR3* and *YOR019W*, but differ in that strain 2.3a contains a mutation in *MIT1* and 2.3b contains a mutation in *MSN2*. Mit1 is a transcription activator necessary for the regulation of haploid invasive

Table 2-6. Candidate causal mutations

Strain	Gene	Causative	Amino Acid Change	Mutant Reads/ Total Reads	Protein Function
1.2	*MOT3*	Yes	N388H	73/74	Transcription factor
	GRR1	Yes	L181stop	67/79	Ubiquitin ligase component
	SGM1	No	L407S	79/95	Required for growth on galactose
	*SGF73**	No	P84L	46/69	Component of SAGA complex
	*TMA108**	No	A458V	42/67	Ribosome biogenesis
	*RIM8**	No	P52L	46/68	Processes a transcription factor
2.3a only	*MIT1*	Yes	H187R	81/81	Transcription factor
2.3b only	*MSN2*	Yes	C652stop	252/252	Transcription factor
Both 2.3a and 2.3b	*PTR3*	Yes	S363stop	68/68 (1.3a) 231/235 (1.3b)	Nutrient sensing
	YOR019W	Yes	N553K	48/97 (1.3a) 62/259 (1.3b)	Unknown function

*Note: *rim8 tma108* and *sgf73* were not determined to be non-causative by allele replacement, but were eliminated as causative during mutant reconstruction. This is described in more detail in the text.

growth and diploid pseudo-hyphal growth (Cain et al., 2012). It is conserved among yeasts, and in *Candida albicans* it is a regulator of white-opaque switching, a fungal morphological change (Zordan et al., 2006). Msn2 is a transcription factor involved in the response to a range of stresses, including heat shock, osmotic stress, and oxidative stress (Martinez-Pastor et al., 1996). Msn2 has a paralog, Msn4, and the two transcription factors are partially redundant (Berry and Gasch, 2008; Martinez-Pastor et al., 1996). Ptr3 is a component of the Ssy1p-Ptr3p-Ssy5p plasma membrane bound sensor. The sensor is involved in signal response to extracellular amino acids (Forsberg et al., 2001; Forsberg and Ljungdahl, 2001). *YOR019W* encodes a protein of unknown function.

Identification of causative mutations in lineage 1

We identified two causative mutations in strain 1.2 by replacement of the mutant allele with the wild-type allele in the mutant strain background. For causative alleles, we expect this change to weaken or abolish the Lda^- phenotype. For both *grr1-531* and *mot3-1162*, replacement with the wild-type alleles causes a reduction in 3-AT resistance, indicating these mutations are causative (Fig. 2-6). In contrast, replacement of *sgm1-1220* with the wild-type allele does not alter 3-AT resistance. This result was surprising because *sgm1-1220* is present at a similar frequency of reads in the mutant pool as *grr1-531*. *SGM1* is linked to *GRR1* by a genetic distance of approximately 20 cM. We expect the two mutations to be linked in the mutant pool, however at this genetic distance we expect crossovers. An alternate explanation for the high frequency of *sgm1-1220* is the

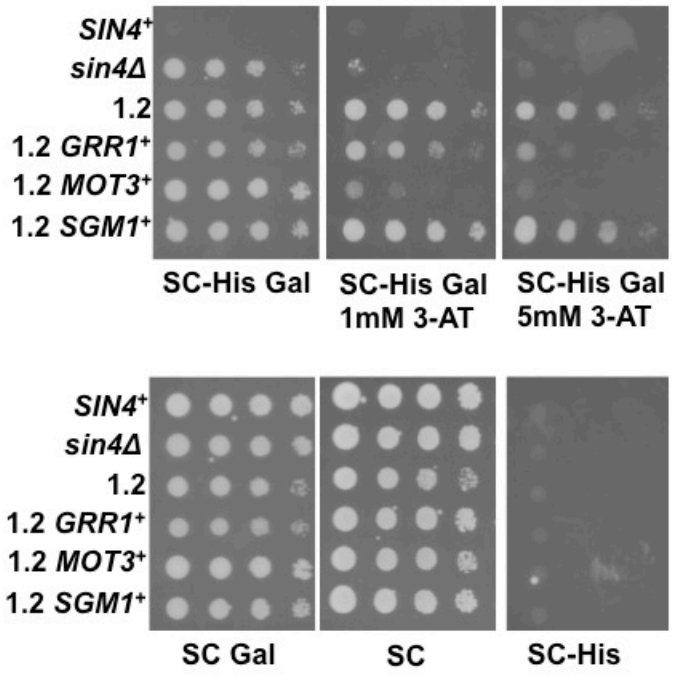


Figure 2-6. Lineage 1 tests for causality by allele replacement. Spot tests of strains in which the candidate causal allele has been replaced with the wild-type allele. Allele replacements were constructed in strain 1.2. Strains are *SIN4*$^{+}$*(*CR214), *sin4 Δ* (CR111), 1.2, 1.2 *GRR1*$^{+}$ (CR200), 1.2 *MOT3*$^{+}$ (CR201), and 1.2 *SGM1*$^{+}$ (CR202). Dilutions were spotted onto the indicated media and incubated for four days at 30°.

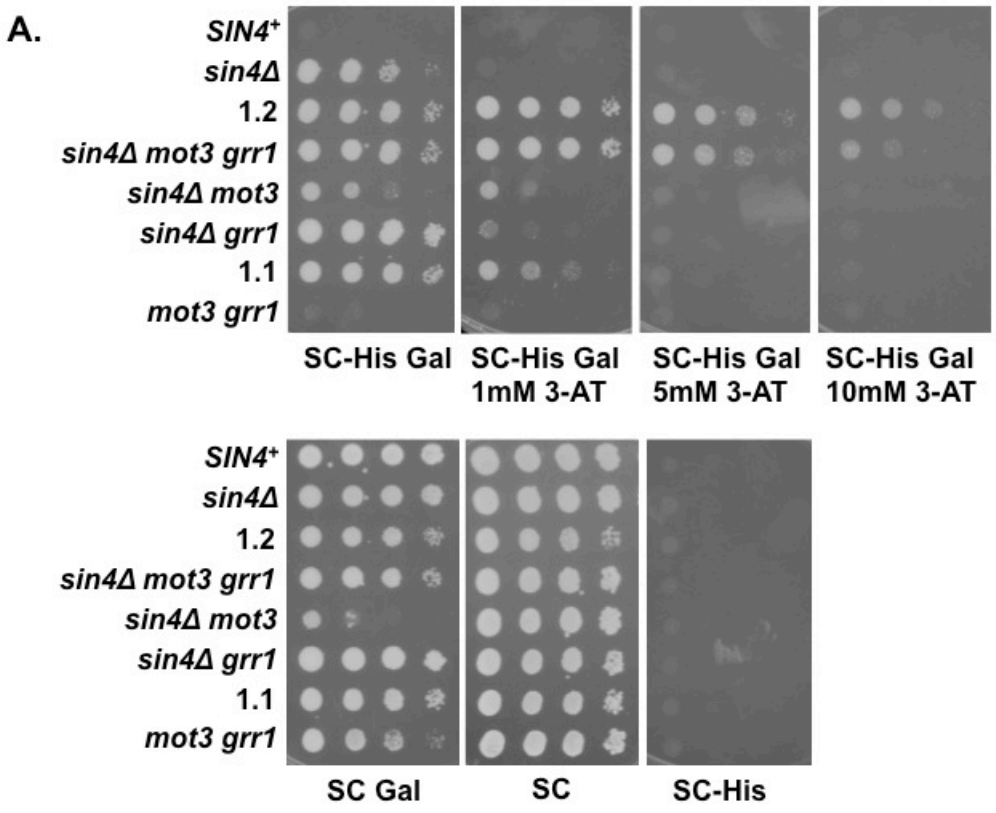
mutation confers a growth benefit unrelated to the Lda$^-$ phenotype. Sanger sequencing of these three genes revealed that the *grr1-531* arose in the first round of selection while *mot3-1162* and *sgm1-1220* arose in the second round.

Reconstruction of lineage 1 mutant strains

As an independent test of the role of each causative mutation in lineage 1, we combined them to test whether together they would recapitulate the original phenotypes of polygenic mutant 1.2. For lineage 1, we constructed the *sin4Δ mot3-1162 grr1-531* triple mutant, as well as all possible double mutants, and analyzed the strains for their Lda- phenotypes by growth and Northern analysis. Significantly, the *sin4Δ0 mot3-1162 grr1-531* triple mutant is resistant to 10 mM 3-AT (Fig. 2-7A) and shows a similar level of *HIS3* mRNA expression as 1.2 (Fig. 2-7B, compare lanes 3 and 4), suggesting that these three causative mutations can account for most or all of the 1.2 mutant phenotype. We note that in the figure, the triple mutant shows slightly less growth than strain 1.2 on 10 mM 3-AT. This growth difference is within the variation normally seen for the triple mutant. Additionally, we observe slight variation in growth of this mutant on galactose-containing media.

We eliminated *rim8-155, tma108-1374,* and *sgf73-251* as candidate causative mutations based on the ability of the *sin4Δ* 0 *mot3-1162 grr1-531* triple mutant to recapitulate the strain 1.2 phenotype. The *sin4Δ*0 *mot3-1162 tma108* and *sin4Δ*0 *mot3-1162 sgf73* triple mutants both display the same 3-AT resistance as the

Figure 2-7. Analysis of strains constructed with lineage 1 causative mutations. Causative mutations were constructed in a *sin4Δ* strain. **A.** Growth of reconstructed mutant strains. Strains are: *SIN4*$^{+}$ (CR214), *sin4Δ* (CR111), 1.2 (CR101), *sin4Δ mot3 grr1* (CR204), *sin4Δ mot3* (CR169), *sin4Δ grr1* (CR194), 2.1 (CR79), and *SIN4*$^{+}$ *mot3 grr1* (CR219). Dilutions were spotted to the indicated media and incubated for four days at 30°. **B.** Northern analysis of reconstructed strains. A Northern blot was hybridized with probes for *HIS3* (top panel) and *SNR190* (bottom panel). Strains are: lane 1 (CR214), lane 2 (CR111) lane 3 (1.2) lane 4 (CR204), lane 5 (CR169) lane 6 (CR194) lane 7(CR79). Strains were shifted from 2% raffinose to 2% galactose.



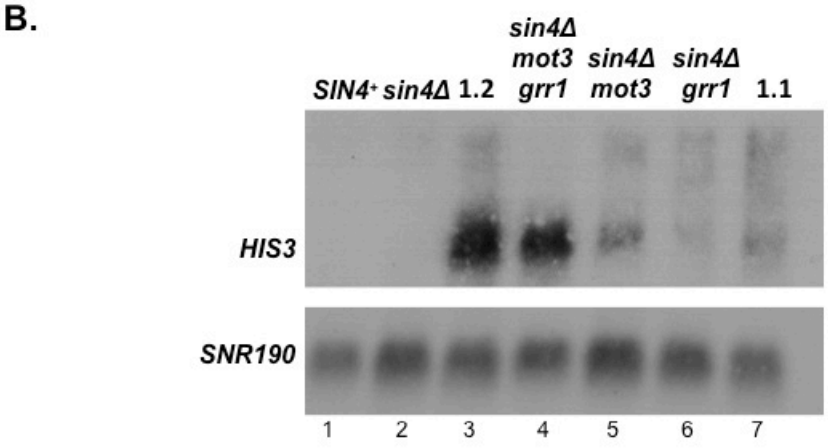


Figure 2-7. (Continued)

*sin4Δ*0 *mot3-1162* double mutant. Additionally, the *sin4Δ*0 *rim8-155 tma108-1374 sgf73-251* quadruple mutant is sensitive to 3-AT, the same phenotype as the *sin4Δ*0 single mutant.

Comparison of the double mutants to strain 1.1, the intermediate strain identified results in an interesting observation. The *sin4Δ0 grr1-531* mutant does not show the same level of 3-AT resistance or *HIS3* mRNA expression as 2.1. Sanger sequencing of strains isolated over the course of the selection revealed *grr1-531* as the only candidate mutation in the strain isolated in the first round. The disparity in phenotype between *sin4Δ 0 grr1-531* and 2.1 suggests the presence of an additional mutation in this strain that was not uncovered by bulk segregant analysis and genome sequencing.

Identification of causative mutations in lineage 2

In strain 2.3a, we individually replaced *mit1-560, ptr3-1088,* and *yor019w-1659* with wild-type alleles to test their possible causality; similarly, in strain 2.3b we replaced *msn2-1956* with the wild-type allele. Our results show that both *mit1-560* and *ptr3-1088* are causal, as replacement of each with its wild-type allele abolishes 3-AT resistance (Fig. 2-8). Our results also indicate that *yor019w-1659* contributes only weakly to the Lda^- phenotype, as replacement with the wild-type allele causes a very modest, but reproducible reduction in growth on 3-AT media. Replacement of *msn2-1956* with the wild-type allele causes a reduction in 3-AT resistance (Fig. 2-8). In addition, this strain grows poorly on media containing galactose. These results indicate the *msn2* mutation

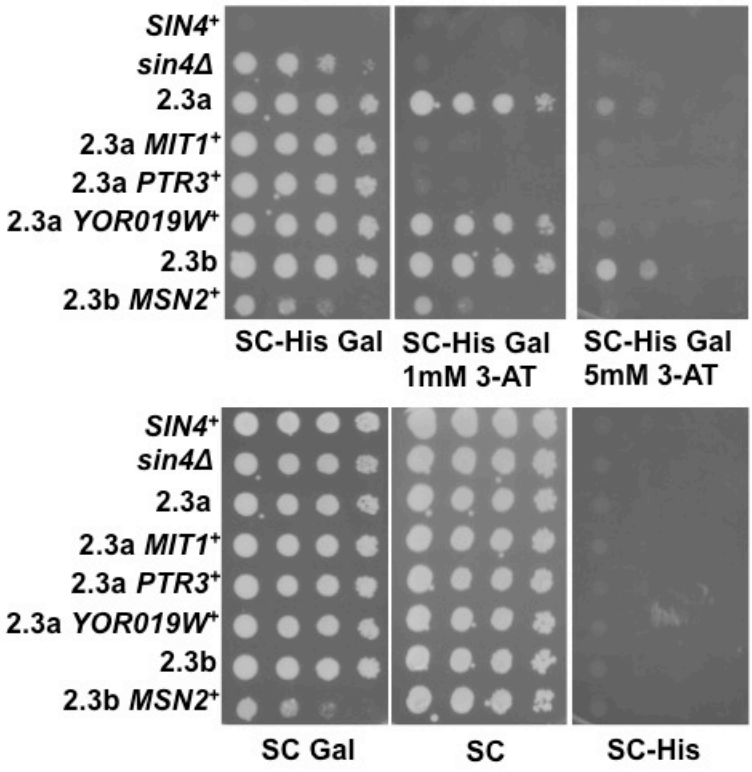


Figure 2-8. Lineage 2 tests for causality by allele replacement. Spot tests of strains in which the candidate causal allele has been replaced with the wild-type allele. Allele replacements were constructed in strain 2.3a or 2.3b. Strains are *SIN4*$^{+}$*(*CR214), *sin4 Δ* (CR111), 1.2, 2.3a *MIT1*$^{+}$ (CR221), 2.3a *PTR3*$^{+}$ (CR171), 2.3a *YOR019W*$^{+}$ (CR198), 2.3b, and 2.3b *MSN2*$^{+}$ (CR224). Dilutions were spotted onto the indicated media and incubated for four days at 30°.

allows increased growth on galactose in the 2.3b background in addition to causing a stronger Lda$^-$ phenotype. To determine the order in which the mutations arose, we used Sanger sequencing of each relevant gene in strains 2.1, 2.2, 2.3a, and 2.3b. This showed that *yor019w-1659* arose in the first round of selection, *ptr3*-1088 in the second round, and *mit1-560* or *msn2-1956* in the third and final round.

Reconstruction of lineage 2 mutant strains

We sought to determine if the identified causal mutations can account for the phenotypes of strains 2.3a and 2.3b. To do this for 2.3a, we generated a strain containing *sin4Δ0*, *mit1-560, ptr3-1088,* and *yor019w-1659*. For 2.3b, we constructed a strain containing *sin4Δ0*, *msn2-1956*, *ptr3-1088* and *yor019w-1659*. Our results show that each reconstructed strain has only modest resistance to 3-AT, weaker than the original 2.3a and 2.3b strains (Fig. 2-9). By Northern analysis, the reconstructed strain *sin4Δ0 mit1-560 ptr3-1088 yor019w-1659* also has a weaker phenotype than 2.3a (Fig. 2-9B, compare lanes 5 and 6). In contrast, the *sin4Δ msn2-1956 ptr3-1088 yor019w-1659* strain shows similar *HIS3* mRNA levels to 2.3b. Given the weak level of 3-AT resistance of this strain, we speculate that this strain may be producing non-functional transcript similar in size to the functional *HIS3* transcript. A similar transcriptional effect was previously reported (Dobi and Winston, 2007). Our results, then, show that we are able to partially recapitulate the phenotype of 2.3a and 2.3b from the known causative mutations,

Figure 2-9. Analysis of strains constructed with lineage 2 causative mutations. Causative mutations were made in a *sin4Δ* strain (CR111). **A.** Growth of reconstructed mutant strains. Strains are WT (CR214), *sin4Δ* (CR111), 2.3a, *sin4Δ mit1 ptr3 yor019w* (CR217), 2.3b, and *sin4Δ msn2 ptr3 yor019w* (CR218). Dilutions were spotted to the indicated media and incubated for four days at 30°. **B.** Northern analysis of reconstructed strains. A Northern blot was hybridized with probes for *HIS3* (top panel) and *SNR190* (bottom panel). Strains are: lane 1 (CR214), lane 2 (CR111) lane 3 (2.3a) lane 4(CR217), lane 5 (2.3b), and lane 6 (CR218). Strains were shifted to 2% galactose.

