Functional genetic analysis reveals intricate roles of conserved X-box elements in yeast transcriptional regulation

by

Sarah Voll

This thesis is submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of
Master of Science in Cellular and Molecular Medicine with Specialization in Bioinformatics

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa

Abstract

Understanding the functional impact of physical interactions between proteins and DNA on gene expression is important for developing approaches to correct disease-associated gene dysregulation. I conducted a systematic, functional genetic analysis of protein-DNA interactions in the promoter region of the yeast ribonucleotide reductase subunit gene *RNR3*. I measured the transcriptional impact of systematically perturbing the major transcriptional regulator, Crt1, and three X-box sites on the DNA known to physically bind Crt1. This analysis revealed interactions between two of the three X-boxes in the presence of Crt1, and unexpectedly, a significant functional role of the X-boxes in the absence of Crt1. Further analysis revealed Crt1-independent regulators of *RNR3* that were impacted by X-box perturbation. Taken together, these results support the notion that higher-order X-box-mediated interactions are important for *RNR3* transcription, and that the X-boxes have unexpected roles in the regulation of *RNR3* transcription that extend beyond their interaction with Crt1.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank Mads Kaern for encouraging me to question things and to expand the depth of my understanding. I also owe many thanks to all members of the Kaern lab for their support and guidance throughout my studies. In particular, I would like to thank Mila Tepliakova for her much needed assistance with data collection as well as her advice and support, Hilary Phenix and Daniel Jedrysiak for mentoring me throughout my studies, and Daniel Charlebois and Nezar Abdennur for their assistance with data analysis. I would also like to thank all lab members including Alex Power, Afnan Azizi and the fourth year honours students for the many insightful discussions about life and science.

Chapter 1

Introduction

Gene regulation defines when, where and by how much individual genes are expressed [6] and is critically important for proper maintenance of cellular functions. Too low expression can lead to deficiencies, while too high expression can be toxic [72]. Although gene expression can be modulated at many levels, a key step in gene regulation occurs during the initiation of gene transcription [54, 58]. At this level, the amount of mRNA transcribed from a gene is determined by the regulated binding of proteins to their sites on the DNA in the promoter region of the gene [10, 73].

There have been many systematic studies elucidating where and when proteins physically interact with promoter DNA (e.g. [35, 52, 65]). However, in these studies, the functional impact of these binding events on gene transcription is typically not measured. When a protein interacts with multiple sites within a promoter, the amount of transcription from that promoter can be determined by the effect of each protein-DNA interaction individually as well as by higher order combinatorial interactions [10]. The objective of my thesis research was to investigate this by functionally mapping how multiple promoter elements known to physically interact with one protein define transcriptional level for a specific gene.

The mechanisms defining transcriptional regulation are quite similar throughout

the eukaryotic kingdom [58, 85]. Therefore, studying gene regulation in the budding yeast *S. cerevisiae* is useful for understanding these conserved pathways and processes [33]. Yeast are faster and less expensive to grow than other eukaryotes [7, 8] and their genomes have been extensively studied, providing resources to target and delete a gene of interest (e.g. Saccharomyces Genome Database (SGD) [15] and Yeastract [1, 60, 87]). Results of these studies can be used to increase understanding of the corresponding conserved protein, DNA element or pathway in higher eukaryotes (e.g. nematodes, fruit flies, or humans), or can enhance the knowledge of differences that have led to evolution [41].

1.1 Regulation of transcription initiation

Transcriptional regulation leading to initiation of DNA polymerization in yeast (reviewed in [33, 84, 85]) involves both trans-factors (proteins) and cis-elements (Figure 1.1). Trans-factors include co-factors (which may be chromatin modifying proteins), transcription factors (TFs) and the general transcription machinery (TATA binding protein (TBP), RNA polymerase II and general transcription factors TFIIB, E, F and H). Cis-elements are typically located upstream of the transcriptional start site (TSS) on the DNA (the promoter region) and include core promoter elements, and upstream activating and repressing sequences [33, 84, 85]. Upstream activating sequences are similar to enhancers in higher eukaryotes, except that enhancers are commonly found both upstream and downstream of the TSS [33, 84, 85].

Important cis-elements are found in the core promoter region which integrates upstream signals to initiate transcription. These include the initiator element (defining the TSS) and the TATA box (binds TBP), usually found between 40-120 base pairs (bp) upstream of the TSS [96]. Less than one fifth of yeast promoters contain a TATA element in the core promoter; the rest are TATA-less and generally display

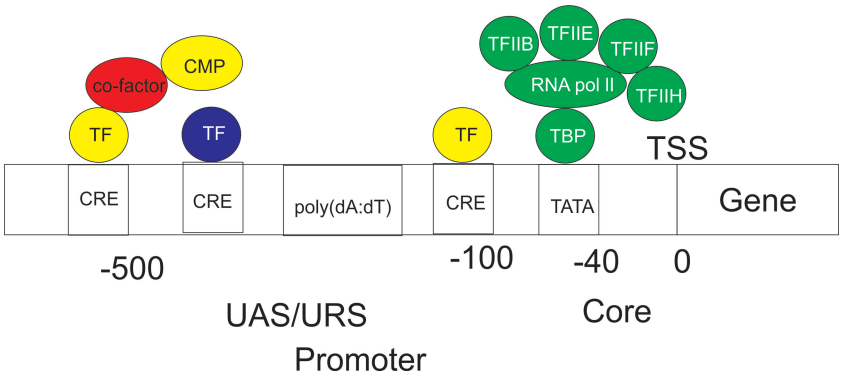


Figure 1.1: Trans-factors and cis-elements important for transcriptional initiation. TF = transcription factor, CRE = cis-regulatory element, TBP = TATA-binding protein, CMP = chromatin modifying protein, TSS = transcriptional start site, UAS/URS = upstream activating sequence/upstream repression sequence.

more constitutive activity [33].

Other cis-elements include upstream activating and repressing sequences. These sequences include binding sites for TFs called cis-regulatory elements (CREs). Activating sequences also include long poly(dA:dT) tracts that prevent repressive chromatin structures [21, 42, 74, 81, 101]. CREs are typically found between 100-500 bp from the TSS [35]. Based on the number and identity of CREs, different promoter architectures exist, ranging from single motif promoters where one TF binds to one CRE to influence downstream expression, to complex co-regulated promoters involving regulation by multiple TFs binding to multiple sites [35].

1.1.1 Binary and higher order TF-CRE interactions

When a TF binds to a CRE, it may affect downstream expression on its own (binary interaction) or in a higher order manner. Binary interactions include one bound TF interacting directly with the general transcriptional machinery, recruiting co-factors or inhibiting the binding of trans-factors [5, 10, 33, 37, 73]. Co-factors include co-activators, which recruit the general transcriptional machinery to the promoter, and co-repressors, which exert their effects by modifying the chromatin (nucleosome

position and histone state) so that the general transcriptional machinery cannot bind to the promoter [33]. These interactions are binary in that any effects on transcription can be attributed to the binding of one TF to one CRE. A higher order interaction is defined when the interaction of multiple TF molecules (same or different) with each other, with co-factors or with the general transcriptional machinery leads to additional effects on transcription that are not available to a TF-CRE combination on its own [10, 25, 37, 49, 50, 59, 62, 63, 70].

Higher order interactions between one TF and multiple CREs are well-documented. Repression of the Lac operon in *E. coli* [62, 63] is a classic example illustrating this. The Lac repressor (LacR) binds to two palindromic operators simultaneously, forming a tetramer. A central operator (O1) with high affinity for LacR interacts with O3 or O2 (weaker operators located ≈100 bp upstream and ≈400 bp downstream of O1, respectively) forming a DNA loop. The DNA loop, which requires O1, acts to increase the local concentration of repressor molecules, making it more likely that weak operators will become occupied. Repression of transcription is therefore increased beyond what would be expected if the Lac-bound operators were acting independently [62].

Thermodynamic models of transcriptional regulation [5, 6, 10, 25, 31, 37, 78] provide a mathematical framework for understanding how much transcription occurs based on binary and higher order TF-CRE interactions. To create a model that closely matches experimental data, the concentration of TFs along a promoter must be accurately estimated. This is related to the binding affinity of the TF for the sequence and the environmental condition [35] if the interactions of TFs with the DNA are binary in nature, but higher order interactions change the concentration of TFs in unexpected ways, as seen above [62]. Therefore, identifying their presence helps to improve predictability.

Gertz and colleagues [31] created a library of synthetic promoters and used the

thermodynamic model in [10] to estimate parameters describing the strength of binary and higher order TF-CRE interactions. They found that the yeast transcriptional repressor Mig1 engaged in binary interactions with two Mig1-specific CREs that resulted in different levels of transcription due to differences in Mig1 binding affinity. Strong and weak Mig1 CREs also interacted in a higher order cooperative manner to create the same amount of repression as two strong sites [31].

1.1.2 Physical mapping of TF-CRE interactions

To gain additional insights into transcriptional regulation that may not be captured in a synthetic promoter environment, it is necessary to investigate TF-CRE interactions in their native context. Prediction of TF concentrations from native promoter sequence often starts with identifying which TFs physically bind to the promoter, where they bind and when [78]. The physical presence of TF-CRE interactions can be mapped with high throughput technologies involving chromatin immunoprecipitation (ChIP). For example, ChIP-on-chip [35], ChIP-seq [65] and mCHIP [52] have been used to dissect where and when hundreds of proteins bind in the yeast genome. Performing ChIP involves cross-linking TFs to the DNA under different environmental conditions, sonicating the chromatin into small fragments and eluting the tagged protein and the DNA physically bound to it with an antibody [65]. The cross-linking is then reversed, and the samples are separated into protein and DNA extracts. mCHIP is a modified ChIP protocol, where gentler sonication is used so that chromatin remodeling factors and other co-factors interacting with the TF at the promoter are not dislodged. These proteins are eluted along with the TF and mass spectrometry can be used to identify all the proteins interacting with the tagged TF at the promoter [52].

As for the DNA eluted from the ChIP procedure, variations on PCR are used to

amplify the sequences bound by the TFs and the products are either hybridized to a microchip containing DNA spots from locations in the genome (ChIP-chip) [35] or are sequenced directly by next generation sequencing (NGS) (ChIP-seq) [65]. The latter allows identification of precise TF binding site locations.

The dissection of where and when a TF binds a CRE is important for establishing the potential for regulation. However, the transcription level from promoters containing CREs for these TFs is not typically measured in ChIP experiments. Thus, it is not known which TF-CRE interactions contribute to downstream expression under a specific environmental condition. Therefore, to gain insight into how cells control and coordinate gene expression, we must decipher the functional impact of TF-CRE interactions that have been physically mapped.

1.1.3 Functional mapping of TF-CRE interactions

The functional effect of a biological component (TF or CRE) on gene expression is commonly deduced by measuring a significant phenotypic effect upon its mutation or ablation (perturbation) (e.g. [63]). The process of perturbation and subsequent phenotype measurement is referred to here as a functional genetic analysis. For large scale functional genetic analysis of many components, a high-throughput approach like the synthetic genetic array (SGA) [89, 90] can be used. Here a query strain is mated with a library of deletions to determine which deletions impact a trait.

Historically, many different traits have been used as a phenotypic readout for deletions [95]. Two examples of univariate, quantitative traits are fitness [3, 55, 89, 90] and reporter expression. Reporter expression can be measured by first fusing a promoter (containing CREs) or gene (encoding a TF) of interest to a fluorescent protein like green fluorescent protein (GFP), and measuring fluorescence by flow cytometry [68] or another fluorescence imaging technique [46].

To identify interactions, the effect of individual component deletions are compared to the effect of combinatorial deletion. If the combined deletion leads to an unexpected effect given the individual deletions, an interaction between components is identified [55, 63, 89, 90].

Neutrality models help to quantitatively identify interactions

Identifying interactions between components becomes consistent with a formal definition of what we expect given that there is no interaction. Independence can be defined by various neutrality functions (null models) [29, 55]. Two such functions are additive and multiplicative. The additive model was originally used to explain genetic interactions, and this model was favoured simply for its statistical tractability [69]. The multiplicative model was used later by population geneticists because it was observed that when two genetic loci have multiplicative fitness, they evolve independently of one another [69]. It has been suggested that certain traits may be better suited for one independence model over another [69].

For a pair of components and a univariate, quantitative trait, the additive neutrality function predicts that the trait observed after deleting both components will be equal to the sum of effects caused by deleting each component individually [29, 69]. The multiplicative neutrality function predicts that the effect of deleting both components is equal to the multiplication of individual effects [29, 55, 69]. Using *APred* and *MPred* to indicate additive and multiplicative neutrality predictions, respectively, x to refer to one component and y to refer to the other, the trait of the double mutant, T_{xy} in the absence of an interaction, is expected to be given by:

$$\begin{aligned} Apred(T_{xy}) &= T_x + T_y, \\ Mpred(T_{xy}) &= T_x \times T_y. \end{aligned} \tag{1.1}$$

Deviations from neutrality can be positive (greater than expected) or negative

(less than expected). Thus, interactions may be classified simply as positive or negative [69]. However, interactions are often classified using a plethora of other descriptive terms [69]. To choose terms that make no assumptions about the beneficial or detrimental effects on the cell, positive interactions (greater than expected deviations) can be called synergistic [55] or cooperative, and negative interactions (less than expected deviations) can be called antagonistic.

There are a couple of other interaction classifications that are useful regardless of the trait or components contributing to it. Extreme synergistic interactions may result in synthetic lethality [89], where cell death and lack of trait measurement is the ultimate response. This indicates that the two deleted components may function within partially or fully redundant parallel pathways mediating an essential process [90]. In other cases, interactions between two components can be further classified as epistatic. "Epistasis" is sometimes used interchangeably with "genetic interaction" [69], but also refers to a case where deleting one of the two genes has an effect only when the other gene is present [68]. The effect of this gene is completely masked in the double mutant because the trait matches the effect of deleting the other gene. Epistatic interactions suggest that one gene requires the other gene for its function, indicating that the two genes may be in the same pathway [2, 68].

Limitations of neutrality models

Application of the additive and multiplicative neutrality models to a particular dataset to identify interactions will be limited by the quality of the data. This is a result, in some cases, of the sensitivity of the approach used to collect the data. In particular, if the predicted trait cannot be accurately measured, then deviations from that prediction are likely and interactions will be falsely identified.

For example, Mani and colleagues [55] observed that when single component deletions had a minor impact on the trait (growth rate in this case), the multiplicative

and additive definitions were equally conservative in the number of interactions they identified. However, when single mutants decreased the trait substantially, the additive definition was less conservative in that it had a tendency to identify positive interactions between two genes. Specifically, summing the growth rates of the single mutants sometimes resulted in an prediction of negative growth. This prediction was meaningless for growth rate calculations that could not account for negative growth. In these cases, any double mutant that grew would signify a greater than expected deviation, and interactions were falsely identified [55].

Based on thermodynamic considerations, the simultaneous independent interaction of two activating TFs with the general transcriptional machinery leads to a multiplicative effect on transcription [5, 25, 37]. However, Bintu and colleagues [5] found no justification that the combined function of two repressors can be multiplicative.

Each null model suffers from its own set of limitations. Therefore, interactions that are identified under multiple null models are less likely to be an artifact of those limitations.

Applications of additive and multiplicative models for identification of interactions

The additive and multiplicative neutrality models have been applied to the functional genetic analysis of biological components to identify interactions. A common application after perturbing components and measuring trait values is to calculate synergy factors. A synergy factor is simply a measure of the deviation from a neutrality model normalized to the expected trait when no interaction is present. For example, [56, 66] calculated synergy between two interacting TFs or protein domains given the null expectation of additivity. The synergy factor (SF) was:

$$SF = T(TF1 + TF2)/T(TF1) + T(TF2), \tag{1.2}$$

where *TF1* and *TF2* represented the effect on the measured trait T attributable to each TF or protein domain. Effects were measured as positive deviations relative to a control strain. Values significantly greater than 1 were considered indicative of synergy between the two TFs or protein domains [66].

Synergy factors were also calculated as a deviation from multiplicative neutrality in [3, 50]. The basic equation used in [50] to measure synergy between pairs of yeast repressor TFs is given below:

$$SF = T(TF1 + TF2)/T(TF1) \times T(TF2), \tag{1.3}$$

where *TF1* and *TF2* were the positive effects, relative to control, of each TF on the measured trait, T [50].

The synergy factors in (1.2) and in (1.3) are limited to identifying interactions between two components, which poses a problem when investigating higher order interactions, for example between a TF and two or more of its CREs. In contrast, multiple linear regression [14] provides a framework to identify interactions involving any number of components (variables). A multiple linear regression model [51, 68] predicts a measured trait (Y_i) by adding the impact of predictors, i.e., variables and variable combinations, to a baseline trait (Y_0). Formally, the regression equation can be written as follows:

$$Y_i = Y_0 + \beta_1 x_{i1} ... + \beta_p x_{ip} + \epsilon_i, \tag{1.4}$$

where $x_{i1} ... x_{ip}$ are the different combinations of predictors, the β coefficients are the magnitude of the effect attributable to each predictor, i is the number of measurements, and ϵ is the error term. The error term is basically the amount of variance in the trait that cannot be accounted for by the predictors and Y_0. A β coefficient of a specific magnitude and sign is estimated for each predictor based on the measured

trait. Regression estimates the subset of predictors that best predict the trait as those that have a coefficient that significantly deviates from zero (a predictor whose coefficient does not deviate significantly from zero has a statistically insignificant effect on the measured trait).

Like the synergy factor, multiple linear regression has been used to identify interactions between two variables [51]. This application involved identification of non-additive interactions between pairs of nucleotides within a CRE [51]. Expression values were measured from promoters containing a 52 bp CRE with single or double nucleotide substitutions. For each pair of nucleotides, two regression models were compared to see which one fit the double nucleotide expression data better. In the first case, the model predicted expression of the double mutant by adding the effects of each individual deletion to the wildtype trait. The second model included an interaction term. In a number of cases, the model with the interaction term predicted the observed effect of the double deletion better, indicating a non-additive interaction between the two nucleotides.

Multiple linear regression has also been applied to analyze interactions between more than two variables [68]. In this application, three Boolean variables were examined: an inducing signal (s) could be either ON or OFF, and each of two TFs (x and z) could either be present or absent (deleted). The downstream gene was fused to a fluorescent reporter gene and log2-transformed mean reporter protein expression values (fluorescence intensity values measured using flow cytometry) were collected for strains with all combinations of TF deletions (x, z, and both x *and* z) and under both signal conditions. Multiple linear regression was then used to identify which variables and interactions significantly contributed to trait variability. A significant interaction term between two variables in the model was defined as a deviation from multiplicative neutrality. This was because adding two log2-transformed expression

values is the same as multiplying them on the linear scale. A significant interaction between x and z for example meant that the effect of deleting both x and z on reporter expression could not be accounted for by multiplying the effects observed by deleting each individually.

The regression equation used in [68] is given by:

$$\begin{aligned} Y(s,x,z) &= Y_0 + \beta_s s + \beta_x x + \beta_z z \\ &+ \beta_{sx} sx + \beta_{sz} sz + \beta_{xz} xz + \beta_{sxz} sxz \\ &+ \epsilon, \end{aligned} \tag{1.5}$$

$$\begin{cases} s = 1, & \text{if signal is ON,} \\ x, z = 1, & \text{if } x,z \text{ is/are deleted,} \\ s, x, z = 0, & \text{otherwise.} \end{cases}$$

This equation is important since it can be extended to analyze the quantitative functional effects of any number of components by adding additional predictors corresponding to individual and combinatorial perturbations.

1.1.4 Research objective

The main objective of my thesis was to investigate how multiple CREs quantitatively define transcriptional level (individually and/or combinatorially) in their native promoter context. To meet this objective, I focused on CREs in the yeast *RNR3* promoter.

The *RNR3* gene encodes a subunit of the highly conserved ribonucleotide reductase enzyme [44], which is essential for DNA replication and repair in all organisms. Dysregulation of this enzyme leads to increased mutation rates in yeast [12], a leading contributing factor to the hallmarks of cancer [34].

An important regulator of the *RNR3* promoter, Crt1 [40], is part of a conserved family of winged-helix proteins called RFX [27]. RFX proteins are found to varying extent throughout the eukaryotic kingdom [16, 40]; mammals including humans have seven identified proteins, whereas *C. elegans* and *S. cerevisiae* have only one. The budding yeast RFX protein is named by the phenotypic effect of its deletion, *c*onstitutive *r*nr *t*ranscription [103], and is referred to throughout this thesis as Crt1.

The *RNR3* gene has low basal expression and is induced highly in the presence of DNA damage [44]. Tight repression is assumed to occur via binding of Crt1 to three conserved X-box CREs identified in the promoter region [40]. Crt1 recruits the co-repressor complex Ssn6-Tup1 [40], which alters the chromatin structure and blocks *RNR3* transcription [53]. Upon DNA damage, Crt1 becomes phosphorylated via the conserved DNA damage checkpoint pathway [26], which leads to dissociation from the X-boxes and de-repression of the promoter [40]. The *RNR3* promoter has the most identified X-boxes of all the *RNR* genes, making it ideal for studying combinatorial transcriptional regulation by multiple CREs.

1.2 Rationale

The X-boxes in the *RNR3* promoter were initially identified by Huang and colleagues [40], who also characterized the functional role of their interaction with Crt1 on *RNR3* transcription. This characterization involved a functional genetic analysis. Each X-box was mutated at two positions known to be important for Crt1 binding. Certain X-boxes were also mutated combinatorially. Additionally, the effect of deleting Crt1 on *RNR3* transcription was compared to the effect observed after perturbation of both Crt1 and the X-boxes. However, the functional genetic analysis in [40] was not systematic, meaning that not all possible perturbations were performed. I wanted to expand on previous work by systematically investigating how

Crt1 and the three X-boxes combinatorially affect *RNR3* expression.

1.2.1 Hypotheses

Examining the results from the Huang study [40] allowed me to generate two testable hypotheses:

Hypothesis 1:

There are higher-order Crt1-mediated interactions between X-boxes in the RNR3 promoter.

The mutation of each X-box individually resulted in a small increase in expression compared to no mutations, but the mutation of all three resulted in a much larger increase than would be expected by simply adding or multiplying the effects of the individual mutations [40]. This suggested higher-order synergism between the three X-boxes defined under both the additive and multiplicative models, but since only certain combinations of sites were mutated, it was not known how much the untested combinations contributed to the interaction.

I also expected higher order Crt1-X-box interactions based on the evidence for higher order RFX-X-box interactions in higher eukaryotes [16, 20, 27]. Although there are differences in the functions performed by RFX protein family members between higher and lower eukaryotes [16, 30, 40, 47, 76], conservation of protein domains is one indication of conserved regulatory mechanisms. Some RFX proteins are known to dimerize via a structurally and functionally independent domain in the C-terminus [24, 76]. Independence between the processes of DNA binding and dimerization allows these proteins to bind as monomers to non-palindromic X-boxes, and then to dimerize, likely by looping of the intervening DNA [76]. The ability for an RFX dimer to simultaneously contact distant sites would have a cooperative effect similar to the binding of Lac dimers to distant operators. There is homology between

the C-terminus of Crt1 and that of dimerizing RFX proteins [24].

Additionally, I expected conserved regulatory mechanisms based on the similarities in the orientation and arrangement of upstream regulatory sequences between higher eukaryotic X-box-containing promoters and the *RNR3* promoter. Like the X-box containing Class II *DRA* promoter [17], the *RNR3* promoter has two X-boxes 5' to a third X-box which are inverted in orientation [40]. Both promoters also have a poly(dA:dT) tract directly upstream of an X-box [15, 17].

Hypothesis 2:

The effects of the X-boxes are mediated solely by Crt1.

This hypothesis came from the observation that in the absence of Crt1, mutating all three X-boxes did not impact expression [40].

1.2.2 Approach

Functional genetic analysis of Crt1 and X-boxes

To test my two hypotheses, I systematically perturbed four variables (Crt1 and each of the three X-boxes) and measured the effect of these perturbations on a trait: transcription from the *RNR3* promoter quantified by the green fluorescent protein (GFP) reporter. My approach involved three steps:

1. **Systematic perturbation of Crt1 and X-boxes in *RNR3* promoter**

   I constructed and validated all possible *RNR3* promoter variants in strains where the Crt1 gene and/or the three X-box elements had been eliminated.

   X-boxes were eliminated because I wanted an approach that ensured elimination of Crt1 binding. In the Huang study [40], the three X-boxes were mutated at two of 13 nucleotide positions. Since mutation of other X-box nucleotides was also found to reduce Crt1 binding [40], mutating the two nucleotides may not have abolished Crt1 binding and the functional impact of the corresponding

TF-CRE interaction may only have been partially captured. To eliminate the X-boxes, I wanted to use replacement over deletion since the latter would change the location of adjacent sites relative to the start site. Positional effects on transcription are known to occur upon deletion of as little as 10 bp [18, 79].

I also wanted to use an X-box replacement approach that would allow for investigation of the level at which cooperativity occurred and the quantitative impact of the factor concentration on transcriptional level. If I identified higher order synergistic interactions with my approach, they could in principle occur at several levels - a Crt1 dimer could interact with multiple X-box elements, or cooperativity could occur at the level of the Crt1-recruited co-repressor Ssn6.

I therefore decided to first replace the 13 bp X-boxes with the 19 bp operator sequence for the bacterial Tet repressor [23]. The rationale was that I could build on previous work by another student in our lab (Daniel Jedrysiak) who created a synthetic system where the bacterial Tet repressor (TetR) sequence was fused to the sequence for Ssn6, and the concentration of this TetR-Ssn6 fusion protein was controlled by a β-estradiol inducible promoter (see [43]). In addition to using this system, we could fuse Crt1 to TetR. We could then examine how much cooperativity Crt1 or Ssn6 was responsible for by recruiting each of them separately to multiple sites in the *RNR3* promoter. Additionally, we could examine how the concentration of Crt1 was related to the transcriptional level of *RNR3* in the absence of different X-boxes by titrating TF concentration with β-estradiol. This would facilitate the development of a thermodynamic model.

2. **Characterization of trait values from *RNR3* promoter variants**

I characterized trait values from the different promoter variants by measuring GFP expression using flow cytometry.

I analyzed the functional impact of interactions between Crt1 and X-box elements in the *RNR3* promoter using a transcriptional reporter assay where GFP was expressed from promoters where Crt1 and/or the three X-boxes were systematically eliminated. Our lab has verified that the level of GFP driven from an integrated *RNR3* promoter correlates well with native promoter mRNA expression (qRT-PCR data from Hilary Phenix). I chose the *ADE2* locus for integration because its disruption results in a red colony colour in the absence of adenine, providing an additional method (aside from drug selection media) for identifying positive transformants.

3. **Quantitative mapping of how Crt1 and X-boxes related to trait by Multiple Linear Regression**

   I investigated the relationship of the varied components on the trait using a multiple linear regression approach that expanded on previous analysis of binary TF-TF interactions in the presence and absence of a drug [68].

   I chose to analyze the four variables of interest (Crt1 and the three X-boxes) in one comprehensive regression model. My rationale for this was that inclusion of more variables known to impact the trait would allow for better trait prediction. I chose multiple linear regression to identify relationships between the four variables over other approaches like synergy factors [50, 66] because, as discussed in Section 1.1.3, regression allows for investigation of higher order interactions between more than two variables. Since the additive [51, 56, 66] and multiplicative [50, 68] neutrality models had each been used in previous studies to identify interactions, but also had potential limitations (see Section 1.1.3), I chose to analyze deviations from both models using linear and log2 transformed expression values respectively.

Based on previous research [40, 53], I initially assumed that the conserved X-box

elements found within the *RNR3* promoter established a fairly straightforward mechanism of transcriptional regulation. This mechanism is depicted in Figure 1.2 and takes into account Crt1 its three CREs. No other factors are included because repression of *RNR3* is thought to depend predominantly on X-box-bound Crt1 [40]. Consequently, higher order interactions caused, for example, by Crt1 itself, by the protein complexes recruited to the promoter by Crt1, or by subsequent changes in chromatin structure, should all be detectable through systematic individual and combinatorial perturbations of Crt1 and the three X-boxes. The idea was that the data generated in these experiments could be used to facilitate the development of a comprehensive thermodynamic model of *RNR3* transcriptional regulation.