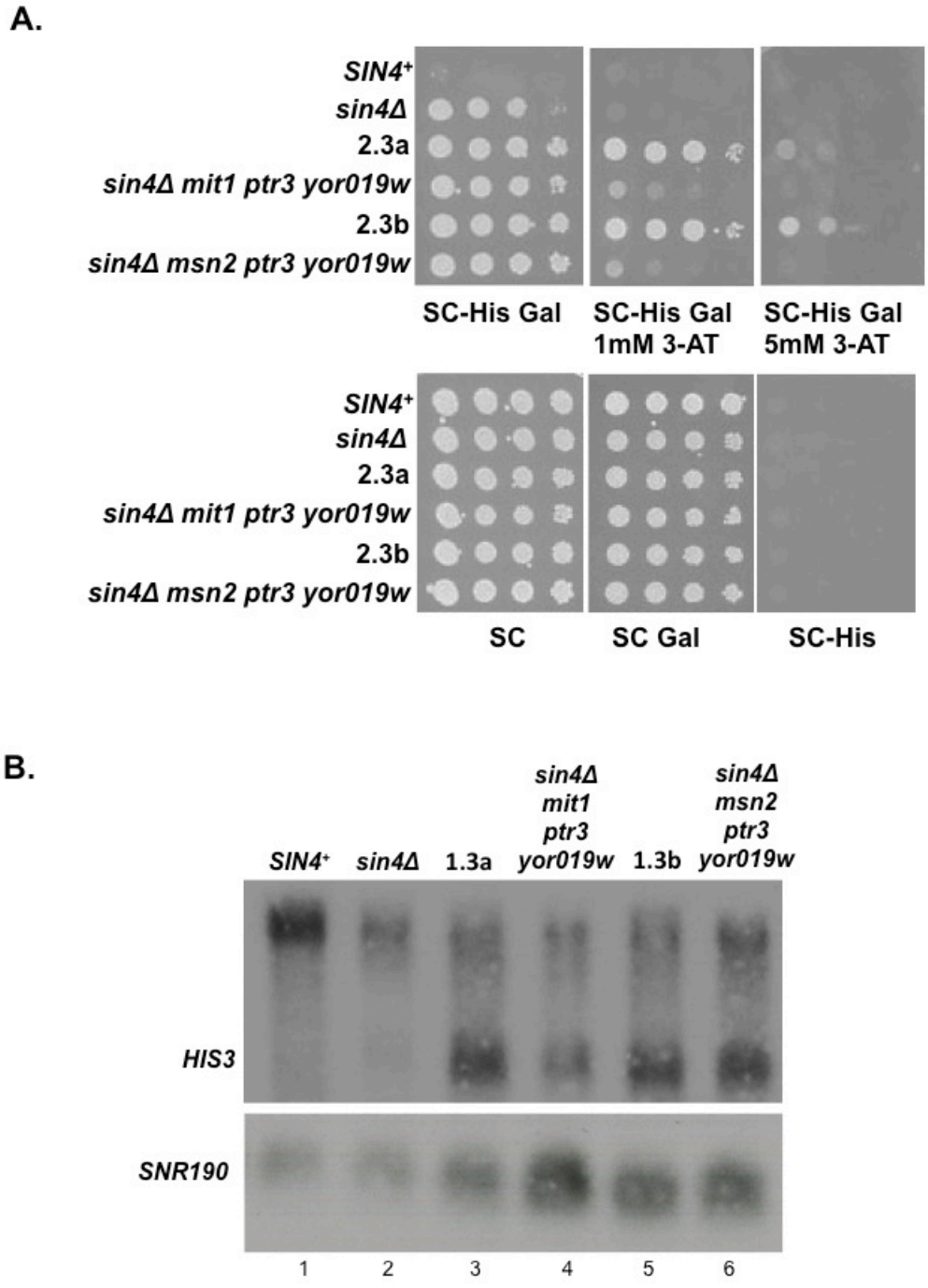


Figure 2-9. (Continued)

although there is still missing heritability. This missing heritability is likely caused by disomy of chromosome III (the location of our reporter), described in a later section.

mit1-560 and *mot3-1162* are not simply loss-of-function mutations

To test whether the mutations we identified confer their phenotypes by loss of function, we deleted each causal gene in the mutant backgrounds. From these results, *grr1Δ0*, *ptr3Δ0*, and *msn2Δ0* display the same level of 3-AT resistance as the original alleles (Fig. 2-10). In contrast, when *mot3Δ100* and *mit1Δ0* replace their original alleles, each results in reduced 3-AT resistance. This result demonstrates the *mit1-560* and *mot3-1162* are not simply loss-of-function alleles. However, the *mot3Δ100* replacement of *mot3-1162* displays a stronger 3-AT resistance than replacement of *mot3-1162* with the wild-type allele (Fig. 2-6). This suggests the *mot3-1162* allele increases the Lda^- phenotype through altered function as well as loss of function. Both mutations are in the DNA binding domains of these transcription factors (Cain et al., 2012; Madison et al., 1998), as *mot3-1162* causes an N-H change in a position that directly contacts DNA (Grishin et al., 1998), while *mit1-560* causes an H-R change at a position that likely causes a conformational change of the DNA binding domain (Lohse and Johnson, personal communication).

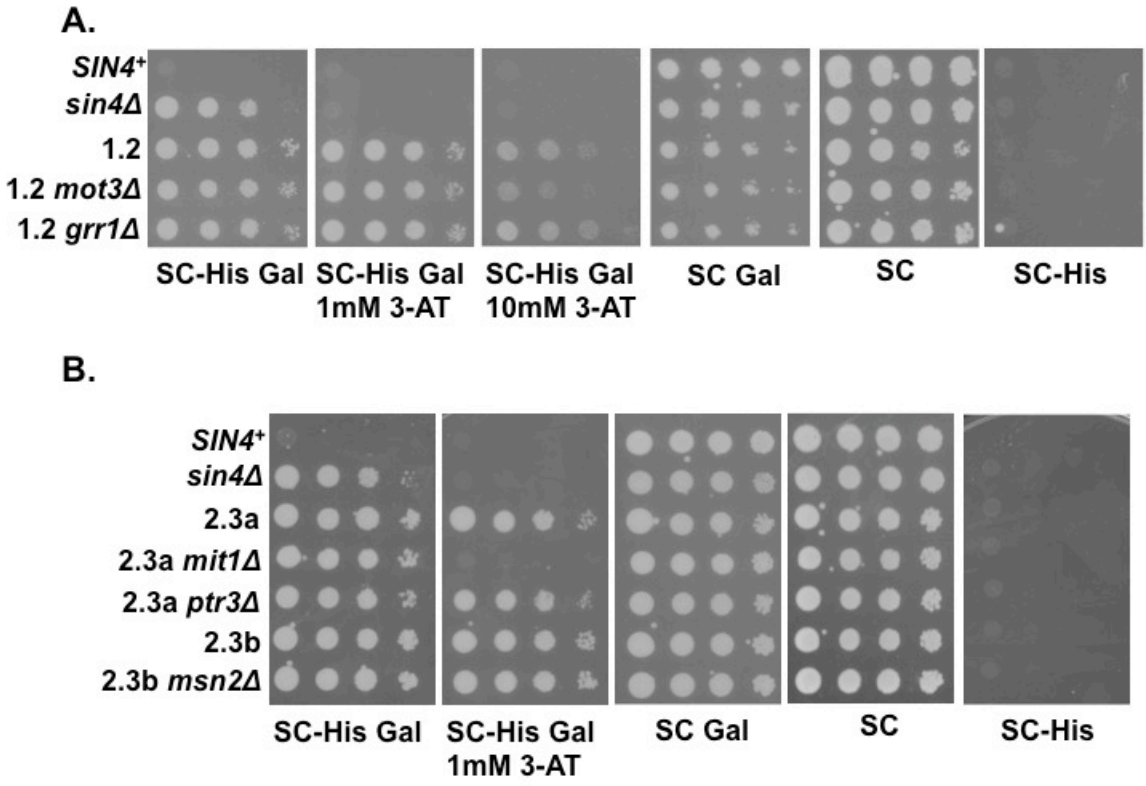


Figure 2-10. Deletion analysis of Lda$^-$ mutants. **A.** Growth of deletions of *MOT3* and *GRR1* in 1.2. Strains are *SIN4*$^+$ (CR214), *sin4Δ* (CR111), 1.2 , 1.2 *mot3Δ* (CR225), and 1.2 *grr1Δ* (CR227). Dilutions were spotted to the indicated media and incubated for four days at 30°. **B.** Growth of deletions of *MIT1* and *PTR3* in 2.3a and *MSN2* in 2.3b. Strains are *SIN4*$^+$ (CR214), *sin4Δ* (CR111), 2.3a *mit1Δ* (CR230), 2.3a *ptr3Δ* (CR231), 2.3b (CR110), and 2.3b msn2*Δ* (CR233). Dilutions were spotted to the indicated media and incubated for four days at 30°.

Lineage 2 strains are unstable for the Lda$^-$ phenotype

During standard analysis of our strains, we noticed that the level of 3-AT resistance was unstable. To test this more quantitatively, we measured stability for all three original polygenic mutants, 2.3a, 2.3b, and 1.2 (Table 2-7). Our results showed that 2.3a and 2.3b are unstable for resistance to 10 mM 3-AT, while 1.2 is stable.

Strains 2.3a and 2.3b are disomic for chromosome III

We suspected strains 2.3a and 2.3b contained a *BPH1* reporter duplication because of the instability of their phenotypes and our inability to fully recapitulate the phenotype of by strain reconstruction. Analysis of the DNA sequencing results for strains 2.3a and 2.3b suggests that they are disomic for chromosome III, the chromosome that contains the *BPH1* reporter. The mutant sequencing pools for these two strains contain approximately twice the number of reads for chromosome III as the corresponding wild-type pools (Fig. 2-11). In addition to chromosome III disomy, strain 2.3a is disomic for chromosome XIV.

The *BPH1* reporter duplication was further tested by transformation. Strains 2.3a and 2.3b were transformed with a fragment encoding *URA3* that, by homologous recombination, should replace the *GAL1* UAS region of the reporter, including the adjacent *kanMX* marker, which confers G418 resistance. However, for these strains, all

Table 2-7. Stability of Lda$^-$ phenotype of 1.2, 2.3a, and 2.3b,

	-His Gal	1 mM 3-AT	5 mM 3-AT	10 mM 3-AT
1.2	100% +/- 0%	100% +/- 0%	100% +/- 0%	100% +/- 0%
2.3a	100% +/- 0%	100% +/- 0%	99.7% +/- 0.5%	65.9% +/- 54.9%
2.3b	100% +/- 0%	100% +/- 0%	99.7% +/- 0.6%	63.9% +/- 50.8%

All percentages are calculated as growth on indicated media relative to growth on SC Gal.

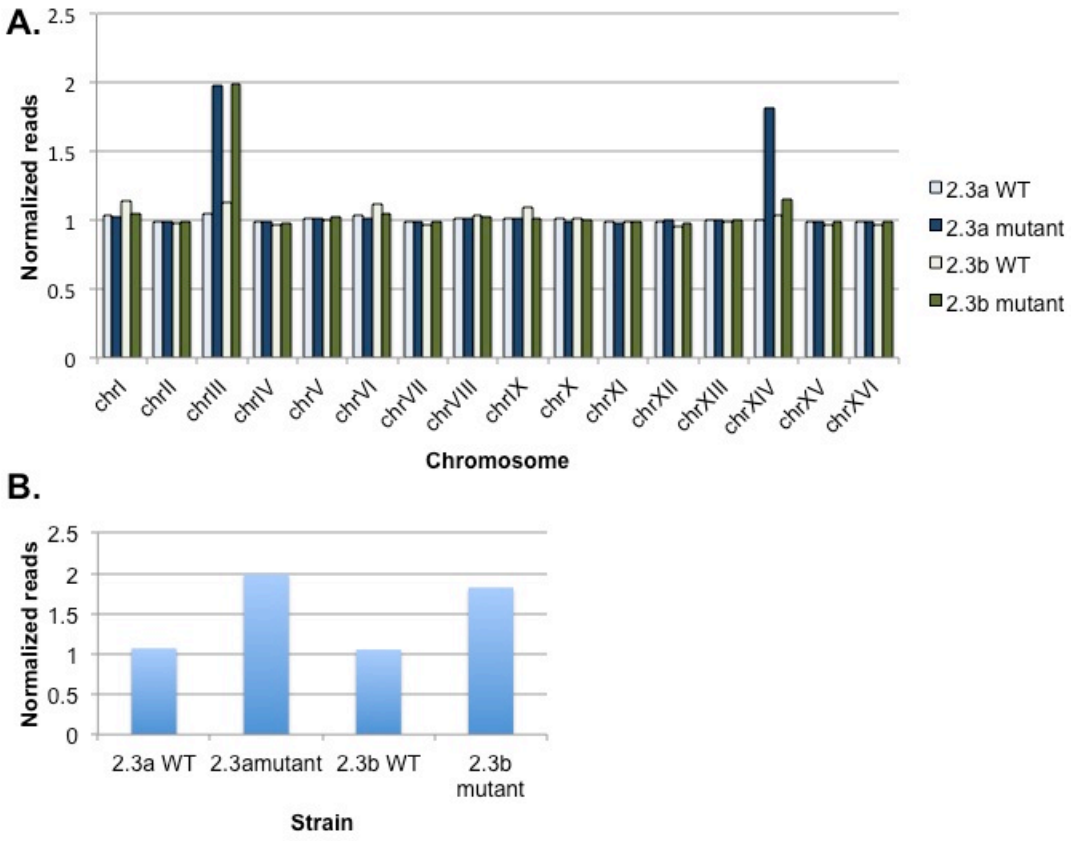


Figure 2-11. Analysis of number of sequence reads. **A.** The average number of sequence reads across each chromosome was normalized to the average number of sequence reads across the genome of the following segregant pools: 2.3a wild-type, 2.3a mutant, 2.3b wild-type, and 2.3b mutant. **B.** The average number of sequence reads across the *HIS3* ORF was normalized to the average number of sequence reads across the genome of the following segregant pools: 2.3a wild-type, 2.3a mutant, 2.3b wild-type, and 2.3b mutant.

Ura^{+} transformants were still resistant to G418. PCR of the reporter region produce two products, consistent with both the *GAL1* UAS region and *URA3* being present (Fig. 2-12). These results are consistent with a duplication of the reporter.

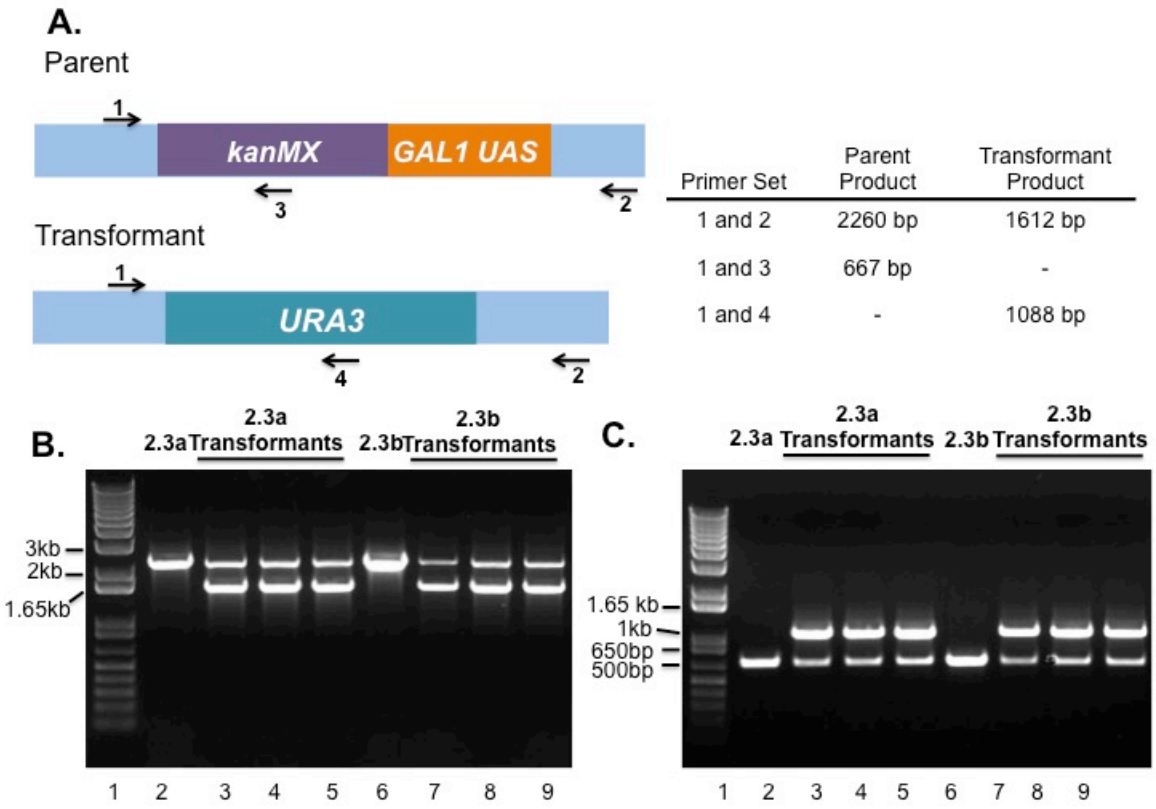


Figure 2-12. Confirmation of reporter duplication. **A.** Diagram of *kanMX-GAL1* UAS cassette in parent strains (2.3a and 2.3b) and predicted transformant. Predicted PCR products for parent or transformant are listed next to reporter diagrams. **B.** PCR products resulting from primer 1 (FO6826) and 2 (FO6829). Lanes are: lane 1 (1 kb ladder), lane 2 (2.3a), lanes 3-5 (2.3a Ura$^+$ transformants), lane 6 (2.3b), and lanes 7-9 (2.3b Ura$^+$ transformants). C. PCR products resulting from triplex PCR with primer 1, 3, and 4. Lanes are: lane 1 (1 kb ladder), lane 2 (2.3a), lanes 3-5 (2.3a Ura$^+$ transformants), lane 6 (2.3b), and lanes 7-9 (2.3b Ura$^+$ transformants).

Discussion

In this study, we have isolated three strains that show an increased Lda$^-$ phenotype compared to the *sin4Δ0* parent. All three strains contained multiple mutations that when combined with *sin4Δ0* contribute to the phenotype. We identified candidate mutations by bulk segregant analysis and genome sequencing, and demonstrated that several of them are causal. We found that the lineage 1 mutant contains mutations in *MOT3* and *GRR1* that contribute to the Lda$^-$ phenotype. The lineage 2a mutant contains mutations in that contribute to the Lda$^-$ phenotype, while lineage 2b contains mutations in *PTR3*, *YOR019W*, and *MSN2*. Identification of causative mutations revealed that the unrelated strains contain a distinct set of mutations.

At the onset of these experiments we hypothesized there would be some overlap in the mutations that arose among the strains, as this has been seen in experimental evolution experiments (Koschwanez et al., 2013; Romano et al., 2010). We might have seen some overlap if we had isolated a larger number of mutant strains. Our selection scheme limited the number of strains we were capable of processing because of the requirement to analyze reporter expression in every strain by Northern analysis.

We have not yet uncovered how each of the mutations specifically affects the ability of these strains to activate transcription at a distance. The previous study of mutants with Lda$^-$ phenotypes resulted in mutations in genes encoding factors that are involved in chromatin structure. It was somewhat surprising that none of the factors we

identified are known to have a role in chromatin structure. We did not expect to isolate mutations in *SPT2*, as the *sin4Δ spt2Δ* double mutant does not show a greater phenotype than the *sin4Δ* single mutant. However, three of the factors we identified, Mit1, Mot3, and Msn2, are transcription factors. Our finding that the *mit1-560* and *mot3-1162* mutations cause amino acid changes within their DNA binding sites and that these mutations do not cause loss of function, suggests that they may cause altered functions. One possibility is that *mit1-560* and *mot3-1162* cause altered binding specificity that in some way increases transcriptional activation distance. The change in transcription factor binding specificity could act indirectly, resulting in altered levels of other factors involved in the regulation of activation distance. Alternatively, these altered transcription factors could function directly by binding the reporter to mediate long-distance activation. The use of two reporters argues against this possibility, unless the transcription factors with altered binding specificity are now binding the *GAL1* UAS directly. The *msn2-1956* mutation affects the Lda$^-$ phenotype through loss of function, suggesting it may regulate other factors involved in regulating activation distance.

Additionally, we identified two factors, Grr1 and Ptr3, that are not known to be directly involved in transcription. Both mutations, *grr1-531* and *ptr3-1088,* cause loss of function mutations. Grr1 is a component of the SCF ubiquitin ligase complex. The complete range of degradation targets of the SCF ubiquitin ligase is unknown; thus, it is possible that loss of Grr1 results in the accumulation of protein(s) that allow long-distance activation. We suspect that *ptr3-1088* may not actually be enhancing the Lda$^-$ phenotype, but rather causes an altered sensing response to media containing 3-AT and

lacking histidine. This would require *ptr3-1088* to have been selected in the same round as an additional mutation that results in increased expression of *URA3* mRNA.

Our original backcrosses of strain 1.2 suggested that the strain contained four mutations, in addition to *sin4Δ0*, that contribute to the Lda$^-$ phenotype. We were surprised to learn that only two mutations, *grr1-531* and *mot3-1162*, are required with *sin4Δ0* to recapitulate the strong Lda$^-$ phenotype of strain 1.2. The mutant parental phenotype (strongly resistant to 10 mM 3-AT) is present at a frequency of 6.8%; however, re-analyzing the segregant data to include progeny that are resistant either strongly or modestly to 10 mM 3-AT results in a frequency of 21%, consistent with the phenotype requiring two causative mutations in addition to *sin4Δ0.* This also fits with the variation in growth on galactose containing media that has been observed for the *sin4Δ0 grr1-531 mot3-1162* triple mutant. We suspect the *sgm1-1220* mutation may contribute to the suppression of this variation, which would explain the high frequency of *sgm1-1220* allele reads in the mutant pool.

Interestingly, we found the *sin4Δ0 grr1-531* has a weaker phenotype than strain 1.1, the first-round mutant in which *grr1-531* arose. This suggests the presence of an additional mutation that was in strain 1.1 that was not identified by bulk segregant analysis. One possibility is that the final strain, 1.2, does not contain this mutation because it was lost during second round of selection. This is probably not the case because replacement of *grr1-531* with the wild-type allele in strain 1.2 results in stronger 3-AT resistance than the *sin4Δ mot3-1162* double mutant. Alternatively, the mutation

does not contribute to the Lda$^-$ phenotype in the context of both *grr1-531* and *mot3-1162*. In this event, this missing mutation would not appear to be causative by bulk segregant analysis. This scenario has precedence as previous studies have highlighted the importance of genetic background on the effect mutations have on a phenotype (Brem et al., 2005; Romano et al., 2010; Sinha et al., 2006).

We suspect that loss of the chromosome III disomy is causing the instability of strains 2.3a and 2.3b. If this is true, loss of the disomy, and therefore the second copy of the reporter, causes strains 2.3a and 2.3b to be resistant to up to 5mM 3-AT. In contrast, the reconstructed strains are only resistant to 1mM 3-AT. This still leaves a missing component of the heritability. Strain 2.3a is also disomic for chromosome XIV; it is possible another missing component is this disomy, although that cannot explain strain 2.3b, which contains only one chromosome XIV. Possibly, the missing heritability is explained by many mutations that each contribute mildly to the phenotype, similar to the contribution of *yor019w-1659*. If different subsets of these mutations are sufficient to confer 10mM 3-AT resistance, the different combinations of causative mutations between segregants could cause causative mutations to have similar frequencies as non-causative mutations.

Strains 2.3a and 2.3b are disomic for chromosome III, resulting in a duplication of the *BPH1* reporter that presumably confers resistance to higher levels of 3-AT. This supports the idea that chromosomal aneuploidies can be advantageous when they confer a selective advantage (Tang and Amon, 2013). A recent study suggests formation of

aneuploidies as a transient method for yeast to respond to evolutionary stress (Yona et al., 2012).

In this chapter, we have identified five additional mutations that affect long-distance activation in yeast. These mutations would not have been identified using a more traditional genetic approach because all mutations require the presence of at least one other mutation in order to affect the Lda$^-$ phenotype. Additionally, by performing multiple rounds of selection on two lineages of strains, we were able to isolate a range of mutations broader than could be identified by simply isolating enhancers through only one round of selection. Two of the mutations identified result in predicted changes in binding specificity of transcription factors. How does this change in binding specificity allow long-distance activation? How does the change in binding specificity affect genome-wide transcription levels?

References

Abramova, N., Sertil, O., Mehta, S., and Lowry, C.V. (2001). Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in Saccharomyces cerevisiae. J Bacteriol *183*, 2881-2887.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds. (1991). Current Protocols in Molecular Biology (New York, Greene Publishing Associates and Wiley-Interscience).

Bailey, R.B., and Woodword, A. (1984). Isolation and characterization of a pleiotropic glucose repression resistant mutant of Saccharomyces cerevisiae. Mol Gen Genet *193*, 507-512.

Barral, Y., Jentsch, S., and Mann, C. (1995). G1 cyclin turnover and nutrient uptake are controlled by a common pathway in yeast. Genes Dev *9*, 399-409.

Ben-Ari, G., Zenvirth, D., Sherman, A., David, L., Klutstein, M., Lavi, U., Hillel, J., and Simchen, G. (2006). Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genet *2*, e195.

Berry, D.B., and Gasch, A.P. (2008). Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. Mol Biol Cell *19*, 4580-4587.

Brauer, M.J., Christianson, C.M., Pai, D.A., and Dunham, M.J. (2006). Mapping novel traits by array-assisted bulk segregant analysis in Saccharomyces cerevisiae. Genetics *173*, 1813-1816.

Brem, R.B., Storey, J.D., Whittle, J., and Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. Nature *436*, 701-703.

Cain, C.W., Lohse, M.B., Homann, O.R., Sil, A., and Johnson, A.D. (2012). A conserved transcriptional regulator governs fungal morphology in widely diverged species. Genetics *190*, 511-521.

Conklin, D.S., Kung, C., and Culbertson, M.R. (1993). The COT2 gene is required for glucose-dependent divalent cation transport in Saccharomyces cerevisiae. Mol Cell Biol *13*, 2041-2049.

Deutschbauer, A.M., and Davis, R.W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet *37*, 1333-1340.

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol Cell Biol *27*, 5575-5586.

Entian, K.D., Schuster, T., Hegemann, J.H., Becher, D., Feldmann, H., Guldener, U., Gotz, R., Hansen, M., Hollenberg, C.P., Jansen, G., *et al*. (1999). Functional analysis of 150 deletion mutants in Saccharomyces cerevisiae by a systematic approach. Mol Gen Genet *262*, 683-702.

Flick, J.S., and Johnston, M. (1991). GRR1 of Saccharomyces cerevisiae is required for glucose repression and encodes a protein with leucine-rich repeats. Mol Cell Biol *11*, 5101-5112.

Forsberg, H., Gilstring, C.F., Zargari, A., Martinez, P., and Ljungdahl, P.O. (2001). The role of the yeast plasma membrane SPS nutrient sensor in the metabolic response to extracellular amino acids. Mol Microbiol *42*, 215-228.

Forsberg, H., and Ljungdahl, P.O. (2001). Genetic and biochemical analysis of the yeast plasma membrane Ssy1p-Ptr3p-Ssy5p sensor of extracellular amino acids. Mol Cell Biol *21*, 814-826.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:12073907.

Gerke, J.P., Chen, C.T., and Cohen, B.A. (2006). Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. Genetics *174*, 985-997.

Grishin, A.V., Rothenberg, M., Downs, M.A., and Blumer, K.J. (1998). Mot3, a Zn finger transcription factor that modulates gene expression and attenuates mating pheromone signaling in Saccharomyces cerevisiae. Genetics *149*, 879-892.

Guarente, L., and Hoar, E. (1984). Upstream activation sites of the CYC1 gene of Saccharomyces cerevisiae are active when inverted but not when placed downstream of the "TATA box". Proc Natl Acad Sci U S A *81*, 7860-7864.

Hongay, C., Jia, N., Bard, M., and Winston, F. (2002). Mot3 is a transcriptional repressor of ergosterol biosynthetic genes and is required for normal vacuolar function in Saccharomyces cerevisiae. EMBO J *21*, 4114-4124.

Koschwanez, J.H., Foster, K.R., and Murray, A.W. (2013). Improved use of a public good selects for the evolution of undifferentiated multicellularity. Elife *2*, e00367.

Landry, B.D., Doyle, J.P., Toczyski, D.P., and Benanti, J.A. (2012). F-box protein specificity for g1 cyclins is dictated by subcellular localization. PLoS Genet *8*, e1002851.

Li, F.N., and Johnston, M. (1997). Grr1 of Saccharomyces cerevisiae is connected to the ubiquitin proteolysis machinery through Skp1: coupling glucose sensing to gene expression and the cell cycle. EMBO J *16*, 5629-5638.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Madison, J.M., Dudley, A.M., and Winston, F. (1998). Identification and analysis of Mot3, a zinc finger protein that binds to the retrotransposon Ty long terminal repeat (delta) in Saccharomyces cerevisiae. Mol Cell Biol *18*, 1879-1890.

Martinez-Montanes, F., Rienzo, A., Poveda-Huertes, D., Pascual-Ahuir, A., and Proft, M. (2013). Activator and Repressor Functions of the Mot3 Transcription Factor in the Osmostress Response of Saccharomyces cerevisiae. Eukaryot Cell *12*, 636-647.

Martinez-Pastor, M.T., Marchler, G., Schuller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996). The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p

are required for transcriptional induction through the stress response element (STRE). EMBO J *15*, 2227-2235.

Romano, G.H., Gurvich, Y., Lavi, O., Ulitsky, I., Shamir, R., and Kupiec, M. (2010). Different sets of QTLs influence fitness variation in yeast. Mol Syst Biol *6*, 346.

Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics 1990: A Laboratory Course Manaual (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Sinha, H., Nicholson, B.P., Steinmetz, L.M., and McCusker, J.H. (2006). Complex genetic interactions in a quantitative trait locus. PLoS Genet *2*, e13.

Struhl, K. (1984). Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. Proc Natl Acad Sci U S A *81*, 7865-7869.

Tang, Y.C., and Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. Cell *152*, 394-405.

Wenger, J.W., Schwartz, K., and Sherlock, G. (2010). Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from Saccharomyces cerevisiae. PLoS Genet *6*, e1000942.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53-55.

Witten, J.T., Chen, C.T., and Cohen, B.A. (2007). Complex genetic changes in strains of Saccharomyces cerevisiae derived by selection in the laboratory. Genetics *177*, 449-456.

Xue, W., Wang, J., Shen, Z., and Zhu, H. (2004). Enrichment of transcriptional regulatory sites in non-coding genomic region. Bioinformatics *20*, 569-575.

Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. Proc Natl Acad Sci U S A *109*, 21010-21015.

Zordan, R.E., Galgoczy, D.J., and Johnson, A.D. (2006). Epigenetic properties of white-opaque switching in Candida albicans are based on a self-sustaining transcriptional feedback loop. Proc Natl Acad Sci U S A *103*, 12807-12812.

Chapter 3

Additional insight into the role of polygenic mutants in long-distance activation

Attribution of experiments in Chapter 3

Combinatorial mutant analysis and reporter distance experiments were performed by Caitlin Reavey. RNA samples for microarrays were prepared by Caitlin Reavey and microarrays were performed by Caitlin Reavey in collaboration with Dr. Patrick Gibney and Dr. David Botstein. Clustering and GO-term analysis were performed by Caitlin Reavey. Statistical analysis of expression change overlap between strains and statistical analysis of tandem adjacent gene pairs were performed by Dr. Burak Alver. Verification of tandem gene pair expression and analysis of *YBR281C/SAF1* were performed by Caitlin Reavey.

Introduction

To understand how transcriptional activation distance is regulated in yeast, we isolated polygenic mutants that are able to activate transcription over a longer distance than is normally non-permissive in yeast. The isolation and initial analysis of these mutants raised several questions, some of which we address in Chapter 3. In these experiments, we further investigate the transcriptional effects of the individual causal mutations isolated in our polygenic mutants. We find that combining mutations isolated from the different lineages gives a range of phenotypes. In addition, we show that several mutant strains are able to activate transcription at distances up to at least 2 kb. Finally, we performed microarray analysis to analyze the global transcriptional affects of mutants. By this analysis, all strains display transcriptional changes and, surprisingly, there is a significant overlap in the changes between strains. Furthermore, the microarray results suggest that long-distance activation may be occurring at endogenous genes in addition to our reporter, as pairs of genes are affected at a significant frequency.

Materials and methods

S. cerevisiae strains

The *S. cerevisiae* strains used in this study (Table 3-1, 3-2) are isogenic with a *GAL2*$^{+}$ derivative of S288C (Winston et al., 1995). As is Chapter 2, strains are distinguished by those that were not subject to selection (Table 3-1) and those that were

(Table 3-2). Rich (YPD) and synthetic complete (SC) dropout media were prepared as previously described (Rose et al., 1990). SC Gal and SC-His Gal contained 2% galactose as the carbon source. YPRaffinose contained 2% raffinose as the carbon source. Specified concentrations of 3-aminotriazole were added to SC-His Gal medium. Strains were constructed by standard methods, either through crosses or transformation (Ausubel et al., 1991).

Northern Analysis

RNA isolation and Northern hybridization experiments were performed as previously described (Ausubel et al., 1991). Strains were grown to mid-log in YP Raffinose and then shifted to 2% galactose for 1 hour. Northern hybridization analysis was conducted with probes to the coding regions of HIS3 (-27 to +376, where +1 is the ATG), *YBR281C*(+2102 to 2450), *SAF1* (+310 to 629), ACT1(+533 to +722) and *SNR190* (+1 to +190) (Table 3-3).

Culturing of strains

Two replicates of wild-type (CR219), 1.2(CR101), 2.3a (CR109), 2.3b (CR110), *sin4Δ mot3-1162 grr1-531* (CR204), and *sin4Δ mot3-1162 mit1-560* (CR205) were grown in YP Raffinose to an OD_{600} of 0.4. At this point a sample was taken and cells were washed and frozen in a dry ice ethanol bath. The remaining culture was split in two and shifted to either 2% glucose or 2% galactose for 1 hour. At the conclusion of the

Table 3-1. Strains used in this study not subjected to selection

Strain	Lineage Alias	Genotype
FY76		*MAT***a** lys2-128δ
FY2574		*MATα his3Δ200 lys2-128d leu2Δ0 trp1Δ63 bph1Δ::kanMX-UASGAL1397-HIS3 sin4Δ0::LEU2*
CR111		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806-URA3-natMX mkc7-1307*
CR113		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806*
CR204		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531*
CR205		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 mit1-560*
CR213		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 mit1-560 msn2-1956*
CR214		*MATα his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806*
CR234		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 msn2-1956*
CR236		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mit1-560 msn2-1956*
CR238		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 put4Δ0*
CR242		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 put4Δ0*
CR243		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$1397*
CR245		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$1397 mot3-1162 grr1-531*
CR246		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$1397 mot3-1162 mit1-560*

Table 3-1. (Continued)

CR247	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-UAS$_{GAL1}$2027*
CR248	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027*
CR250	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 mot3-1162 grr1-531*
CR251	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 mot3-1162 mit1-560*

Table 3-2. Selected strains and their derivatives used in this study

Strain	Lineage Alias	Genotype
CR101	1.2	*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251*
CR109	2.3a	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 mit1-560 ptr3-1088 mkc7-1307 yor019w-1659*
CR110	2.3b	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 msn2-1956 ptr3-1088 mkc7-1307 yor019w-1659*
CR240		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251 put4Δ0*
CR241		*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$799 ybr281c Δ::TRP1-UAS$_{GAL1}$806 mit1-560 ptr3-1088 mkc7-1307 yor019w-1659 put4Δ0*
CR244		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$1397 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251*
CR249		*MATα his3Δ200 ura3Δ0 leu2Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{GAL1}$2027 ybr281cΔ::TRP1-UAS$_{GAL1}$806 mot3-1162 grr1-531 sgm1-1220 rim8-155 tma108-1374 sgf73-251*

Table 3-3. Oligonucleotides used in this study

Oligo	Purpose	Sequence (5' to 3')
FO201	Northern probe of *HIS3*	TGAGCAGGCAAGATAAAC
FO609	Northern probe of *HIS3*	GCCTCATCCAAAGGCGC
FO481	Northern probe of *URA3*	CCATGGAGGGCACAGTTAAGCCGC
F0483	Northern probe of *URA3*	CCCTTCCCTTTGCAAATAGTCCTC
FO1324	Northern probe of *SNR190*	GGCCCTGATGATAATG
FO1325	Northern probe of *SNR190*	GGCTCAGATCTGCATG
FO961	Northern probe of *ACT1*	TGTCACCAACTGGGACGATA
FO962	Northern probe of *ACT1*	GGCTTGGATGGAAACGTAGA
CRO416	Replace *kanMX-GAL1* with *URA3*	ATTCACAACTTTGGTCAAACGCCTTTACAAA TATTTCAGGAGATTGTACTGAGAGTGCAC
CRO417	Replace *kanMX-GAL1* with *URA3*	AAGATTGTCTTCTCAAATATTGGCTTCATTG GAACCTTACCTGTGCGGTATTTCACACCG
FO6666	Integrate *kanMX-GAL1* at 2027 bp	CGCAAGAATCACGGGGATATGACGGTTAGC TGAATTCGAGCTCGTTTAAAC
FO6667	Integrate *kanMX-GAL1* at 2027 bp	AGTTTCCAAACAAAGACTTCGTGCTTTAGG TCATCGCTTCGCTGATTAATTACCC
CRO430	Delete *PUT4* promoter with *URA3*	GATTGAAGGGTGTAAAGTGCGTGTGGTGGC GTTCTTTCCAAGATTGTACTGAGAGTGCAC
CRO431	Delete *PUT4* promoter with *URA3*	TGTGTCTATTGTTCTTGTGGAAGGGCAGTAT ATTTACCATCTGTGCGGTATTTCACACCG
CRO440	*TDA6* RT primer	GGTATTCCGCCTTCAAGTCA

Table 3-3. (Continued)

CRO441	*TDA6* RT primer	CCGAATACGGAACAGGCTAC
CRO444	*CUR1* RT primer	TCCACCCCTTCGAGAGAATA
CRO445	*CUR1* RT primer	TCTGCAATGAGTTGGCATGT
CRO448	Clean deletion of *PUT4* promoter	TATTGTTCATGATTGAAGGGTGTAAAGTGCG TGTGGTGGCGTTCTTTCCAATGGTAAATA
CRO449	Clean deletion of *PUT4* promoter	ACTCCCGCGCTGTGTCTATTGTTCTTGTGGA AGGGCAGTATATTTACCATTGGAAAGAAC
FO6145	*ACT1* RT primer	TTTTGTCCTTGTACTCTTCC
FO6146	*ACT1* RT primer	CTGAATCTTTCGTTACCAAT
CRO450	*PUT4* RT primer	CGAGCCGCACAAACTAAAAC
CRO451	*PUT4* RT primer	ATGAAGCGTGGATGAAGTCC
CRO454	*PYK2* RT primer	GTTATCGTTCCGGGGAGATT
CRO455	*PYK2* RT primer	TAACCCGAGTTTACCGCTTG
CRO460	*BDS1* RT primer	TAGGGAAAGCTGCCTCTCAC
CRO495	*NIT1* RT primer	CAAAGTTCGATCCCTTTGGA
CRO496	*YIL165C* RT primer	AGATTATTGCAGGGCCATTG
CRO497	*YIL165C* RT primer	AATGTCCGACAGGGTCAAGA
CRO498	*DLD3* RT primer	CCCATTGGATCTGCCTTCTA
CRO499	*DLD3* RT primer	ATCTCACCGTTGGGTAGCAC
CRO500	*DSF1* RT primer	GAAAAGATGGCCAATCCAGA
CRO501	*DSF1* RT primer	GCTTTTCTGGGTGGTTCAAA
CRO506	TIP1 RT primer	TCATCATCTGCCGAATCATC
CRO507	TIP1 RT primer	AACAACAGCACCGAAAGAGG

Table 3-3 (Continued)

CRO508	NRG2 RT primer	GGGCTGTGGACAGAGGTTTA
CRO509	NRG2 RT primer	CTGCTAGCCTCCCTCCTCTT
CRO510	Northern probe of *YBR281C*	CCGAGCTTGCAAATATCGAC
CRO511	Northern probe of *YBR281C*	GGTTCTACGTCCCATGCAGT
CRO504	Northern probe of *SAF1*	TTACGGGCATATGATGCAAA
CRO512	Northern probe of *SAF1*	GATCCGCTGCTGTAAAGGTT

hour, cells were washed and frozen. All three samples were then used for RNA extraction.

Microarrays

RNA was extracted from cells grown as described above, labeled, and hybridized to Agilent expression microarrays as previously described (McIsaac et al., 2011). RNA was extracted using standard phenol-chloroform extraction and purified with RNEasy columns (Qiagen). mRNA was converted to cDNA, then cRNA, using Agilent Quick-Amp Labeling Kit (Part No. 5190-0424). Reference RNA was labeled with Cy3 and experimental RNA was labeled with Cy5. Labeled cRNA was hybridized to Agilent 8x15k custom printed yeast arrays (AMADID 017566) for 17h at 65° on a rotisserie at 20 rpm. Each gene on the array contains 2-3 identical probes. RNA from CR219 cultured in YP Raffinose was used for the reference for all arrays. After washing, arrays were scanned using Agilent Feature Extractor Software version 9.5. Genes that had flagged features marked as unreliable were excluded from data analysis; this resulted in the analysis of 5610 genes. The raw signal intensity values were floored to a value of 350 (values <350 were set to 350). Flooring the data makes genes with very low signal intensity less sensitive to small fluctuations in reference signal intensity. The log2 ratio was calculated using the floored data values.

For each microarray, an experimental sample was compared to RNA from the wild-type strain CR219. Each microarray was performed twice for wild-type (CR219),

1.2 (CR101), 2.3a (CR109), 2.3b (CR110), *sin4Δ mot3-1162 grr1-531* (CR204), and *sin4Δ mot3-1162 mit1-560* (CR205). The wild-type array values were used as the reference array and the log2 values for each mutant array were normalized to the reference array. Analysis of the log2 ratios revealed that one of the 1.2 replicates and one of the 2.3a replicates were unusable, so these were discarded from analysis. These replicates varied from the usable replicates and more specifically, these replicates did not show induction of any of the GAL genes in the galactose samples. For the remaining replicates, the log2 ratios were averaged and these averages were used for subsequent analysis. The complete list of Log2 ratios normalized to the wild-type raffinose reference array are found electronically in Table 3-9. The Log2 ratios of the mutant samples shifted to glucose were normalized to wild-type glucose and the Log2 ratios of the mutant samples shifted to galactose were normalized to wild-type galactose; these values can be found electronically in Table 3-10.