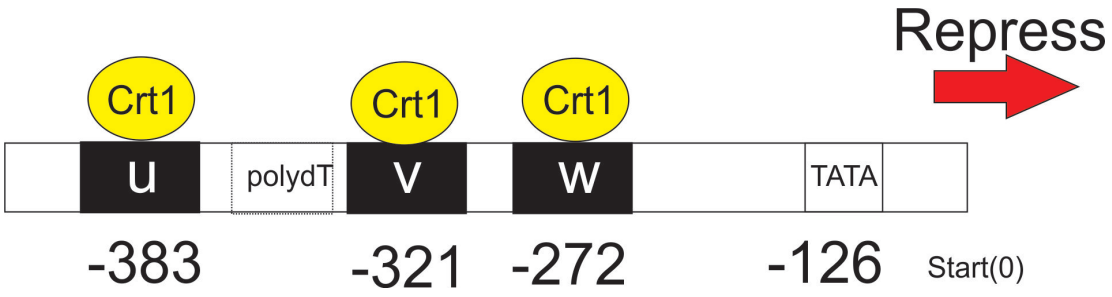


Figure 1.2: A model for Crt1-X-box mediated repression of the *RNR3* promoter. The X-boxes located at 383 bp, 321 bp and 272 bp upstream from the transcriptional start site are labeled u, v and w respectively. The TATA box is located 126 bp upstream of the start site. A poly(21)T region is located 5 bp upstream of the X-box v. Sequence information is from [40] and SGD [15]. The simplest model is shown, which is that Crt1 binds independently to each X-box. Only Crt1 binding is shown for simplicity. While other factors are involved in X-box-mediated repression of *RNR3* transcription, their functions are thought to be dependent on Crt1 binding (see text).

Chapter 2

Methods

2.1 List of strains

A complete list of yeast strains used in this thesis is given in Table 2.1.

2.2 Functional genetic analysis of Crt1 and X-boxes

2.2.1 Strain construction and validation

The steps for *RNR3* promoter variant construction are shown in Figure 2.1. To generate Construct 1 (used to eliminate Crt1), a forward and reverse primer were designed with 40 bp homology to the region upstream and downstream of the *CRT1* ORF and 20 bp homology to the beginning and end of the URA3 selection cassette. These primers were used to amplify the URA3 selection cassette from a PRS plasmid.

There were eight variations of Construct 2 (used to measure *RNR3* promoter-driven GFP) containing none, one, combinations of two, or all three X-box replacements. Labels u, v and w are used to refer to the X-boxes located respectively 383 bp, 321 bp and 272 bp upstream from the transcription start site (see Figure 1.2).

Table 2.1: List of Strains. See text for details on strain construction including explanation of constructs and abbreviations. All strains were haploid, and had a genomic background of: a, b, or c. a = YBA15, a BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) [9] variant created by Daniel Jedrysiak containing the construct *GAL4::HIS3-GEV*. See [57] for details on the GEV construct and uses. b = Y8205, obtained from Boone lab at University of Toronto. The genotype for this strain was MATα, can1Δ ::*pSTE2*-Sp his5 (complements *HIS3* deletion), lyp1Δ ::*pSTE3*-LEU2, his3Δ1 leu2Δ0 ura3Δ0. c = BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) [9].

Strain name	variants	Genomic background	Genotype
autofl	1	a	
wt	1	a	*ADE2::KAN-pRNR3-GFP*
x	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP*
u	2	a	*ADE2::KAN-pRNR3*-GFP*, u::Tet operator or u::scrambled sequence
v	2	a	*ADE2::KAN-pRNR3*-GFP*, v::Tet operator or v::scrambled sequence
w	2	a	*ADE2::KAN-pRNR3*-GFP*, w::Tet operator or w::scrambled sequence
uv	2	a	*ADE2::KAN-pRNR3*-GFP*, u,v::Tet operators or u,v::scrambled sequences
uw	2	a	*ADE2::KAN-pRNR3*-GFP*, u,w::Tet operators or u,w::scrambled sequences
vw	2	a	*ADE2::KAN-pRNR3*-GFP*, v,w::Tet operators or v,w::scrambled sequences
uvw	2	a	*ADE2::KAN-pRNR3*-GFP*, u,v,w::Tet operators or u,v,w::scrambled sequences
xu	2	a	*CRT1::URA3*, *ADE2::KAN-pRNR3*-GFP*, u::Tet operator or u::scrambled sequence
xv	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, v::Tet operator or v::scrambled sequence
xw	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, w::Tet operator or w::scrambled sequence
xuv	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, u,v::Tet operators or u,v::scrambled sequences
xuw	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, u,w::Tet operators or u,w::scrambled sequences
xvw	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, v,w::Tet operators or v,w::scrambled sequences
xuvw	2	a	*CRT1::URA3,ADE2::KAN-pRNR3*-GFP*, u,v,w::Tet operators or u,v,w::scrambled sequences
query strain	1	b	*CRT1::BFP-pACT1-NATMX-pRNR3-GFP*
DMA library strains	370	c	*FO::KANMX*
SGA library strains (named by FO deleted)	≈380	progeny of b+c	*CRT1::BFP-pACT1-NATMX-pRNR3-GFP, FO::KANMX*
neutral control strains	3	c	*HIS3::KANMX* or *URA3::KANMX* or *HO::KANMX*
x+Ace2	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP, ACE2::NATMX*
x+Aft1	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP, AFT1::NATMX*
x+Spt10	1	a	*CRT1::URA3,ADE2::KAN-pRNR3-GFP, SPT10::NATMX*
x+Met18	1	a	*CRT1::URA3,ADE2::KAN-pRNR3-GFP, MET18::NATMX*
x+Spt3	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP,SPT3::NATMX*
x+Spt8	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP,SPT8::NATMX*
x+Ada2	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP,ADA2::NATMX*
x+Mot3	1	a	*CRT1::URA3,ADE2::KAN-pRNR3-GFP, MOT3::NATMX*
x+Rox1	1	a	*CRT1::URA3, ADE2::KAN-pRNR3-GFP,ROX1::NATMX*
xuvw+Spt3	1	a	*CRT1::URA3, ADE2::KAN-pRNR3*-GFP* u,v,w::scrambled sequences, *SPT3::NATMX*
xuvw+Spt8	1	a	*CRT1::URA3, ADE2::KAN-pRNR3*-GFP* u,v,w::scrambled sequences, *SPT8::NATMX*
xuvw+Mot3	1	a	*CRT1::URA3, ADE2::KAN-pRNR3*-GFP* u,v,w::scrambled sequences, *MOT3::NATMX*
xuvw+Rox1	1	a	*CRT1::URA3, ADE2::KAN-pRNR3*-GFP* u,v,w::scrambled sequences, *ROX1::NATMX*

To generate variants of Construct 2 one at a time, the *RNR3* promoter was varied by replacing the X-box(es) with a Tet operator or a scrambled sequence (a and b in Figure 2.1). The sequence for the 19 bp bacterial Tet operator was obtained from [23]. The original X-box sequences were scrambled using the function Randperm in MATLAB. X-boxes u and w were randomly shuffled, and the same shuffled sequence used for u was used to replace v. This was done to keep the base composition fairly consistent. Specifically, X-box u differed from X-box w substantially (5 bp differed). However, X-box u and v only differed by one base (A for u and G for v) at position six of 13, a position that has not been reported to be important for Crt1 binding [40]. Since it was shown in [40] that changing the last base on the 3' end from a C to an A resulted in the greatest disruption of Crt1 binding, it was made certain that all shuffled sequences did not have a C at the 3' end. A summary of the original and replaced sequences is given in Table 2.2.

In order to replace the 13 bp X-boxes with the 19 bp Tet operators, primers were designed (with the help of Daniel Jedrysiak) where the operator was added as an overhang in both directions (Figure 2.2). The distance from the end of each operator to the transcriptional start site after replacement was kept consistent with the distance from the end of the X-box to the start site in the unaltered promoter. This was done to maintain the distance to the transcriptional start site as well as the distance between different X-boxes. Six additional bp were thus replaced upstream of each X-box.

The middle X-box was situated directly next to the poly(21)T region. With the standard primer design, the 5' end of the reverse primer could anneal anywhere along the poly(dA:dT) tract. To prevent the primer from annealing improperly, a longer linker reverse primer (spanning the poly(21)T region and the flanking region) was designed to replace this X-box along with a standard length forward primer (Figure

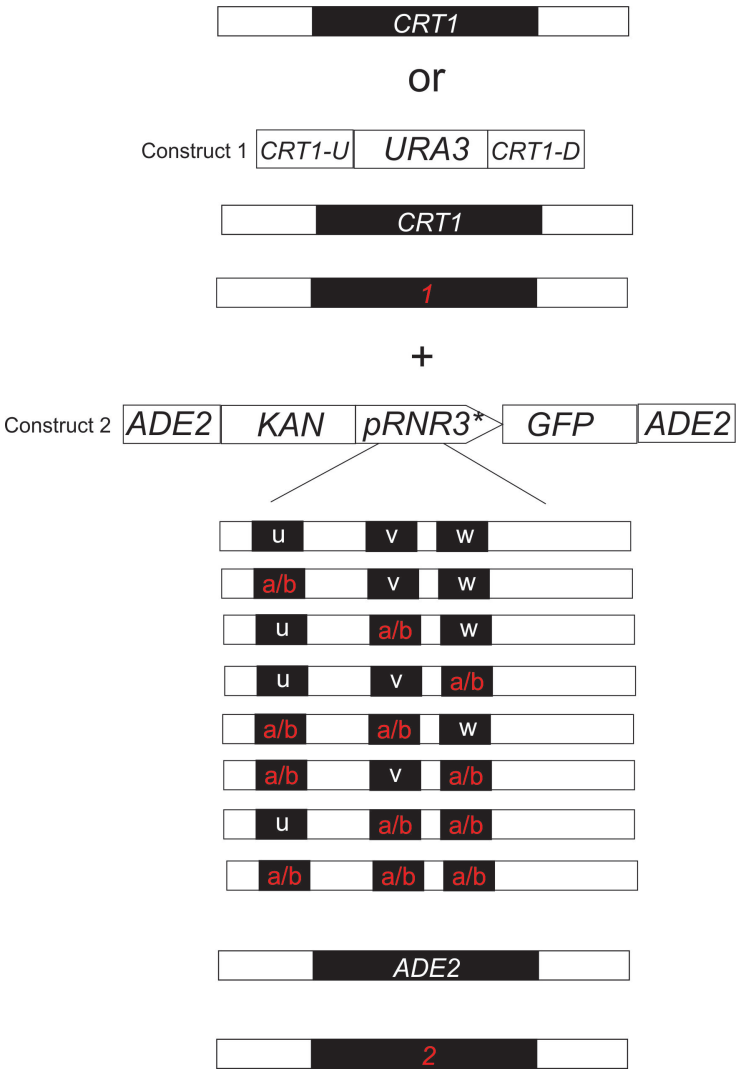


Figure 2.1: Approach for systematic genetic perturbation of Crt1 and X-boxes u, v and w. Construct 1 was used to replace the *CRT1* ORF. It contained a URA3 selection cassette flanked by the sequences *CRT1−U* and *CRT1−D*, which were 40 bp long and homologous to the regions immediately upstream and downstream of the Crt1 gene. Construct 2 was integrated into the *ADE2* locus in strains that had the Crt1 gene intact or replaced with Construct 1. Construct 2 contained different variants of the *RNR3* promoter (*pRNR3*) driving the expression of GFP as well as a *KAN* expression cassette, which contained the Kanamycin resistance gene that enabled selection by growth on plates with the drug G418. The GFP and KAN expression cassettes were flanked by two 200 bp fragments that directed integration of the construct to the *ADE2* locus (the extra long regions of homology are referred to as "landing pads"). The variants of Construct 2 contained all the possible combinations of the three X-boxes labeled u, v and w. * indicates the variable component. X-boxes were replaced using either Tet operator sequences (a) or scrambled X-box sequences (b).

Table 2.2: Summary of X-box replacement elements. TetO = Tet operator replacement approach, Scrambling = Scrambling replacement approach.

Strain	Original element	Replaced element
TetO		
u	ctacggttgtcacagcaac	tccctatcagtgatagaga
v	tcgtggttgtcgcagcaac	tccctatcagtgatagaga
w	aacccggttgccatggcga	tccctatcagtgatagaga
Scrambling		
u	ttgtcacagcaac	gaacccatatgca
v	ttgtcgcagcaac	gaacccatatgca
w	gttgccatggcga	accatgaggtggt

Figure 2.2: Example of primer set used to replace an X-box with the Tet operator sequence. The red box represents the sequence for replacement and the black box represents the sequence complementary to the DNA to amplify.

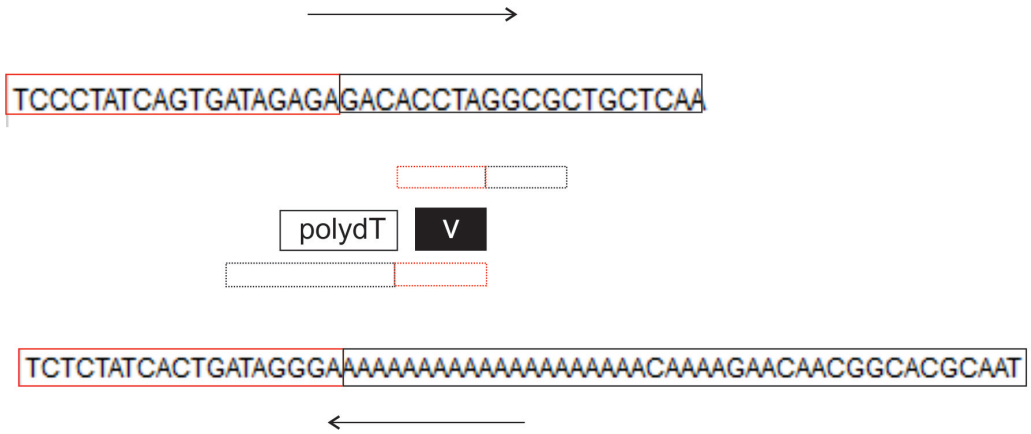


Figure 2.3: Example of primer set used to replace an X-box adjacent to a poly(dA:dT) region. The red box represents the sequence for replacement and the black box represents the sequence complementary to the DNA to amplify.

2.3).

The X-boxes were replaced with scrambled sequences using a similar approach (Figure 2.4).

Each construct for X-box replacement was then pieced together by consecutive PCR reactions. The *RNR3* promoter variant pieces were first joined to the rest of the *RNR3* promoter, and then to two other pieces of DNA amplified from other strains in the lab (*ADE2-KAN* and *GFP-ADE2*) to create a full construct for integration into the yeast *ADE2* locus. PCR reactions were carried out using an authorized thermocycler and Phusion polymerase enzyme.

After each PCR reaction, the product of interest was separated from non-specific bands by gel electrophoresis at 90V for 30 minutes on a 1% agarose gel and purified using a gel extraction kit (BioBasic, Inc. Molecular Biology Kits - EZ-10 spin column DNA gel extraction kit). Alternatively, the sample was purified directly by using a PCR clean-up kit (BioBasic, Inc. Molecular Biology Kits - EZ-10 spin column PCR products purification kit). This was done if the reaction was known to result in a product with very little non-specific bands. In this case, a small amount of the product (2 μl) was set aside for size verification by gel electrophoresis. UV visualization of gel

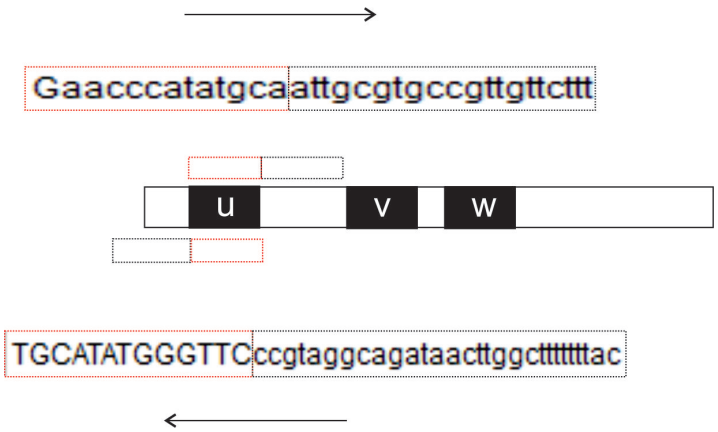


Figure 2.4: Example of primer set used to replace an X-box with a scrambled sequence. The red box represents the sequence for replacement and the black box represents the sequence complementary to the DNA to amplify.

bands (by a hand-held light or using an ImageQuant 300 machine and software) was possible by the use of 2 μl of RedSafe^{TM} nucleic acid staining solution for every 50 μl gel solution.

After the product of interest had been purified, the concentration of DNA was measured in ng/μL using the NanoDrop spectrophotometer (ND-1000, Thermo Scientific) from the Blais lab. This was done after each consecutive PCR reaction, including the final reaction forming the full construct.

The full constructs were then integrated into the yeast genome using the yeast transformation protocol in [32] to generate the Crt1 and *RNR3* promoter variant strains.

A modification to the amount of DNA was made to the transformation protocol in [32]. In cases where the construct had 40 bp homology to the integration locus (e.g. the construct for *CRT1* replacement), approximately 1500 ng of DNA was used. In cases where the construct had landing pads (extra long regions of homology to the locus for integration, i.e. the construct for *ADE2* replacement), approximately 1000 ng of DNA was used for transformation. Cells were also recovered for up to one day before plating on selection media.

The *CRT1* deletion was performed first. Then, after confirmation of this deletion by PCR (see next section), the wild-type promoter construct or constructs with X-box replacements were introduced to the confirmed Crt1 deletion strain and the wild-type YBA15 strain. Single CRE eliminations were introduced first, followed by double eliminations and finally triple eliminations. After each round of transformations and selection on plates with G418, high yield, high quality genomic DNA was isolated using the organic extraction protocol in [39]. This protocol produced a high yield of DNA because cells were grown to a high density in media overnight and the quality was high due to an enzymatic RNA degradation step. The fragments containing the X-box deletions were amplified from this genomic DNA and combined by PCR to generate the constructs for the next round of eliminations. For most strains, three biological replicates were stored at -80°C in 15% glycerol.

PCR confirmation of Crt1 deletion

After transformation of the Crt1 deletion construct and selection on plates lacking uracil, a few transformants were selected and patched onto separate plates to create a source for storing at -80°C in 15% glycerol. Genomic DNA was isolated from these patches using an organic extraction protocol for quick extraction straight from a colony. This protocol was a condensed version of the high quality extraction protocol described above [39]. The original protocol was modified by using one tenth of the amount of each reagent and eliminating the RNA degradation step. *CRT1* deletion was then confirmed by PCR. This involved two sets of primers for two PCR reactions (Figure 2.5). Amplification of a band with primers that bound inside the *URA3* cassette and downstream of the *CRT1* ORF was a positive result for deletion, in combination with lack of amplification of a band using a primer set that bound inside *CRT1*. Band presence and size for each reaction was determined after 1% gel electrophoresis at 90V for 30 minutes.

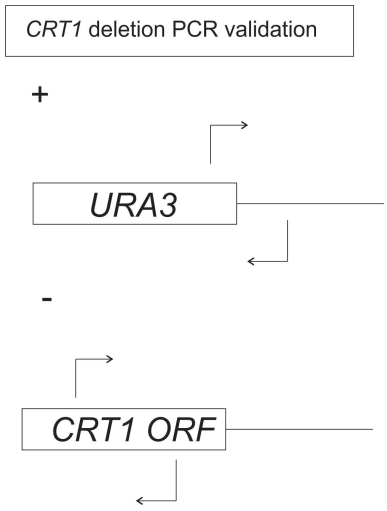


Figure 2.5: Primer set used to verify replacement of *CRT1* ORF with URA3 selection cassette. + = positive primer set. Visualization of a product using this set was a positive result for deletion, - = negative primer set. If the deletion was successful, visualization of a product using this set was not expected.

Sequencing

Samples were shipped to and sequenced by the TCAG sequencing facility in Toronto. High quality sequencing data was difficult to obtain when reading products amplified from genomic DNA, so plasmid products were prepared. The *RNR3* promoter was first amplified from genomic DNA, integrated into a bacterial A3 (Ampicillin resistance) plasmid backbone, transformed into competent Mach1^{TM} *E. coli* cells (available from Invitrogen) and miniprepped.

For the amplification step, primers were used to amplify the *RNR3* promoter with overhangs for the sequence recognized by the endonucleases EcoR1 and Pst1. The promoter DNA was digested for one hour with the enzymes. The A3 backbone was digested for 40 minutes, and was then de-phosphorylated with Antarctic Phosphatase for 20 minutes. The enzymes were deactivated by incubating the samples at 80°C for 20 minutes. A PCR clean-up (BioBasic, Inc. Molecular Biology Kits - EZ-10 spin column PCR products purification kit) was then performed on the promoter DNA sample to eliminate any enzymes that could interfere with the transformation. The

backbone (≈33ng) and promoter (≈100ng) were then ligated together with T4 DNA ligase for 10 minutes at room temperature, and the enzyme was deactivated for 20 minutes at 80°C.

For the transformation step, plasmid preparations (containing the *RNR3* promoter variant amplified from the yeast genome and ligated to the A3 backbone) were added to competent Mach1^{TM} cells and after half an hour on ice, the cells were heat shocked for one minute at 42°C. The cells were put back on ice for 5 minutes, and then allowed to recover for one hour with shaking at 37°C in enriched SOC media. The resultant transformed cells were plated onto Ampicillin plates.

The next day, transformants were selected, inoculated into 3 mL Luria Broth media and incubated with shaking at 37°C overnight. The following day, the samples were miniprepped using a kit (BioBasic, Inc., EZ-10 spin column plasmid DNA miniprep kit). Water instead of elution buffer was used to elute the plasmid DNA during the last step as per the sequencing facility's requirements (buffers were said to interfere with sequencing enzyme efficiency). A diagnostic digest was done on each miniprepped plasmid to ensure that ligation and transformation were successful. The resultant plasmid DNA was then sent for sequencing at the appropriate concentration specified by the facility (approximately 200 ng of plasmid DNA in a 7.7 μl preparation with water and diluted sequencing primer).

2.2.2 Characterization of trait values

Measuring fluorescence of technical replicates by flow cytometry

Seventeen strains were assayed for fluorescence. These were: Autofl (the strain background used to measure autofluorescence), wt (the strain containing the *RNR3* promoter with no deletions), x (the strain lacking Crt1), and the 14 promoter variant strains (see Table 2.1). Strains were grown overnight in synthetic media (with double

the amount of adenine normally used due to deficiency from the *ADE2* knockout) to stationary phase and released in the morning for three or four hours (to ensure at least one cell division) at an optical density (OD) of 0.04. OD was measured using a Wallac Victor^{3TM} 1420 multichannel counter plate reader (PerkinElmer) set at an absorbance wavelength of 600 nm. After growth for three or four hours, cells were diluted 1:7 or 1:10 in the same media. This dilution was to ensure that by four hours later, the samples would be at an OD indicative of growth in the exponential phase (assumed to be approximately 0.2 by the method used for measuring OD). Approximately four hours later, GFP expression was assayed by flow cytometry using a 488 nm solid state blue laser on the Cyan-ADP 9 flow cytometer from Beckman Coulter. Expression values were either collected by tubes or by 96 shallow well plates using a HyperCyt® Plate Loader (Intellicyt, Beckman Coulter) sampling robot connected to the flow cytometer. Testing the samples in plates allowed for more rapid data collection, whereas testing the samples in tubes ensured that a consistent number of events was collected for each sample.

The samples were diluted 1:10 in non-sterile water or buffer (phosphate-buffered saline (PBS) or sodium citrate buffer) immediately prior to collection of fluorescence measurements. In cases where tubes were used, 30,000 events were collected per sample. In cases where plates were used, a constant sip time was used, resulting in a minimum of 3,000 and a maximum of 15,000 events per well depending on the strain.

Measuring fluorescence of biological replicates by flow cytometry

Three biological replicates from each strain (with the exception of autofl, wt, x, and xuvw, where only a single colony was available) were picked and grown overnight in synthetic media. In the morning, the cells were released to an OD of 0.05 for three hours and GFP expression was assayed via a 96 well plate with the robot connected to the flow cytometer. The minimum number of events collected per well was 21,556

and the maximum was 44,397.

Converting fluorescence data to plots of mean expression values

The raw fluorescence data collected from the flow cytometer was either a series of flow cytometry standard (fcs) files (one for each sample) when samples were collected in tubes, or a large fcs file containing all samples when a plate was used. After running a plate, the file was separated into one file for each well using the Hypercyte® software for the sampling robot. The mean expression values were then extracted from the raw data in these individual sample files using a series of customized MATLAB scripts. See Appendix A for details.

After obtaining expression values, MATLAB scripts were designed with the help of Hilary Phenix and Nezar Abdennur to load the data file, combine replicate data, and plot bar graphs of mean linear GFP expression over multiple experiments using the function Errorbar in MATLAB. Alternatively, plots were generated directly from raw data using another set of customized MATLAB scripts (see Appendix A).

2.2.3 Data analysis

Multiple linear regression model

The regression model used to capture and quantify single variable influences and interactions between all combinations of variables is shown below:

$$\begin{aligned} Y(x,u,v,w) &= Y_0 + \beta_x x + \beta_u u + \beta_v v + \beta_w w \\ &+ \beta_{uv} uv + \beta_{vw} vw + \beta_{uw} uw + \beta_{uvw} uvw \\ &+ \beta_{xu} xu + \beta_{xv} xv + \beta_{xw} xw \\ &+ \beta_{xuv} xuv + \beta_{xvw} xvw + \beta_{xuw} xuw + \beta_{xuvw} xuvw + \epsilon, \\ &\begin{cases} x,u,v,w = 1, & \text{if } x,u,v,w \text{ is/are deleted,} \\ x,u,v,w = 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{2.1}$$

where Y indicates the trait, x is Crt1, and u,v and w are the three X-boxes. Trait values were linearized to test an additive null model, and subsequently converted to the log2 scale to test a multiplicative null model.

Model implementation

Linear expression values were first extracted from MATLAB. Autofluorescence was then subtracted from each mean expression value. To obtain the log2 transformed values, the expression values minus autofluorescence were log2 transformed using the command log2(x) in MATLAB. Autofluorescence was eliminated on a linear scale before log2 transformation because it was unclear how ratios could be interpreted in a regression analysis (ratios relative to autofluorescence would be obtained by subtracting log values).

The regression analysis was then performed on the resulting values. Linear regression was performed using the function Regstats in MATLAB. The y variable was a single column of expression values (linearly or log2 transformed) for each promoter variant over three separate experiments. The z variable was an n by p matrix, with n being the number of expression values and p being the number of predictors. For the expression value associated with each promoter variant, ones and zeros were assigned to each predictor according to the variables deleted in that particular strain. The y variable was then regressed against the z variable. The Regstats function outputs a series of statistics based on what the user specifies. The output selected in this case was t-stat, which includes the β coefficient values, their t-tests, and p-values. The p-value was for a t-test of whether the β coefficient was significantly different than 0. Significant predictors were identified at a p-value of ≤ 0.05.

Diagnostic plots

The following diagnostic statistics were extracted using the Regstats function in MATLAB: yhat (fitted values of the trait data) and sres (standardized residuals).

Plots were generated using two functions in MATLAB. The first plotting function used was qqplot(x), where x was sres. This function compares values to a theoretical normal distribution. The second plotting function was scatter(x,y) where x was yhat and y was sres. This function creates a scatter plot of the data points in x and y.

Identifying, quantifying and characterizing interactions

Interactions were identified by a significant p-value, quantified by the magnitude of the β coefficient, and characterized based on deviations from neutrality as shown in Table 2.3. Significant predictors u, v or w indicated important binary TF-CRE interactions and significant predictors uv, uw, vw or uvw indicated higher order TF-CRE interactions. Significance of these terms when x was deleted further indicated that the sites or interactions were important when the TF was absent. Higher order interactions containing more than two variables which had opposing effects were not characterized.

Interactions were further characterized based on significance under multiple null models as shown in Table 2.4. Regardless of characterization, significance under both null models indicated an important interaction.

T-tests

Expression values were first corrected for autofluorescence. Independent-sample t-tests were conducted in MATLAB at an alpha level of 0.05 using the function Ttest2, a function that calculates a 2-tailed t-test for comparing two means. Equal variances between the means were not assumed. The p-values were not corrected for multiple comparisons.