Microarray analysis

Analysis of RNA levels was performed for all carbon sources. Changes in the mutants in glucose and galactose were normalized to wild-type expression changes for glucose and galactose, respectively. We used a two-fold cutoff for calling genes as up or downregulated (log2 value of 1 or -1). Hierarchical clustering was performed using Cluster 3.0 with average linkage using the Pearson correlation distance as the metric of similarity between genes (de Hoon et al., 2004). K-means clustering was performed with MultiExperiment Viewer using a setting of 10 clusters and using Euclidean distance as

the metric of similarity between genes (Saeed et al., 2006; Saeed et al., 2003). GO-term enrichment was performed using the Saccharomyces Genome Database Gene Ontology Term Finder. Overlap of upregulated and downregulated genes between different strains was compared by Venn diagrams, which were created using the Whitehead Venn diagram generator (http://jura.wi.mit.edu/bioc/tools/). Fisher's exact test was used to calculate the p-value for the Venn diagram analysis. All p-values were lower than the numerical precision of the hypergeometric function implementation in Microsoft Excel, which is 10^{-14}.

Analysis of tandem adjacent gene pairs

To identify possible cases of long-distance activation in the genome, we examined the microarray data to look for possible cases where the activator for one gene was able to activate an adjacent, tandem gene. We considered two possible circumstances. In the first circumstance (case 1), the expression of both genes in the pair is elevated in the mutant compared to wild-type. In the second circumstance (case 2), the 5' gene is activated in either glucose or galactose (relative to raffinose) in wild-type and, in the mutant, the 3' gene is also activated.

Statistical analysis of the tandem adjacent gene pairs was performed on log2 values that had been normalized to wild-type by respective carbon source. Neighboring genes represented on the array with no other annotated transcripts in between were annotated as gene pairs, with the one with the smaller genomic coordinate denoted the

first gene and the one with the larger coordinate denoted the second gene. For all gene pairs in tandem the null model expected number of co-upregulated pairs (both >two fold change) was calculated as the number of upregulated first genes, multiplied by the number of upregulated second genes divided by the number of tandem adjacent gene pairs. The number of tandem adjacent gene pairs was assumed to display a Poisson distribution, with the expected number of co-upregulated pairs the mean. The p-value for the actual observation was calculated based on this Poisson distribution with parameter lambda given by the null expectation.

Quantitative real time PCR (qRT-PCR) analysis

RNA was converted to cDNA using Supercript III reverse transcriptase (Invitrogen) and oligo dT primer (Invitrogen). cDNA was analyzed using the primers listed in table 3-. and Brilliant III SYBR green reagent (Agilent). qRT-PCR was performed on a Stratagene MX3000P. The following gene pairs were analyzed after being shifted to glucose: *TDA6/CUR1*, *DLD3/DSF1*, *BDS1/YOL163W, GPD2/ARG1, NIT1/YIL165C, TIP1/NRG2*. The *PUT4/PYK2* gene pair was analyzed after being shifted to galactose. The values presented are the average of two experiments.

Results

Combinatorial analysis of transcription factor mutations

To test for genetic interactions between mutations isolated in the different polygenic mutants, we constructed a new series of mutant strains and tested them for activation of the long-distance reporter. We focused on combinations of the *mot3, mit1*, and *msn2* mutations, each of which arose in a separate lineage. Our results (Figure 3-1) show that combining the mutations in a *sin4Δ* background increases 3-AT resistance, ranging from a modest effect, for *sin4Δ mit1-531 msn2-1956,* to an intermediate effect for *sin4Δ mot3-1162 mit1-531*, to a strong effect for *sin4Δ mot3-1162 msn2-1967* and *sin4Δ mot3-1162 mit1-560 msn2-1967*. We compared *HIS3* mRNA levels of *sin4Δ mot3-1162 mit1-531* and *sin4Δ mot3-1162 mit1-560 msn2-1967* by Northern analysis*;* both mutants show increased HIS3 mRNA levels compared to the *sin4Δ* single mutant (Fig. 3-1).

Effect of increasing reporter distance

To test if the mutant strains are able to activate transcription at a distance greater than in our reporter, 799 bp, we constructed and tested longer reporters of 1397 bp and 2027 bp, testing these differences in three of the multiple mutants. All three strains tested show a stronger Lda$^-$ phenotype relative to *sin4Δ* at these greater distances, as shown

Figure 3-1. Lda$^-$ phenotypes of mutant combinations. **A.** Growth of combined mutants. The strains are: *SIN4*$^+$ (CR214), *sin4Δ* (CR111), *sin4Δ mot3 grr1* (CR204), *sin4Δ mot3 mit1* (CR205), *sin4Δ mot3 msn2* (CR234), *sin4Δ mit1 msn2* (CR236), and *sin4Δ mot3 mit1 msn2* (CR213). 10-fold serial dilutions were made of saturated YPD cultures that had been normalized for cell number by $OD6_{00}$. The dilutions were spotted to the specified media and grown for four days at 30°. **B.** Northern analysis of combined mutants. A Northern blot was hybridized with probes for *HIS3* (top panel) and *SNR190* (bottom panel). Lanes are: lane 1 (FY76), lanes 2 and 3 (CR214), lanes 4 and 5 (1.2), lanes 6 and 7 (2.3a), lanes 8 and 9 (2.3b), lanes 10 and 11 (CR204), lanes 12 and 13 (CR205), and lanes 14 and 15 (CR213). Strains were shifted from growth in 2% raffinose to 2% galactose.

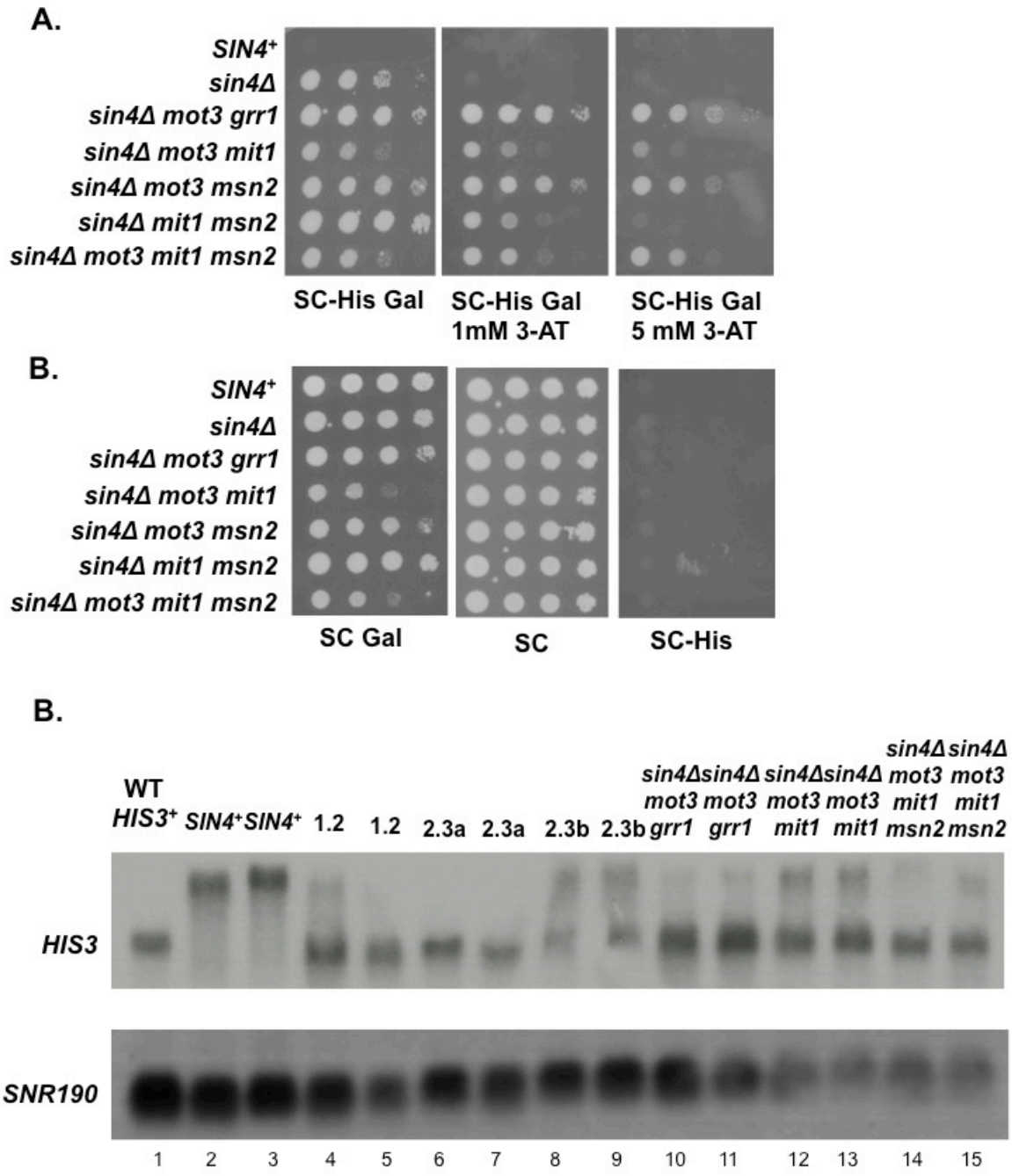


Figure 3-1. (Continued)

by growth on SC-His Gal for both the 1397 and 2027 bp reporters (Fig. 3-2). Surprisingly, strains 1.2, *sin4Δ mot3-1162 grr1-531,* and *sin4Δ mot3-1162 mit1-560* display a stronger Lda$^-$ phenotype with the 2027 reporter strain than with the 1397 reporter strain. This is in contrast to the *sin4Δ* single mutant, which shows a weaker Lda$^-$ phenotype with increased reporter distances.

Genome-wide expression analysis of the polygenic mutants

To determine possible changes in transcription genome-wide, we performed microarrays. For these studies, we included three classes of polygenic strains. First, we analyzed the three original polygenic mutants, 1.2, 2.3a, and 2.3b. In addition, we analyzed the reconstructed strain *sin4Δ mot3-1162 grr1-531*, which recapitulates the 1.2 phenotypes with our long-distance reporter. Finally, we included *sin4Δ mot3-1162 mit1-560* to analyze a strong polygenic mutant that combines mutations from different lineages. To increase the scope of the analysis, we determined changes in transcription for these strains grown in three carbon sources: raffinose, glucose, and galactose.

Our microarray analysis revealed extensive transcriptional changes in the mutant strains in all three carbon sources (Table 3-4). For all strains, a large number of genes had increased RNA levels and a large number had decreased RNA levels, demonstrating that the Lda$^-$ phenotype is not a general upregulation of transcription.

Figure 3-2. Polygenic mutants strengthen activation at distances up to 2 kb. **A.** Growth of strains with the 799 bp reporter. The strains are: *SIN4*$^{+}$ 799(CR214), *sin4Δ* 799 (CR111), 1.2 (CR101), *sin4Δ mot3 grr1* 799(CR204), and *sin4Δ mot3 mit1* 799(CR205). Ten-fold serial dilutions of saturated YPD cultures were spotted to the specified media and grown for four days at 30°. **B.** Growth of strains with the 1397 bp reporter. The strains are: *SIN4*$^{+}$ 799(CR214), *sin4Δ* 799 (CR111), *SIN4*$^{+}$ 1397(CR243), *sin4Δ* 1397 (FY2574), 1.2 1397 (CR244) *sin4Δ mot3 grr1* 1397 (CR245), and *sin4Δ mot3 mit1* 1397(CR246). Plates were incubated for four days or six days (the one indicated) at 30°. **C.** Growth of strains with 2027 bp reporter. The strains are: *SIN4*$^{+}$ 799(CR214), *sin4Δ* 799 (CR111), *SIN4*$^{+}$ 2027 (CR247), *sin4Δ* 2027 (CR248), 1.2 2027 (CR249) *sin4Δ mot3 grr1* 2027 (CR250), and *sin4Δ mot3 mit1* 2027 (CR251). Plates were incubated for four or six days at 30°.

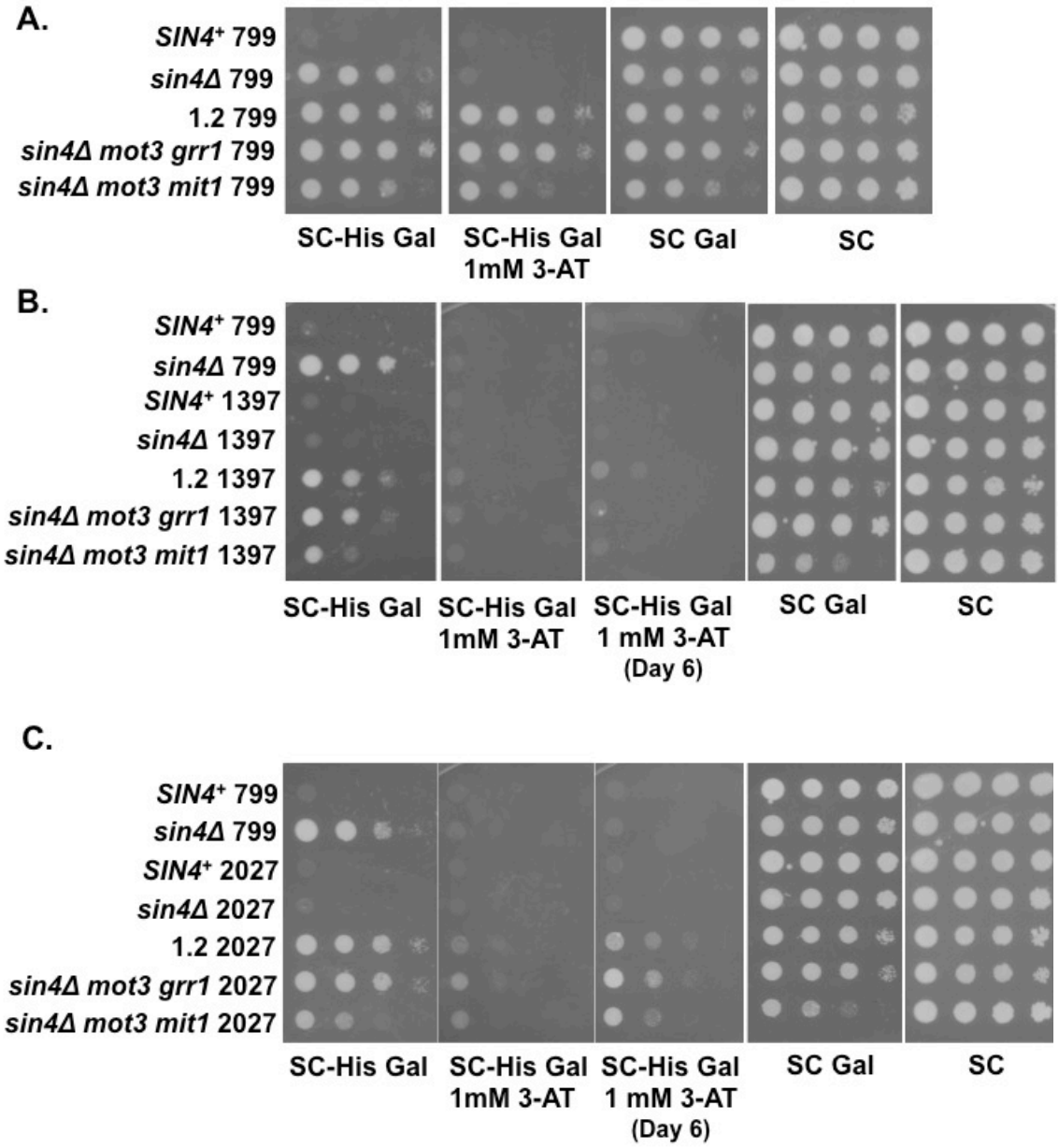


Figure 3-2. (Continued)

Table 3-4. Numbers of genes displaying a two-fold or greater change

Carbon Source	1.2	2.3a	2.3b	*sin4Δ mot3-1162 grr1-531*	*sin4Δ mot3-1162 mit1-560*
Raffinose	313 up 135 down	180 up 123 down	172 up 141 down	152 up 84 down	86 up 50 down
Glucose	416 up 158 down	145 up 62 down	167 up 82 down	182 up 79 down	38 up 37 down
Galactose	474 up 426 down	146 up 78 down	184 up 133 down	133 up 139 down	82 up 40 down

Hierarchical clustering of the microarray results revealed several general trends among the data sets. First, there is a noticeable overlap in the expression changes of genes for the mutants between carbon sources (Fig. 3-3). As expected, strains 2.3a and 2.3b, which arose from the same lineage, display similar profiles of expression changes. However, there are also clusters of genes with similar expression changes between most of the mutants.

We performed k-means clustering to identify clusters of genes that are either upregulated or downregulated in the majority of mutants (Fig. 3-4). Using these two clusters, we performed GO-term enrichment to determine if any of the GO-terms could give insight into particular classes of genes that are affected, perhaps identifying regulators now able to activate at an increased distance. The upregulated genes (Table 3-5) are enriched for genes involved in sulfur metabolism and a variety of biosynthetic processes. The downregulated genes are enriched for genes in serine catabolism and amino acid transport (Table 3-6). Although not enriched by GO-term, we did notice the mutants are upregulated for multiple arginine biosynthetic genes and proline utilization genes. A recent proteomic study indicated proteins in these pathways are upregulated in response to various stress conditions (Grady, 2013). Unfortunately the GO-terms did not give any mechanistic insight to long-distance activation.

Lineage 1 and lineage 2 mutations were isolated independently of one another, yet have similar effects on the Lda$^-$ phenotype, so we were interested in transcriptional changes common among 1.2, 2.3a, and 2.3b. We compared genes that showed a two-fold

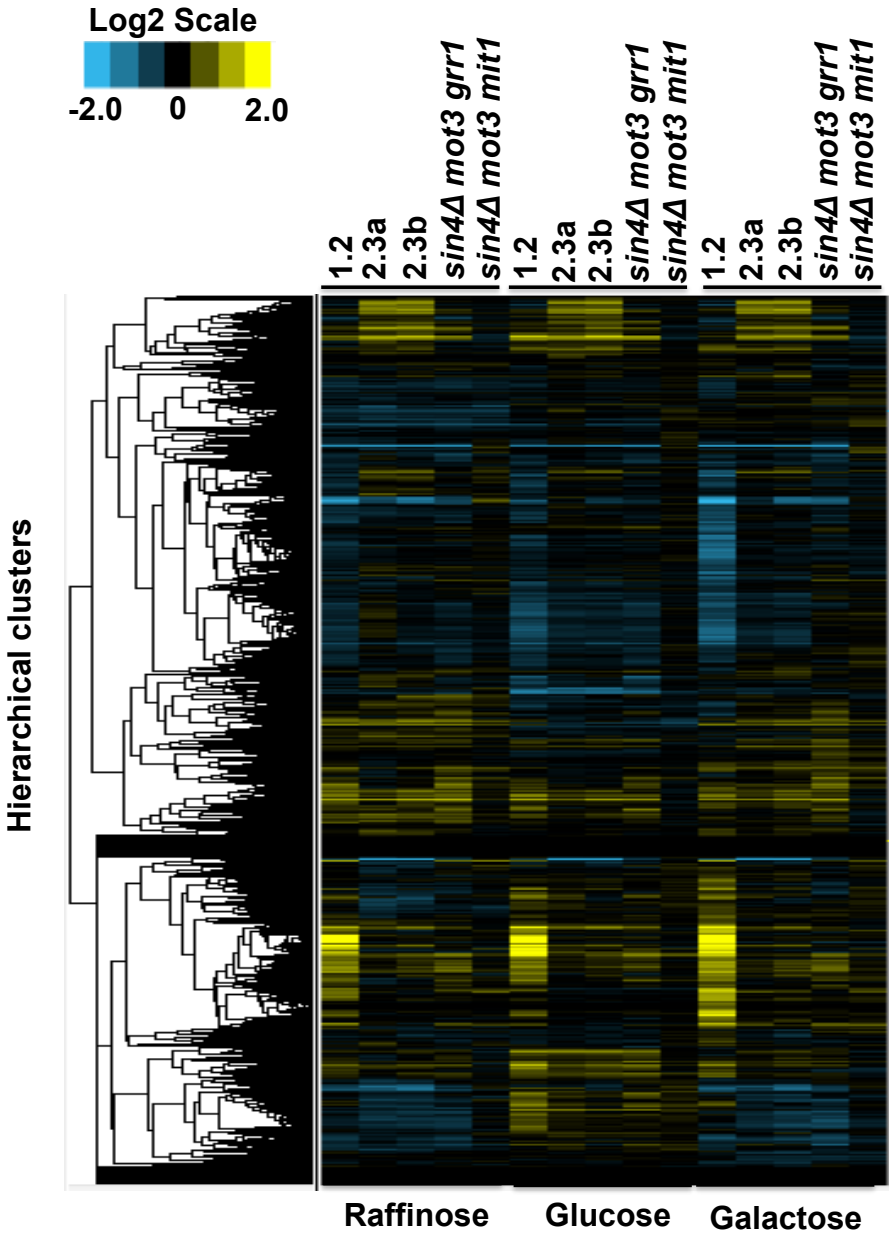


Figure 3-3. Analysis of transcriptional changes by microarray. Hierarchical clustering of gene expression changes was performed on the microarray datasets from mutant strains cultured in raffinose, glucose, and galactose. The strains are 1.2 (CR101), 2.3a (CR109), 2.3b (CR110), *sin4Δ mot3 grr1* (CR204), and *sin4Δ mot3 mit1* (CR205). Values used for clustering were normalized to expression of $SIN4^{+}$ (CR214).