Table 2.3: Characterization of interactions using multiple linear regression: example for two variables x and u. D = deviation (positive + or negative -)

Sign of βx	Sign of βu	Sign of βxu	D	Type of interaction
+	+	+	+	Synergistic
-	-	-	+	Synergistic
+	+	-	-	Antagonistic
-	-	+	-	Antagonistic
+	-	+		
+	-	-		

Table 2.4: Further characterization of interactions using multiple linear regression: example for two interacting variables x and u. D = Deviation (positive + or negative -), A = additive model, M = multiplicative model.

D under A	D under M	Type of interaction	Important interaction
+		Greater than additive	
-		Less than additive	
	+	Greater than multiplicative	
	-	Less than multiplicative	
+	+	Synergistic	✓
+	-	Greater than additive, less than multiplicative	✓
-	+	Less than additive, greater than multiplicative	✓
-	-	Antagonistic	✓

2.3 Synthetic genetic array

2.3.1 Library construction

General approach

The general approach for SGA library construction from the query strain and deletion mutant array (DMA) library is given in Figure 2.6 and based on [90]. The DMA library was first condensed from four 96 well plates to 384 density agar format so that the whole library would fit on one plate. The query strain (modified from the standard by deleting Crt1) was then mated with the DMA library on YPD agar, the diploids were pinned onto enriched sporulation (SPO) agar and were subsequently incubated at 25°C for eight days. After confirming the presence of tetrads for four random strains under a light microscope, the sporulated cells were subjected to several rounds of selection to select for haploid cells containing both Crt1 and DMA mutations.

Standard query strain

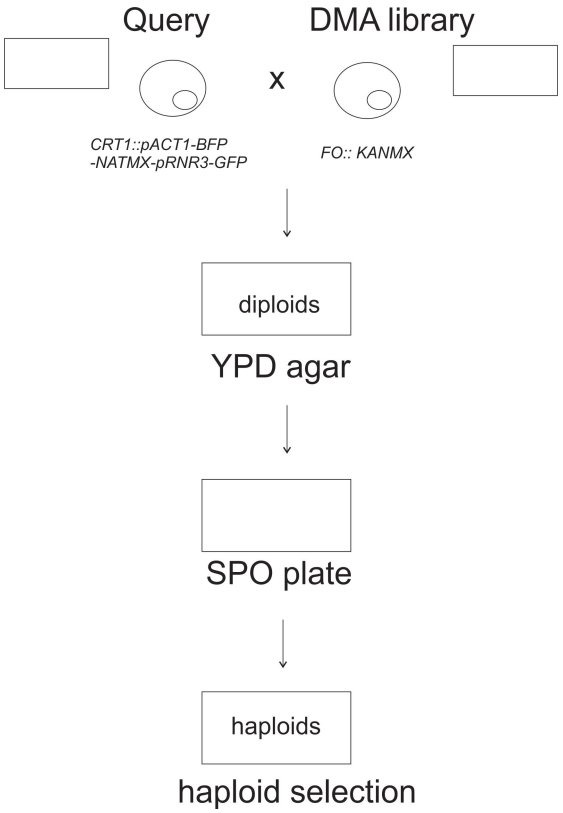


Figure 2.6: General approach for synthetic genetic array (SGA) library construction. Adapted from [90]. *FO* refers to a factor ORF with a GO annotation for transcription. Each rectangle represents a 384 density agar plate. The query strain (modified from the standard, described in more detail below) was mated with each deletion mutant array (DMA) library strain on YPD agar, and the resulting diploids were sporulated on sporulation (SPO) agar. These sporulated strains were then subjected to several rounds of haploid selection to select for haploid cells containing both the Crt1 and FO mutations.

The standard query strain, Y8205, was obtained from the Boone lab at the University of Toronto (see Table 2.1). It contained two modifications that enabled selection of the correct MATa haploids after mating and sporulation [90]. First, deletions at the *CAN1* and *LYP1* loci prevented the drugs canavanine and thialysine from entering and killing the cell. Therefore, selection on this drug media should have eliminated any cells with the *CAN1* and *LYP1* genes (diploids would contain a copy so this selected for haploids).

Second, the *CAN1* gene was replaced with a MATa specific promoter (*pSTE2*)

driving expression of *HIS5* and the *LYP1* gene was replaced with a MATα specific promoter driving *LEU2* [90]. Only MATa cells should have expressed *HIS5*, which would allow growth on plates lacking the amino acid histidine. MATα cells should have repressed the expression of *HIS5*, and expressed *LEU2* instead [90].

Modification of the standard query strain

The construct introduced to the standard query strain is shown in Figure 2.7. This construct was designed to simultaneously introduce the *RNR3*-GFP promoter-reporter and eliminate the Crt1 gene. To generate this construct, a forward and reverse primer were designed with 40 bp overhangs to the regions upstream and downstream of the *CRT1* ORF, and 20 bp homology to the beginning and ends of a dual colour promoter-reporter construct generated by Hilary Phenix. The dual-colour promoter-reporter construct contained the *RNR3* promoter fused to the sequence for GFP and the constitutive *ACT1* promoter fused to the sequence for blue fluorescent protein (BFP). This constitutive promoter-reporter was used so that strains showing abnormal increases or decreases in constitutive reporter expression relative to others could be eliminated. The dual colour promoter-reporter construct was amplified from a plasmid obtained from Hilary Phenix using the designed primers with Crt1 overhangs, resulting in the complete construct.

The construct was integrated into the standard query strain using the yeast transformation protocol discussed previously [32]. The correct location and functionality of the integrated construct after yeast transformation was verified via flow cytometry. Since deletion of Crt1 causes constitutively high *RNR3* expression [40] and BFP was driven by a constitutive promoter, increased GFP and BFP fluorescence relative to an autofluorescence control indicated correct location and expression of the construct. Three transformants tested showed between 49-55-fold increase in GFP expression relative to the autofluorescence control, which was similar to the deletion

Figure 2.7: Construct used to modify the standard SGA query strain. *CRT1−U* and *CRT1−D* refer to 40 bp regions of homology immediately upstream and downstream of *CRT1*. *pRNR3* = *RNR3* promoter sequence, *GFP* = sequence for green fluorescent protein, *pACT1* = *ACT1* promoter sequence, *BFP* = blue fluorescent protein sequence, *NATMX* = expression cassette that conferred clonNAT resistance.

of Crt1 with the *URA3* cassette in strain x (53-fold). BFP levels were also increased between 31 and 32-fold compared to autofluorescence, indicating functionality of the constitutive promoter-reporter.

Specifics of the deletion mutant array library

The DMA library was derived from a combination of genes from the strain library of 6,000 yeast ORF knockouts (ORFs deleted by replacement with *KANMX* resistance) and the decreased abundance of messenger RNA perturbation (dAmP) library for essential factors (ORFs knocked down 4-10 fold by adding the *KANMX* gene immediately after the ORF). Both libraries are available from Open Biosystems but were obtained from the Baetz lab. The DMA library was constructed by Hilary Phenix. 370 factor ORFs (FOs) encoding factors with GO annotations for transcription were chosen. The DMA library was spread over four 96 well plates, leaving 14 empty wells. To some of these empty wells, one of three neutral control strains (generated by Hilary Phenix) were added. In these strains, a neutral gene (*HIS3*, *URA3* or *HO*) was replaced with *KANMX*. These replacements were not thought to affect transcription of the reporter construct.

Specific approach

To select for MATa haploids with the Crt1 and the DMA mutations after mating and sporulation, agar media was used with the drugs G418 and clonNAT (to select for KANMX and NATMX respectively), canavanine and thialysine (to select for haploids) and a lack of histidine (to select for MATa cells). All steps in generating the SGA

library were performed using a RoTor HDA pinning robot (Singer Instruments). Solid (agar) pinning steps were carried out on RoTor Plusplate© dishes and liquid steps were carried out in 384 shallow well plates.

2.3.2 Solid phase testing protocol

A solid phase testing protocol developed by Hilary Phenix was used. The library of mutant promoter-reporter strains to be tested was thawed from -80°C and pinned to YPD + G418 agar. The spotted cells were grown for two days at 30°C. The cells were then pinned to synthetic complete media agar and grown in the same conditions for 16 hours. Following this, the cells were pinned to 50 μl of citrate buffer and diluted again approximately 1:10 to achieve an appropriate cell density for flow cytometry. Fluorescence (BFP measured from the 405 nm laser and GFP measured from the 488 nm laser) was assayed using a 384 shallow well plate and the sampling robot connected to the flow cytometer. This procedure was repeated on three different occasions to obtain replicate data in triplicate.

2.3.3 Statistical analysis

Z-test for outliers

Mean linear GFP fluorescence values were first normalized to the BFP signal. Z-scores were then calculated for each SGA strain based on the normalized mean of the sample population (each strain's GFP fluorescence distribution amalgamated over three technical replicates) and an estimated control population of normalized medians. A Bonferroni adjusted p-value of ≤ 0.01 divided by 367 (the number of strains for which data was collected) ($p \leq 2.72 \times 10^{-5}$) was used to preliminarily identify hits.

2.3.4 Validation

Re-streaking

Spotted colonies from the master plate were dabbed with a pipette tip, re-streaked onto agar plates, and incubated for one to two days at 30°C.

YPD+G418+clonNat agar plates: Even with the additional round of haploid selection, spots could have been a mixture of different types of cells, some being an inappropriate genotype that were able to grow on top of other cells and survive selection. Re-streaking onto plates with both drugs ensured that selection cassettes for both gene deletions were present and functional. If strains that were streaked from the master plate grew on these plates, this indicated that the strains had both *KANMX* and *NATMX* resistance cassettes. The strains selected for re-streaking were: Ace2, Aft1, Ckb1, Spt10, Spt3, Sfp1, Spt8, Mot2, Ada2, Mot3, and Rox1.

SM-His and SM-Leu agar plates: These plates took advantage of the mating type specific promoters in the standard query strain [90] discussed earlier. Since the correct MATa haploid would express the *Sp-his5* cassette, the strains should have been able to grow on -His plates. The incorrect MATα strains should have expressed *LEU2*, allowing them to grow on -Leu plates. Thus, the correct strains should have grown on -His and not on -Leu. Strains selected for re-streaking onto -His and -Leu plates were: Spt10, Spt3, Spt8, Mot2, Ada2, and Rox1.

PCR confirmation

CRT1 deletion with NATMX was confirmed first because all strains should have had this deletion, enabling the same primer set to be used for all confirmations. Genomic DNA was isolated from strains directly from the master plate or from a re-streaked plate using the protocol for quick preparations discussed earlier. PCR reactions were then performed with two sets of primers to test for *CRT1* deletion (Figure 2.8). The first primer set was designed so that the forward primer bound

inside of *GFP* and the reverse primer bound downstream of *CRT1*. A band of the correct size using this primer set was a positive result for *CRT1* deletion. The other primer set was the same set used to verify *CRT1* replacement with *URA3*: designed so that the forward and reverse primer bound inside of the *CRT1* ORF. A band using these primers indicated that *CRT1* was still present. Thus, *CRT1* deletion was validated when input DNA from the same strain was amplified in the positive reaction, and not in the negative one.

Once *CRT1* deletion was confirmed, the *KANMX*-associated mutation was confirmed. For confirmation of these mutations, the forward primer for the positive reaction bound inside the *KANMX* cassette and the reverse primer bound downstream of the factor ORF (Figure 2.8). For the negative reaction, the forward primer bound inside the factor ORF and the reverse primer bound downstream.

Mating experiments

Selected SGA library strains were re-streaked from the master plate and were mated with a MATαhis1- tester and MATahis1- tester strain (obtained from the Baetz lab) on minimal synthetic media agar plates lacking 6 amino acids (SC-6). The protocol used was modified from the original mating test protocol in [92]. Since the tester strains were lacking histidine, they could not grow on their own on minimal media. Additionally, the SGA strains were all missing the *URA3* gene and could not grow on minimal media. If the tester strains mated with the SGA strains, the missing amino acids would compliment each other and the resulting diploids would be able to grow on the SC-6 plates. The mating experiment was performed so that the testers were streaked down the left and right length of the plate and the SGA strain was streaked width-wise across both strains multiple times. The plates were incubated for three days and were inspected for the presence of newly formed colonies which indicated that mating had occurred.

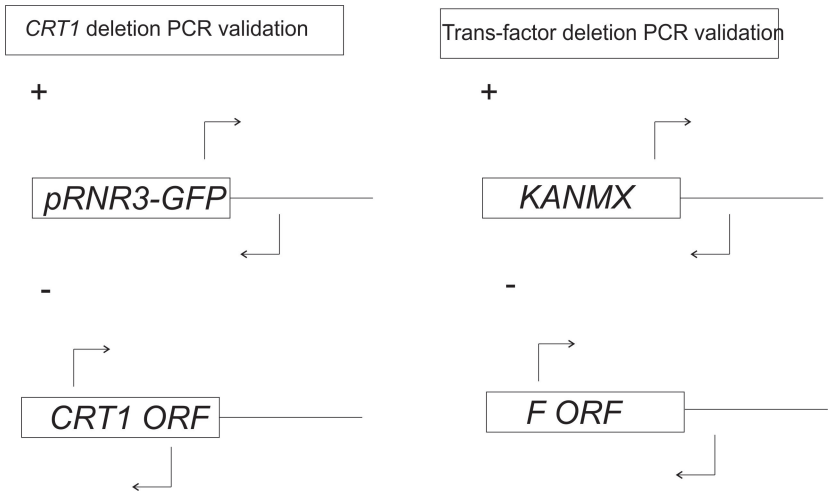


Figure 2.8: Primer sets used for PCR validation of SGA library hits. *F* refers to factor. + = positive primer set. Visualization of a product using this set was a positive result for deletion, - = negative primer set. If the deletion was successful, visualization of a product using this set was not expected.

Sporulation

Selected SGA library strains were plated onto enriched sporulation medium and kept at 25°C for 11 days. A small sample of each strain was then observed under the light microscope to determine whether tetrads could be visualized. A positive result indicated diploids in the original population.

Hit regeneration and testing

The construct for hit regeneration is shown in Figure 2.9. Primers were designed with 40 bp overhangs for ≈ the beginning and the end of the FO and 20 bp overhangs to the *NATMX* resistance cassette. The construct was then amplified from a pFA6a-NATMX plasmid using these primers and was integrated into the genome of strain x as per the transformation protocol previously described [32], interrupting the FO of interest. For validated Crt1-independent regulators, the factor was subsequently deleted in strain xuvw.

To validate trans-factor deletion after transformation, quick genomic preparations were made and positive transformants were identified with a PCR reaction. The

Figure 2.9: Construct used to regenerate hits from the SGA procedure.*FO-U* and *FO-D* refers to 40 bp homology to the beginning and end of the factor ORF (FO) of interest, used to direct construct integration. *NATMX* is the selection cassette that confers resistance to the antibiotic clonNAT.

forward primer for this reaction bound inside the *NATMX* cassette and the reverse primer bound downstream of the factor ORF (Figure 2.10). Amplification of a band indicated correct location of the *NATMX* cassette and thus factor deletion.

Strains were then assayed for GFP expression by flow cytometry. Strains were grown overnight in synthetic liquid media with double adenine and were released to an OD of 0.04 for four hours in the morning. Strains were then diluted 1:10 in the same media and grown for another four hours before being assayed for fluorescence. Since some of the factor deletions significantly impaired the growth rate, all strains were assayed in flow tubes so that a set number of events (30,000) could be collected. Fluorescence data was converted to plots of mean expression values as described in Section 2.2.2.

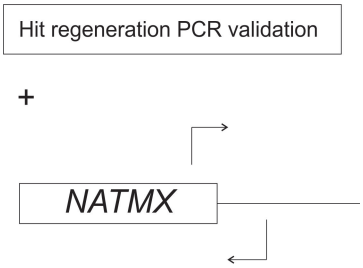


Figure 2.10: Primers used for PCR validation of regenerated SGA factor deletions. + = positive primer set. Visualization of a product using this set was a positive result for deletion.

Chapter 3

Results

3.1 Higher order interactions and Crt1-independent effects revealed by replacing X-boxes with Tet operators

Following construction of all strains with perturbations to Crt1 (present/deleted) and/or the X-boxes (present or replaced with Tet operators), I had a total of 16 strains (see Methods). As mentioned earlier, strains are labeled by the components deleted. Crt1 is referred to as x, and the three X-boxes are referred to as u, v and w. I confirmed replacement of the X-boxes with Tet operators by sequencing (see Methods) at least one biological replicate (a clone from a single transformation) for each of the 16 strains. For representative sequencing chromats, see Appendix B Figures B.1 and B.2.

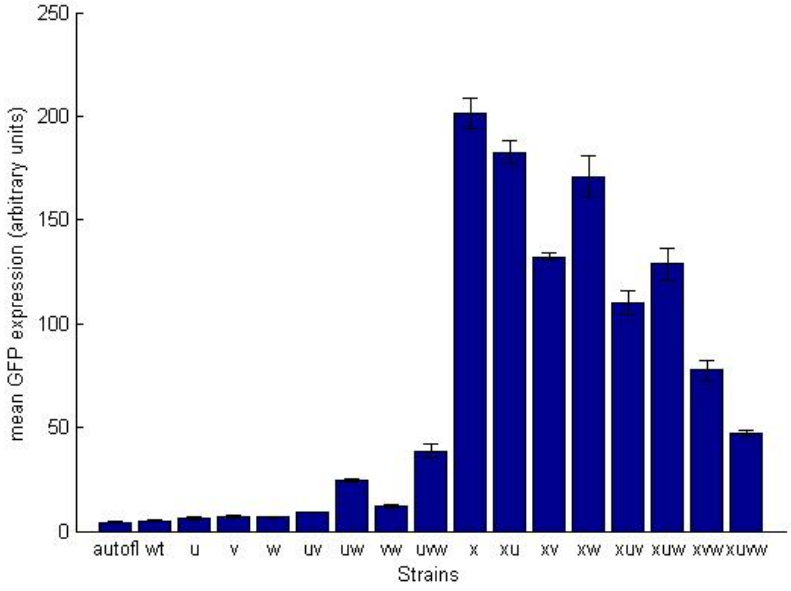
I characterized reporter expression (the trait) for each strain by flow cytometry. To account for experimental error, I took three measurements (technical replicates) for each strain (see Methods). Mean linear GFP expression values are shown in Figure 3.1. I then performed multiple linear regression (see Methods) on the trait

values to test both an additive and multiplicative model of independence between the four biological components perturbed (Crt1 and each X-box).

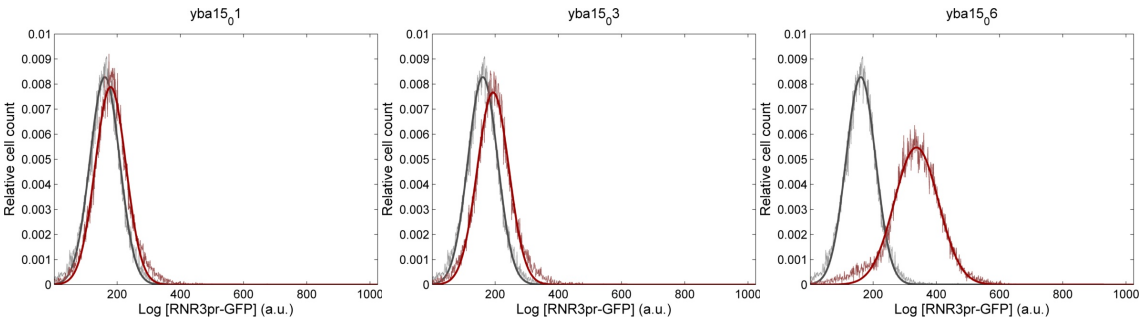
In order to determine whether the regression models I used, additive and multiplicative, were appropriate for my dataset, I tested whether three assumptions of linear regression [14, 64] had been satisfied. I generated diagnostic plots (see Methods) of predicted trait values versus standardized residuals to test the assumptions of linearity between variables and homogeneity of variance, and standardized residuals versus a theoretical normal distribution to test the assumption of normally distributed error. The results are shown in Figures 3.2 and 3.3.

The plots of predicted trait values versus the standardized residuals (Figures 3.2 a. and 3.3 a.) did not show a bow-shaped pattern, which suggested that the variables had a linear rather than curvilinear relationship. Figure 3.2 a. also shows that the additive model (predicting linear expression values) made systematic errors when predicting large values, indicating a violation of homogeneity of variance. This systematic error was reduced when the expression values were log2-transformed (testing a multiplicative model). However, the multiplicative model made systematic errors when predicting small values (Figure 3.3 a.). Lastly, the plots in Figures 3.2 b. and 3.3 b. show that the residuals deviated from the line specifying normality in both models, indicating a violation of normally distributed error. This was not surprising since the predicted values shown in Figures 3.2 a. and 3.3 a. were larger in some cases than two standard deviations from the observed response.

Overall, the results suggested that depending on the magnitude of the observed trait, either the additive or multiplicative model produced more accurate trait predictions. To move forward with data analysis in a way that minimized errors in trait prediction, I therefore chose to focus on only those interactions that were identified as important regardless of neutrality model. See Methods for an in-depth description of

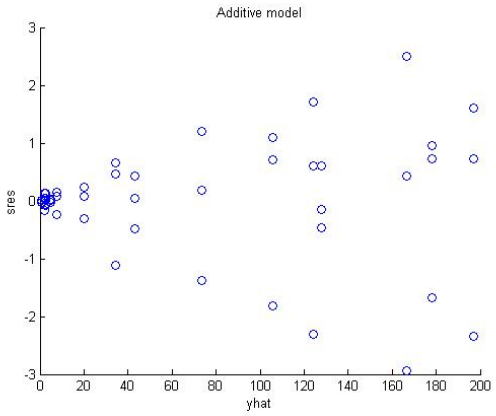


(a) Linearized trait values for all strains.

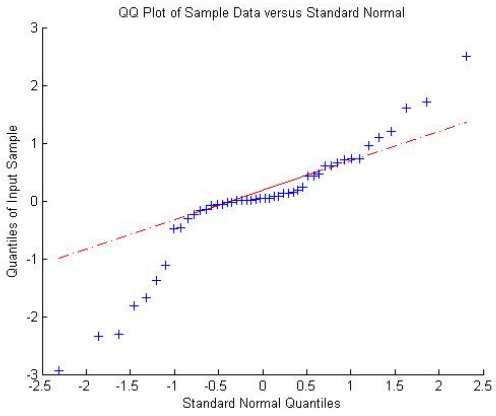


(b) wt vs u (c) wt vs w (d) wt vs uw

Figure 3.1: Trait values for *RNR3* promoter variants where X-boxes were replaced with Tet operators. Strains (Table 2.1): autofl = the strain background used to measure autofluorescence, wt = the strain containing the *RNR3* promoter with no deletions, and the 15 remaining strains = labeled by the elements replaced (Crt1 = x, X-boxes = u, v, and w). a. Linearized mean GFP expression values for all strains. Error bars represent the standard deviation of the mean linear GFP expression for three technical replicates. b-d. Log expression distributions for a representative sample of strains. Each distribution is for one experimental measurement (one technical replicate). The wt strain (grey), which contained an *RNR3* promoter lacking any X-box replacements, is plotted along with strain u (b), w (c) and uw (d) in red (darker than wt if viewing without colour) on the same plot for distribution comparison. Statistical analyses of trait values included multiple linear regression and t-tests, shown in Tables 3.1, 3.2, and 3.4.

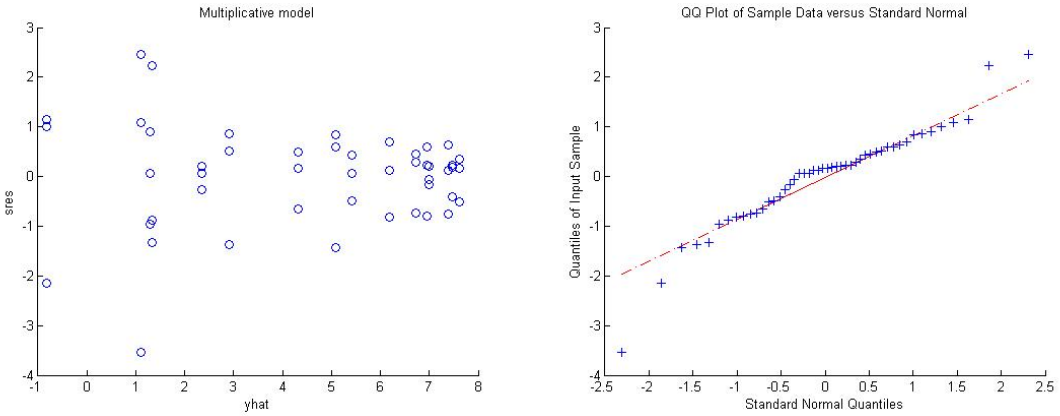


(a) scatter plot of yhat vs. sres



(b) qqplot of sres

Figure 3.2: Plots to test three assumptions of multiple linear regression for additive model, X-boxes replaced with Tet operators. a. yhat = predicted trait values, sres = standardized residuals. b. qqplot = quantile-quantile plot, comparing values of standardized residuals to a theoretical normal distribution.



(a) scatter plot of yhat vs. sres

(b) qqplot of sres

Figure 3.3: Plots to test three assumptions of multiple linear regression for multiplicative model, X-boxes replaced with Tet operators. a. yhat = predicted trait values, sres = standardized residuals. b. qqplot = quantile-quantile plot, comparing values of standardized residuals to a theoretical normal distribution.

the approach I took to identify, quantify and characterize interactions in this manner.

The results of the additive and multiplicative linear regression analysis for the Tet operator replacement approach are shown in Tables 3.1 and 3.2 respectively.

Regardless of neutrality model, x was identified as a significant predictor with a large positive value. This meant that deleting Crt1 resulted in a significant increase in expression. A summary of the significant interaction terms and the resulting classification is given in Table 3.3.

There was a significant interaction between variables u and w that was identified regardless of neutrality model and classified as synergistic. This meant that replacing X-boxes u and w together resulted in an increase in expression that could not be accounted for by adding or multiplying the effects of replacing u and w individually. Replacing the X-boxes with Tet operators also resulted in additional effects on transcription in the absence of Crt1 (significant predictors *xu, xv, xw, xuw*). The *xu*, *xv* and *xw* interactions were classified as antagonistic. This meant that deleting Crt1 along with any one of the X-boxes resulted in an opposing impact on expression that

Table 3.1: Multiple Linear Regression results, additive model, X-boxes replaced with Tet operators. C = coefficient, p = p-value, Significant = β C significantly different from 0 at p<0.05.

Predictor	term tested	β C	p	Significant
1	x	196.4	$1.3\times10^{-}32$	Y
2	u	1.63	0.67	
3	v	1.96	0.61	
4	w	1.89	0.62	
5	xu	-20.3	$5.6\times10^{-}4$	Y
6	xv	-71.0	$1.1\times10^{-}14$	Y
7	xw	-32.1	$9.0\times10^{-}07$	Y
8	xuv	-4.40	0.56	
9	xuw	-39.6	$8.6\times10^{-}06$	Y
10	xvw	-27.5	$8.8\times10^{-}4$	Y
11	$xuvw$	7.20	0.50	
12	uv	0.90	0.87	
13	uw	16.0	$5.0\times10^{-}3$	Y
14	vw	3.16	0.55	
15	uvw	8.20	0.28	

Table 3.2: Multiple Linear Regression results, multiplicative model, X-boxes replaced with Tet operators. C = coefficient, p = p-value, Significant = β C significantly different from 0 at p<0.05.

Predictor	term tested	β C	p	Significant
1	x	8.44	$6.4\times10^{-}36$	Y
2	u	1.93	$3.3\times10^{-}16$	Y
3	v	2.14	$1.5\times10^{-}17$	Y
4	w	2.11	$2.3\times10^{-}17$	Y
5	xu	-2.07	$5.8\times10^{-}13$	Y
6	xv	-2.77	$2.1\times10^{-}16$	Y
7	xw	-2.36	$1.8\times10^{-}14$	Y
8	xuv	0.78	$4.4\times10^{-}03$	Y
9	xuw	-1.38	$5.1\times10^{-}06$	Y
10	xvw	-0.04	0.87	
11	$xuvw$	-0.27	0.45	
12	uv	-0.91	$1.6\times10^{-}05$	Y
13	uw	1.10	$6.6\times10^{-}07$	Y
14	vw	-0.52	$6.4\times10^{-}3$	Y
15	uvw	0.06	0.83	

Table 3.3: Interactions identified and characterized for the Tet operator replacement approach. D = Deviation (positive + or negative -), A = additive model, M = multiplicative model. Important interaction = Identified under both MLR models and characterization.