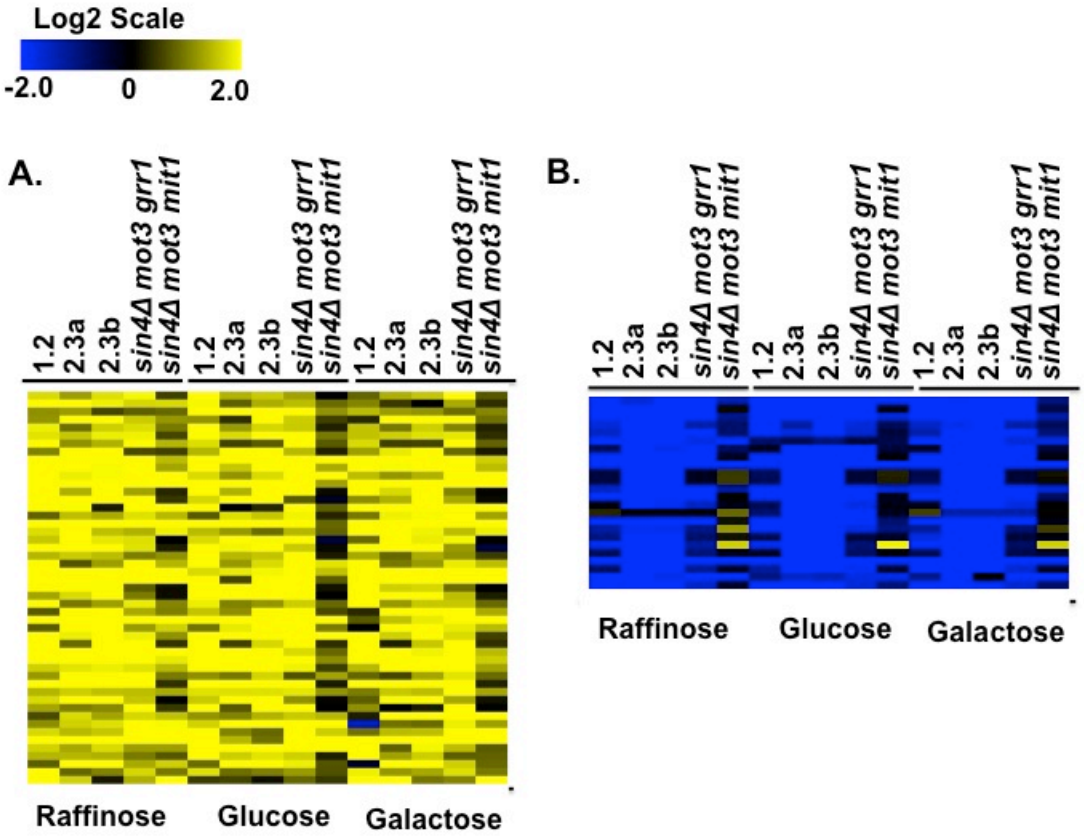


Figure 3-4. Identification of genes upregulated or downregulated in mutants. K-means clustering was performed using a parameter of 10 clusters. The strains are: 1.2 (CR101), 2.3a (CR109), 2.3b (CR110) *sin4Δ mot3 grr1* 799(CR204), and *sin4Δ mot3 mit1* 799(CR205). Values used for clustering were normalized to expression of $SIN4^{+}$ 799(CR214). **A.** Clustering of 47 genes upregulated two-fold or more across majority of mutants. **B.** Clustering of 23 genes downregulated two-fold or more across majority of mutants.

Table 3-5. GO-Term enrichment for genes upregulated in mutants

GO-Term	Cluster Frequency*	Background Frequency*	P-value
Sulfur amino acid metabolic process	7.4%	0.5%	6.06 x 10-09
Sulfur compound metabolic process	21.7%	1.3%	3.46 x 10-08
Sulfate assimilation	10.9%	0.1%	3.23 x 10-07
Organic acid metabolic process	32.6%	4.9%	3.60 x 10-07
Organonitrogen compound metabolic process	37.0%	7.5%	2.01 x 10-06
Carboxylic acid metabolic process	30.4%	4.8%	2.19 x 10-6
Cellular amino acid metabolic process	26.1%	3.4%	3.71 x 10-6
Sulfate reduction	6.5%	0.0%	3.73 x 10-5
Small molecule metabolic process	34.8%	9.3%	0.00028
Sulfur amino acid biosynthetic process	8.7%	0.3%	0.00081
Organic acid biosynthetic process	17.4%	2.4%	0.00155
Carboxylic acid biosynthetic process	17.4%	2.4%	0.00155
Single-organism metabolic process	39.1%	13.4%	0.00178
Cellular amino acid biosynthetic process	15.2%	1.8%	0.00259
Serine family amino acid metabolic process	8.7%	0.4%	0.0041
Small molecule biosynthetic process	19.6%	3.9%	0.00775
Single-organism biosynthetic process	19.6%	3.9%	0.0082

*Cluster frequency is defined as the percentage of genes in the cluster with the indicated function. Background frequency is the percentage of genes in the genome with the indicated function

Table 3-6. GO-Term enrichment for genes downregulated in mutants

GO-Term	Cluster Frequency	Background Frequency	P-value
Serine catabolic process	18.8	0.1	1.03×10^{-5}
Amino acid transport	25.0	0.7	1.90×10^{-4}

Mating factor genes were excluded from analysis, as they were down as a result of mating type difference between reference and experimental strains.

or greater change relative to wild-type for each of the carbon sources in 1.2, 2.3a, and 2.3b (Fig. 3-5). Each strain displays a significant overlap (p-value $<10^{-14}$) of upregulated and downregulated genes with the other strains for all carbon sources. The overlap between 2.3a and 2.3b is greater than the overlap between 1.2 and 2.3a. The overlap between 1.2 and 2.3b is more dramatic than the overlap between 1.2 and 2.3a. This may be a result of overlapping functions of Mot3 and Msn2, as both are involved in activation of osmotic stress response genes (Martinez-Montanes et al., 2013).

We compared the genes that show two-fold or greater changes in RNA levels between the original polygenic mutant, 1.2, and the strain reconstructed from its causal mutations, *sin4Δ mot3-1162 grr1-531* (Fig. 3-6). Our results show that a larger number of genes have changed RNA levels in 1.2 than in the reconstructed strain, suggesting that 1.2 contains additional mutations that contribute to transcriptional changes, but that do not necessarily affect expression of our long-distance reporter. In spite of the difference in the number of genes affected, there is a significant (p-value $<10^{-14}$) overlap between 1.2 and the reconstructed strain. Interestingly, a subset of genes is changed in *sin4Δ mot3-1162 grr1-531* that remains unchanged in 1.2. This suggests the presence of genetic modifiers in 1.2 that are not present in the reconstructed strain.

We also asked whether known targets of Mot3 and Mit1 regulation are altered in our mutant strains. The ergosterol biosynthetic genes *ERG2*, *ERG6*, and *ERG9* are increased in *mot3Δ*, but remain unchanged in *mot3-1162* strains (Hongay et al., 2002).

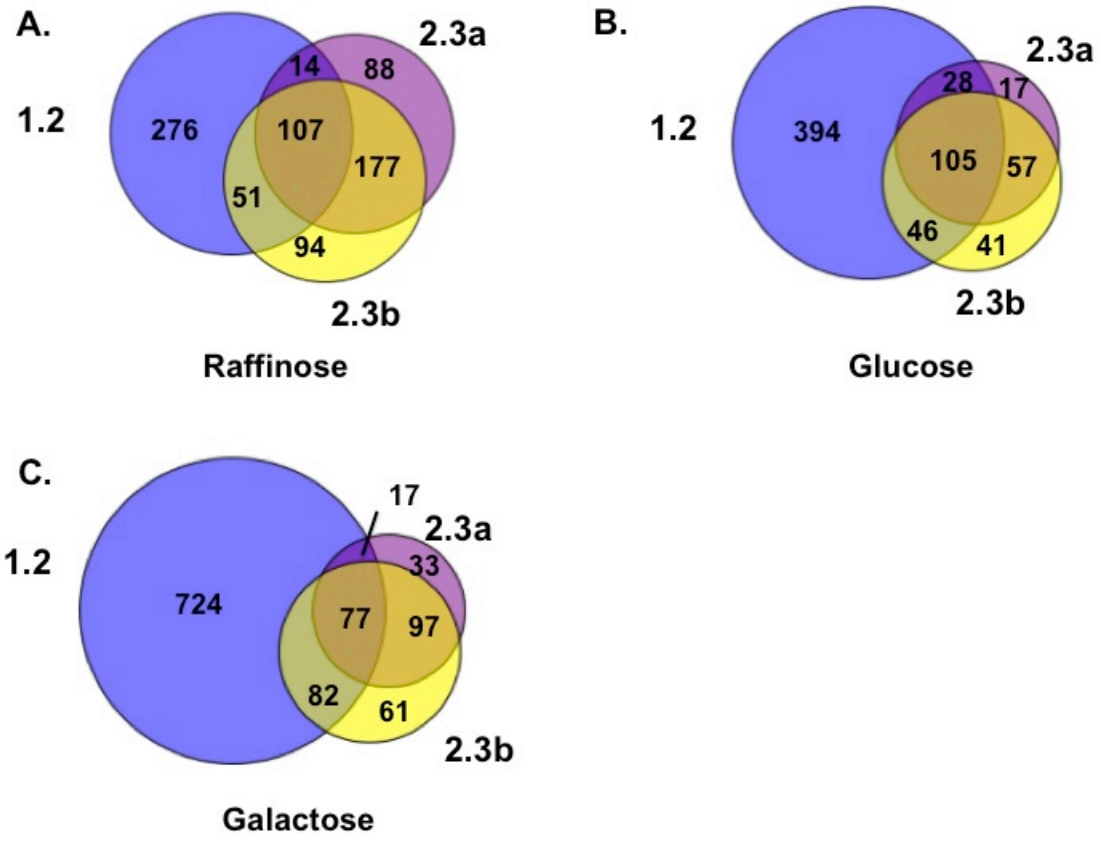


Figure 3-5. Comparison of genes changed two-fold or more in original mutant isolates. Venn diagrams depict the overlap in the genes that are upregulated and downregulated in the original mutant isolates. The strains are in blue 1.2 (CR101), in purple 2.3a (CR109), and in yellow 2.3b (CR110). Analysis was performed for raffinose **(A.)**, glucose **(B.)**, and galactose **(C.)**.

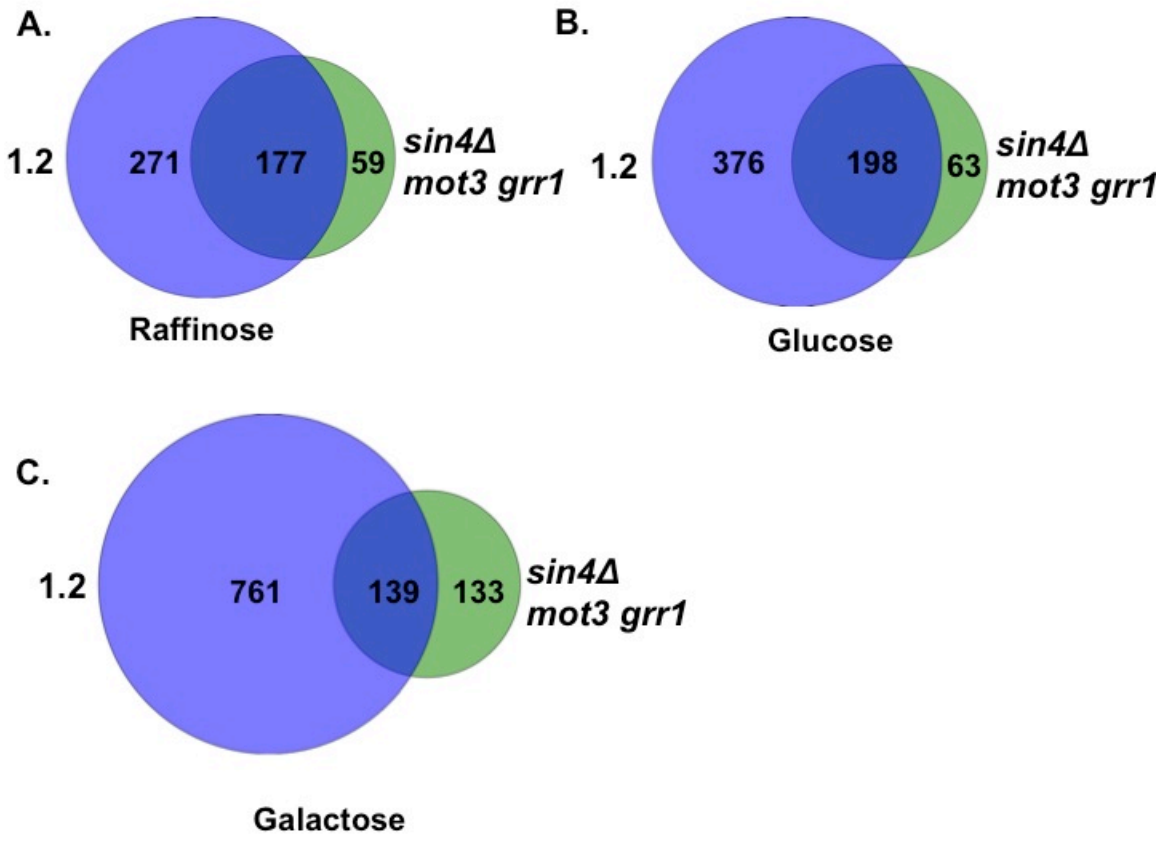


Figure 3-6. Comparison of genes changed two-fold or more in the original mutant (1.2) isolate and the reconstructed strain, *sin4Δ mot3 grr1*. Venn diagrams depict the overlap in the total number of genes that are either upregulated or downregulated in the mutants. Analysis was performed for raffinose **(A.)**, glucose **(B.)**, and galactose **(C.)**.

FLO11, which is downregulated in the *mit1Δ* strain is significantly up in all mutant strains, including those containing *mit1-560* (Cain et al., 2012). All other genes displaying a two-fold or greater change in the *mit1Δ* remain unchanged in *mit1-560* strains. These results suggest the *mot3-1162* and *mit1-560* mutant proteins allow normal regulation of some of their known targets.

We were also curious if the genes containing mutations themselves display changes in expression, as several of the mutations identified are loss-of function-mutations. *GRR1*, *MIT1*, *PTR3*, *MSN2*, and *YOR019W* RNA levels remain unchanged in the mutant strains. *MOT3* expression is down two to three-fold in both 1.2 and 2.3b in all carbon sources as well as in *sin4Δ mot3-1162 grr1-531* in galactose. *MOT3* expression remains unaffected in *sin4Δ mot3-1162 mit1-560*. The lowered *MOT3* expression supports the hypothesis that *mot3-1162* affects the Lda^- phenotype through loss-of-function as well as altered function. We speculate loss of *MOT3* expression in strain 2.3b may contribute to the Lda^- phenotype.

Analysis of adjacent gene pairs

One of the goals of the microarray experiments was to determine if long-distance activation occurs in the genome beyond the reporters. We first looked specifically at genes downstream of galactose-activated genes, but did not see aberrant galactose-dependent activation. Then, to identify candidate regions for long-distance activation, we identified tandem adjacent gene pairs whose RNA levels are increased comparison to

wild-type. We looked for gene pairs where both genes have increased levels (called class 1) and for gene pairs where the upstream gene is normally elevated in glucose or galactose in wild-type, but the downstream gene is only activated under the same condition in the mutant (class 2). Of the 70 gene pairs identified, 66 fell into class 1.

A total of 70 tandem adjacent gene pairs are elevated two-fold or more in comparison to wild-type in at least one of the mutants, for at least one carbon source. A subset of these pairs is presented in Table 3-7 and the complete set is presented in Table 3-11. The subset in Table 3-7 fits two requirements: the gene pairs are up in at least two mutants, one of which is a reconstructed mutant. The number of upregulated tandem adjacent gene pairs is significantly greater than one would expect by chance, with p-values lower than 0.05 (Table 3-8). The most striking p-values were observed for strains 2.3a, 2.3b, 1.2, and *sin4Δ mot3-1162 grr1-531.* There were no significant increases in convergently or divergently transcribed gene pairs. These results suggest long-distance activation may be occurring at other genomic loci beyond the reporter.

Verification of adjacent gene pair expression by RT PCR

We chose a set of eight tandem adjacent gene pairs based on the microarrays to verify by quantitative RT-PCR. For this set of eight gene pairs, we compared RNA levels in 2.3a, 2.3b, and *sin4Δ mot3-1162 grr1-531* to the wild-type strain (Fig. 3-6). We found that the qRT-PCR of *DLD3/DSF1, BDS1/YOL163C, GPD2/ARG1, TIP1/NRG2,* and *PUT4/PYK2* follows the same trends of expression as the microarray, although there

Table 3-7. Tandem adjacent gene pairs identified from microarray analysis

+ Indicates gene pair belongs to class 1
++Indicates gene pair belongs to class 2

Gene Pair	Carbon Source	1.2	2.3a	2.3b	*mot3 grr1 sin4Δ*	*mot3 mit1 sin4Δ*
TIP1/NRG2	Raffinose	+	+	+	+	
	Glucose	+			+	
	Galactose			+		
YBR281C/SAF1	Raffinose					
	Glucose					
	Galactose	++	++	++	++	++
GPD1/GPM2	Raffinose	+				
	Glucose					
	Galactose	+			+	
DLD3/DSF1	Raffinose	+	+		+	
	Glucose	+			+	
	Galactose			+	+	
PAU13/ARN2	Raffinose	+			+	
	Glucose	+			+	+
	Galactose	+			+	
*NIT1/YIL165C**	Raffinose	+		+		
	Glucose	+				
	Galactose	+	+	+		
DAL4/DAL2	Raffinose					
	Glucose	+	+	+	+	
	Galactose					
DAL3/DAL7	Raffinose	+	+	+		
	Glucose	+	+	+	+	
	Galactose	+	+	+	+	
YKL151C/GPM1	Raffinose	+				
	Glucose					
	Galactose			+		+
*GPD2/ARG1***	Raffinose			+		
	Glucose	++	++	++	++	
	Galactose					
BDS1/YOL163W (Incomplete array data)	Raffinose					
	Glucose	+	+		+	
	Galactose					
PUT4/PYK2	Raffinose	+	+	+	+	+
	Glucose	+			+	
	Galactose		+	+	+	+

Table 3-7. (Continued)

*TDA6/CUR1**	Raffinose					
	Glucose	++	++	++	++	++
	Galactose	+				

* These gene pairs were removed from consideration after qRT-PCR analysis revealed they are not both up

**qRT-PCR validation resulted in the reclassification of this gene pair as class 1

Table 3-8. Statistical analysis of adjacent gene pairs

Strains	Carbon Source	Expected number of tandem pairs	Observed number of tandem pairs	P-value
1.2	Raffinose	6.06	22	1.2×10^{-7}
2.3a	Raffinose	1.85	11	6.3×10^{-7}
2.3b	Raffinose	1.91	9	3.2×10^{-5}
sin4Δ mot3-1162 grr1-531	Raffinose	0.94	5	4.4×10^{-4}
sin4Δ mot3-1162 mit1-560	Raffinose	0.46	2	0.012
1.2	Glucose	11.89	24	6×10^{-4}
2.3a	Glucose	1.39	8	1.6×10^{-5}
2.3b	Glucose	2.77	10	2.7×10^{-6}
sin4Δ mot3-1162 grr1-531	Glucose	2.24	9	1.2×10^{-4}
sin4Δ mot3-1162 mit1-560	Glucose	0.09	1	0.004
1.2	Galactose	14.99	35	2.9×10^{-6}
2.3a	Galactose	1.40	11	1.1×10^{-8}
2.3b	Galactose	2.08	16	1×10^{-10}
sin4Δ mot3-1162 grr1-531	Galactose	0.75	4	0.0011
sin4Δ mot3-1162 mit1-560	Galactose	0.42	2	0.0089

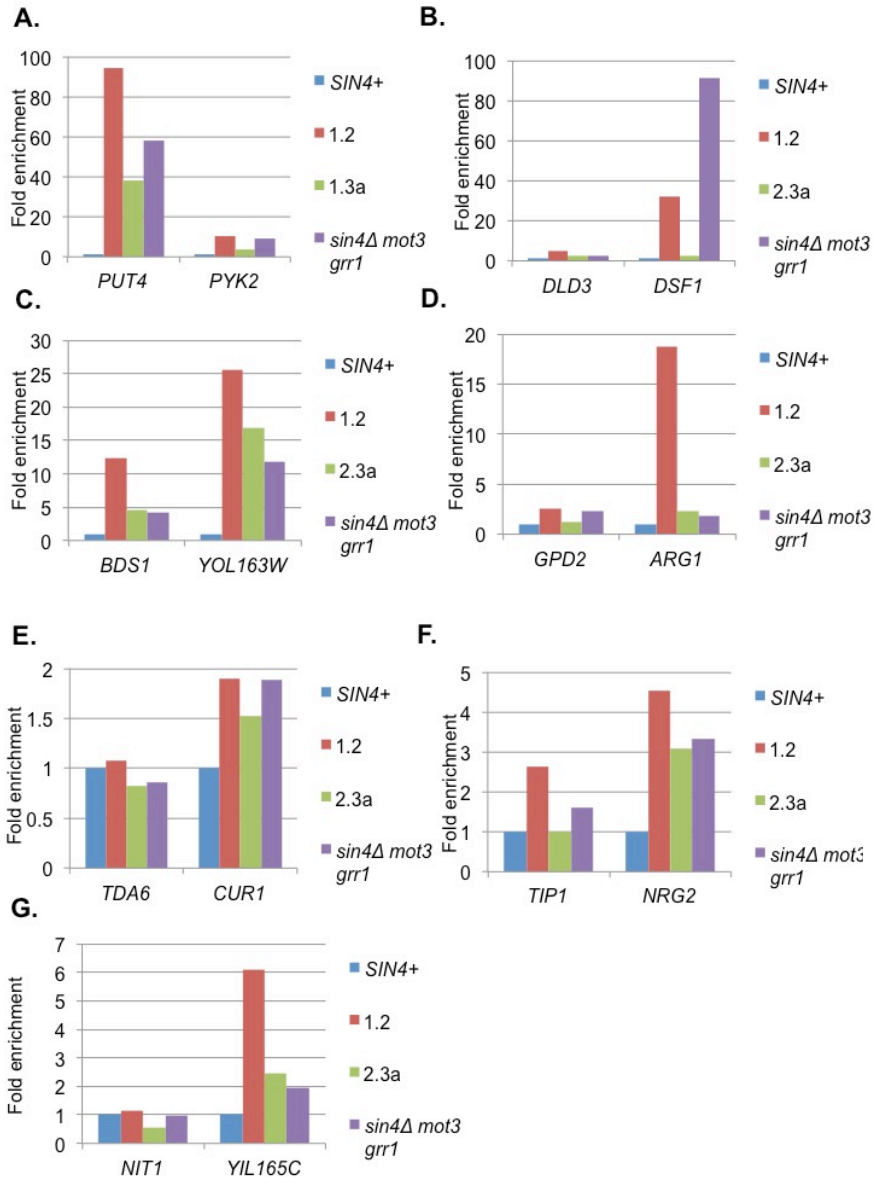


Figure 3-7. Verification of microarray results of tandem gene pairs. qRT-PCR was performed to measure RNA levels using the primers listed in Table 3-3. Shown are the fold increases for the mutant strains normalized to the wild-type value. The values are an average of two experiments. The wild-type strain is CR214. The mutants are as follows: 1.2, 2.3a, and *sin4Δ mot3 grr1* (CR204). In each panel, the upstream gene is shown on the left.

are some differences in fold changes. *NIT1*/*YIL165C* and *TDA6*/*CUR1* do not show increased expression levels above the two-fold cutoff at both genes in the mutant strains.

To test whether the activation of the downstream gene is dependent upon that of the upstream gene, we will delete the UAS of the upstream gene. If downstream activation is dependent, this deletion should abolish the activation we observe in the mutants. To date, we have performed this test for the *PUT4*/*PYK2* pair. Our results show that deletion of the *PUT4* promoter abolishes *PUT4* expression as expected, but does not impair *PYK2* expression (Figure 3-8), indicating long-distance activation does not occur at *PUT4*/*PYK2*.

Analysis of the *YBR281C*/*SAF1* adjacent gene pair

Analysis of the microarray data revealed galactose dependent expression of *YBR281C,* the site of the remnant of one of our long-distance reporters, and of the tandem adjacent gene *SAF1*. In the strains used for the microarrays, *YBR281C* no longer contains the *URA3* reporter, but it still contains the *GAL1* UAS followed by the 3' 699 bp of the *YBR281C* ORF and the normal *YBR281C* termination sequence. The activation of *SAF1* downstream in the mutants suggests long-distance activation. However, we wanted to determine if the *SAF1* transcript in our mutants is a wild-type length transcript or if it is a read-through transcript from *YBR281C*, similar to the long transcript seen at the reporter (Dobi and Winston, 2007). We performed Northern analysis and probed for *YBR281C* and *SAF1* to determine transcript sizes (Fig. 3-9). The *YBR281C* probe hybridizes to an

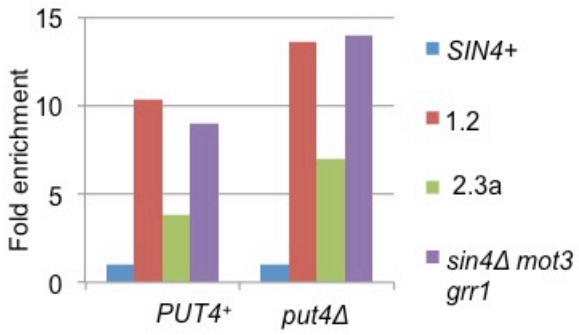


Figure 3-8. Effect of a *PUT4* promoter deletion on *PYK2* RNA levels. *PYK2* expression was measured by qRT-PCR using the primers in Table 3-3. The strains are *PUT4*$^+$ *SIN4*$^+$ (CR214), *PUT4*$^+$ 1.2 (CR101), *PUT4*$^+$ 2.3a (CR109), *PUT4*$^+$ *sin4Δ mot3 grr1* (CR204), *put4Δ SIN4*$^+$ (CR214), *put4Δ* 1.2 (CR240), *put4Δ* 2.3a (CR241), and *put4Δ sin4Δ mot3 grr1* (CR242). *PUT4*$^+$ strain values are normalized to *PUT4*$^+$ *SIN4*$^+$ and presented as fold enrichment. *put4Δ* strain values are normalized to *put4Δ SIN4*$^+$ and presented as fold enrichment.

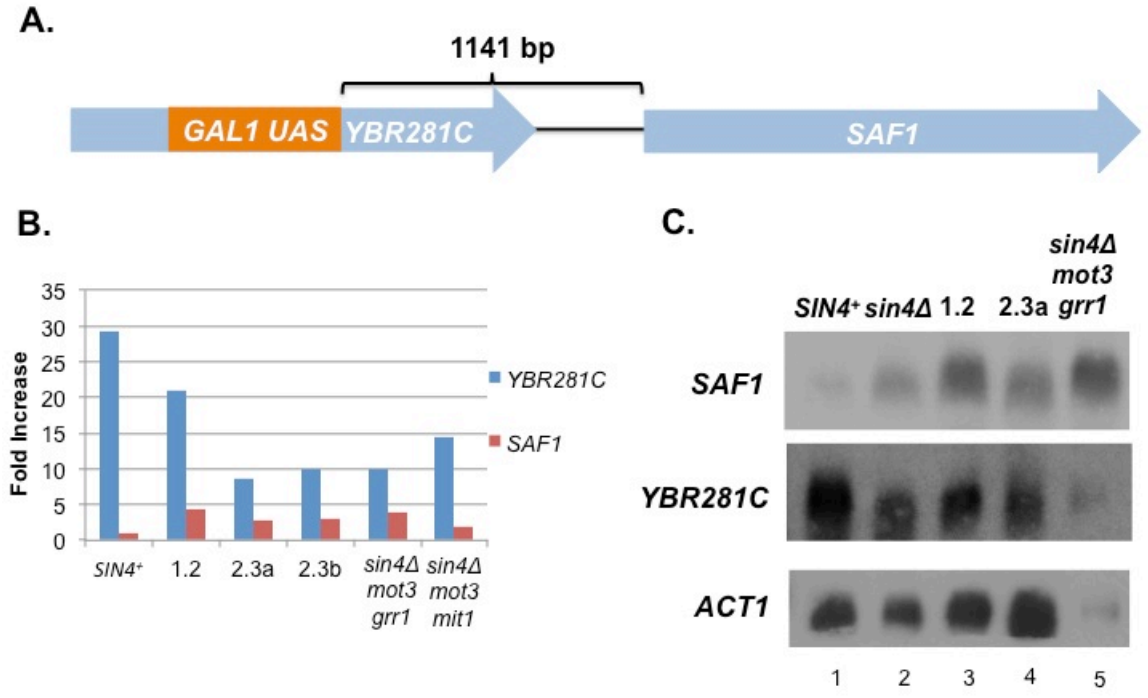


Figure 3-9. Northern analysis of *YBR281C/SAF1* gene pairs. **A.** Schematic diagram of the *GAL1* UAS in *YBR281C* and downstream *SAF1* ORF. **B.** Fold changes in expression of *YBR281C* and *SAF1*. *YBR281C* values are normalized to WT raffinose. SAF1 values are normalized to WT galactose. **C.** Northern analysis of mutant strains. Northerns were probed for *SAF1* (top) *YBR281C* (middle) and *ACT1* (bottom) as a loading control. Lanes are: lanes 1 and 6 (CR214), lanes 2 and 7 (CR111), lanes 3 and 8 (1.2), lanes 4 and 9 (2.3a), and lanes 5 and 10 (CR204). Strains were shifted from growth in 2% raffinose to 2% galactose.

mRNA approximately 700 bp in size, consistent with transcription initiating proximal to the UAS and terminating at the endogenous *YBR281C* termination sequence. The *SAF1* probe hybridized to an RNA of approximately 2 kb, which is consistent with the 1914 bp *SAF1* ORF. The *sin4Δ*, 1.2, 2.3a, and sin*4Δ mot3-1162 grr1-531* strains all have increased *SAF1* RNA levels relative to wild-type. This is the first clear example of long-distance activation in our mutant strains of a gene outside of our reporter.

Discussion

The data presented in this chapter provide greater insight into the transcriptional effects of the mutations isolated in our selections for mutants that allow long-distance activation. We found that mutations within the genes encoding the transcription factors Mot3, Mit1, and Msn2 interact to give stronger Lda^- phenotypes. These mutations are also able to increase the Lda^- phenotype in strains carrying reporters with increased distance between the UAS and reporter gene. Microarray analysis revealed genome-wide transcriptional changes in the mutants; interestingly, there is significant overlap between mutants that were isolated independently. Statistical analysis of the number of tandem adjacent gene pairs that are upregulated in the mutants suggests that long-distance activation may be occurring at other locations throughout the genome in addition to the reporter. Finally, we have strong evidence that the endogenous *SAF1* gene is activated from the *GAL1* UAS present in the upstream *YBF281C* ORF.

When we tested for activation of long reporters, we were surprised to discover that these strains displayed stronger activation over 2027 bp than over 1397 bp. One possible explanation for this result is that the 2027 bp reporter might contain transcription factor binding sites that are not present in the 1397 bp reporter. In fact, the 2027 reporter contains one additional Mot3 binding site not present in the 1397 reporter. However, we do not know what binding sequence the altered Mot3 recognizes, so we do not yet know if the presence of this binding site is relevant to the phenotype. An alternate explanation is that activation is occurring in the mutant strains via looping of the UAS to the core promoter. By this model, loop formation would be more favorable at a distance of 2027 bp because of DNA flexibility or chromatin structure. 3C experiments will be used to directly test whether loops are formed at the reporter, and if so whether loop formation is more favorable for certain distances.

We analyzed the number of tandem adjacent gene pairs that are upregulated in the mutant strains in an effort to determine if long-distance activation is now occurring throughout the genome in the mutant strains. We did not see increases in genes downstream of Gal4 regulated genes, so we decided to look at tandem adjacent gene pairs. We saw that all the mutant strains analyzed by microarray show a greater number of tandem gene pairs upregulated in the mutants than one would expect by chance. Greater numbers of adjacent gene pairs were up in 1.2, 2.3a, and 2.3b than in the reconstructed strains, *sin4Δ mot3-1162 grr1-531* and *sin4Δ mot3-1162 mit1-560*. We had expected *sin4Δ mot3-1162 grr1-531* would have a greater amount of overlap with 1.2 for upregulated adjacent gene pairs because the reconstructed strain recapitulates *HIS3*

mRNA expression of the reporter seen in 1.2. Data presented in chapter 2 suggests the presence of a mutation in strain 1.2 that was not identified by bulk segregant analysis. Perhaps long-distance activation occurs at each UAS only if a specific set of mutations is present. *sin4Δ* does not allow long-distance activation in a reporter where the *CUP1* UAS is the regulatory sequence, indicating there is not one genome-wide mechanism for regulating activation distance (Dobi, 2007).

Although our analysis of upregulated tandem adjacent gene pairs suggests long-distance activation may be occurring in the mutant strains, there may be other cases that would be missed by this analysis. In isolating mutants that more closely resemble metazoans in their ability to activate over longer distances, we have created a problem also faced by metazoans: how to match the regulatory element with the gene it regulates? It is conceivable that in the mutant strains, long-distance activation does not result in activation of the adjacent gene, but instead results in activation of genes further away. We have not observed activation farther than 2 kb for our reporter strains, but that does not mean activation can't occur over greater distances for other regulatory elements. Additionally, in the mutant strains, distant UASs could act to enhance transcription in combination with more proximal UASs. They may act similarly to transcription of the *HO* promoter, which contains a regulatory region at -1000 to -1800 bp that is required for regulation in addition to the proximal regulatory region (McBride et al., 1997).

There are several potential mechanisms to explain how the *MOT3*, *MIT1*, and *MSN2* mutations could be allowing long-distance activation. The first possibility is the

transcription factors directly bind the reporters; the mutations prevent binding and therefore allow long-distance activation. Neither reporter contains the Mit1 or Msn2 binding site, although both reporters contain one Mot3 binding site. We think this mechanism is unlikely because it would predict the mutations are loss-of function. A second possibility for the *MOT3* and *MIT1* mutations is that a change in binding specificity now allows binding of the transcription factor to the reporter. The use of two reporters makes this mechanism less likely; although, it is possible the reporter spacer regions could both contain binding sites recognized by the altered transcription factors. The third possibility is that the mutations act indirectly to allow long-distance activation by causing altered expression of factors required to regulate activation distance. Analysis of genes showing a two-fold or greater expression change did not give any insight into what these potential factors could be. However, it is possible an unpredictable factor may play a role in regulation of activation distance in yeast. Cohesin is involved in regulating sister chromatid separation during cell division, but it also has a secondary role in the formation of chromatin loops to mediate long-distance interactions in metazoans (Hadjur et al., 2009; Mishiro et al., 2009). Perhaps the mutations result in altered expression of a protein in yeast that has dual functions, one of which is regulating activation distance.

It still remains unclear if the long-distance activation at the reporter occurs through the looping or scanning mechanism. Northern analysis of *YBR281C* and *SAF1* show long-distance activation past a terminator, which suggests looping could be mediating this activation. The only way to directly test for looping is 3C of the region. The increased activation of mutants in the 2027 bp reporter relative to the 1397 bp

reporter is consistent with looping of the reporter. Previous studies demonstrate that in yeast when loop formation is forced, long-distance activation does occur (de Bruin et al., 2001; Petrascheck et al., 2005). If long-distance activation occurs at other regions of the genome, as the microarray results suggest, the activation could be occurring by different mechanisms at different locations.

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds. (1991). Current Protocols in Molecular Biology (New York, Greene Publishing Associates and Wiley-Interscience).

Cain, C.W., Lohse, M.B., Homann, O.R., Sil, A., and Johnson, A.D. (2012). A conserved transcriptional regulator governs fungal morphology in widely diverged species. Genetics *190*, 511-521.

de Bruin, D., Zaman, Z., Liberatore, R.A., and Ptashne, M. (2001). Telomere looping permits gene activation by a downstream UAS in yeast. Nature *409*, 109-113.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open Source Clustering Software. Bioinformatics *20*, 1453-1454.

Dobi, K.C. (2007). Upstream Activation Sequence Function in *S. cerevisiae*: Long-distance Transcriptional Activation and the Role of Sin4. In Genetics (Cambridge, MA, Harvard University), pp. 215.

Grady, J.T.W. (2013). Adapting Quantitative Protein and Phosphorylation Analyses to a Proteome-Wide Scale. In Cell Biology (Cambridge, MA, Harvard University), pp. 258.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.g., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Hongay, C., Jia, N., Bard, M., and Winston, F. (2002). Mot3 is a transcriptional repressor of ergosterol biosynthetic genes and is required for normal vacuolar function in Saccharomyces cerevisiae. EMBO J *21*, 4114-4124.

Martinez-Montanes, F., Rienzo, A., Poveda-Huertes, D., Pascual-Ahuir, A., and Proft, M. (2013). Activator and Repressor Functions of the Mot3 Transcription Factor in the Osmostress Response of Saccharomyces cerevisiae. Eukaryot Cell *12*, 636-647.

McBride, H.J., Brazas, R.M., Yu, Y., Nasmyth, K., and Stillman, D.J. (1997). Long-range interactions at the HO promoter. Mol Cell Biol *17*, 2669-2678.

McIsaac, R.S., Silverman, S.J., McClean, M.N., Gibney, P.A., Macinskas, J., Hickman, M.J., Petti, A.A., and Botstein, D. (2011). Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. Mol Biol Cell *22*, 4447-4459.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J *28*, 1234-1245.
Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W., and Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res *33*, 3743-3750.

Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics 1990: A Laboratory Course Manaual (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., and Quackenbush, J. (2006). TM4 microarray software suite. Methods Enzymol *411*, 134-193.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., *et al*. (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques *34*, 374-378.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53-55.

Chapter 4

Summary and perspectives

While a few studies have demonstrated that activation distance is regulated in yeast (Dobi and Winston, 2007; Guarente and Hoar, 1984; Struhl, 1984), only one has attempted to understand how this regulation occurs (Dobi and Winston, 2007). Following up on this study, we sought to gain a more complete understanding of how activation distance is regulated by identifying the factors involved. To do this, we isolated polygenic mutants that display increased activation of a long-distance reporter. Using bulk segregant analysis and whole-genome sequencing, we identified mutations in *MOT3*, *GRR1*, *PTR3*, *MIT1*, *MSN2*, and *YOR19W* that each contribute to the Lda$^-$ phenotype.