Interacting variables	D under A	D under M	Important interaction
uv		-	
uw	+	+	✓synergistic
vw		-	
uvw			
xu	-	-	✓antagonistic
xv	-	-	✓antagonistic
xw	-	-	✓antagonistic
xuv		✓	
xuw	✓	✓	✓
xvw	✓		
xuvw			

could not be accounted for by deleting each component individually.

One possibility for the findings of the regression analysis was that the effects observed could be accounted for by some other sequence variation. While inspecting the chromat traces from the portion of the *RNR3* promoter sequenced for these strains, I observed that the number of thymine bases in the poly(dA:dT) tract located 5 bp upstream of X-box v (Figure 1.2) deviated in a number of cases from the length of 21 bp documented in SGD [15]. Results are shown in Appendix B Table B.1. As mentioned earlier, several studies have noted the importance of these sequences for activation, with length of the tract positively associated with strength of activation [21, 42, 74, 81, 101].

The tract length varied between strains. Therefore, the variable length of the tract among different *RNR3* promoter variants could have potentially explained the functional effects of replacing the X-boxes. If the length of the tract was proportional

to the amount of activation possible, I expected that strains with shorter tracts would have lower expression than strains with longer tracts, regardless of the X-box(es) that had been altered. However, some strains in the Tet operator replacement approach with shorter tracts (uw) showed higher activation than strains with longer tracts (w) (Figure 3.1 and Appendix B Table B.1), so tract length could not be directly correlated to amount of activation.

The tract length also varied between biological replicates. Therefore, biological variability could have been an explanation for the effects observed. In order to determine whether the trends in trait values were general or an effect of some unknown variability in a specific colony, I collected trait values from biological replicates (see Methods) and compared the trends in mean values to the technical replicate data in Figure 3.1. The results are shown in Figure 3.4.

I expected that if the biological variation accounted for the effects observed, then the trait values between strains would regress toward a common value when measurements from multiple colonies were taken. However, this was not the case. Strains where Crt1 had been deleted (strains containing x) showed varying levels of expression, similar to the technical replicate data. Additionally, replacing X-boxes u and w together still had a much larger impact on expression than replacing either site on its own. Therefore, the higher order interaction between u and w and the effects observed in the absence of Crt1 could not be accounted for by biological variability.

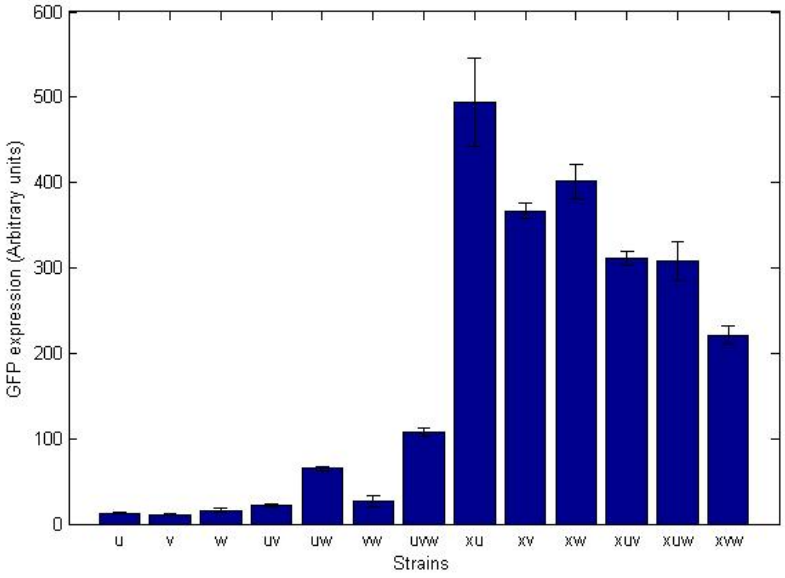
Interpreting the observation of significant regression coefficients containing the x variable required further investigation. If the X-boxes acted solely through Crt1, I expected the effect of deleting the sites to be masked by deleting the factor, i.e. I expected the expression values for strains with a Crt1 deletion to be identical regardless of X-box replacement. In the regression analysis, this would translate to an interaction between x and the site variable (u,v or w) equal in magnitude, but opposite in

sign to the effect of the site on its own. To investigate whether the effects observed in the absence of Crt1 could be accounted for by masking, I performed t-tests (see Methods) to see whether the expression from strains xu, xv and xw differed significantly from strain x. Results of the t-tests are given in Table 3.4. It was difficult to interpret the significant *xuw* predictor using t-tests since it involved more than two variables, so I did not further consider this interaction.

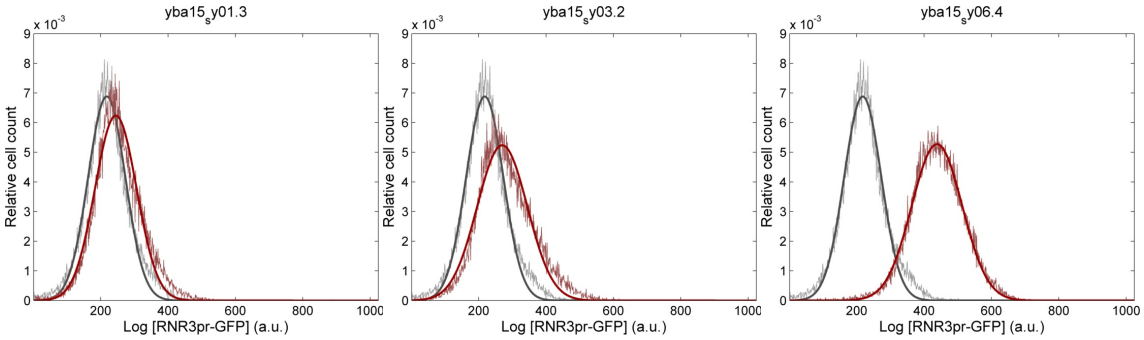
Regardless of the scale used to represent the data, the mean of each strain was significantly less than the mean of strain x, indicating that the effect of deleting X-boxes u, v and w in the absence of Crt1 resulted in a significant decrease in expression that could not be accounted for by masking effects.

There were two other possible explanations for observing an impact on expression upon X-box deletion in the absence of Crt1. One was that replacing the X-boxes with a longer Tet operator resulted in disruption of functional TF binding sites. To investigate this possibility, I created a map of all known TF consensus sequences found in the *RNR3* promoter (Figure 3.5).

Results revealed consensus sequences for Mot3 and Swi4 in the replaced region. Multiple Mot3 and Swi4 binding sites were present in the *RNR3* promoter, and the functionality of these particular interrupted CREs on downstream expression was not known. The second possibility for the effects observed was that the X-boxes had important functional roles in the absence of Crt1. To distinguish between these possibilities, I employed a second X-box replacement approach: scrambling the original 13 bp X-boxes. In this approach, the consensus sequences for Mot3 and Swi4 were left intact. If this approach revealed a significant impact on transcription of deleting the X-boxes in the absence of Crt1, then the hypothesis that the effects were due to disruption of Mot3 and Swi4 consensus sequences could be ruled out. Another advantage to scrambling the original X-box sequences was that possible effects on



(a) Linearized trait values for a subset of strains.



(b) wt vs u (c) wt vs w (d) wt vs uw

Figure 3.4: Biological replicate data for strains where X-boxes were replaced with Tet operators. Strains (Table 2.1) labeled by the elements replaced: Crt1 = x, X-boxes = u,v, and w. a. Mean linear GFP expression values for strains where three biological replicates were available. Error bars represent the standard deviation of the mean linear GFP expression for three biological replicates. b-d. Log expression distributions for a representative sample of strains. Each distribution is for one biological replicate. The wt strain (grey), which contained an *RNR3* promoter lacking any X-box replacements, is plotted along with strain u (b), w (c) and uw (d) in red (darker than wt if viewing without colour) on the same plot for distribution comparison. A different biological replicate is shown compared to the one that was used for the technical replicates. Statistics were not performed on these trait values since the goal was to examine whether the effects observed for the technical replicates could be accounted for by biological variation.

Table 3.4: T-tests comparing strains xu, xv and xw to strain x, X-boxes replaced with Tet operators. Log2 = Log2-transformed expression values used for comparisons, Linear = linear expression values used for comparisons, mean = mean GFP expression over three technical replicates, SD = standard deviation of the mean, Significant = mean expression values significantly different at $p<0.05$.

Strain or comparison	Log2 mean (SD)	Linear mean (SD)	Log2 t-test (p-value)	Significant	Linear t-test (p-value)	Significant
x	7.62 (0.06)	196.9 (7.75)				
xu	7.48 (0.04)	178.3 (5.45)				
xv	7.00 (0.02)	127.9 (2.06)				
xw	7.38 (0.09)	166.7 (10.3)				
xu vs x			-3.4 (0.029)	Y	-3.4 (0.032)	Y
xv vs x			-17.4 (0.00085)	Y	-14.9 (0.0025)	Y
xw vs x			-3.9 (0.023)	Y	-4.07 (0.018)	Y

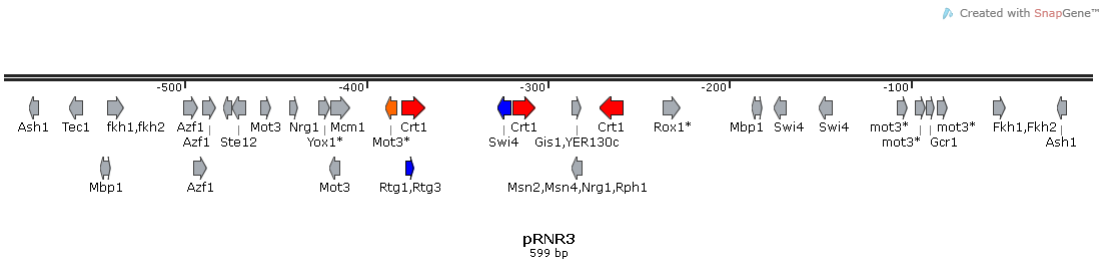


Figure 3.5: Map of all known consensus CREs in the *RNR3* promoter. Sequence information obtained from Yeastract [1, 60, 87]. * indicates additional sequence information obtained from [50] or SGD [15]. The orange and blue sites were disrupted by replacement of the red X-boxes with Tet operators.

transcription of using a bacterial sequence with different G/C content could also be eliminated.

3.2 Higher order interactions and Crt1-independent effects confirmed by second X-box replacement approach

Following construction of *RNR3* promoter variants using the scrambling X-box replacement approach, I had a total of 14 strains (see Table 2.1, strains wt and x did not have to be constructed again because they did not contain any X-box replacements). I confirmed replacement of the X-boxes with scrambled sequences by sequencing of at least one biological replicate for the 14 strains. For representative sequencing chromats, see Appendix B Figures B.3 and B.4.

While inspecting the chromat traces from the *RNR3* promoter for these strains, I again observed variability in the poly(dA:dT) tract length (Appendix B Table B.1). Therefore, this phenomenon was not unique to the sequence used for replacement. Even though I concluded that biological variation alone could not account for the effects on expression of replacing the X-boxes, I wanted to further investigate the

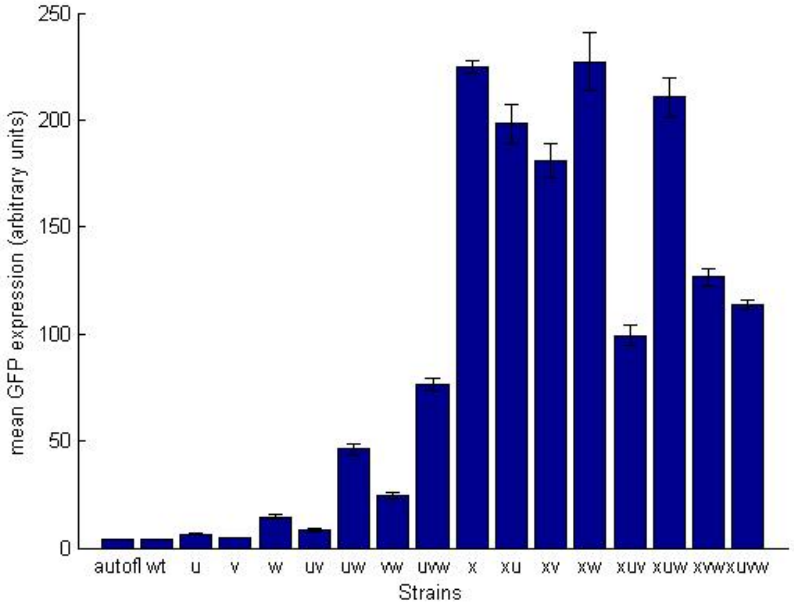
source of tract variability to see if it could be eliminated in the future. The variation in tract length could have been an artifact of enzyme slippage during PCR, bacterial transformation, or sequencing read and not reflective of the number of Ts present in the yeast genome. To examine this possibility, I sequenced a few different preparations of the same clone (technical sequencing replicates). Tabulation of ploy(dA:dT) tract length for these strains is shown in Appendix B Table B.1. There was variability between technical replicates of strains uw and xuvw in the scrambling approach, making the steps leading to preparing DNA for sequencing a source of variability. Therefore, it was impossible to determine the source of apparent variability in tract length, and the observed variability was not necessarily reflective of the actual genomic sequence.

The linearized trait values for strains generated by the systematic replacement of the X-boxes with scrambled sequences are shown in Figure 3.6.

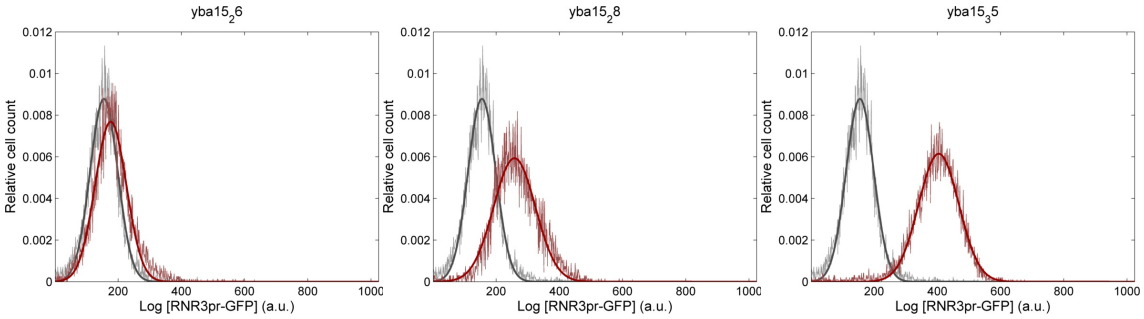
One again, to determine whether the additive and multiplicative regression models were appropriate for my dataset, I tested whether the three assumptions of linear regression had been satisfied. The results are shown in Figure 3.7 and 3.8. The pattern in a. and the deviation from the straight line in b. for both Figures 3.7 and 3.8 again suggested that depending on the magnitude of the trait, either the additive or multiplicative model produced more accurate predictions. Therefore, I once again focused on only those interactions that were identified as important regardless of neutrality model.

The results of the additive and multiplicative regression analysis for the X-box scrambling approach are shown in Tables 3.5 and 3.6.

Like the first replacement approach, x was identified as a significant individual predictor of the trait regardless of neutrality model. This was expected since strain x was the same for both replacement approaches. In contrast to the Tet operator

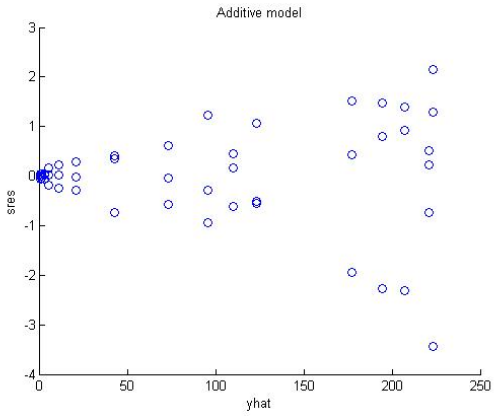


(a) Linearized trait values for all strains.

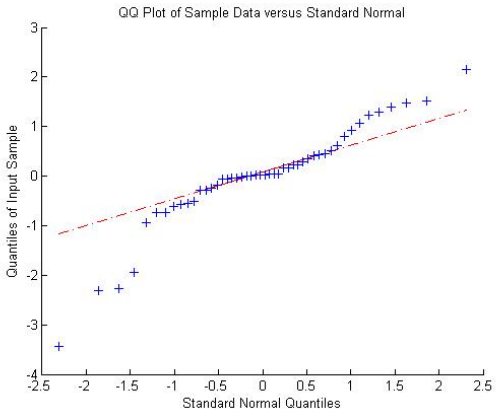


(b) wt vs u (c) wt vs w (d) wt vs uw

Figure 3.6: Trait values for *RNR3* promoter variants where X-boxes were replaced with scrambled sequences. Strains (Table 2.1): autofl = the strain background used to measure autofluorescence, wt = the strain containing the *RNR3* promoter with no deletions, and the 15 remaining strains = labeled by the elements replaced (Crt1 = x, X-boxes = u, v, and w). a. Linearized mean GFP expression values for all strains. Error bars represent the standard deviation of the mean linear GFP expression for three technical replicates. b-d. Log expression distributions for a representative sample of strains. Each distribution is for one experimental measurement (one technical replicate). The wt strain (grey), which contained an *RNR3* promoter lacking any X-box replacements, is plotted along with strain u (b), w (c) and uw (d) in red (darker than wt if viewing without colour) on the same plot for distribution comparison. Statistical analyses of trait values included multiple linear regression and t-tests, shown in Tables 3.5, 3.6 and 3.8.

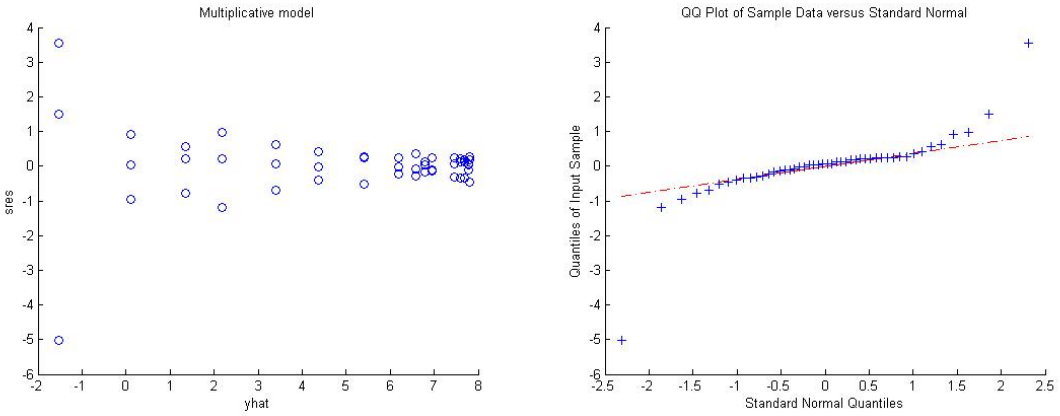


(a) scatter plot of yhat vs. sres



(b) qqplot of sres

Figure 3.7: Plots to test three assumptions of multiple linear regression for additive model, X-boxes replaced with scrambled sequences. a. yhat = predicted trait values, sres = standardized residuals. b. qqplot = quantile-quantile plot, comparing values of standardized residuals to a theoretical normal distribution.



(a) scatter plot of yhat vs. sres

(b) qqplot of sres

Figure 3.8: Plots to test three assumptions of multiple linear regression for multiplicative model, X-boxes replaced with scrambled sequences. a. yhat = predicted trait values, sres = standardized residuals. b. qqplot = quantile-quantile plot, comparing values of standardized residuals to a theoretical normal distribution.

Table 3.5: Multiple Linear Regression results, additive model, X-boxes replaced with scrambled sequences. C = coefficient, p = p-value, Significant = β C significantly different from 0 at p<0.05.

Predictor	term tested	β C	p	Significant
1	x	220.76	$8.6\times10^{-}32$	Y
2	u	2.18	0.63	
3	v	0.68	0.88	
4	w	10.14	0.03	Y
5	xu	-28.93	$6.7\times10^{-}05$	Y
6	xv	-44.54	$5.4\times10^{-}08$	Y
7	xw	-7.95	0.22	
8	xuv	-56.36	$4.5\times10^{-}07$	Y
9	xuw	-19.28	0.04	Y
10	xvw	-66.06	$2.1\times10^{-}08$	Y
11	$xuvw$	39.62	$3.7\times10^{-}3$	Y
12	uv	1.35	0.83	
13	uw	29.52	$5.1\times10^{-}05$	Y
14	vw	9.41	0.15	
15	uvw	19.00	0.04	Y

Table 3.6: Multiple Linear Regression results, multiplicative model, X-boxes replaced with scrambled sequences. C = coefficient, p = p-value, Significant = β C significantly different from 0 at p<0.05.

Predictor	term tested	β C	p	Significant
1	x	9.32	$1.8\times10^{-}29$	Y
2	u	2.89	$2.8\times10^{-}14$	Y
3	v	1.63	$2.6\times10^{-}08$	Y
4	w	4.92	$6.4\times10^{-}21$	Y
5	xu	-3.08	$4.2\times10^{-}11$	Y
6	xv	-1.95	$6.4\times10^{-}07$	Y
7	xw	-4.91	$1.8\times10^{-}16$	Y
8	xuv	0.10	0.83	
9	xuw	0.96	0.04	Y
10	xvw	0.12	0.79	
11	$xuvw$	0.04	0.95	
12	uv	-0.80	0.02	Y
13	uw	-0.89	$8.5\times10^{-}03$	Y
14	vw	-0.66	0.045	Y
15	uvw	0.61	0.18	

approach, however, w was also identified as a significant individual predictor regardless of neutrality model. The coefficient was positive, which meant that deleting w resulted in a significant increase in expression. A summary of the significant interaction terms and the resulting classification for the scrambling approach is given in Table 3.7.

Once again, there was an important higher order interaction identified between sites u and w. However, the classification of this interaction was different. In contrast to the synergistic classification from the Tet operator replacement approach, the scrambling replacement approach resulted in a classification of greater than additive, less than multiplicative. Therefore, deleting u and w together resulted in an impact on expression that was greater than that expected by adding the effects of u and w together, but less than that expected by multiplying their individual effects.

There were also, once again, significant effects of deleting the X-boxes in the

Table 3.7: Interactions identified and characterized for the X-box scrambling replacement approach. D = Deviation (positive + or negative -), A = additive model, M = multiplicative model. Important interaction = Identified under both MLR models and characterization.

Interacting variables	D under A	D under M	Important interaction
uv		-	
uw	+	-	✓> additive, < multiplicative
vw		-	
uvw	+		
xu	-	-	✓antagonistic
xv	-	-	✓antagonistic
xw		-	
xuv	✓		
xuw	✓	✓	✓
xvw	✓		
xuvw	✓		

absence of Crt1. These were evident by the significant predictors *xu, xv* and *xuw*. For *xu* and *xv*, the classification was the same as for the Tet operator approach: antagonistic. This meant that deleting Crt1 along with X-box u or v resulted in an opposing impact on expression that could not be accounted for by deleting each component individually.

Once again, to rule out masking, I compared the expression from strains xu and xv to the expression from strain x. The results are shown in Table 3.8. Deleting u and v in the absence of Crt1 significantly decreased *RNR3* expression. The finding of Crt1-independent effects of the X-boxes using two different replacement approaches strongly suggested that the X-boxes had additional roles beyond Crt1. Therefore, I developed an approach to further investigate these functional roles.

Table 3.8: T-tests comparing strains xu and xv to strain x, X-boxes replaced with scrambled sequences. Log2 = Log2-transformed expression values used for comparisons, Linear = linear expression values used for comparisons, mean = mean GFP expression over three technical replicates, SD = standard deviation of the mean, Significant = mean expression values significantly different at $p<0.05$.

Strain or comparison	log2 mean (SD)	Linear mean (SD)	Log2 t-test (p-value)	Significant	Linear t-test (p-value)	Significant
x	7.79 (0.02)	221.2 (2.88)				
xu	7.60 (0.07)	194.4 (8.92)				
xv	7.47 (0.06)	177.3 (7.88)				
xu vs x			-4.65 (0.032)	Y	-4.94 (0.026)	Y
xv vs x			-8.23 (0.0089)	Y	-9.06 (0.0055)	Y

3.3 Preliminary identification of Crt1-independent regulators of *RNR3* by synthetic genetic array

In order to investigate which factors could bind to the X-boxes and mediate the functional roles observed in the absence of Crt1, I first systematically identified important factors for *RNR3* transcription in the absence of Crt1. To do this, I used a synthetic genetic array (SGA) [90] screening approach. The end result of performing this screen (described in detail in the Methods) was a library of approximately 380 strains (see Table 2.1) containing a Crt1 deletion and another factor deletion associated with transcription. The SGA library strains are referred to by the name of the factor targeted for deletion, aside from Crt1.

I collected trait values for the SGA library using a solid phase testing protocol (see Methods). To identify mutations with the most severe impact on transcription from the *RNR3* promoter (hits), a Z-test for outliers was performed (see Methods). Hits were classified as activators when deleting the factor decreased reporter expression compared to a Crt1 deletion strain, and repressors when deletion of the factor increased expression. A heat map of GFP expression and the Z-scores for all SGA library strains are given in Figure 3.9 and Appendix C Table C.1 respectively. Nineteen activators and 10 repressors were identified as significant hits (Appendix C Table

C.1).

I expected Crt1-independent activators rather than repressors to bind the X-boxes since a decrease in expression was observed upon deletion of X-box u and v in the absence of Crt1. Therefore, I focused more on activator hits than repressors in subsequent analyses. The log GFP distributions compared to control for 19 hits (top 16 activators and top three repressors) are shown below in Figure 3.10. I assessed these hits for functional annotations, links to *RNR3*, and links to DNA damage. Results are shown in Appendix C Tables C.2 and C.3. Many of the FOs deleted had previously been implicated in the regulation of *RNR3* (Aft1, Spt3, Spt8, Mot2, Mot3, Rox1). Several others showed links to the DNA damage response (Bdf1, Spt10, Met18, Sfp1, TDA9).

3.4 Four Crt1-independent regulators of *RNR3* validated

Validation of the SGA procedure was required to ensure that the strains generated were haploid MATa and that they had both deletions (Crt1 and the factor) present. I chose several representative hits from the top 19 for each validation step (described in detail in Methods).

Re-streaking

All strains re-streaked grew on plates that selected for *KANMX* and *NATMX* resistance cassettes and the MATa specific *HIS3* reporter in the query strain, suggesting successful trans-factor and Crt1 deletions and haploid cells of the correct mating type.