In Chapter 2, we isolated independent lineages of strains that display long-distance activation by the *GAL1* UAS. Previous selections for yeast polygenic mutants have been performed, though these experiments select for growth under a specific environmental condition by in lab evolution (Koschwanez et al., 2013; Romano et al., 2010). Our selection differs in that we are using a reporter-based system to select for mutants. The use of a reporter system allows for study of processes beyond environmental growth conditions, in our case long-distance activation. The reporter system was constructed with two reporters to reduce the number of *cis*-acting mutations. The second reporter used in this study requires Northern analysis to look for increases in *URA3* mRNA, which limits the number of mutants that can be processed. A potential adaptation of this reporter would be to use the *Cryptococcus neoformans ILV2* gene, which makes *C. neoformans* highly resistant to sulfometuron methyl (SM), an acetolactate synthase inhibitor (Kingsbury et al., 2004). Expression of *C. neoformans*

ILV2 in *S. cerevisiae* also confers resistance to high levels of SM. If increasing expression of *C. neoformans ILV2* can be correlated to increasing SM resistance, a second reporter could be constructed where increased resistance to SM was the readout. Adaptation of the reporter system could allow for the isolation of more mutations that increase long-distance activation. More broadly, the dual reporter system could be adapted to study other aspects of gene regulation as polygenic traits.

Bulk segregant analysis and whole-genome sequencing of strains 1.2, 2.3a, and 2.3b resulted in identification of five mutations that cause a marked increase in long-distance activation. Additional analysis of the sequence data for strains 2.3a and 2.3b revealed these strains are disomic for chromosome III. This disomy is likely contributing to the Lda$^-$ phenotype. However, there is still an unaccounted for component of heritability in strains 2.3a and 2.3b. Additionally, as described in Chapter 2, strain 1.2 appears to contain a causative mutation that was not identified by bulk segregant analysis, most likely because it does not enhance the Lda$^-$ phenotype of the *sin4Δ mot3-1162 grr1-531* triple mutant. One potential method for identifying these missing components of heritability is to perform bulk segregant analysis on strains 1.1 and 2.2, the intermediate strains isolated in each lineage.

In Chapter 3, we identified transcriptional changes in the original polygenic mutants and some reconstructed polygenic mutants by microarray. From these data, we attempted to identify other instances of long-distance activation occurring in the genome by analyzing tandem adjacent gene pairs where both genes had increased mRNA levels.

A greater number of tandem adjacent gene pairs were up than are expected by random chance. We focused on gene pairs that were found in at least one of the reconstructed mutants to increase the chance that they are affected by the identified causative mutations. While one of the genes pairs tested, *PUT4/PYK2,* did not prove to be an example of long-distance activation, there are several other candidates to test. *BDS1/YOL163W* and *DLD3/DSF1* are two additional candidate pairs worth testing by deletion of the 5' gene promoter, as the mutants show the greatest increases in expression for both genes in each pair. Additional analysis of the tandem adjacent gene pairs will reveal if they are upregulated as a result of long-distance activation. Along with this analysis, it could be informative to perform expression analysis on the mutants using different environmental conditions, particularly stress conditions. These experiments could reveal other candidate sites of long-distance activation as well as identify how stress-response differs in the *mot3-1162* mutant.

Two of the mutations isolated, *mot3-1162* and *mit1-560*, may cause altered DNA binding specificity, it would be of interest to determine the binding patterns of the mutant Mot3 and Mit1 proteins. We looked for expression changes at several genes known to be regulated by Mot3 or Mit1 and found these genes are unaffected in the mutants based on our microarray data, so we speculate these mutant proteins are still able to bind to their known binding sites. ChIP-seq experiments would reveal if the factors are binding normally across the genome, including the possibility that they are acting at our reporters.

The mechanism for how the mutations allow long-distance activation is currently unknown; however, an attractive model is that in the mutants, the *GAL1* UAS is able to activate the reporter via looping. Currently the strongest evidence for looping in the polygenic mutants is galactose-dependent activation of *SAF1*. At *SAF1,* long-distance activation occurs past a terminator, which is inconsistent with a scanning mechanism of activation REF. The method of 3C can be used to determine if looping is occurring either at this locus or at our long-distance reporters. Analysis of increased reporter distances revealed that in the polygenic mutants, the *GAL1* UAS activates the reporter more strongly at a distance of 2027 bp than 1397 bp. We speculate that the distance of 2027 bp may be more permissive to loop formation than 1397 bp, thus resulting in stronger activation. When performing these experiments it is important to keep in mind that 3C experiments are only semi-quantitative. If loops form at a low frequency that is sufficient to allow long-distance activation, we may not detect them by 3C. While we hope to determine if chromatin loops form in the mutants, a negative 3C result does not conclusively rule looping out as a possible mechanism.

Taken together, our results provide a more comprehensive view of long-distance activation in yeast. Our method of isolation allowed us to identify mutations that would not have been found using traditional screening and selection experiments. Analysis of genome-wide transcriptional changes suggests long-distance activation occurs at regions other than the reporter. Additional analysis will determined if candidate regions of long-distance activation are truly regions of long-distance activation or if these regions reflect other transcriptional changes. Finally, 3C analysis of the *BPH1* reporter as well as

YBR281C/SAF1 will examine the possibility long-distance activation is occurring by looping of the UAS to the core promoter.

References

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol Cell Biol *27*, 5575-5586.

Guarente, L., and Hoar, E. (1984). Upstream activation sites of the CYC1 gene of Saccharomyces cerevisiae are active when inverted but not when placed downstream of the "TATA box". Proc Natl Acad Sci U S A *81*, 7860-7864.

Kingsbury, J.M., Yang, Z., Ganous, T.M., Cox, G.M., and McCusker, J.H. (2004). Cryptococcus neoformans Ilv2p confers resistance to sulfometuron methyl and is required for survival at 37 degrees C and in vivo. Microbiology *150*, 1547-1558.

Koschwanez, J.H., Foster, K.R., and Murray, A.W. (2013). Improved use of a public good selects for the evolution of undifferentiated multicellularity. Elife *2*, e00367.

Romano, G.H., Gurvich, Y., Lavi, O., Ulitsky, I., Shamir, R., and Kupiec, M. (2010). Different sets of QTLs influence fitness variation in yeast. Mol Syst Biol *6*, 346.

Struhl, K. (1984). Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. Proc Natl Acad Sci U S A *81*, 7865-7869.

Appendix 1

Isolation and identification of *sin4* and *rgr1* mutations that allow activation over a UAS-TATA distance of 1397 basepairs

Introduction

To understand how activation distance is regulated in yeast, we isolated mutants that allow activation over a UAS-TATA distance that is normally non-permissive in yeast. A previous study resulted in isolation of mutations in *SIN4*, *RGR1*, *SPT2*, *SPT10,* and *HTA1-HTB1* that allowed activation of a reporter with a UAS-TATA distance of 799 bp (Dobi and Winston, 2007, Dobi and Leeman, unpublished). In an attempt to isolate stronger mutants, we selected additional mutants that allow long-distance activation using a reporter with a spacer distance of 1397 bp. We found that every strain with a single recessive mutation contained a mutation in either *SIN4* or *RGR1*.

Materials and methods

S. cerevisiae strains

The *S. cerevisiae* strains used in this study (Table 5-1) are isogenic with a *GAL2*$^{+}$ derivative of S288C (Winston et al., 1995). Rich (YPD) and synthetic complete (SC) dropout media were prepared as previously described (Rose et al., 1990). SC Gal and SC-His Gal media contained 2% galactose as the carbon source. Strains were constructed by crosses (Ausubel et al., 1991).

Table 5-1. Strains used in this study

Strain	Genotype
CR1	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397*
CR2	*MATα his3Δ200 ura3Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397*
lda5	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397 rgr1-5*
lda6	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397 rgr1-6*
lda12	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397 rgr1-12*
lda13	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397 sin4-13*
lda25	*MATα his3Δ200 ura3Δ0 trp1Δ63 lys2-128δ bph1Δ::kanMX-*UAS_{GAL1}*1397 rgr1-25*

Selection of mutants

Five independent cultures of CR1 and CR2 were grown up grown in YPD overnight at 30°C. From each culture, 200 μL of cells were plated onto SC-His Gal and the plates were UV irradiated at 5000μJ/cm^3. Plates were incubated at 30° and colonies that grew were purified on SC-His Gal and then on YPD.

Plasmid complementation

Mutant strains were transformed with pM1387 (*SIN4 URA3 CEN4 ARS1*) and pM2597 (*RGR1*, *URA3*, 2 micron). Transformants were tested for growth on SC-Ura His Gal.

Diploid complementation

*MAT***a** mutants were crossed by *MAT* α mutants and diploids were selected on SC-Leu Trp media. Diploids were tested for complementation by growth on selective SC-His Gal media.

Results

A total of 52 mutants were isolated after UV mutagenesis. UV mutagenesis enhanced the mutation frequency; plates that had been UV mutagenized contained 4-5

fold more colonies than plates containing spontaneous mutants. Of these mutants, 29 were recessive single mutants and showed a similar phenotype to the *sin4Δ*. Mutant strains were tested for mutations in *SIN4* or *RGR1* by plasmid complementation. These strains were transformed with three plasmids: a *SIN4* plasmid, an *RGR1* plasmid, and a vector control. Complementation of the Lda$^-$ phenotype with the *SIN4* or *RGR1* plasmid is indicated by loss of growth on selective SC-Ura His Gal media. Examples of mutants complemented by either *RGR1* or *SIN4* plasmid are shown in Figure 5-1. Using plasmid complementation, we identified 16 strains with *sin4* mutations and 13 strains with *rgr1* mutations.

We performed diploid complementation tests on the *rgr1* mutant strains. We found that while some of the diploids resulted in non-complementation, as would be expected for mutations in the same gene, a subset of the diploids displayed complementation of the Lda$^-$ phenotype. We hypothesized that the complementation could be the result of mutations in different domains of the protein. We were particularly interested in this hypothesis because all previously isolated *rgr1* mutations result in changes to the C-terminus of the protein (Sakai et al., 1990; Wang and Michels, 2004). We chose three mutants to sequence, two that fell into one of the *rgr1* complementation groups and one that fell into the other *rgr1* complementation group. All three strains contain nonsense mutations in *RGR1* that result in truncation of the C-terminal domain of the protein (Table 5-2).

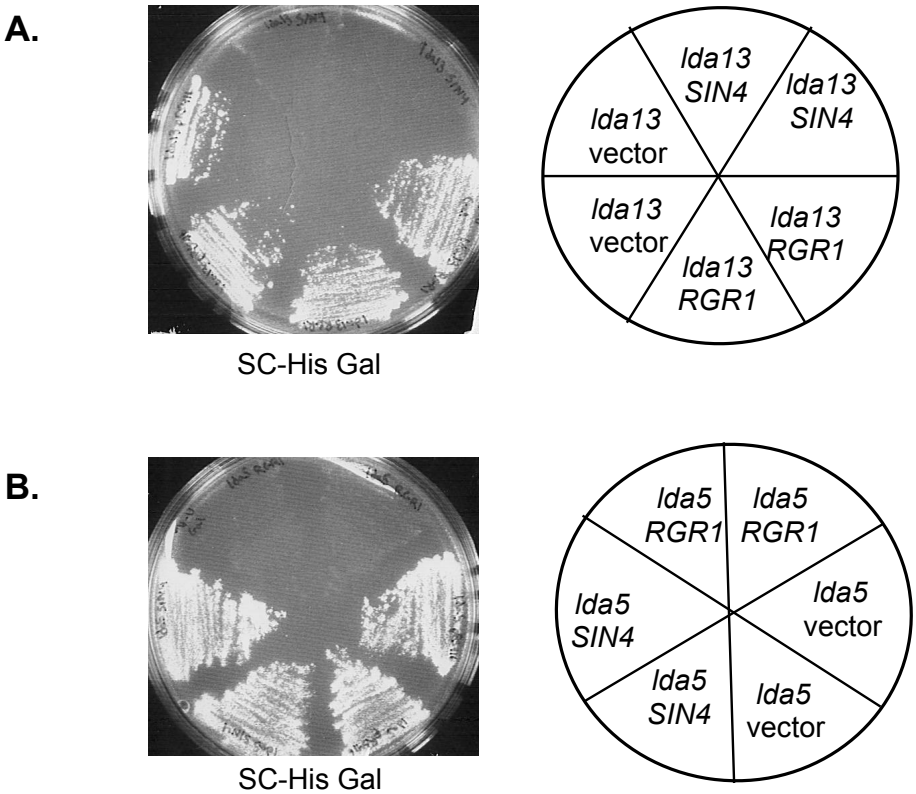


Figure 5-1 Plasmid complementation of *sin4* and *rgr1* mutants. Strains were transformed with *SIN4*, *RGR1*, and vector only plasmids. Transformants were purified on SC-Ura media and replica plated to SC-Ura His Gal. Plates were incubated at 30° for 2 days. A. *lda13 SIN4*, *RGR1*, and vector transformants. B. *lda5 SIN4*, *RGR1*, and vector transformants.

Table 5-2. *rgr1* mutant sequencing

Mutant	DNA change	Protein change
lda6	A2202AT	Q734H, premature stop 737
lda12	C2690T	Q897stop
lda25	C2179T	K726stop

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds. (1991). Current Protocols in Molecular Biology (New York, Greene Publishing Associates and Wiley-Interscience).

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol Cell Biol *27*, 5575-5586.

Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics 1990: A Laboratory Course Manaual (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Sakai, A., Shimizu, Y., Kondou, S., Chibazakura, T., and Hishinuma, F. (1990). Structure and molecular analysis of RGR1, a gene required for glucose repression of Saccharomyces cerevisiae. Mol Cell Biol *10*, 4130-4138.

Wang, X., and Michels, C.A. (2004). Mutations in SIN4 and RGR1 cause constitutive expression of MAL structural genes in Saccharomyces cerevisiae. Genetics *168*, 747-757.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53-55.

Appendix 2

Identification of polygenic mutant strains containing reporter duplications

Introduction

We decided to study the regulation of activation distance as a polygenic trait based on preliminary results seen with several mutant selection experiments. In order to isolate mutants with stronger long distance activation phenotypes, we performed selection experiments using a reporter with a spacer distance of 2027 bp. This 2027 reporter was chosen because *sin4Δ* strains show a very weak Lda$^-$ phenotype at this distance. We found that several of the mutants isolated contained a partial or complete chromosomal duplication that resulted in duplication of the reporter. Isolation of strains containing chromosomal aneuploidies led to the decision to use strains with two reporters for the multiple mutant selections.

Materials and methods

S. cerevisiae strains

The *S. cerevisiae* strains used in this study (Table 6-1) are isogenic with a *GAL2*$^+$ derivative of S288C (Winston et al., 1995). Rich (YPD) and synthetic complete (SC) dropout media were prepared as previously described (Rose et al., 1990). SC Gal and SC-His Gal media contained 2% galactose as the carbon source. Strains were constructed by crosses (Ausubel et al., 1991).

Table 6-1. Strains used in this study

Strain	Genotype
FY23	*MATα his3Δ200 ura3Δ0 trp1Δ63 arg4-12 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027*
CR24	*MAT***a** *his3Δ200 ura3Δ0 trp1Δ63 leu2Δ0 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027*
CR31	*MATα his3Δ200 ura3Δ0 trp1Δ63 arg4-12 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh1*
CR32	*MATα his3Δ200 ura3Δ0 trp1Δ63 arg4-12 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh2*
CR33	*MATα his3Δ200 ura3Δ0 trp1Δ63 arg4-12 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh3*
CR34	*MATα his3Δ200 ura3Δ0 trp1Δ63 arg4-12 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh4*
CR35	*MAT***a** *his3Δ200 ura3Δ0 trp1Δ63 leu2Δ0 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh5*
CR36	*MAT***a** *his3Δ200 ura3Δ0 trp1Δ63 leu2Δ0 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh6*
CR37	*MAT***a** *his3Δ200 ura3Δ0 trp1Δ63 leu2Δ0 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh7*
CR38	*MAT***a** *his3Δ200 ura3Δ0 trp1Δ63 leu2Δ0 lys2-128δ sin4Δ0::TRP1 bph1Δ::kanMX-UAS$_{GAL1}$2027 enh8*

Selection of polygenic mutants

Starting with strains CR23 and CR24, each of which contains *sin4Δ*, we selected for mutants with stronger Lda$^-$ phenotypes. 10 independent cultures of each strain were grown in YPD overnight at 30°C. From each culture, 200 μL of cells were plated onto two SC-His Gal 1 mM 3-AT, one of which was UV irradiated for 5000μJ/cm^3. Plates were incubated at 30° and colonies that grew were purified on SC-His Gal 1 mM 3-AT. Spot tests were also used to test whether the His$^+$ phenotype was dependent on galactose.

CHEF gel and Southern blot analysis

CHEF gel analysis was performed as previously described (Libuda and Winston, 2006). Southern hybridization analysis was conducted with a probe to the coding regions of *HIS3* (-27 to +376, where +1 is the ATG).

Comparative genome hybridization (CGH) analysis

Genomic DNA was extracted from each strain as previously described (Rose et al., 1990). CGH analysis was performed as previously described (Torres et al., 2007). Reference RNA was labeled with Cy3 and experimental RNA was labeled with Cy5. Genomic DNA from CR23 was used as a reference.

Results

Eight strains were isolated that showed growth on SC-His Gal 1 mM 3-AT, on this media the *sin4Δ* parent strains do not grow. These strains were retested for growth on SC-His Gal 1 mM 3-AT by replica plating. Replica plating revealed that some single colonies lost the ability to grow on selective media. We suspected this instability could be due to changes in copy number of the reporter.

We first tested for altered copy number of the reporter by CHEF gel, which separates chromosomes by size, and Southern blot analysis. An increase in reporter copy number could be caused by chromosomal rearrangements, which may be visible by CHEF gel. Probing the Southern for *HIS3* should result in hybridization to chromosome III, the site of the reporter. Interestingly, the *HIS3* probe hybridizes to a larger than expected chromosome in the mutant strains CR34 and CR36 (Fig. 6-1). Additionally, the *HIS3* probe hybridizes to two different sized chromosomes in a third mutant strain (Fig. 6-1).

We performed CGH analysis to determine if the chromosomal rearrangements resulted in changes in copy number to the reporter. We found that the three strains that showed abnormal *HIS3* hybridization also displayed a partial or complete duplication of chromosome III that results in duplication of the reporter (Fig. 6-2). Additionally, we performed CGH analysis on a strain that had lost resistance to 1 mM 3-AT.

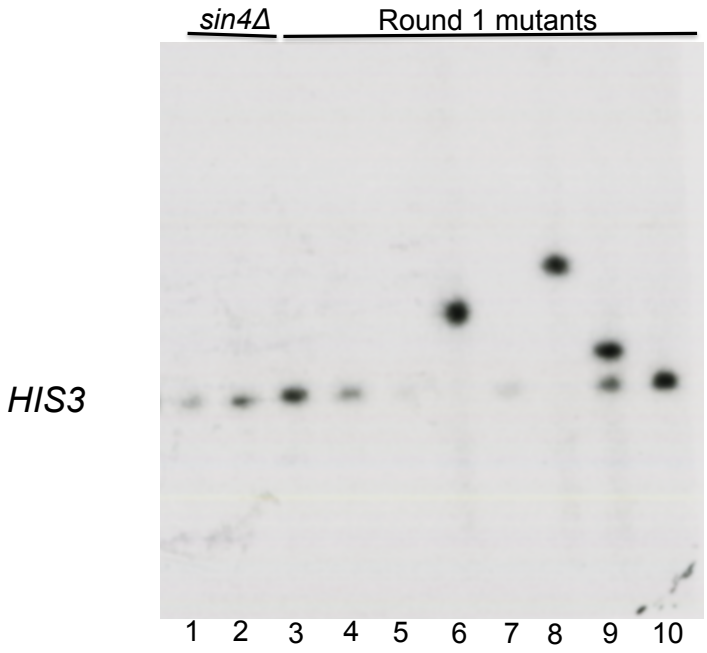


Figure 6-1. Southern blot of CHEF gel analysis. Southern blot was probed for *HIS3*, which is expected to hybridize to chromosome III, roughly 316 kb in size. Strains are as follows: CR23 (lane 1), CR24 (lane 2), CR31 (lane 3), CR32 (lane 4), CR33 (lane 5), CR34 (lane 6), CR35 (lane 7), CR36 (lane 8), CR37 (lane 9), and CR38 (lane 10).