PCR Confirmation

CRT1 deletion was confirmed for five targeted SGA library strains: Spt3, Spt8, Met18, Mot3 and Rox1. Trans-factor deletion was confirmed for four library strains:

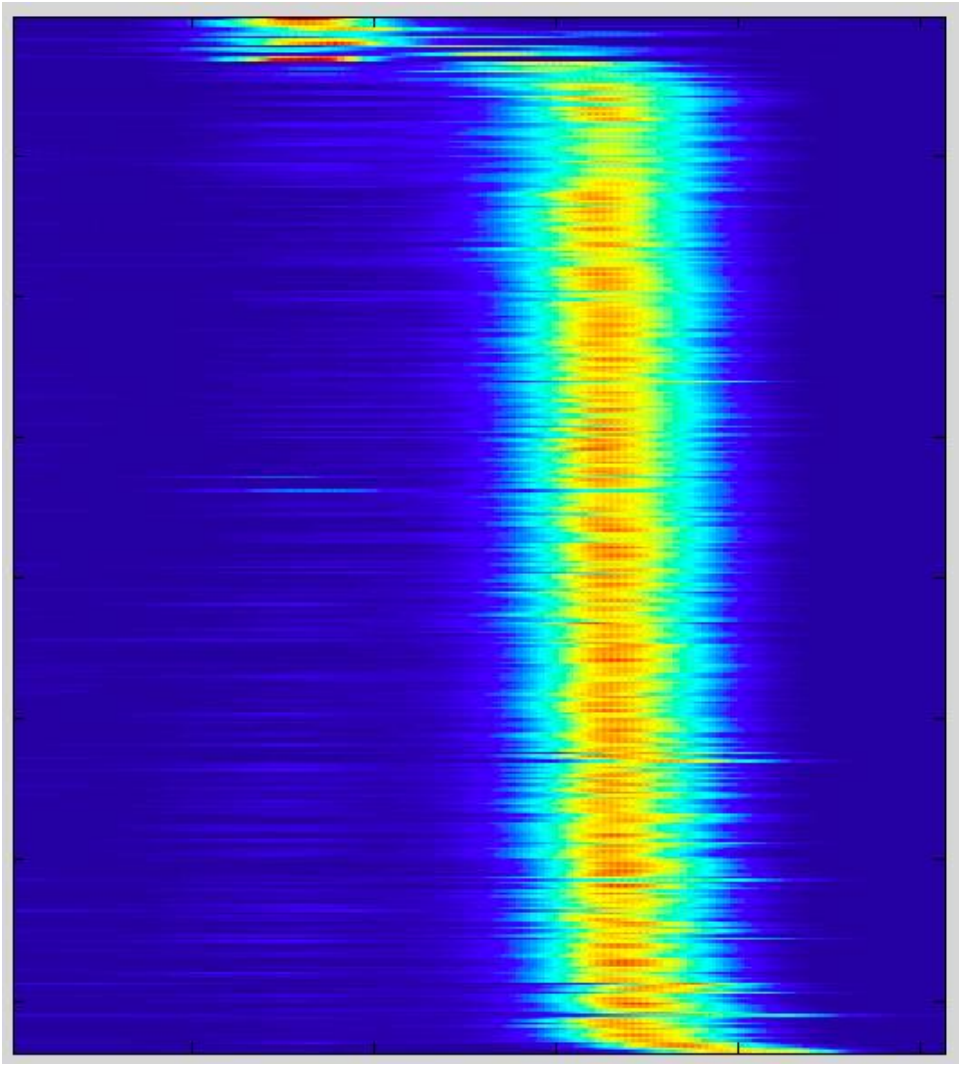


Figure 3.9: Heat map of mean linear GFP expression for all strains in the SGA library after amalgamation over three technical replicates and normalization for BFP. GFP is plotted on the x-axis for each strain (y-axis) in the library. Strains are ordered based on median fluorescence intensity.

Spt3, Spt8, Mot3 and Rox1. For six hits tested, including the top two activators, *CRT1* deletion was ambiguous. For these deletion strains: Ace2, Aft1, Ckb1, Sfp1, Spt10 and Mot2, bands were present for both the positive and the negative PCR confirmation reactions. The result that *CRT1* was both present and absent for these strains suggested that they had a copy of *CRT1* present and so were diploids.

Mating Experiments

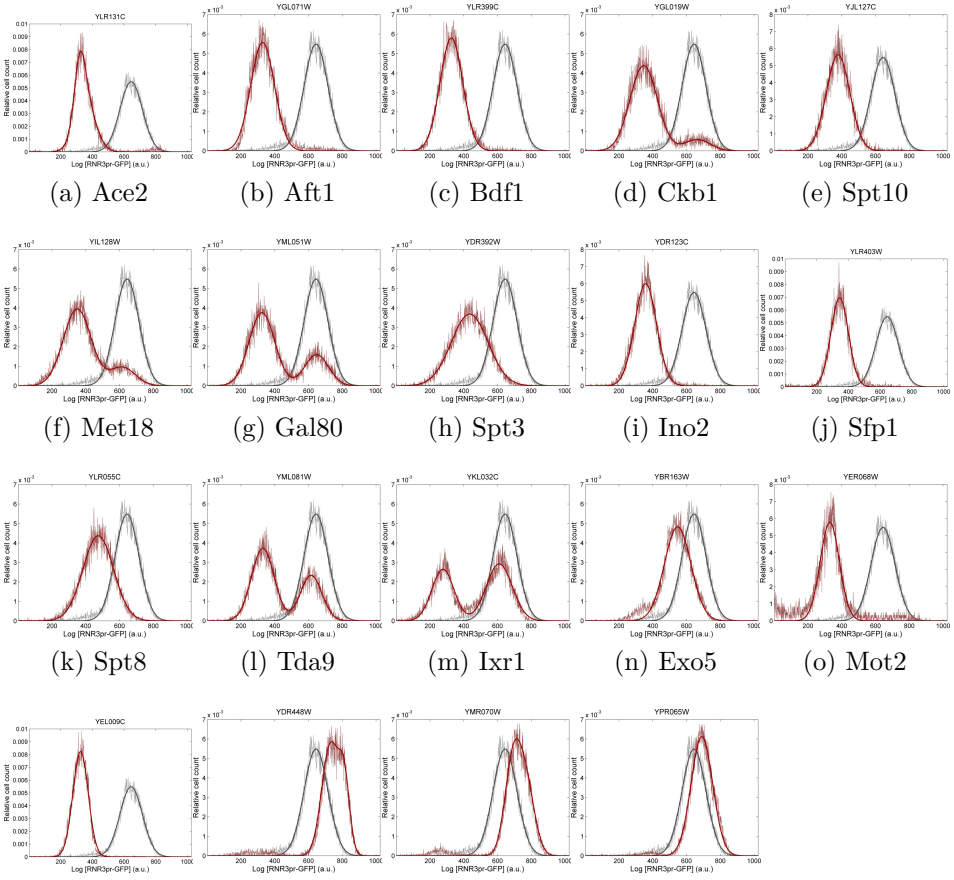
Due to the ambiguity regarding *CRT1* deletion for some of the strains, I performed mating experiments to determine whether the SGA strains were haploid or diploid, and if haploid, which mating type was predominant (see Methods). The results of the mating experiment are shown in Table 3.9. All strains with ambiguous Crt1 deletion mated with the MATα tester strain and not with the MATa tester, suggesting that they were haploid MATa. Spt3 and Spt8 were included in this experiment as controls since they both passed the PCR validation steps, but showed very little (Spt3) or no mating at all (Spt8). This was not of major concern because based on GO annotations (Appendix C Table C.3), Spt3 is involved in sporulation and the deletion of Spt3 and Spt8 have been shown in previous research to impair the mating ability of cells [22, 38].

Sporulation

As an alternative way to determine whether a particular SGA population was made up of diploids, I attempted to sporulate the ambiguous strains (see Methods). The results are shown in Table 3.10. The six SGA library strains with ambiguous *CRT1* deletion did not show any tetrads, also suggesting haploidy.

Hit regeneration

The results of the previous validation steps were inconclusive since PCR suggested that some strains were diploid, but plating, mating and sporulation suggested otherwise. Therefore, I attempted to regenerate 14 out of the top 19 deletions identified as



(a) Ace2 (b) Aft1 (c) Bdf1 (d) Ckb1 (e) Spt10

(f) Met18 (g) Gal80 (h) Spt3 (i) Ino2 (j) Sfp1

(k) Spt8 (l) Tda9 (m) Ixr1 (n) Exo5 (o) Mot2

(p) Gcn4 (q) Ada2 (r) Mot3 (s) Rox1

Figure 3.10: Log GFP fluorescence distributions for the top 19 SGA library hits, all identified as outliers by Z-scores with p values $\leq 2.72 \times 10^{-5}$. Each distribution is for one experimental measurement (one technical replicate). a-s. The strain labeled in grey contains a Crt1 deletion and a neutral control deletion. The strains labeled in red (or the darker distribution if viewing without colour) contain a Crt1 deletion along with a factor deletion of interest (labels are the names of the factors deleted).

Table 3.9: Results of Mating experiment. * indicates only one colony was observed.

SGA library strain	Mated with MATα tester	Mated with MATa tester	Mating type inferred
Ace2	Y	N	MATa
Aft1	Y	N	MATa
Ckb1	Y	N	MATa
Sfp1	Y	N	MATa
Spt10	Y	N	MATa
Mot2	Y	N	MATa
Spt3	Y*	N	MATa
Spt8	N	N	undetermined

Table 3.10: Results of Sporulation experiment.

SGA library strain	Tetrads observed
Ace2	N
Aft1	N
Ckb1	N
Sfp1	N
Spt10	N
Mot2	N

hits in haploids of known genotype (strain x) (see Methods) to see which were true hits and which were false positives.

Of the 14 hits chosen for regeneration, deletion of nine were confirmed by PCR (see Methods). These were Ace2, Aft1, Spt10, Met18, Spt3, Spt8, Ada2, Mot3, and Rox1. Trans-factor deletions that were performed but not able to be verified included Bdf1, Ckb1, Sfp1, Mot2 and Gcn4. For these strains, the PCR validation reaction for multiple transformants yielded a negative result, suggesting that the *NATMX* cassette had not been successfully integrated.

I then assayed the verified strains for GFP expression (see Methods). Of the nine hits verified by PCR, only four resulted in a similar effect on *RNR3* expression as I observed in the original screen (compare Figure 3.11 with Figure 3.10). *RNR3* expression was decreased upon deletion of Spt3 and Spt8, validating their effect as activators. Additionally, *RNR3* expression was increased upon deletion of Mot3 and Rox1, verifying their effect as repressors.

The rest of the PCR-validated regenerated hits did not impact *RNR3* expression as predicted based on the SGA. Aft1, one of the ambiguous Crt1 deletion hits, did not appear to have any impact on *RNR3* expression upon deletion (Figure 3.11). The effect on expression of deleting of Met18 and Ada2 was also very similar to the level of expression without the deletion.

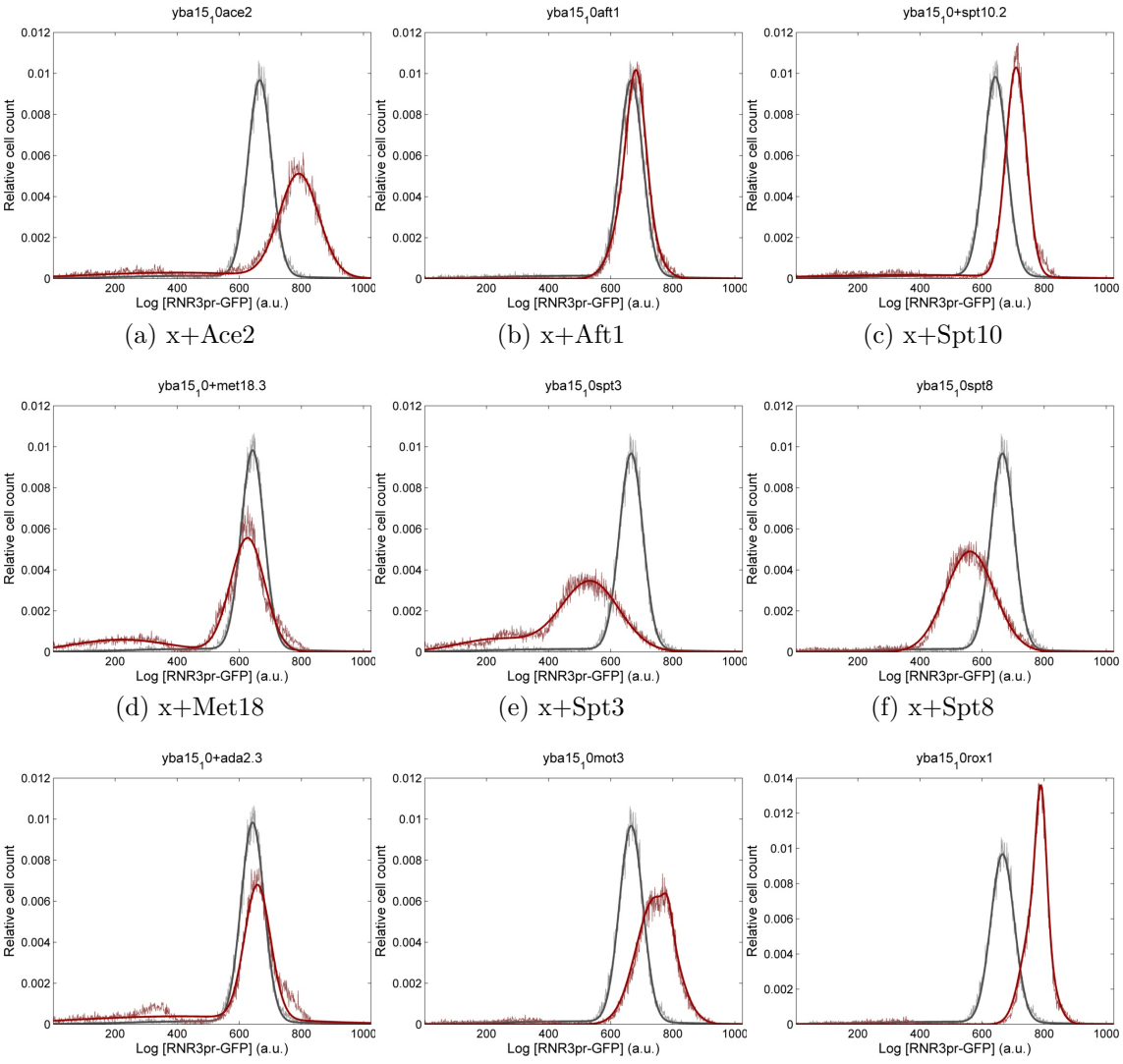
Ace2 and Spt10 deletion resulted in increased *RNR3* expression rather than the expected decrease, presumably due to morphological effects which are known to confound data because they cause changes in autofluorescence [77].

Ace2 is responsible for degrading the septum after cytokinesis (Appendix C Table C.2) and I observed that cells lacking this protein had an abnormal phenotype under the light microscope. Ace2 and Spt10 also showed an increase in the forward/side scatter distribution collected via flow cytometry (Figure 3.12). Forward scatter is indicative of cell size while side scatter is indicative of cell complexity. Morphology changes to the cell would therefore be expected to change the plot of the log of these properties. Thus, the increase in GFP was expected to represent a general effect on the cell (i.e. a change in size or shape that allowed for abnormal amounts of protein accumulation), rather than an impact on *RNR3* expression.

3.5 Validated Crt1-independent regulators of *RNR3* affected by X-box perturbations

For the four hits that were validated by regeneration, I assessed the dependency of their effects on the X-boxes. Performing the factor deletion in strains with and without X-box perturbations (strains xuvw and x respectively) (see Methods) allowed me to determine the dependency of the factor on the X-boxes under these conditions. If the X-boxes did not have a role in mediating regulation by the factor, I expected the percentage change in GFP expression upon factor deletion in a strain containing X-boxes replacements to match the percentage change upon factor deletion in a strain containing no X-box replacements.

Results are shown in Figures 3.13 and 3.14. Deleting Spt3 and Spt8 had approximately 10% less of an impact on expression when the X-boxes were deleted.

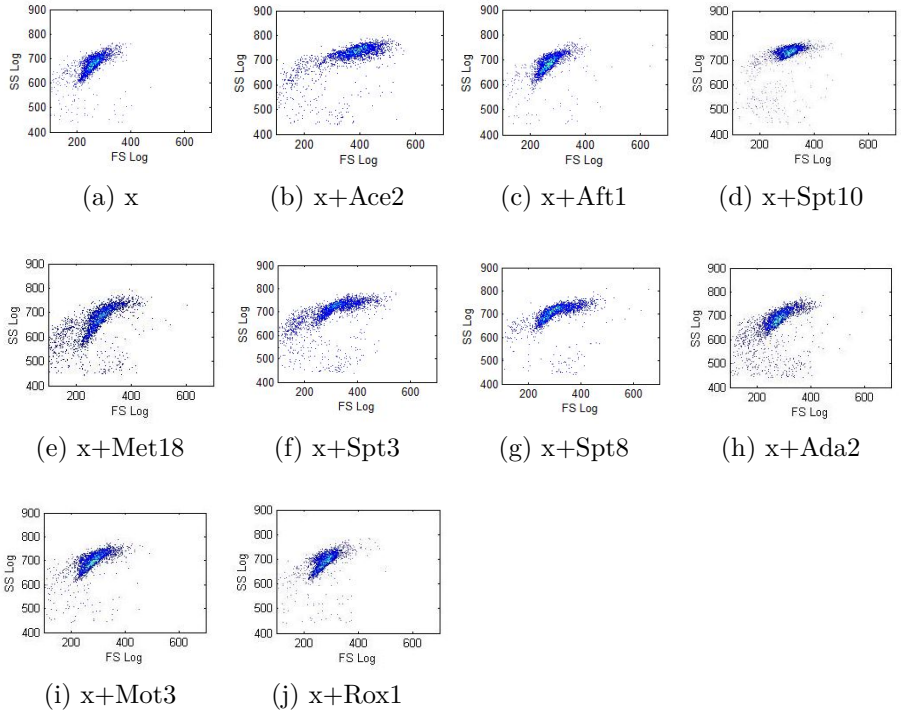


(a) x+Ace2 (b) x+Aft1 (c) x+Spt10

(d) x+Met18 (e) x+Spt3 (f) x+Spt8

(g) x+Ada2 (h) x+Mot3 (i) x+Rox1

Figure 3.11: Log GFP fluorescence distributions for the nine PCR-validated regenerated SGA library hits. a-i. The parental strain x is labeled in grey and the regenerated SGA library strain is overlayed in red (or darker shade if viewing without colour).



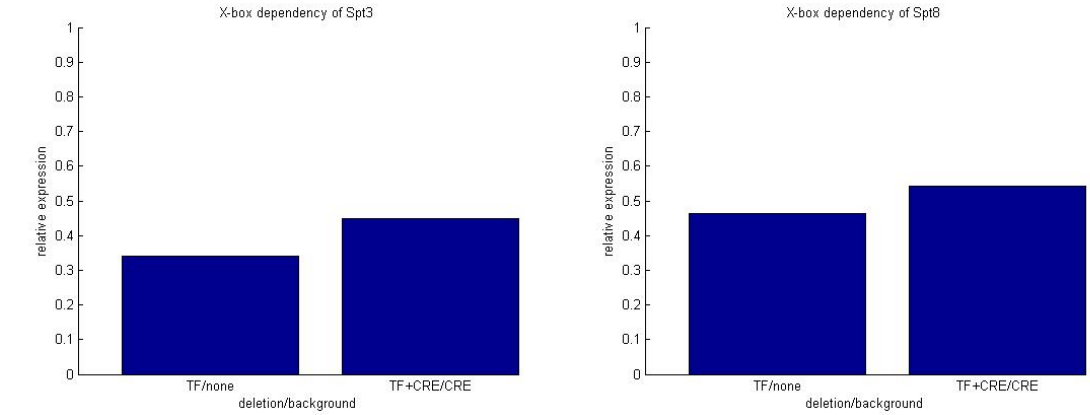
(a) x (b) x+Ace2 (c) x+Aft1 (d) x+Spt10

(e) x+Met18 (f) x+Spt3 (g) x+Spt8 (h) x+Ada2

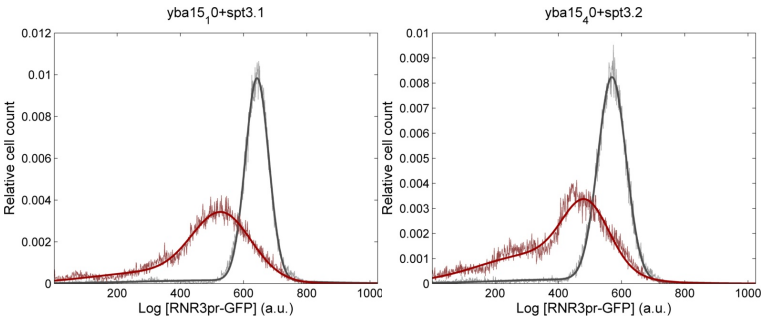
(i) x+Mot3 (j) x+Rox1

Figure 3.12: Plots of forward scatter/side scatter distributions for a. the parental control strain x and b-j. the nine regenerated SGA library strains.

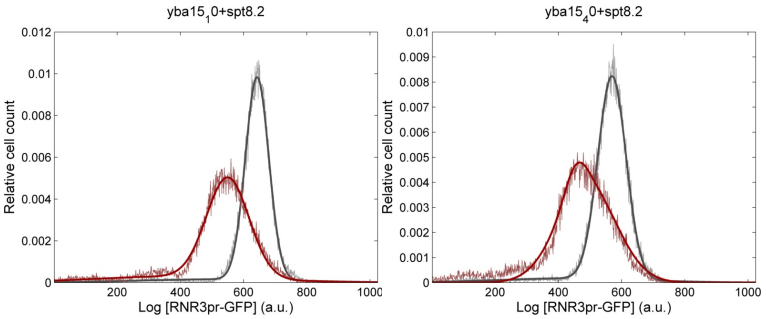
Therefore, there was blunting of the effect of deleting the factor when the X-boxes were also deleted. This partial masking suggested that the X-boxes could play a small role in activation by Spt3 and Spt8. In contrast, deleting Mot3 and Rox1 had approximately 75% more impact on expression when the X-boxes had been replaced. This suggested a potential role for the X-boxes in inhibiting the binding of Mot3 and Rox1.



(a) Spt3 (b) Spt8

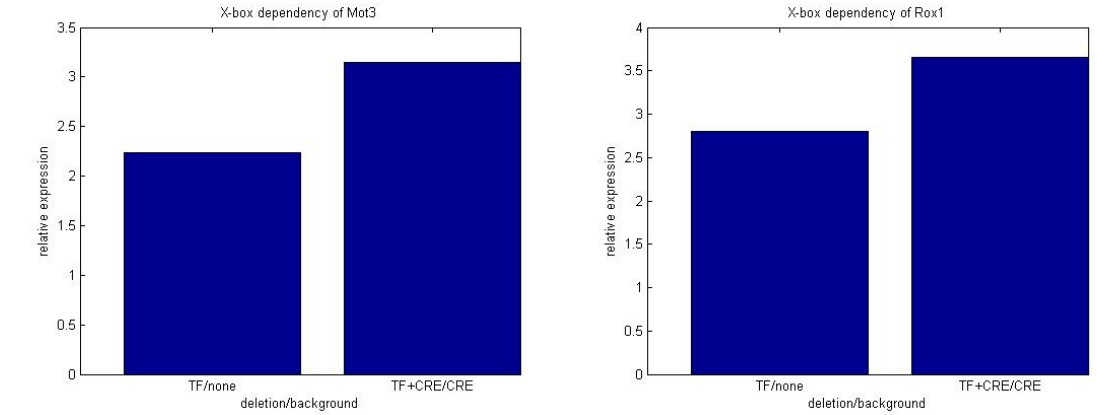


(c) x+spt3 (d) xuvw+Spt3

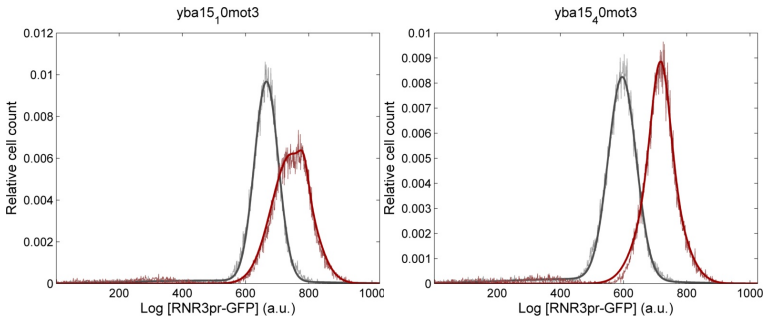


(e) x+Spt8 (f) xuvw+Spt8

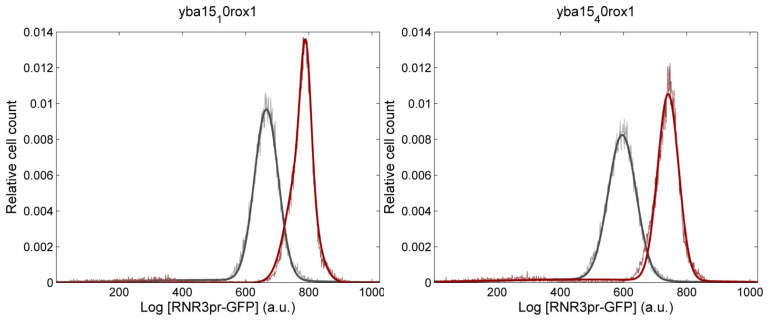
Figure 3.13: X-box dependencies of Spt3 and Spt8. a-b. Mean linear GFP expression is plotted for each strain, relative to the background strain (x or xuvw) after subtraction of autofluorescence. none= no deletions, TF = factor deleted, CRE = u,v,w deleted. c-f. log GFP expression distribution of the background strain in grey overlayed by the additional effect of factor deletion in red (red appears darker if viewing without colour).



(a) Mot3 (b) Rox1



(c) x+mot3 (d) xuvw+mot3



(e) x+rox1 (f) xuvw+rox1

Figure 3.14: X-box dependencies of Mot3 and Rox1. a-b. Mean linear GFP expression is plotted for each strain, relative to the background strain (x or xuvw) after subtraction of autofluorecence. none= no deletions, TF = factor deleted, CRE = u,v,w deleted. c-f. log GFP expression distribution of the background strain in grey overlayed by the additional effect of factor deletion in red (darker than the background strain if viewing without colour).

Chapter 4

Discussion

The results of my thesis support the first hypothesis investigated. A higher order interaction between X-box elements u and w mediates *RNR3* transcription in the presence of Crt1. In contrast to my second hypothesis, however, my results suggest that X-box elements u and v have important functional roles for *RNR3* transcription in the absence of Crt1. Spt3, Spt8, Mot3 and Rox1 all affect *RNR3* transcription in the absence of Crt1, and replacement of the X-boxes appears to alter the functional impact of these factors on transcription.

4.1 Functional roles of X-box elements in the *RNR3* promoter

My systematic perturbation of the three X-boxes revealed important information about the roles of these CREs in the repression of *RNR3* transcription. Notably, in the presence of Crt1, I observed a higher-order interaction that was independent of the replacement approach and neutrality model; the interaction between X-boxes u and w, which are the farthest spaced sites in the *RNR3* promoter, and are found in opposite relative orientation. This interaction was classified as synergistic (greater

than additive and greater than multiplicative) in the Tet operator replacement approach but greater than additive, less than multiplicative in the scrambling approach. Regardless of replacement approach, the additive model identified this interaction as synergistic.

Interestingly, X-box u and w have been previously identified as weak and strong sites respectively with regard to how strongly they bind Crt1 [40]. Cooperativity between a strong and a weak site in transcriptional repression has also been observed in previous work with the mechanistically similar TF Mig1 [31, 93]. In this study [31], the combined effect of a strong and a weak binding site was the same as that of two strong binding sites. Such a synergistic interaction is consistent with a mechanistic model where a TF bound to the strong site interacts with the weak site to establish a DNA loop with one of the weaker sites to increase the strength of repression. It is possible that I did not detect an interaction between the strong X-box w and X-box v (also identified as weak in [40]), because X-box v has the lowest Crt1 binding affinity [40] and may also be too close to X-box w to stabilize loop formation.

If DNA looping does occur between u and w, one way for it to happen would be through Crt1 dimerization. As discussed earlier, RFX proteins are known to dimerize in higher eukaryotes through a dimerization domain in the C-terminus [24], and a model of DNA looping has been proposed to account for the fact that dimers can bind to non-palindromic X-boxes [76]. Some work suggests that Crt1 cannot dimerize or repress through the C-terminus, and that Ssn6 took over these functional roles early in evolution [48]. However, Zhang and colleagues [102] have found that Crt1 has a repression function in the C-terminus that is separate from the N-terminus known to interact with Ssn6.

Although my data cannot distinguish between Crt1 dimerization and Ssn6-mediated cooperativity between elements, both scenarios are consistent with looping of the

DNA. Formation of a DNA loop between u and w would bring the Mot3 binding site next to u and the Rox1 binding site next to w in close proximity to each other (see Figure 3.5). This model would be consistent with the previous observation that Mot3, Rox1 and Crt1 synergistically interact to establish repression, likely through Ssn6 [50].

My systematic approach also revealed challenges in identifying expected interactions. Based on previous work [40], I expected that each X-box would mediate a significant amount of repression on its own with the effects of X-box u and v being quite weak. Scrambling X-box w resulted in a significant increase in expression (regardless of neutrality model), verifying its importance in repression. However, I could not detect any significant individual effect of scrambling X-box u or v using multiple linear regression. This may indicate that flow cytometry is not a sensitive enough assay to detect the small effects of u and v. Alternative reporter genes (see [45] for review) should be investigated and compared to GFP in the future. The *LACZ* reporter, which can be measured using a sensitive enzymatic assay, may be better suited to capture the effect of weak binding sites on expression.

Another reason for failure to identify important interactions may have been my data analysis method. I analyzed all the variables at once and identified a subset of interactions that were significant under both the additive and multiplicative regression models. Restricting my analysis was necessary to reduce type I error due to systematic errors in prediction for both models. However, including all the variables and focusing on only these interactions could have increased type II error. This problem could be resolved in the future by testing the impact of subsets of variables in several smaller models as opposed to one large model, similar to [51].

Unexpectedly, my analysis of systematic X-box perturbations in the absence of Crt1 suggested that these regulatory elements have functions that are independent of

this repressor. This represents a significant deviation from the commonly accepted model of *RNR3* transcriptional regulation and X-box function [40, 53]. Deleting Crt1 significantly increased *RNR3* reporter expression, as expected [40, 103]. However, X-box perturbations in the absence of Crt1 revealed significant effects on expression of eliminating the weak X-boxes (predictors *xu* and *xv*). The antagonistic interactions between x and u as well as x and v could not be accounted for by Crt1-dependent masking of the effects of the X-boxes. This in turn suggests that unidentified TFs interact in a binary manner with u and v to mediate *RNR3* activation. Since the sequences of X-box u and v are nearly identical, it is possible that they associate with the same activator.

It is worth considering why, in contrast to my results, Huang and colleagues [40] did not observe a significant impact upon X-box mutation in the absence of Crt1. It may be due to the fact that I changed the entire 13 bp X-box sequence. Huang and colleagues only mutated two nucleotides, leaving the functional impact of the rest of the nucleotide positions in the site unperturbed. The X-boxes share 9 of 14 bases with other species [40]. This leaves five variable nucleotide positions between species. There is additional variability within yeast as the sequence composition between the three sites is not identical. These sources of variability are likely functionally important since single nucleotides at a specific position in a consensus CRE can impart significant function to a promoter [86].

It is as a final note worth pointing out that I chose an approach that completely eliminated the X-box elements from the *RNR3* promoter. This approach is not without limitations, since this could also eliminate functional binding sites for TFs other than Crt1, or even introduce new ones. Although Tet operator sequences have been used in many synthetic yeast promoters [23], the sequence contains a consensus sequence associated with the transcriptional repressor Ash1 (Yeastract [1, 60, 87]) and

it is possible that this site contributes to the repression observed in the absence of Crt1. The observation that scrambling the X-box sequences, which contain no known consensus sequences (Yeastract [1, 60, 87]), yields a similar repressing effect alleviates this concern in the present context.

Even though the scrambled X-box sequences contain no known consensus sequences, it is still possible that they function as weak TF binding sites. Indeed, with a single nucleotide substitution, all sequences used to replace the X-boxes encode consensus Mot3 binding sites (Yeastract [1, 60, 87]). It is therefore possible that Mot3 could bind to these sequences weakly and is responsible for the effects on transcription observed in the absence of Crt1. One way to eliminate the effect of the random introduction of binding sites in future experiments would be to test a library of random sequences for replacing each X-box instead of just one. This way, the library expression as a whole would eliminate the impact of random site insertions and enable assessment of the functional impact of deleting the CRE in question.

4.2 Crt1-independent *RNR3* regulators

Because of the evidence I had suggesting additional functional roles of the X-box elements in the transcriptional regulation of *RNR3*, I systematically assessed the impact of transcription-related genes on Crt1-independent *RNR3* expression. My analysis identified four factors that affected *RNR3* reporter expression in the absence of Crt1. Two activators (Spt3 and Spt8) and two repressors (Mot3 and Rox1) were initially identified using the high-throughput screening approach and subsequently validated. Spt3 and Spt8 are part of the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex which is known to play important roles in RNA polymerase II-dependent transcription in yeast (Appendix C Tables C.2 and C.3) [83], and is thought to help deliver core transcriptional machinery to the *RNR3* promoter [100]. Mot3 and Rox1

have also been implicated in *RNR3* repression [50].

The dependencies of Spt3, Spt8, Mot3 and Rox1 on the X-boxes for their regulation of *RNR3* in the absence of Crt1 had not been previously mapped. My results suggest that the X-boxes may be involved in blocking the repressive effects of Mot3 and Rox1. Deleting these factors had a greater impact on expression when the X-boxes had been eliminated. Since binding sites for Mot3 and Rox1 are found relatively close to the X-boxes (Figure 3.5), it would not be hard to imagine a co-factor that binds to the X-box sites during activation, preventing the repressors from interacting with their binding sites by steric hindrance.

My analysis also suggests that the X-boxes may play a small role in activation of *RNR3* by Spt3 and Spt8. Deleting the X-boxes and Spt3 or Spt8 allowed for some blunting of the effect of deleting the activator on its own. However, the difference in relative expression was small, i.e. nowhere close to 100% which would indicate complete masking/ epistasis, meaning that Spt3 and Spt8 have X-box independent functions as well. This is consistent with previous observations indicating that both the upstream regulatory sequence (which includes the X-boxes) and the core promoter region are important for Spt3 binding and SAGA-mediated activation [100].

The X-box dependencies of the Crt1-independent regulators must be investigated in more detail in the future due to the possibility of random introduction of binding sites and the small amount of data collected. One way to do this would be to combine the deletion of the regulators with elimination of the X-boxes using a library of random sequences. Further investigations will also be required to determine if the suspected X-box dependent regulatory functions are direct or indirect. This could be accomplished by determining whether the factor binds directly to the X-boxes using a low-throughput ChIP assay or electrophoretic mobility assay [67].

My identification of Spt3, Spt8, Mot3 and Rox1 as putative X-box dependent reg-

ulators highlights some of the complications associated with current high-throughput strain generation technologies. Specifically, in the SGA procedure, there are many steps that can potentially introduce artifacts leading to false-positive identifications. In my case, false identification of activators occurred because one copy of a transcriptional repressor that should have been deleted was retained. As a result, transcription was low in these strains, mimicking the effect expected by the loss of function of an activator. Because of the needed regeneration step due to false positives, I had to restrict my focus to a subset of the identified hits. Thus, factors affecting *RNR3* activation and binding to the X-boxes may have been missed.

Interestingly, my examination of genes identified in the initial screen revealed several potential activators already known to be associated with *RNR3* or DNA damage (Appendix C Tables C.2 and C.3). However, my further examination by PCR amplification using *CRT1* specific primers suggested that many of these strains retained an intact copy of *CRT1*. Curiously, these strains turned out not to be diploids according to mating and sporulation experiments.

Two possible routes by which a copy of *CRT1* can be retained are meiotic non-disjunction and chromosomal recombination. Meiotic non-disjunction involves the retention of an extra chromosome by unequal segregation during anaphase 1. This effect has been reported to occur with SGA previously [82]. Alternatively, homologous recombination of *CRT1* into the genome at an alternative location during the diploid stage could explain the results. Recombination is more common in a DNA-damage deficient background [28] so it is expected to be higher than usual in my particular experiment. In addition, these effects are expected to be even more common when two genes (*CRT1* and the gene encoding another trans-factor) involved in the DNA damage pathway are deleted. This would explain the curious observation that many of the hits that had an extra copy of *CRT1* have also been shown to be related to

DNA damage.

Overall, my results suggest that X-boxes contribute to activation, possibly by co-factor-mediated blocking of Mot3 and Rox1. The simplest explanation would be that a factor related to Crt1 mediates effects observed in its presence and absence. Ssn6 is a prime candidate for consideration.

The first possibility is that Ssn6 on its own mediates activation through the X-boxes. Ssn6-Tup1 acts as a repressor when recruited to *RNR3* by Crt1 [53] but is known to remain at the promoters of active genes [98]. Tup1 and Crt1 are known to dissociate from *RNR3* upon activation [102]. The C-terminus of Ssn6 which is not required for repression has an activation function at some promoters through its interaction with the activator TF Rtg3 [19]. Interestingly, there is an Rtg3 consensus binding site located within X-box u (Figure 3.5). Rtg3 was not identified as a hit in my screen, but this may represent a false negative result.

Alternatively, Crt1 and Ssn6 may both contribute to activation. Human RFX proteins are involved in repression and activation of transcription [47]. Some work suggests that Crt1 is both a repressor and a transient activator of *RNR3* transcription [102]. It has been argued that since Crt1 is required for repression, deleting it masks its role in activation [102]. Crt1 may be involved in reversing the repressed state, which is not necessary when it is deleted [102]. Maintenance of an activated state once the repressed state is reversed [102] may require Ssn6. When Ssn6 is deleted, Crt1 does not dissociate from *RNR3* and very little activation is possible [53, 102]. However, when Crt1 is deleted and Ssn6 is present, full activation is possible [102]. This supports a multi-step activation model and that Crt1 acts upstream of Ssn6 in the earlier step. My results are not inconsistent with a model where Ssn6 dislodges Crt1 and occupies the X-boxes during the second step.

It is further possible that Crt1 and Ssn6 play redundant roles in activating essential

RNR genes, and that each factor on its own causes an amount of activation that prevents cell death. My SGA did not produce a colony for the Δssn6/crt1 double mutant. It turns out that the Reese lab also had difficulty generating the mutant (correspondence with Joseph Reese via email), providing further support that the two factors share a redundant function in activating an essential process. Since the RNR enzyme is required to synthesize DNA before replication, lack of activation would lead to cell death. Even though *RNR3* is not essential, *RNR2* and *RNR4* are (SGD [97]) and they both have multiple X-box sequences in their promoters [40].

The notion that Ssn6 is a dual repressor-activator of *RNR3* is an important hypothesis to investigate in the future. This would best be accomplished in a similar manner to [102], where the functional activation and repression domains of Crt1 were discovered by systematic perturbation in the absence of Ssn6. The C-terminus of Ssn6 has been previously implicated in promoter activation [19]. Therefore, portions of its C-terminal domain could be eliminated systematically in a background where Crt1 has been deleted. This way, we could see which portion of Ssn6 is important for activation of *RNR3* in the absence of Crt1. The dependency of an important region of Ssn6 on the X-boxes could then be mapped in a similar manner to how I mapped the X-box dependencies of Spt3, Spt8, Mot3 and Rox1.

4.3 New models of *RNR3* repression and activation

Results from my work and the literature can be amalgamated to produce models of *RNR3* promoter function. These models can be used to drive further research.

Repression and de-repression of the *RNR3* promoter involves coordinated action among a variety of factors and complexes [50, 53, 80, 88, 100, 101, 102] which also

depends on the shape of the DNA, similar to the enhanceosome (a protein complex composed of transcriptional activators) found in higher eukaryotes [11, 53].

A model for repression is several TFs acting together with one co-repressor complex to keep the chromatin in a repressed state (Figure 4.1). Under baseline conditions, Crt1 interacts with its three X-boxes and the Ssn6-Tup1 co-repressor complex [40]. The interaction between X-box u and w I identified is consistent with both Crt1 dimerization and flexible interactions with the Ssn6 co-factor. Thus, both are included in the model. Crt1 dimerization is represented by a dashed line between the two molecules bound to u and w. The interaction is illustrated by looping of the DNA, allowing Crt1 molecules to interact via distant sites. Mot3 and Rox1 are thought to act synergistically with Crt1 through Ssn6 [50]. Ssn6, in turn, interacts with nucleosomes and Crt1 to bend the promoter DNA and repress transcription [53].

De-repression/activation of *RNR3* is complex and understanding may be facilitated by thinking of it as involving at least two steps: the first step involves chromatin modification and depends on Crt1, and the second step involves maintenance of the chromatin state and PIC formation [53, 80, 88, 100, 101, 102]. Since my work examined *RNR3* activation in the absence of Crt1, my results add to a model of the later step.

A model for the first step of activation of *RNR3* is shown in Figure 4.2. Opening the chromatin requires the general transcription factor complex TFIID and the chromatin remodeling complex SWI/SNF [100, 101]. Crt1 is known to interact with TFIID and SWI/SNF [102]. The repressor/activator TF Rap1 also interacts with TFIID and SWI/SNF [88]. It does not have consensus sequences in the *RNR3* promoter but binds to the upstream regulatory sequence [88]. In addition to what has been established, our lab has data suggesting that the transcriptional activator Swi4 may play a role in this process. Swi4 is a documented TF for *RNR3* in Yeastract [1, 60, 87], there is

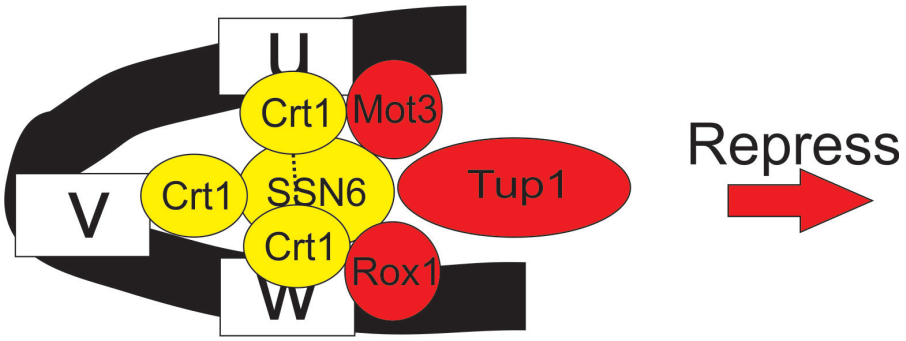


Figure 4.1: Updated model for repression of *RNR3*.

a consensus sequence for this activator directly upstream of X-box v, and the effect of deleting Swi4 is masked by deleting Crt1 (GFP fluorescence expression data from Hilary Phenix, not shown). The effect of another activator that acts during the first step of activation (Rap1) is also masked by deleting Crt1 [80, 88]. These epistatic relationships suggest that Swi4 and Rap1 act downstream of Crt1, meaning they are activated by or dependent on Crt1 for their function.

The second stage of activation is less well defined. The existing evidence seems to suggest that there may be an intermediate activation step between the first and second step that depends on a Crt1-defined chromatin structure, but not Crt1 itself. The chromatin modifying protein Rpd3L is recruited to the upstream regulatory sequence of *RNR3* after TFIID and SWI/SNF modify the chromatin, and its recruitment does not depend on Crt1 [80]. However, the effect of Rpd3L on activation is masked by deleting Crt1 [80]. This epistatic relationship suggests that Rpd3 is somehow dependent on Crt1 [2]. Taken together, the existing evidence suggests that the activating effects of Rpd3L depend on a chromatin configuration at *RNR3* that is absent when Crt1 is deleted. Swi4 may also be involved in this intermediate step, since our laboratory methods are unable to separate these preliminary steps.

A model of the second step of *RNR3* activation is shown in Figure 4.3. Once the chromatin is opened up, there is no longer a requirement for co-activators like

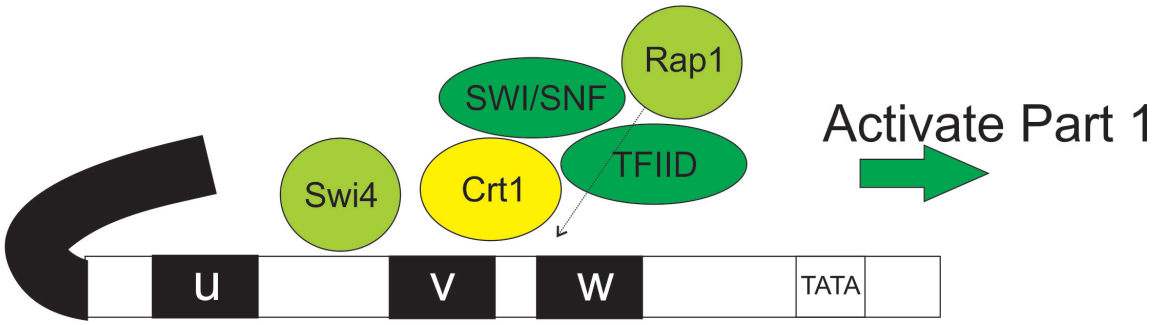


Figure 4.2: First Step of activation of *RNR3*.

SWI/SNF and TFIID [101], and Crt1 dissociates from the promoter [40, 102]. Evidence suggests that the final stage (second step) of *RNR3* activation requires only factors mediating PIC formation and another factor [88]. For most TATA-containing promoters, PIC formation includes the co-activator complex SAGA [33]. My work shows consistency with [100] in that Spt3 and Spt8 (part of SAGA) are important factors for *RNR3* activation when Crt1 is absent. My data further suggest that X-boxes u and v individually contribute to activation in the absence of Crt1. Further, I observed that repressors bound better when the X-box sequences were replaced. Consistent with all of these results is a model where Crt1 dissociation (by deletion or DNA damage in a native context [40]) disrupts the conditions for repression. Dissociation results in a change in the conformation of the Ssn6 protein [19] while it remains at the X-boxes. This altered conformation inhibits Tup1, Mot3 and Rox1 from binding. Since Ssn6 is not known to bind DNA directly [85], it is possible that another activator TF (e.g. Rtg3) is yet to be identified that interacts with Ssn6 to mediate the activation.

4.4 Conclusion

The goal of this thesis was to investigate how multiple promoter elements combinatorially define transcriptional regulation. My results suggest that the conserved X-box

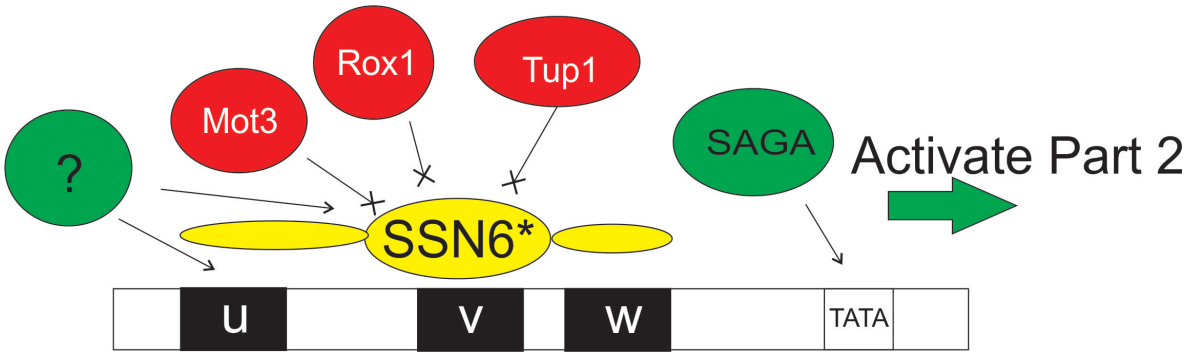


Figure 4.3: Updated model for second step of activation of *RNR3*.

sequence elements play an intricate role in *RNR3* transcriptional regulation. Specifically, higher-order X-box-mediated interactions are important for *RNR3* transcription and the role of the X-boxes extends beyond their interaction with Crt1.

Overall, the results in this thesis illustrate the utility of systematic perturbation and multiple linear regression for quantitative identification of functional interactions. This is a useful step towards the creation of predictive models of transcriptional regulation. However, my investigations also highlight that significant challenges must be overcome to develop comprehensive thermodynamic models of transcriptional regulation through native eukaryotic promoters. Even for a promoter supposedly regulated by a simple repression mechanism in a relatively simple microorganism, developing such models will likely be highly resource intensive. Firstly, highly sensitive transcriptional assays must be employed in order to accurately characterize the function of individual CREs and the interactions between them. Current flow cytometry technologies may be insufficient to fully characterize the functions of relatively weak CREs, and other, more sensitive assays are not easily performed in a high-throughput manner.

The absence of accurate systematic perturbation data inevitably limits the accuracy and scope of the thermodynamic models that can be supported. Secondly, as my research suggests, CREs may have unexpected functions that extend beyond what is

generally captured by binary interactions in thermodynamic models. Closer analysis of the existing literature on the interaction between Crt1 and Ssn6 further supports this notion. Correspondingly, it will be necessary to employ an iterative approach, similar to the one employed here, where these additional roles are identified and characterized. The basic problem here is that the scope of systematic perturbation approaches scales exponentially with the number of components analyzed. Correspondingly, the generation of complete perturbation datasets for natural promoters will likely exceed what can be accomplished as a thesis research project.

References

[1] D Abdulrehman, P T Monteiro, M C Teixeira, N P Mira, A B Lourenço, S Costa dos Santos, T R Cabrito, A P Francisco, S C Madeira, R S Aires, A L Oliveira, I Sá-Correia, and A T Freitas. YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in saccharomyces cerevisiae through a web services interface. *Nucleic Acids Research*, 39(suppl 1):D136–D140, January 2011. PMID: 20972212.

[2] L Avery and S Wasserman. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends in genetics: TIG*, 8(9):312–316, September 1992. PMID: 1365397.

[3] C Batenchuk, L Tepliakova, and M Kaern. Identification of response-modulated genetic interactions by sensitivity-based epistatic analysis. *BMC Genomics*, 11(1):493, September 2010. PMID: 20831804.

[4] S Berthelet, J Usher, K Shulist, A Hamza, N Maltez, A Johnston, Y Fong, L J Harris, and K Baetz. Functional genomics analysis of the saccharomyces cerevisiae iron responsive transcription factor aft1 reveals iron-independent functions. *Genetics*, 185(3):1111–1128, July 2010. PMID: 20439772.

[5] L Bintu, N E Buchler, H G Garcia, U Gerland, T Hwa, J Kondev, T Kuhlman, and R Phillips. Transcriptional regulation by the numbers: applications. *Cur-*

rent opinion in genetics & development, 15(2):125–135, April 2005. PMID: 15797195.

[6] L Bintu, N E Buchler, H G Garcia, U Gerland, T Hwa, J Kondev, and R Phillips. Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–124, April 2005. PMID: 15797194.

[7] D Botstein and G R Fink. Yeast: an experimental organism for modern biology. *Science*, 240(4858):1439–1443, June 1988. PMID: 3287619.

[8] D Botstein and G R Fink. Yeast: An experimental organism for 21st century biology. *Genetics*, 189(3):695–704, November 2011. PMID: 22084421 PMCID: PMC3213361.

[9] C B Brachmann, A Davies, G J Cost, E Caputo, J Li, P Hieter, and J D Boeke. Designer deletion strains derived from saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)*, 14(2):115–132, January 1998. PMID: 9483801.

[10] N E Buchler, U Gerland, and T Hwa. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, 100(9):5136–5141, April 2003. PMID: 12702751.

[11] M Carey. The enhanceosome and transcriptional synergy. *Cell*, 92(1):5–8, January 1998. PMID: 9489694.

[12] A Chabes, B Georgieva, V Domkin, X Zhao, R Rothstein, and L Thelander. Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell*, 112(3):391–401, February 2003.

[13] A Chatr-Aryamontri, B-J Breitkreutz, S Heinicke, L Boucher, A Winter, C Stark, J Nixon, L Ramage, N Kolas, L O'Donnell, T Reguly, A Breitkreutz, A Sellam, D Chen, C Chang, J Rust, M Livstone, R Oughtred, K Dolinski, and M Tyers. The BioGRID interaction database: 2013 update. *Nucleic acids research*, 41(Database issue):D816–823, January 2013. PMID: 23203989.

[14] S Chatterjee and J S Simonoff. Multiple linear regression. In *Handbook of Regression Analysis*, pages 1–21. John Wiley & Sons, Inc., 2012.

[15] J M Cherry, E L Hong, C Amundsen, R Balakrishnan, G Binkley, E T Chan, K R Christie, M C Costanzo, S S Dwight, S R Engel, D G Fisk, J E Hirschman, B C Hitz, K Karra, C J Krieger, S R Miyasato, R S Nash, J Park, M S Skrzypek, M Simison, S Weng, and E D Wong. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Research*, 40(D1):D700–D705, November 2011.

[16] J S C Chu, M Tarailo-Graovac, D Zhang, J Wang, B Uyar, D Tu, J Trinh, D L Baillie, and N Chen. Fine tuning of RFX/DAF-19-regulated target gene expression through binding to multiple sites in caenorhabditis elegans. *Nucleic acids research*, 40(1):53–64, January 2012. PMID: 21908398.

[17] J P Cogswell, P V Basta, and J P Ting. X-box-binding proteins positively and negatively regulate transcription of the HLA-DRA gene through interaction with discrete upstream w and v elements. *Proc Natl Acad Sci U S A*, 87(19):7703–7707, October 1990. PMID: 2120707 PMCID: PMC54816.

[18] R S Cohen and M Meselson. Periodic interactions of heat shock transcriptional elements. *Nature*, 332(6167):856–858, April 1988. PMID: 2833709.

[19] R S Conlan, N Gounalaki, P Hatzis, and D Tzamarias. The tup1-cyc8 protein

complex can shift from a transcriptional co-repressor to a transcriptional co-activator. *J. Biol. Chem.*, 274(1):205–210, January 1999.

[20] F Cornille, P Emery, W Schuler, C Lenoir, B Mach, B P Roques, and W Reith. DNA binding properties of a chemically synthesized DNA binding domain of hRFX1. *Nucleic Acids Research*, 26(9):2143–2149, May 1998. PMID: 9547272 PMCID: PMC147513.

[21] M Dadiani, D van Dijk, B Segal, Y Field, G Ben-Artzi, T Raveh-Sadka, M Levo, I Kaplow, A Weinberger, and E Segal. Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. *Genome Research*, 23:966–976, February 2013.

[22] D M Eisenmann, C Chapon, S M Roberts, C Dollard, and F Winston. The saccharomyces cerevisiae SPT8 gene encodes a very acidic protein that is functionally related to SPT3 and TATA-binding protein. *Genetics*, 137(3):647–657, July 1994. PMID: 8088510.

[23] T Ellis, X Wang, and J J Collins. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotech*, 27(5):465–471, May 2009.

[24] P Emery, B Durand, B Mach, and W Reith. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res*, 24(5):803–807, March 1996. PMID: 8600444 PMCID: PMC145730.

[25] T D Frank, A M Carmody, and B N Kholodenko. Versatility of cooperative transcriptional activation: A thermodynamical modeling analysis for greater-than-additive and less-than-additive effects. *PLoS ONE*, 7(4):e34439, April 2012.

[26] Y Fu, L Pastushok, and W Xiao. DNA damage-induced gene expression in saccharomyces cerevisiae. *FEMS microbiology reviews*, 32(6):908–926, November 2008. PMID: 18616603.

[27] K S Gajiwala and S K Burley. Winged helix proteins. *Curr. Opin. Struct. Biol.*, 10(1):110–116, February 2000. PMID: 10679470.

[28] D J Galgoczy and D P Toczyski. Checkpoint adaptation precedes spontaneous and damage-induced genomic instability in yeast. *Mol Cell Biol*, 21(5):1710–1718, March 2001. PMID: 11238908 PMCID: PMC86717.

[29] H Gao, J M Granka, and M W Feldman. On the classification of epistatic interactions. *Genetics*, 184(3):827–837, March 2010. PMID: 20026678.

[30] C W Garvie, J R Stagno, S Reid, A Singh, E Harrington, and J M Boss. Characterization of the RFX complex and the RFX5(L66A) mutant: implications for the regulation of MHC class II gene expression. *Biochemistry*, 46(6):1597–1611, February 2007. PMID: 17279624.

[31] J Gertz, E D Siggia, and B A Cohen. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457(7226):215–218, January 2009.

[32] R D Gietz and R H Schiestl. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, 2(1):31–34, 2007. PMID: 17401334.

[33] S Hahn and E T Young. Transcriptional regulation in saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, 189(3):705–736, November 2011. PMID: 22084422.

[34] D Hanahan and R A Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, March 2011.

[35] C T Harbison, D B Gordon, T I Lee, N J Rinaldi, K D Macisaac, T W Danford, N M Hannett, J-B Tagne, D B Reynolds, J Yoo, E G Jennings, J Zeitlinger, D K Pokholok, M Kellis, P A Rolfe, K T Takusagawa, E S Lander, D K Gifford, E Fraenkel, and R A Young. Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431(7004):99–104, September 2004. PMID: 15343339.

[36] J L Hartman 4th and N P Tippery. Systematic quantification of gene interactions by phenotypic array analysis. *Genome biology*, 5(7):R49, 2004. PMID: 15239834.

[37] X He, Md. A H Samee, C Blatti, and S Sinha. Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput Biol*, 6(9):e1000935, 2010.

[38] J N Hirschhorn and F Winston. SPT3 is required for normal levels of a-factor and alpha-factor expression in saccharomyces cerevisiae. *Mol. Cell. Biol.*, 8(2):822–827, February 1988.