Figure 6-2. CGH analysis of mutant strains. Reference DNA (CR23) and query DNA were labeled and hybridized to an array. Reads across each chromosome are presented. Duplications of query DNA are represented by a two fold increase in red signal intensity. A. CGH analysis of CR24 compared to CR23. B. CGH analysis of CR24 compared to CR31. C. CGH analysis of CR24 compared to CR36. D. CGH analysis of CR24 compared to CR37. E. CGH analysis of CR24 compared to CR37 strain that is no longer resistant to 1 mM 3-AT.

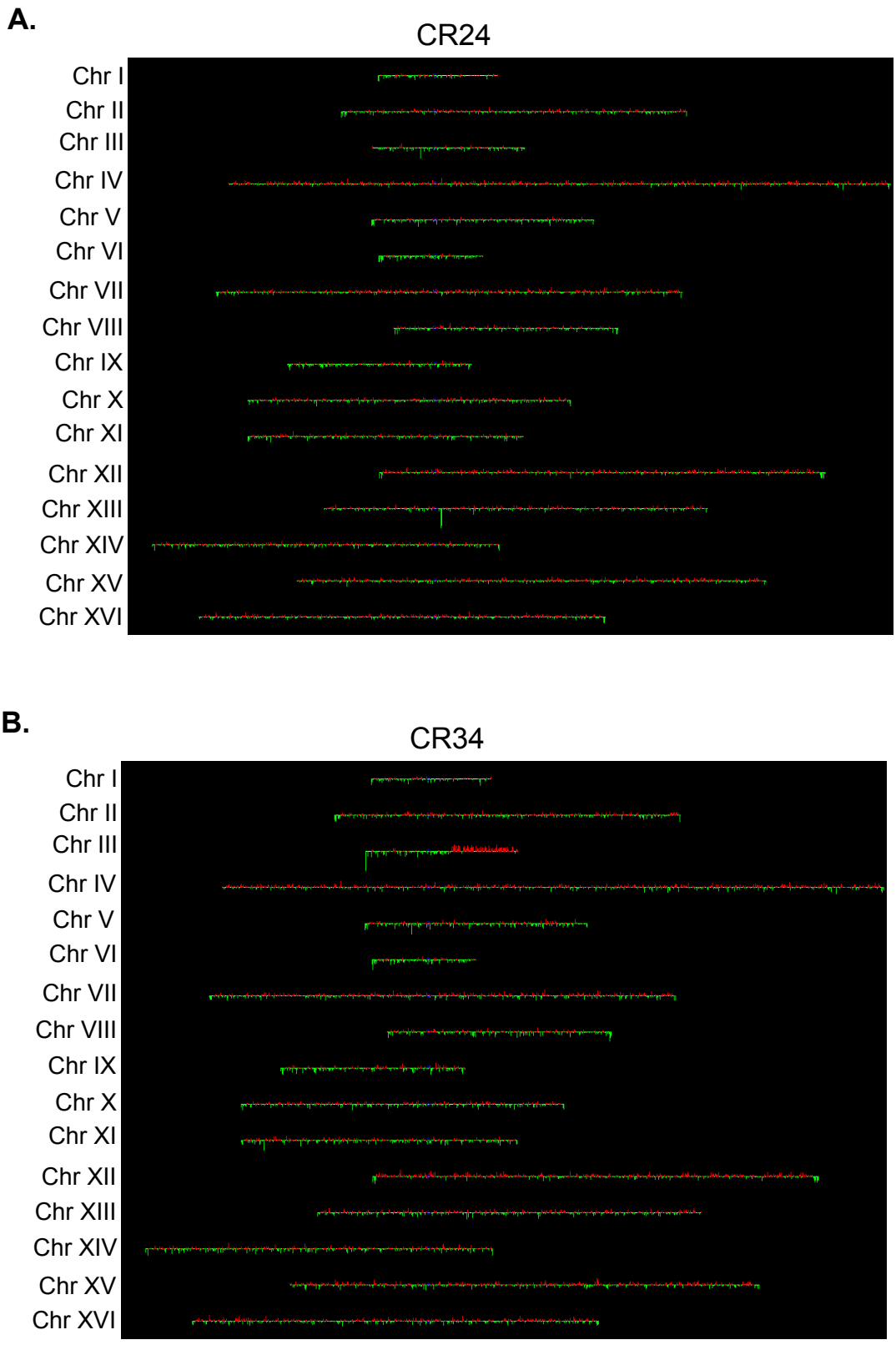


Figure 6-2. (Continued)

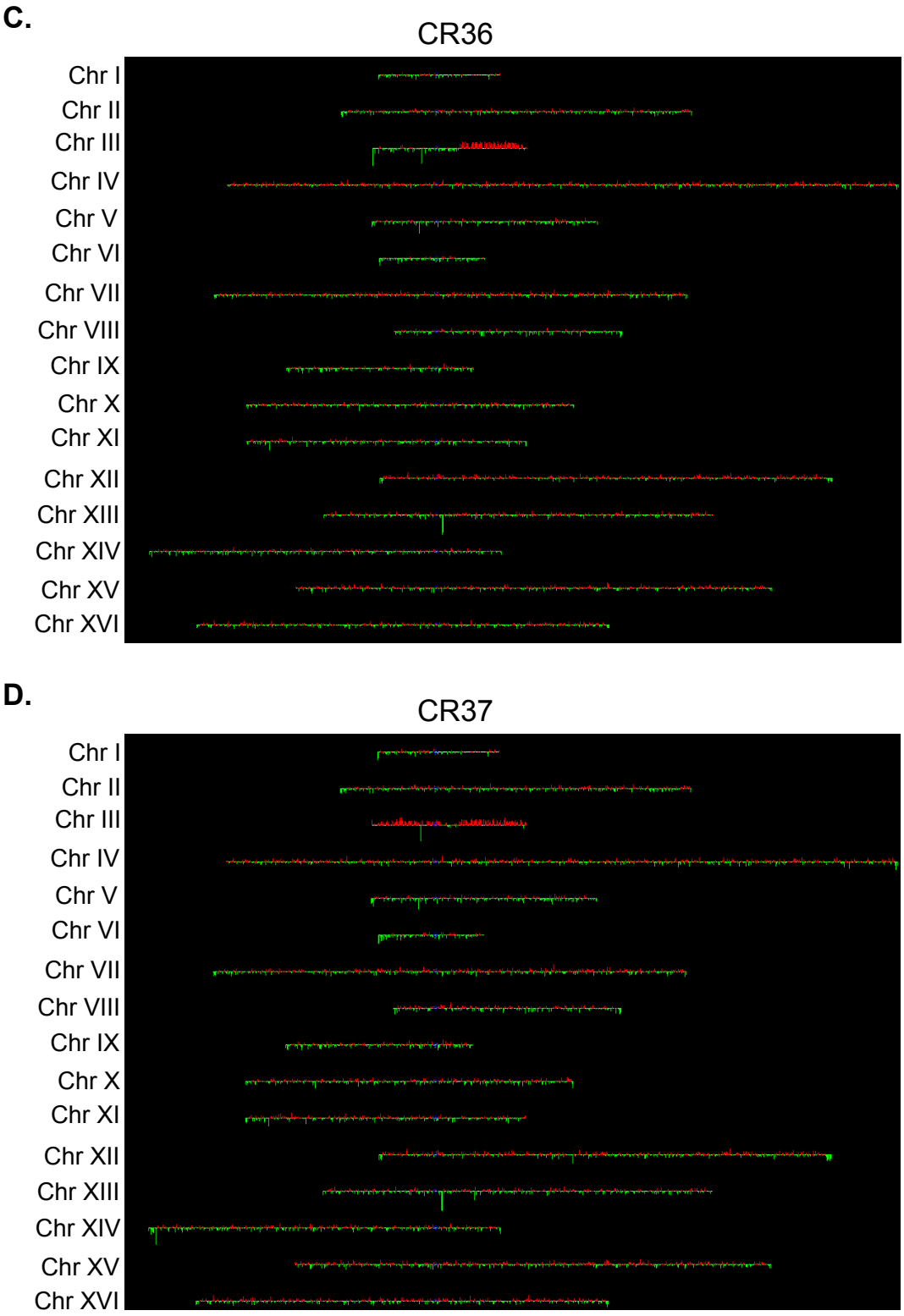


Figure 6-2. (Continued)

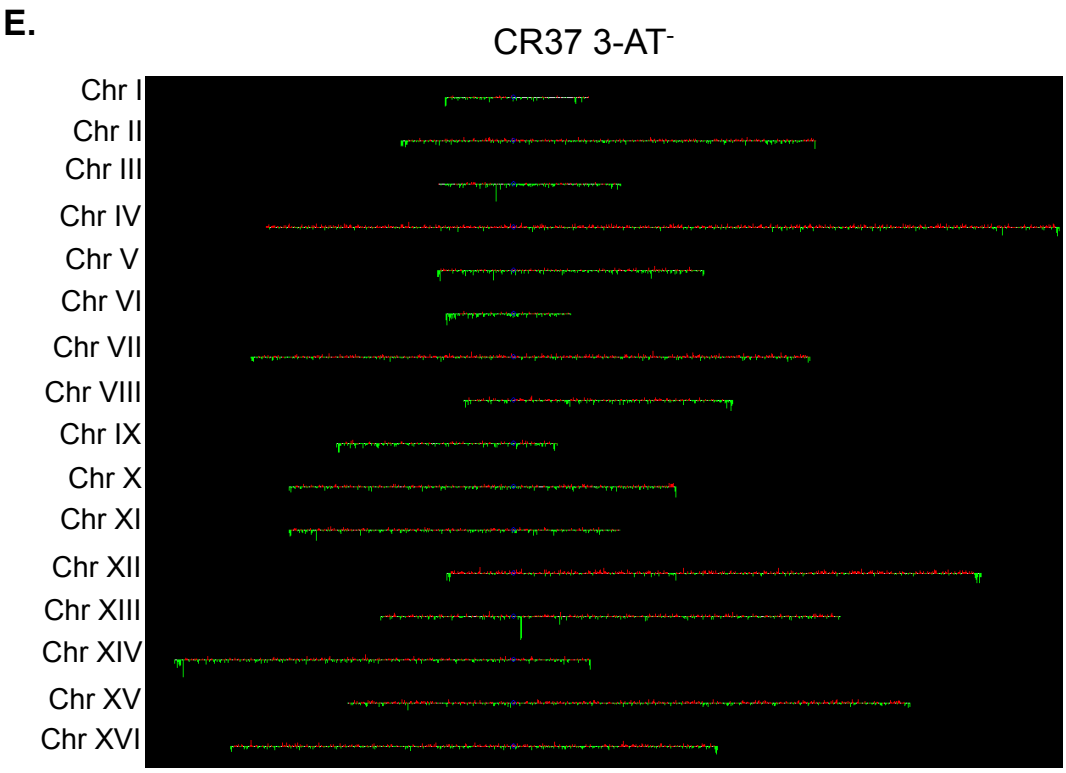


Figure 6-2. (Continued)

This strain no longer contained the chromosome III duplication, suggesting that the reporter duplication causes the 3-AT resistance phenotype. These results indicate that using only one reporter will result in a high proportion of cis-acting elements if the selection is conducted in a strain with only one reporter. These results led us to use a strain with two reporters for the polygenic mutant selection.

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds. (1991). Current Protocols in Molecular Biology (New York, Greene Publishing Associates and Wiley-Interscience).

Libuda, D.E., and Winston, F. (2006). Amplification of histone genes by circular chromosome formation in Saccharomyces cerevisiae. Nature *443*, 1003-1007.

Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics 1990: A Laboratory Course Manaual (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916-924.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53-55.

Appendix 3

Isolation and identification of long-distance activation mutants in a *CUP1* reporter strain

Introduction

The majority of the work done on the regulation of transcriptional activation distance has been done using a reporter containing the *GAL1* UAS; however, we are also interested in how activation distance is regulated with other regulatory elements. Previous results suggest that selection for Lda$^-$ mutants using a *CUP1* UAS reporter may identify different factors than those identified using the *GAL1* UAS reporter (Dobi, 2007). Using the *GAL1* UAS 799 reporter, *sin4Δ* and *spt2Δ* were found to cause the strongest Lda$^-$ phenotype. However, *sin4Δ* and *spt2Δ* do not confer an Lda$^-$ phenotype with respect to the *CUP1* UAS 326 bp reporter. We therefore performed a selection to identify factors that allow long distance activation of the *CUP1* UAS; these mutants were placed into three different complementation groups. Two of the complementation groups have been identified as containing mutations in *SPT10* or *HTA1-HTB1*. Interestingly mutations in these genes also result in long distance activation of the *GAL1* UAS reporter. The mutation present in the third complementation group remains unidentified.

Materials and methods

S. cerevisiae strains

The *S. cerevisiae* strains used in this study (Table 7-1) are isogenic with a *GAL2*$^+$ derivative of S288C (Winston et al., 1995). Rich (YPD) and synthetic complete (SC)

Table 7-1. Strains used in this study

Strain	Genotype
FY76	*MAT***a** *lys2-128 δ*
CR12	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{CUP1}$326*
CR13	*MATα his3Δ200 ura3Δ0 trp1Δ63 lys2-128δ sin4Δ0::LEU2 bph1Δ::kanMX-UAS$_{CUP1}$326*
lda201	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{CUP1}$326 lda201*
lda211	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{CUP1}$326 lda211*
lda228	*MAT***a** *his3Δ200 ura3Δ0 leu2Δ0 lys2-128δ bph1Δ::kanMX-UAS$_{CUP1}$326 lda228*
CR16	*MAT α ybr281cΔ::hphMX-UAS$_{CUP1}$-HIS3 his3Δ200 leu2Δ0 trp1Δ63 can1Δ::STE2pr-LEU2 lda211*
CR17	*MAT α ybr281cΔ::hphMX-UAS$_{CUP1}$-HIS3 his3Δ200 leu2Δ0 trp1Δ63 can1Δ::STE2pr-LEU2 lda201*
CR18	*MAT α ybr281cΔ::hphMX-UAS$_{CUP1}$-HIS3 his3Δ200 leu2Δ0 trp1Δ63 can1Δ::STE2pr-LEU2 lda211*

dropout media were prepared as previously described (Rose et al., 1990). SC Gal and SC-His Gal media contained 2% galactose as the carbon source. Strains were constructed by crosses (Ausubel et al., 1991).

Selection of *CUP1* mutants

Five independent cultures of CR12 and CR13 were grown up grown in YPD overnight at 30°C. From each culture, 200 µL of cells were plated onto two SC-His $CuSO_4$ plates, one of which was UV irradiated for two minutes at 5000µJ/cm^3. Plates were incubated at 30° and colonies that grew were purified on SC-His $CuSO_4$.

Synthetic genetic array (SGA) analysis

SGA analysis was performed as previously described (Tong and Boone, 2005). SGA analysis was performed with CR16 to look for additional mutants that allow long distance activation of the *CUP1* UAS. Additionally, SGA analysis was performed on group 1 and group 3 mutants to look for deletion set mutations that fail to complement these mutations.

High-throughput sequencing of yeast segregant pools

A yeast segregant pool was created for the group 3 mutant *lda211* as described in Chapter 2. Similar analysis was used to identify the mutation present in this strain.

Results

Selection for *CUP1* Lda$^-$ mutants resulted in the isolation of 80 recessive mutants that were placed into three different complementation groups (Table 7-2). The Lda$^-$ phenotype of a representative mutant from each group is shown in Figure 7-1. Complementation groups 2 and 3 both display Spt$^-$ phenotypes in addition to the Lda$^-$ phenotype. We originally attempted to clone these strains by transformation with a *CEN* plasmid library; however, these attempts were unsuccessful.

We next attempted to clone groups 2 and 3 by testing if these strains contained mutations in known *SPT* genes. We found that transformation of group 2 mutants with a plasmid containing wild-type *SPT10* results in loss of the Lda$^-$ phenotype. Additionally, crosses revealed the group 2 mutant is linked to *SPT10*. These results led us to conclude the group 2 strains contain mutations in *SPT10*. Similar techniques were unsuccessful in identifying the identity of the group 3 mutants.

In the effort to identify the mutations present in complementation groups 1 and 3, we crossed a mutant from each complementation group to a copy of the deletion set containing the *CUP1* reporter. We then screened the resulting diploids for strains that failed to complement the mutations, which would result in diploids with an Lda$^-$ phenotype. All of the strains in the deletion set complemented the group 1 and group 3

Table 7-2. *CUP1* complementation groups

	Number of mutants	Additional phenotypes	Gene
Group 1	64	n/a	Unknown
Group 2	15	Spt^-	*SPT10*
Group 3	11	Spt^-	*HTA1-HTB1*[*]

* This result requires further verification

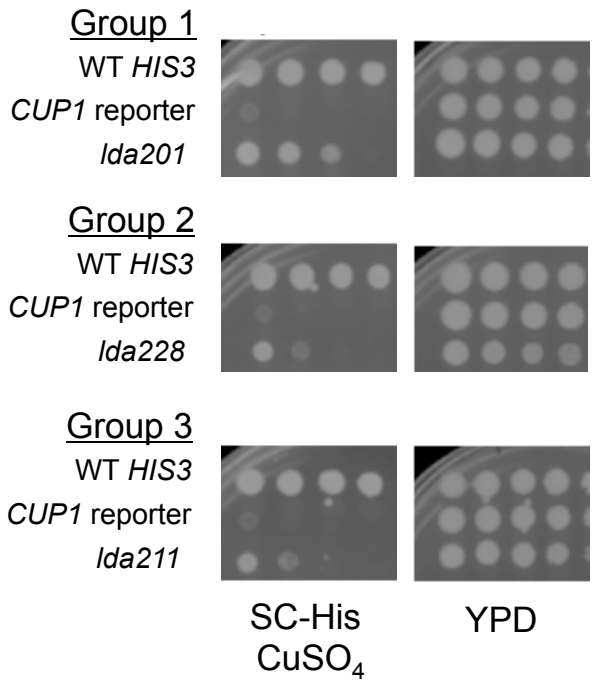


Figure 7-1 Lda$^-$ phenotypes of *CUP1* mutants. Ten-fold serial dilutions were made of saturated YPD cultures. The dilutions were spotted to the specified media and incubated for four days at 30°. Strains are as follows: WT *HIS3* (FY76), *CUP1* reporter (CR12), *lda201*, *lda228*, and *lda211*.

mutants, making this cloning attempt unsuccessful. This result could indicate the group 1 and group 3 mutations are in essential genes.

We also screened the haploid deletion set for additional mutants that allow long-distance activation of the *CUP1* reporter. We identified *mrc1Δ*, *tof1Δ*, *hst3Δ*, and *hst4Δ* strains as allowing long distance activation. We were interested to see a different set of mutations than those identified using the *GAL1* UAS reporter. Further study of these mutants indicated the Lda$^-$ phenotypes were not stable, so we decided not to follow up on any of these mutants.

We used whole-genome sequencing to identify a candidate mutation for complementation group 3. This mutation is in *HTA1* and is present in 100% of the reads for the mutant sequencing pool, giving strong support for this mutation as the causative mutation. This result should be verified by sequencing *HTA1* in other members of group 3. Additionally, these strains should be transformed with a wild-type *HTA1-HTB1* plasmid.

Discussion

The results presented in this appendix suggest that some of the factors involved in regulating activation distance have this role at multiple regulatory elements. *spt10* and *hta1-htb1* mutants allow long-distance activation at both the *GAL1* and *CUP1* UASs. However, some mechanisms of long-distance activation appear to be UAS specific, as the

sin4Δ mutant does not allow activation of the *CUP1* reporter. Further work is required to determine the different requirements for long-distance activation between UASs.

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds. (1991). Current Protocols in Molecular Biology (New York, Greene Publishing Associates and Wiley-Interscience).

Dobi, K.C. (2007). Upstream Activation Sequence Function in *S. cerevisiae*: Long Distance Transcriptional Activation and the Role of Sin4. In Genetics (Cambridge, MA, Harvard University), pp. 215.

Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics 1990: A Laboratory Course Manaual (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Tong, A.H., and Boone, C. (2005). Synthetic Genetic Array (SGA) Analysis in *Saccharomyces cerevisiae*. Yeast Protocols, Methods in Molecular Biology, 171-92.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *11*, 53-55.