[39] C S Hoffman. Preparation of yeast DNA. In F M Ausubel, R Brent, R E Kingston, D D Moore, J G Seidman, J A Smith, and K Struhl, editors, *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, May 2001.

[40] M X Huang, Z Zhou, and S J Elledge. The DNA replication and damage checkpoint pathways induce transcription by inhibition of the crt1 repressor. *Cell*, 94(5):595–605, September 1998.

[41] T Ideker, T Galitski, and L Hood. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2:343–372, 2001. PMID: 11701654.

[42] V Iyer and K Struhl. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.*, 14(11):2570–2579, June 1995. PMID: 7781610.

[43] D Jedrysiak. Construction and Characterization of Gene Regulatory Networks in Yeast. Master's thesis, University of Ottawa, Ottawa, Ontario, Canada, 2013.

[44] A Jordan and P Reichard. Ribonucleotide reductases. *Annual review of biochemistry*, 67:71–98, 1998. PMID: 9759483.

[45] G R S Jouzani and I V Goldenkova. A new reporter gene technology: Opportunities and perspectives. *Iranian Journal of Biotechnology*, 3(1):1–15, January 2005.

[46] P Kainth, H E Sassi, L Peña Castillo, G Chua, T R Hughes, and B Andrews. Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. *Methods*, 48(3):258–264, July 2009.

[47] Y Katan, R Agami, and Y Shaul. The transcriptional activation and repression domains of RFX1, a context-dependent regulator, can mutually neutralize their activities. *Nucl. Acids Res.*, 25(18):3621–3628, September 1997.

[48] Y Katan-Khaykovich, I Spiegel, and Y Shaul. The dimerization/repression domain of RFX1 is related to a conserved region of its yeast homologues crt1 and sak1: a new function for an ancient motif. *J. Mol. Biol.*, 294(1):121–137, November 1999. PMID: 10556033.

[49] L G Klinkenberg, T A Mennella, K Luetkenhaus, and R S Zitomer. Combinatorial repression of the hypoxic genes of saccharomyces cerevisiae by DNA binding proteins rox1 and mot3. *Eukaryotic Cell*, 4(4):649–660, April 2005. PMID: 15821125.

[50] L G Klinkenberg, T Webb, and R S Zitomer. Synergy among differentially regulated repressors of the ribonucleotide diphosphate reductase genes of saccharomyces cerevisiae. *Eukaryotic cell*, 5(7):1007–1017, July 2006. PMID: 16835445.

[51] J C Kwasnieski, I Mogno, C A Myers, J C Corbo, and B A Cohen. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47):19498–19503, November 2012. PMID: 23129659.

[52] J-P Lambert, L Mitchell, A Rudner, K Baetz, and D Figeys. A novel proteomics approach for the discovery of chromatin-associated protein networks. *Molecular & cellular proteomics: MCP*, 8(4):870–882, April 2009. PMID: 19106085.

[53] B Li and J C Reese. Ssn6-tup1 regulates RNR3 by positioning nucleosomes and affecting the chromatin structure at the upstream repression sequence. *J. Biol. Chem.*, 276(36):33788–33797, September 2001. PMID: 11448965.

[54] H Lodish, A Berk, S L Zipursky, P Matsudaira, D Baltimore, and J Darnell. Regulation of transcription initiation. In *Molecular Cell Biology, 4th Ed.* W. H. Freeman and Company, 2000.

[55] R Mani, R P St Onge, J L Hartman 4th, G Giaever, and F P Roth. Defining genetic interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3461–3466, March 2008. PMID: 18305163.

[56] M Markus, Z Du, and R Benezra. Enhancer-specific modulation of e protein activity. *Journal of Biological Chemistry*, 277(8):6469–6477, February 2002. PMID: 11724804.

[57] R S McIsaac, S J Silverman, M N McClean, P A Gibney, J Macinskas, M J Hickman, A A Petti, and D Botstein. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in saccharomyces cerevisiae. *Molecular Biology of the Cell*, 22(22):4447–4459, November 2011. PMID: 21965290 PMCID: PMC3216669.

[58] S L McKnight. *Transcriptional regulation*. Cold Spring Harbor Laboratory Press, 1992.

[59] L A Mirny. Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22534–22539, December 2010. PMID: 21149679.

[60] P T Monteiro, N D Mendes, M C Teixeira, S d′Orey, S Tenreiro, N P Mira, H Pais, A P Francisco, A M Carvalho, A B Lourenço, I Sá-Correia, A L Oliveira, and A T Freitas. YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in saccharomyces cerevisiae. *Nucleic Acids Research*, 36(suppl 1):D132–D136, January 2008. PMID: 18032429.

[61] K W Mulder, G S Winkler, and H T Timmers. DNA damage and replication stress induced transcription of RNR genes is dependent on the ccr4-not complex. *Nucleic Acids Res.*, 33(19):6384–6392, 2005. PMID: 16275785.

[62] S Oehler, M Amouyal, P Kolkhof, B von Wilcken-Bergmann, and B Müller-Hill. Quality and position of the three lac operators of e. coli define efficiency

of repression. *EMBO J*, 13(14):3348–3355, July 1994. PMID: 8045263 PMCID: 395232.

[63] S Oehler, E R Eismann, H Krämer, and B Müller-Hill. The three operators of the lac operon cooperate in repression. *EMBO J.*, 9(4):973–979, April 1990. PMID: 2182324.

[64] J W Osbourne and E Waters. Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2), January 2002.

[65] P J Park. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10(10):669–680, October 2009.

[66] M H Parker, R L S Perry, M C Fauteux, C A Berkes, and M A Rudnicki. MyoD synergizes with the e-protein HEB beta to induce myogenic differentiation. *Molecular and cellular biology*, 26(15):5771–5783, August 2006. PMID: 16847330.

[67] P Perez-Romero and M J Imperiale. Assaying protein-DNA interactions in vivo and in vitro using chromatin immunoprecipitation and electrophoretic mobility shift assays. *Methods in molecular medicine*, 131:123–139, 2007. PMID: 17656780.

[68] H Phenix, K Morin, C Batenchuk, J Parker, V Abedi, L Yang, L Tepliakova, T J Perkins, and Mads Kaern. Quantitative epistasis analysis and pathway inference from genetic interaction data. *PLoS Comput Biol*, 7(5):e1002048, May 2011.

[69] P C Phillips. Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. *Nature reviews. Genetics*, 9(11):855–867, November 2008. PMID: 18852697 PMCID: PMC2689140.

[70] K J Polach and J Widom. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *Journal of molecular biology*, 258(5):800–812, May 1996. PMID: 8637011.

[71] L Prakash and S Prakash. Isolation and characterization of MMS-sensitive mutants of saccharomyces cerevisiae. *Genetics*, 86(1):33–55, May 1977. PMID: 195865.

[72] G Prelich. Gene overexpression: Uses, mechanisms, and interpretation. *Genetics*, 190(3):841–854, March 2012. PMID: 22419077.

[73] M Ptashne. Regulated recruitment and cooperativity in the design of biological regulatory systems. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1807):1223–1234, June 2003. PMID: 12816608.

[74] T Raveh-Sadka, M Levo, U Shabi, B Shany, L Keren, M Lotan-Pompan, D Zeevi, E Sharon, A Weinberger, and E Segal. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet.*, 44(7):743–750, July 2012. PMID: 22634752.

[75] R J D Reid, S González-Barrera, I Sunjevaric, D Alvaro, S Ciccone, M Wagner, and R Rothstein. Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase i-induced DNA damage. *Genome research*, 21(3):477–486, March 2011. PMID: 21173034.

[76] W Reith, C Herrero-Sanchez, M Kobr, P Silacci, C Berte, E Barras, S Fey, and B Mach. MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. *Genes & development*, 4(9):1528–1540, September 1990. PMID: 2253877.

[77] S Renggli, W Keck, U Jenal, and D Ritz. The role of auto-fluorescence in flow-cytometric analysis of escherichia coli treated with bactericidal antibiotics. *Journal of bacteriology*, July 2013. PMID: 23836867.

[78] E Segal, T Raveh-Sadka, M Schroeder, U Unnerstall, and U Gaul. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature*, 451(7178):535–540, January 2008. PMID: 18172436.

[79] R Segal and A J Berk. Promoter activity and distance constraints of one versus two sp1 binding sites. *The Journal of biological chemistry*, 266(30):20406–20411, October 1991. PMID: 1939095.

[80] V M Sharma, R S Tomar, A E Dempsey, and J C Reese. Histone deacetylases RPD3 and HOS2 regulate the transcriptional activation of DNA damage-inducible genes. *Mol. Cell. Biol.*, 27(8):3199–3210, April 2007.

[81] E Sharon, Y Kalma, A Sharp, T Raveh-Sadka, M Levo, D Zeevi, L Keren, Z Yakhini, A Weinberger, and E Segal. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature biotechnology*, 30(6):521–530, June 2012. PMID: 22609971.

[82] I Singh, R Pass, S O Togay, J W Rodgers, and J L Hartman. Stringent mating-type-regulated auxotrophy increases the accuracy of systematic genetic interaction screens with saccharomyces cerevisiae mutant arrays. *Genetics*, 181(1):289–300, January 2009. PMID: 18957706 PMCID: PMC2621176.

[83] D E Sterner, P A Grant, S M Roberts, L J Duggan, R Belotserkovskaya, L A Pacella, F Winston, J L Workman, and S L Berger. Functional organization of the yeast SAGA complex: distinct components involved in structural integrity,

nucleosome acetylation, and TATA-binding protein interaction. *Molecular and cellular biology*, 19(1):86–98, January 1999. PMID: 9858534.

[84] K Struhl. Molecular mechanisms of transcriptional regulation in yeast. *Annual Review of Biochemistry*, 58(1):1051–1077, 1989. PMID: 2673007.

[85] K. Struhl. Yeast transcriptional regulatory mechanisms. *Annual Review of Genetics*, 29(1):651–674, December 1995.

[86] K B S Swamy, C-Y Cho, S Chiang, Z T-Y Tsai, and H-K Tsai. Impact of DNA-binding position variants on yeast gene expression. *Nucleic acids research*, 37(21):6991–7001, November 2009. PMID: 19767613.

[87] M C Teixeira, P Monteiro, P Jain, S Tenreiro, A R Fernandes, N P Mira, M Alenquer, A T Freitas, A L Oliveira, and I Sá-Correia. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. *Nucleic Acids Research*, 34(suppl 1):D446–D451, January 2006. PMID: 16381908.

[88] R S Tomar, S Zheng, D Brunke-Reese, H N Wolcott, and J C Reese. Yeast rap1 contributes to genomic integrity by activating DNA damage repair genes. *EMBO J*, 27(11):1575–1584, June 2008. PMID: 18480842 PMCID: PMC2426730.

[89] A H Tong, M Evangelista, A B Parsons, H Xu, G D Bader, N Pagé, M Robinson, S Raghibizadeh, C W Hogue, H Bussey, B Andrews, M Tyers, and C Boone. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)*, 294(5550):2364–2368, December 2001. PMID: 11743205.

[90] A H Y Tong and C Boone. Synthetic genetic array analysis in saccharomyces

cerevisiae. *Methods in molecular biology (Clifton, N.J.)*, 313:171–192, 2006. PMID: 16118434.

[91] K Tounekti, M Aouida, A Leduc, J Poschmann, X Yang, O Belhadj, and D Ramotar. Deletion of the chromatin remodeling gene SPT10 sensitizes yeast cells to a subclass of DNA-damaging agents. *Environmental and molecular mutagenesis*, 47(9):707–717, December 2006. PMID: 17078097.

[92] D A Treco and F Winston. Growth and manipulation of yeast. In Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl, editors, *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, April 2008.

[93] M A Treitel and M Carlson. Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proceedings of the National Academy of Sciences of the United States of America*, 92(8):3132–3136, April 1995. PMID: 7724528.

[94] O Tsaponina, E Barsoum, S U Aström, and A Chabes. Ixr1 is required for the expression of the ribonucleotide reductase rnr1 and maintenance of dNTP pools. *PLoS genetics*, 7(5):e1002061, May 2011. PMID: 21573136.

[95] N Van Driessche, J Demsar, E O Booth, P Hill, P Juvan, B Zupan, A Kuspa, and G Shaulsky. Epistasis analysis with global transcriptional phenotypes. *Nature genetics*, 37(5):471–477, May 2005. PMID: 15821735.

[96] B J Venters and B F Pugh. A canonical promoter organization of the transcription machinery and its regulators in the saccharomyces genome. *Genome research*, 19(3):360–371, March 2009. PMID: 19124666.

[97] E A Winzeler, D D Shoemaker, A Astromoff, H Liang, K Anderson, B Andre, R Bangham, R Benito, J D Boeke, H Bussey, A M Chu, C Connelly, K Davis,

F Dietrich, S W Dow, M El Bakkoury, F Foury, S H Friend, E Gentalen, G Giaever, J H Hegemann, T Jones, M Laub, H Liao, N Liebundguth, D J Lockhart, A Lucau-Danila, M Lussier, N M'Rabet, P Menard, M Mittmann, C Pai, C Rebischung, J L Revuelta, L Riles, C J Roberts, P Ross-MacDonald, B Scherens, M Snyder, S Sookhai-Mahadeo, R K Storms, S Véronneau, M Voet, G Volckaert, T R Ward, R Wysocki, G S Yen, K Yu, K Zimmermann, P Philippsen, M Johnston, and R W Davis. Functional characterization of the s. cerevisiae genome by gene deletion and parallel analysis. *Science (New York, N.Y.)*, 285(5429):901–906, August 1999. PMID: 10436161.

[98] K H Wong and K Struhl. The cyc8-tup1 complex inhibits transcription primarily by masking the activation domain of the recruiting protein. *Genes Dev.*, 25(23):2525–2539, December 2011. PMID: 22156212.

[99] Z Xu and D Norris. The SFP1 gene product of saccharomyces cerevisiae regulates G2/M transitions during the mitotic cell cycle and DNA-damage response. *Genetics*, 150(4):1419–1428, December 1998. PMID: 9832520.

[100] H Zhang, J A Kruk, and J C Reese. Dissection of coactivator requirement at RNR3 reveals unexpected contributions from TFIID and SAGA. *J Biol Chem*, 283(41):27360–27368, October 2008. PMID: 18682387 PMCID: PMC2562059.

[101] H Zhang and J C Reese. Exposing the core promoter is sufficient to activate transcription and alter coactivator requirement at RNR3. *Proc Natl Acad Sci U S A*, 104(21):8833–8838, May 2007. PMID: 17502614 PMCID: PMC1885588.

[102] Z Zhang and J C Reese. Molecular genetic analysis of the yeast repressor Rfx1/Crt1 reveals a novel two-step regulatory mechanism. *Molecular and cellular biology*, 25(17):7399–7411, September 2005. PMID: 16107689.

[103] Z Zhou and S J Elledge. Isolation of crt mutants constitutive for transcription of the DNA damage inducible gene RNR3 in saccharomyces cerevisiae. *Genetics*, 131(4):851–866, August 1992.

Appendix A

Customized MATLAB scripts

Converting fluorescence data to mean expression values

Before any analysis was done, an information file was created for each experiment which included the file name and the strain information for that particular file. This information file was saved as a file recognized by MATLAB (file name extension .txt or .mat) and loaded during the analysis so that the data would be stored along with information that could later be accessed and plotted. Files obtained from the flow cytometer were analyzed by three scripts (file name extension .m) written by Mads Kaern to: 1) join the files together into one amalgamated file (fcs_join.m), 2) gate the data to eliminate the top 10% of outliers in the population (representing dead, compromised cells, etc.) (fcs_gating.m) and 3) convert the probe data into linear mean fluorescence values for plotting and statistical analysis (singlecolour_analysis.m). Scripts specific for each data set were then written to plot the data in a meaningful way.

An example of the output for fcs_join.m is shown in Figure A.1. The left plot shows the forward scatter (FS) by side scatter (SS) log plot of the cell population representing all files joined together. FS and SS show the size and complexity of the cells in the population respectively. It was important that there was minimal variability in this plot. The second plot shows an amalgamation of all .fcs files and

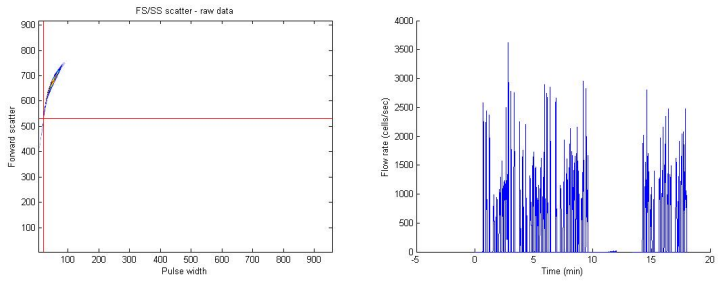
their flow rate over time for one experiment. In this example, there was a space between some of the files because I ran two plates several minutes apart. During this step it was important to look for huge discrepancies in flow rate and to make sure the number of peaks in the plot represented the number of files one wished to join together.

An example of the output for the fcs_gating.m script is shown in Figure A.2. FL1 is flow probe 1, which is the probe detecting GFP fluorescence expression.

After the gating script, the singlecolour_analysis.m was run. An example of the output for one file is given in Figure A.3. This script was written to first normalize each cell population to constitutive BFP expression values. This was only relevant if a population of cells contained a reporter where a constitutive promoter was used to drive BFP expression. This was used as another measure to eliminate sick or abnormal cells. Next, the script was written to fit population distributions for each file to single and double Gaussian distributions, and plot the log GFP values for the corresponding sub-populations generated for each file according to these models. Then, a linear transformation on the gated probe data was performed. The flow cytometer probes collect log expression values. These values were linearly transformed by computing 10 to the power of the log expression value divided by the FL1 probe range defined earlier in the script, which was 10 to the power of 1:1024 divided by 256. The resulting output of this script was a data file that included mean linearized GFP expression values that could be plotted, normalized for autofluorescence, tranformed or analyzed further with statistics.

Generating Plots Directly from Raw Data

Alternatively, the files obtained from the flow cytometer were analyzed with a series of scripts written by Hilary Phenix. The first script allowed the user to visualize the raw data from each .fcs file by plotting raw FS/SS distributions as well as raw and



(a) Variability in samples (b) Samples over time

Figure A.1: Examples of output for running the joining script in MATLAB on a series of fcs files obtained after flow cytometry. The first figure shows an amalgamated FS/SS log plot for all files. The second figure shows a plot of all files and their flow rate over time.

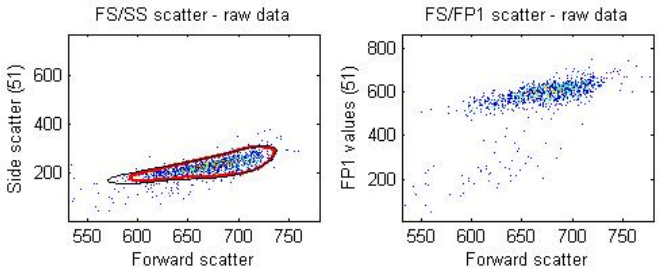


Figure A.2: Example of output while running the gating script in MATLAB on a series of fcs files obtained after flow cytometry. The left figure shows a FS/SS log plot of an individual file (file 51 in this example) and the right plot shows the FS by FL1 (GFP fluorescence) log plot. The red and black lines represent the 90% gating for this file, eliminating some of the outliers.

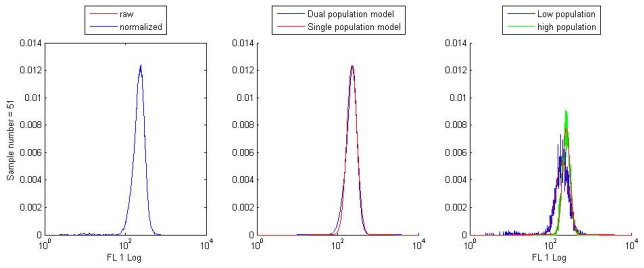


Figure A.3: Example of the output for one file while running the single colour analysis script in MATLAB. The first plot is not relevant since this particular cell population did not have a constitutive promoter-reporter construct. The middle plot shows fits to a single (single population model) and double (dual population model) Gaussian distribution. The rightmost plot is a histogram of the log GFP distribution for one sample population and its corresponding sub-populations according to the dual population model (sample 51).

relative probe distributions (check_files.m). An example of the output from this script is shown in Figure A.4. FL6 is the violet probe detecting BFP fluorescence, which is only meaningful if the strain being analyzed has a reporter driving BFP. Even without the reporter, abnormal fluctuations in this distribution represent autofluorescence fluctuations, which can be typical of sick or dying cells. This script also allowed the user to define a minimum flow rate and bimodal estimation threshold, where files below these thresholds were discarded or estimated as unimodal, respectively. Based on the plot visualization, files indicative of experimental error were also discarded at this stage (i.e. with abnormal GFP or FS/SS distributions). Next, a script could be run to load an amalgamated data file composed of individual checked .fcs files from a series of experiments. Then, a script was run that allowed the user to define the wild-type file and plotted each file's FL1 log population distribution on top of the wildtype distribution on the same graph (Cmp_MutWT_FLDist_sameplot.m). Plots of interest could be exported directly from the script in jpeg format.

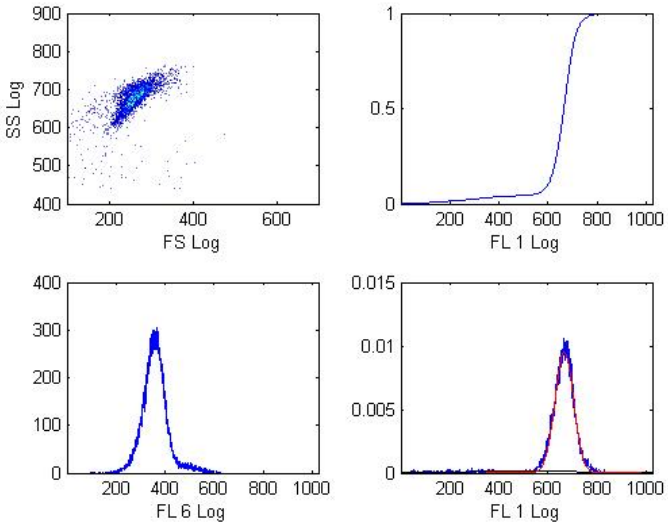


Figure A.4: Example of the output for one file while running the check files script in MATLAB. From top left, the plots are the FS/SS log distribution, the FL1 values relative to maximal FL1 expression, the FL6 log distribution and the FL1 log distribution.

Appendix B

Sequencing

Table B.1: Tabulation of the number of Ts in the poly(dA:dT) region for each *RNR3* promoter variant colony sequenced. TetO refers to strains from the Tet operator replacement approach, while scramb refers to strains from the scrambling X-box replacement approach. Biological replicate counts for the same strain are separated by commas, and technical replicate counts of one biological replicate are given in brackets.

Strain	number of Ts
wt	17
x	17
TetO	
u	16,18
v	20,21
w	21
uv	22,17
uw	20
vw	17,18
uvw	20
xu	27
xv	22
xw	14
xuv	20
xuw	15
xvw	21
xuvw	16
Scramb	
u	21
v	(19,19)
w	19
uv	13
uw	26,(22,23),(17,20)
vw	(14,14)
uvw	15
xu	24
xv	16
xw	20
xuv	(16,16)
xuw	22
xvw	15
xuvw	(14,16)

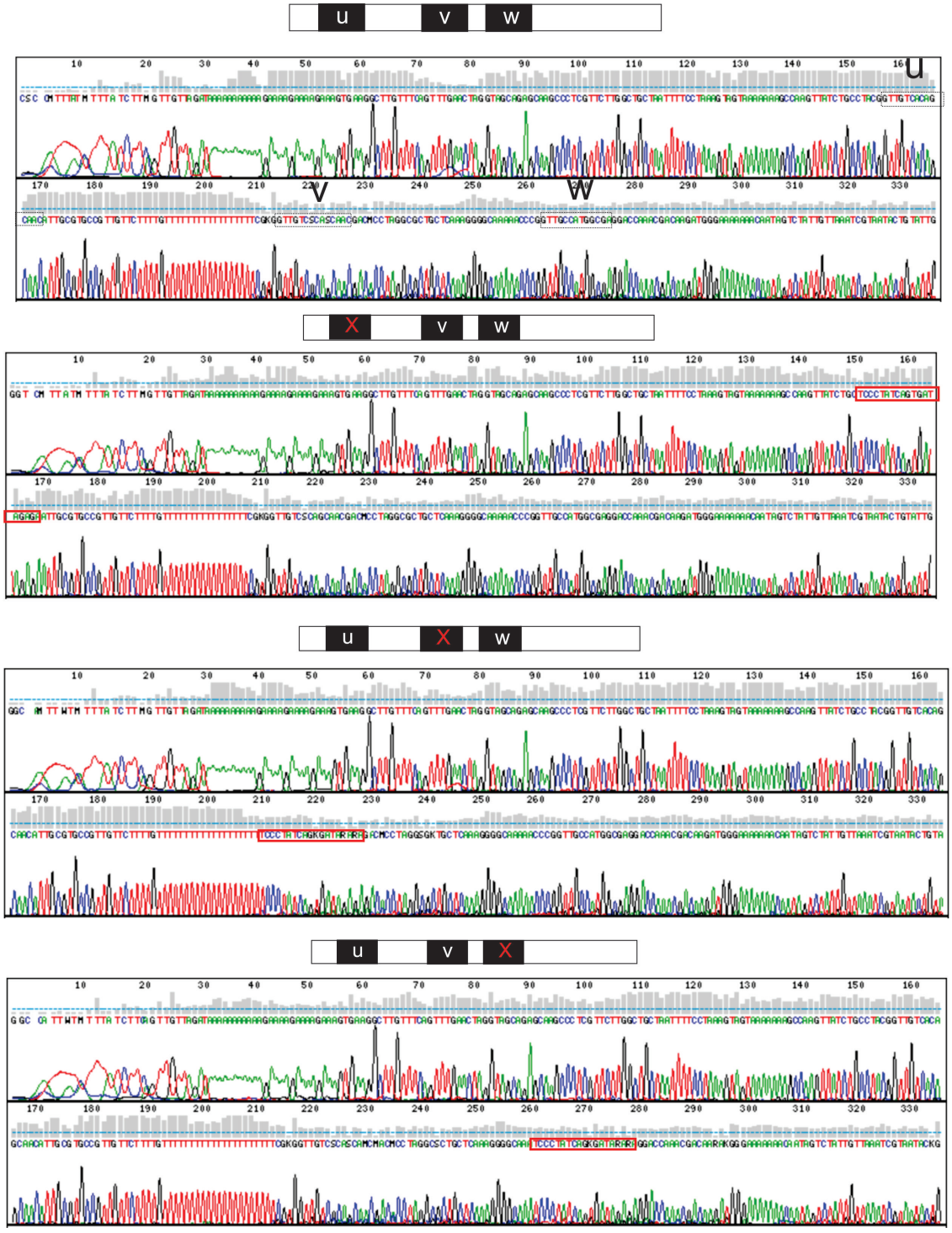


Figure B.1: Sequencing results: X-boxes replaced with Tet operators. Chromat traces are shown for sequencing of the wild-type *RNR3* promoter (X-boxes boxed in black and labeled), and three strains where single X-boxes were replaced (boxed in red). Chromat traces are only shown for one strain background (Crt1+), but X-box replacement in Crt1- strains was also confirmed.

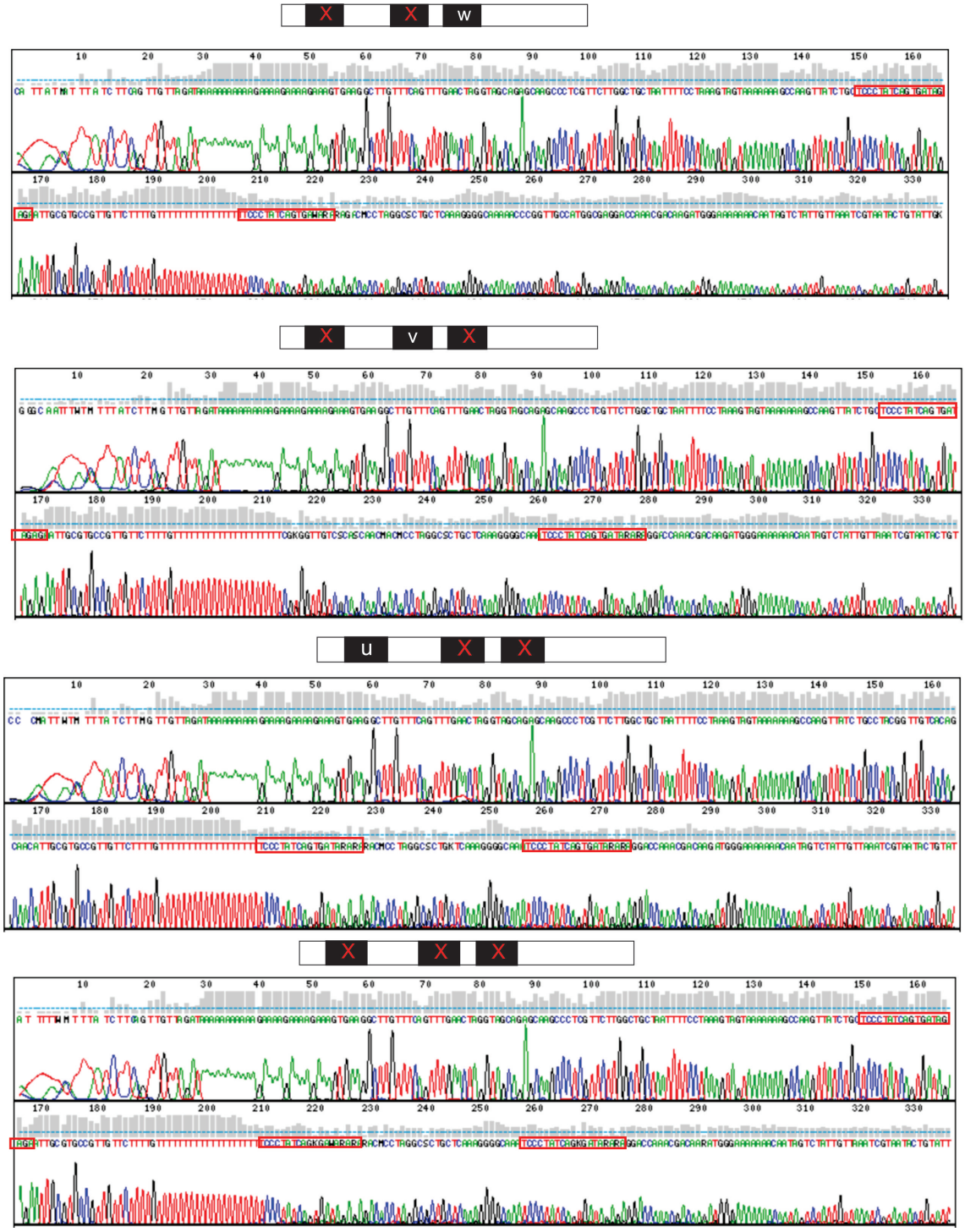


Figure B.2: Sequencing results continued: X-boxes replaced with Tet operators. Chromat traces are shown for sequencing of strains where two or three X-boxes were replaced (shown in red). Chromat traces are only shown for one strain background (Crt1+), but X-box replacement in Crt1- strains was also confirmed.

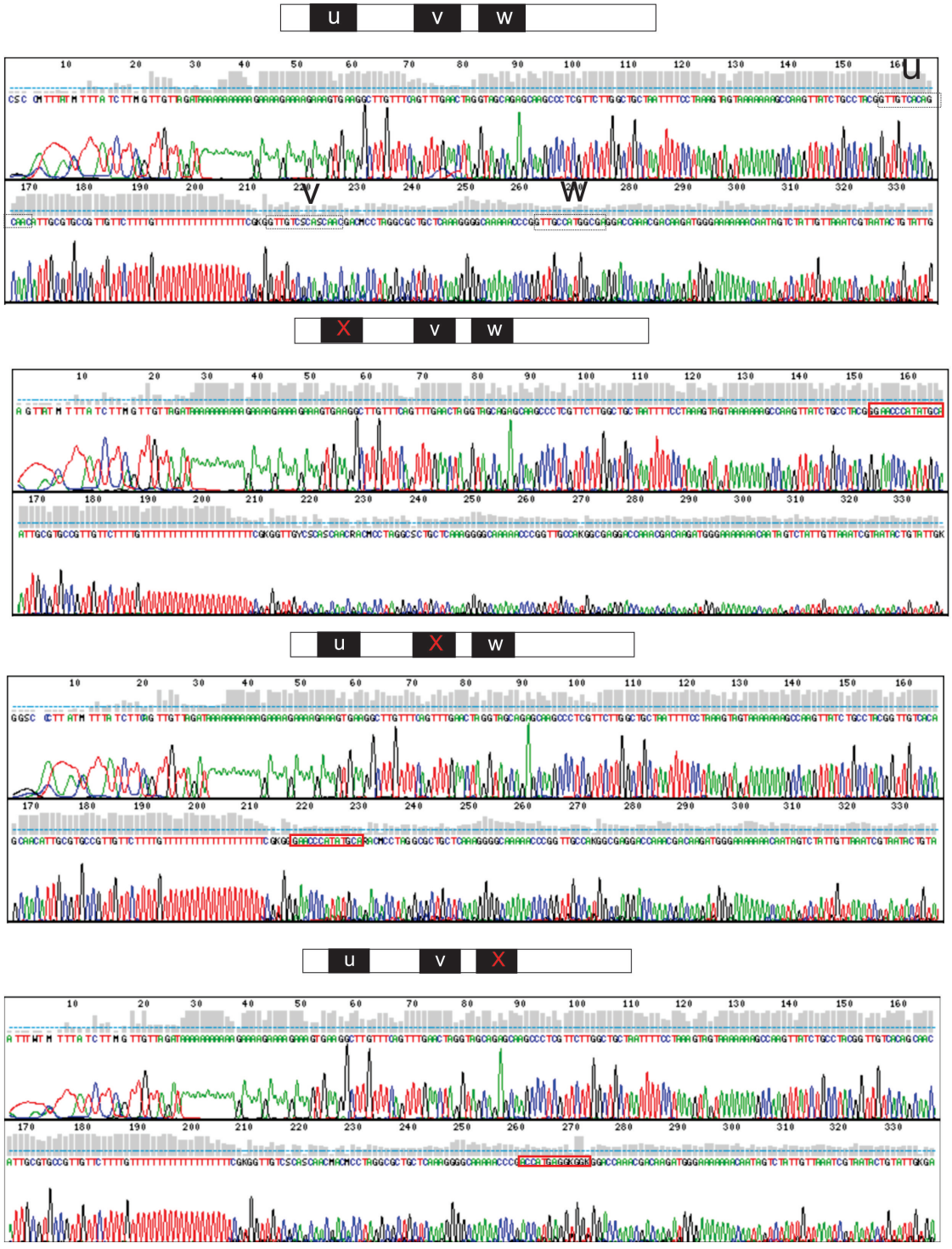


Figure B.3: Sequencing results: X-boxes replaced with scrambled sequences. Chromat traces are shown of the wild-type *RNR3* promoter, and three strains where single X-boxes were replaced. Chromat traces are only shown for one strain background (Crt1+), but X-box replacement in Crt1- strains was also confirmed.

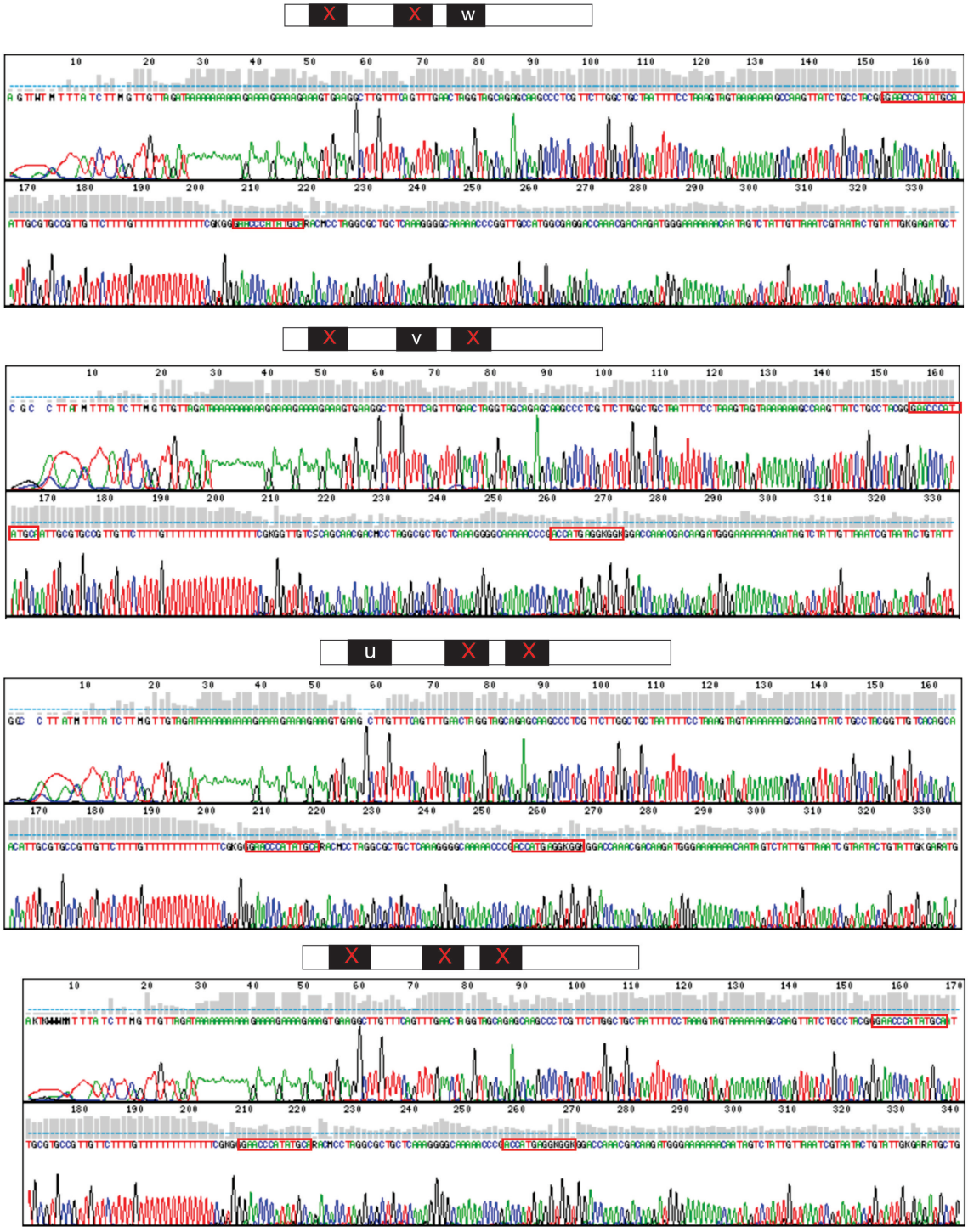


Figure B.4: Sequencing results continued: X-boxes replaced with scrambled sequences. Chromat traces are shown for sequencing of strains where two or three X-boxes were replaced. Chromat traces are only shown for one strain background (Crt1+), but X-box replacement in Crt1- strains was also confirmed.

Appendix C

Synthetic Genetic Array analysis

Table C.1: SGA strains and Z-test for outliers. Strains labeled by the gene name encoding the factor that was deleted aside from Crt1.

Strain	z-score	p-val
ACE2	-9.80248	1.09858×10^{-22}
AFT1	-9.2911	1.52699×10^{-20}
BDF1	-9.16922	4.76474×10^{-20}
CKB1	-9.10142	8.91586×10^{-20}
SPT10	-9.00145	2.22753×10^{-19}
MET18	-8.95429	3.41916×10^{-19}
GAL80	-8.85207	8.5908×10^{-19}
SPT3	-7.76582	8.11179×10^{-15}
INO2	-7.43636	1.03499×10^{-13}
SFP1	-7.40716	1.29033×10^{-13}
SPT8	-7.3372	2.181×10^{-13}
TDA9	-7.2202	5.19126×10^{-13}
IXR1	-6.89094	5.54248×10^{-12}
EXO5	-5.89375	3.77533×10^{-09}
MOT2	-5.42984	5.64045×10^{-08}
GCN4	-4.93496	8.01676×10^{-07}
DEP1	-4.50679	6.58167×10^{-06}
CTI6	-4.44231	8.89974×10^{-06}
PHO23	-4.39	1.13352×10^{-05}
MAC1	-3.67901	0.000234139
YPR013C	-3.40022	0.000673306
BUR6	-3.22815	0.00124593
RIM101	-2.8898	0.00385489

Table C.1 – continued from previous page

Strain	z-score	p-val
HAL9	-2.82275	0.00476132
EMI1	-2.56533	0.0103079
MTC7	-2.54229	0.011013
SAS4	-2.34612	0.01897
FLO8	-2.30484	0.0211757
RML2	-2.28185	0.0224984
HMO1	-2.26686	0.0233987
PGD1	-2.16888	0.0300921
STB4	-2.14265	0.0321415
CRZ1	-1.93304	0.0532309
URE2	-1.92959	0.0536572
SKO1	-1.86848	0.061695
EDS1	-1.84689	0.0647631
UME1	-1.779	0.0752392
MUK1	-1.73997	0.0818639
GAT2	-1.72912	0.0837882
CRF1	-1.68678	0.0916454
DOT6	-1.64744	0.0994672
HAP2	-1.6117	0.107027
CDC36	-1.60879	0.107661
RDS2	-1.60762	0.107919
PHD1	-1.6045	0.108604
MSN4	-1.60394	0.108728
HAP4	-1.56033	0.118682
SFL1	-1.54262	0.122924
DAL82	-1.53817	0.124007
SYN8	-1.51313	0.130246
AHC2	-1.51208	0.130514
SAS5	-1.50741	0.131706
YNL155W	-1.49974	0.133682
ITT1	-1.45383	0.145994
CKA2	-1.404	0.16032
SAS2	-1.36497	0.172263
YLR278C	-1.35799	0.174468
RLM1	-1.33865	0.180685
UPC2	-1.33419	0.182141
BAS1	-1.3253	0.185072
SNF6	-1.32261	0.185965
SUB1	-1.29756	0.194439
ARR1	-1.29078	0.196781
YPR015C	-1.26223	0.206867
DIG1	-1.24519	0.213063

Table C.1 – continued from previous page

Strain	z-score	p-val
TOS8	-1.19662	0.231453
YPR022C	-1.19327	0.232764
SEF1	-1.16961	0.24216
CAF130	-1.16928	0.24229
ASH1	-1.14653	0.251575
YPR127W	-1.12433	0.260875
FSH3	-1.11057	0.266752
RPI1	-1.10449	0.269383
PFK27	-1.08976	0.27582
VHR1	-1.06903	0.285058
EMI2	-1.06353	0.287542
ECM22	-1.05519	0.291341
MED6	-1.04972	0.293847
GAC1	-1.03245	0.30186
YER184C	-1.02118	0.307167
SGV1	-1.0187	0.308344
HSF1	-0.99032	0.322018
HAP3	-0.950965	0.341622
YER130C	-0.946716	0.343784
MSN1	-0.944095	0.345121
HTL1	-0.916389	0.359463
ACA1	-0.914168	0.360628
USV1	-0.90536	0.365275
YFL052W	-0.897831	0.369276
GAT3	-0.892258	0.372255
YIL024C	-0.883094	0.377185
FHL1	-0.881932	0.377814
MED8	-0.85079	0.394886
YTH1	-0.848001	0.396438
ZIM17	-0.845546	0.397806
RRN7	-0.837645	0.40223
BDP1	-0.805431	0.420571
PRR1	-0.802931	0.422015
SNG1	-0.781603	0.434448
GAL3	-0.779558	0.435651
RPD3	-0.769503	0.441595
SUT2	-0.764052	0.444836
GLN3	-0.760459	0.44698
CKA1	-0.758489	0.448159
ARP7	-0.75145	0.452382
MBF1	-0.739535	0.459582
MKS1	-0.738199	0.460394

Table C.1 – continued from previous page

Strain	z-score	p-val
LOT6	-0.73318	0.463449
HMS2	-0.727579	0.466872
MED7	-0.722197	0.470173
SUA7	-0.711969	0.476484
SNF11	-0.710093	0.477646
MGA2	-0.695159	0.486955
TFG1	-0.679128	0.497057
NTE1	-0.675087	0.49962
HAA1	-0.674503	0.499991
NCB2	-0.673057	0.500911
CIT2	-0.665738	0.505579
SRB4	-0.665651	0.505634
RSF2	-0.657408	0.510919
YNR063W	-0.641854	0.520968
RVB1	-0.638872	0.522906
NRG1	-0.623228	0.533135
CNA1	-0.616457	0.537593
FZF1	-0.615107	0.538484
CAT8	-0.569467	0.569039
MAL13	-0.544073	0.586391
TEC1	-0.532932	0.59408
HST4	-0.529256	0.596628
CTS1	-0.525816	0.599016
ADR1	-0.514824	0.606676
IME1	-0.508803	0.61089
RGM1	-0.50538	0.613292
NOT5	-0.50123	0.616209
SUT1	-0.475023	0.634771
RGT1	-0.472906	0.636281
MIG3	-0.469729	0.638549
GTS1	-0.465456	0.641605
SRB7	-0.458949	0.646271
RDS1	-0.452333	0.651029
YLL054C	-0.443267	0.657573
CLN1	-0.435956	0.662869
HCM1	-0.428754	0.668103
CBF1	-0.41418	0.678743
MAL33	-0.401181	0.688287
SPS18	-0.40056	0.688744
STP4	-0.378743	0.704879
URC2	-0.377398	0.705878
RTG1	-0.371227	0.710469

Table C.1 – continued from previous page

Strain	z-score	p-val
BDF2	-0.366664	0.713869
MSN2	-0.364386	0.715569
PZF1	-0.363341	0.71635
RSC58	-0.361692	0.717583
TFB1	-0.358295	0.720122
SRB2	-0.357976	0.720361
IFH1	-0.356214	0.72168
LUC7	-0.353782	0.723502
YRR1	-0.352712	0.724305
MCM1	-0.344304	0.730618
SIR1	-0.328416	0.742597
YAK1	-0.315033	0.752737
HUG1	-0.29685	0.766581
YMR181C	-0.295741	0.767428
SFG1	-0.289008	0.772575
YIR014W	-0.285766	0.775057
DAT1	-0.277836	0.781138
ASG1	-0.254505	0.799105
HIR2	-0.254129	0.799396
PDR1	-0.252256	0.800843
SWI5	-0.245551	0.80603
OAF1	-0.222865	0.823641
PDE1	-0.21626	0.828785
UGA3	-0.213652	0.830819
FKH2	-0.212678	0.831578
RSF1	-0.199176	0.842125
FRA2	-0.194627	0.845685
BRE4	-0.182531	0.855166
YAP6	-0.172247	0.863243
MED4	-0.146711	0.88336
DIG2	-0.144517	0.885092
AFT2	-0.111712	0.911052
KIN28	-0.109241	0.913011
RIF1	-0.0966581	0.922998
TOD6	-0.0836947	0.933299
GIS1	-0.0819201	0.93471
CUP9	-0.0639962	0.948973
WTM2	-0.0390315	0.968865
MET28	-0.0308753	0.975369
RVB2	-0.0167791	0.986613
BUR2	-0.00401479	0.996797
RDR1	0.00344446	0.997252

Table C.1 – continued from previous page

Strain	z-score	p-val
FKH1	0.0101568	0.991896
STB3	0.0424255	0.96616
RTG3	0.0523041	0.958286
FRA1	0.0620393	0.950532
YLR346C	0.0666572	0.946855
SNT2	0.0697049	0.944429
NRM1	0.0742733	0.940793
CMP2	0.0791357	0.936925
YGR035C	0.0906909	0.927738
ERT1	0.106195	0.915427
FRE8	0.112685	0.91028
LEU3	0.147731	0.882555
NRG2	0.153409	0.878075
MIG1	0.155611	0.87634
MKT1	0.163598	0.870048
CAD1	0.167622	0.86688
YMR102C	0.171881	0.863531
TYE7	0.173596	0.862183
STP3	0.174537	0.861444
YOR059C	0.176419	0.859965
YAP5	0.185991	0.852452
HAP5	0.196399	0.844298
CRT10	0.199184	0.842119
SIP4	0.222274	0.824101
RFX1	0.240571	0.809887
REH1	0.266054	0.790198
KAR4	0.268798	0.788085
PRR2	0.288154	0.773229
YDR026C	0.289337	0.772323
RIF2	0.290028	0.771795
YLR040C	0.295594	0.76754
CNB1	0.298372	0.765419
STB1	0.298372	0.765419
SWI1	0.308729	0.757528
CTH1	0.309807	0.756708
RAD14	0.314346	0.753259
ZPS1	0.316979	0.75126
CEP3	0.331469	0.740291
ARG81	0.334813	0.737766
HAC1	0.341947	0.732391
BYE1	0.350557	0.725921
SIN4	0.35911	0.719512

Table C.1 – continued from previous page

Strain	z-score	p-val
AZF1	0.362319	0.717114
ZRT1	0.37635	0.706657
HIR3	0.405017	0.685465
YLR046C	0.41601	0.677403
ZRT2	0.419846	0.674598
TAF11	0.433781	0.664447
YGR067C	0.458892	0.646312
BUD27	0.465714	0.64142
CIN5	0.46703	0.640478
PDR8	0.472907	0.63628
YAP3	0.47535	0.634538
GAT4	0.482396	0.629524
YPR196W	0.491253	0.623248
HUA1	0.493606	0.621584
MET31	0.50595	0.612891
CHA4	0.513112	0.607873
ECO1	0.533959	0.59337
YRM1	0.556897	0.577598
RPH1	0.560303	0.575273
PHO2	0.569493	0.569021
RRN11	0.572862	0.566738
YJL206C	0.575245	0.565126
OPI1	0.57964	0.562157
YER156C	0.581001	0.56124
THO2	0.595417	0.551565
MIG2	0.59671	0.550701
HTB1	0.634465	0.525777
PIP2	0.647519	0.517296
GCR1	0.652679	0.513963
HOT1	0.685181	0.49323
RRN6	0.685357	0.493119
SWI3	0.693745	0.487842
SIP3	0.705005	0.480807
RRN5	0.707988	0.478953
HMS1	0.727991	0.466619
ABF1	0.745676	0.455863
RME1	0.745897	0.45573
SOK2	0.746576	0.45532
JEN1	0.746706	0.455241
SMP1	0.75761	0.448685
YKL222C	0.758096	0.448394
GZF3	0.762047	0.446032

Table C.1 – continued from previous page

Strain	z-score	p-val
TBS1	0.763213	0.445336
YBP1	0.777475	0.436878
SAP1	0.791893	0.428423
GAL4	0.795295	0.426442
WAR1	0.796387	0.425807
ECL1	0.802974	0.42199
MSS11	0.807324	0.41948
OAF3	0.808309	0.418913
IWR1	0.81195	0.41682
PDR3	0.824816	0.409476
MSA2	0.834702	0.403885
STB5	0.837643	0.402231
YPL088W	0.84668	0.397173
ARO80	0.85608	0.391953
YOX1	0.887883	0.374604
THI2	0.907721	0.364025
YAP7	0.908805	0.363453
PLM2	0.916426	0.359444
ZAP1	0.924178	0.355394
YLL056C	0.93556	0.3495
CAF120	0.94492	0.3447
ETP1	0.949545	0.342344
CST6	0.959361	0.337377
ARG80	0.963914	0.335089
STP2	0.97097	0.331563
CUP2	0.979923	0.327124
RAP1	1.02174	0.306905
GIS2	1.02625	0.304774
XBP1	1.02737	0.304247
RTG2	1.03255	0.301812
PHR1	1.03433	0.300984
SRB5	1.04662	0.295276
YML037C	1.0533	0.292201
CDC39	1.07059	0.284352
SSN2	1.10767	0.268006
MET32	1.12225	0.261755
MGA1	1.13635	0.255811
NDT80	1.14977	0.250238
ISW1	1.15908	0.246422
CCR4	1.17552	0.239788
MSA1	1.19835	0.23078
YIR016W	1.19882	0.230599

Table C.1 – continued from previous page

Strain	z-score	p-val
SPT2	1.20634	0.227688
GSM1	1.21377	0.224837
TOM1	1.2149	0.224404
WTM1	1.21959	0.222621
PFK26	1.22094	0.222109
RRN3	1.23612	0.216413
ELF1	1.23762	0.215856
MBP1	1.25868	0.208145
GAT1	1.26458	0.206023
NOT3	1.30153	0.193077
LYS14	1.30467	0.192006
TOS4	1.34074	0.180006
PUT3	1.36299	0.172886
HTB2	1.3972	0.162353
SPT23	1.42494	0.154175
HST3	1.43164	0.152248
DAL80	1.44977	0.147123
STD1	1.48897	0.136495
MET4	1.51626	0.129453
SSL1	1.51628	0.129448
CKB2	1.56057	0.118625
OTU1	1.67481	0.0939719
WHI5	1.70353	0.0884692
YHP1	1.77518	0.0758681
HYR1	1.79288	0.0729922
CKS1	1.875	0.0607932
CLN2	1.87675	0.0605528
IMP2'	1.88404	0.0595599
GCR2	2.20887	0.0271835
PHO4	2.25676	0.024023
HIR1	2.31118	0.0208229
PPR1	2.38103	0.0172641
STP1	2.42965	0.0151135
YAP1	2.78286	0.00538825
SUM1	2.90307	0.00369521
HTZ1	3.19839	0.00138196
SWI4	3.60757	0.000309083
RPN4	4.04999	5.12198×10^{-05}
UME6	4.12253	3.74741×10^{-05}
DAL81	4.13395	3.56583×10^{-05}
SGF11	4.26305	2.01656×10^{-05}
SKN7	4.51816	6.23791×10^{-06}

Table C.1 – continued from previous page

Strain	z-score	p-val
CHD1	4.52134	6.145×10^{-06}
ISW2	4.98557	6.1778×10^{-07}
TAF13	6.32567	2.52128×10^{-10}
SUS1	7.30307	2.81273×10^{-13}
NGG1	7.7745	7.57435×10^{-15}
ROX1	7.81448	5.51894×10^{-15}
MOT3	10.2143	1.71175×10^{-24}
ADA2	27.8812	0

Table C.2: Functional analysis of SGA library hits. SGA Hit functions from SGD [15], whether they are linked to DNA damage or regulate *RNR3*, and interesting interactions from the BioGRID database [13].

hit	SGD description (copied directly)	RNR3	BioGRID
Ace2	Transcription factor required for septum destruction after cytokinesis		Aft1
Aft1	Transcription factor involved in iron utilization and homeostasis	Y [4]	Spt3,Rox1
Bdf1	Protein involved in transcription initiation at TATA-containing promoters	Y [36]	Spt8
Ckb1	Beta regulatory subunit of casein kinase 2 (CK2), a Ser/Thr protein kinase with roles in cell growth and proliferation		Bdf1
Spt10	Putative histone acetylase with a role in transcriptional silencing	Y [91]	Swi4,Mot2
Met18	Component of cytosolic iron-sulfur protein assembly (CIA) machinery	Y [71]	Rfx1,Spt3,TAF1
Gal80	Transcriptional regulator involved in the repression of GAL genes in the absence of galactose		Ino2
Spt3	Subunit of the SAGA and SAGA-like transcriptional regulatory complexes	Y [100]	Mot3,Spt8,Met18, Aft1,Ada2
Ino2	Component of the heteromeric Ino2p/Ino4p basic helix-loop-helix transcription activator that binds inositol/choline-responsive elements (ICREs)		Gal80,Swi4
Sfp1	Regulates transcription of ribosomal protein and biogenesis genes	Y [99]	Tor1
Spt8	Subunit of the SAGA transcriptional regulatory complex but not present in SAGA-like complex SLIK/SALSA	Y [100]	
Tda9	DNA-binding protein	Y [75]	
Ixr1	Protein that binds DNA containing intrastrand cross-links formed by cisplatin	Y [94]	
Exo5	Mitochondrial 5'-3' exonuclease and sliding exonuclease		
Mot2	Ubiquitin-protein ligase subunit of the CCR4-NOT complex	Y [61]	Spt10
Gcn4	Basic leucine zipper (bZIP) transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation		Ada2,Spt3,Rap1
Ada2	Transcription coactivator		
Mot3	Transcriptional repressor and activator with two C2-H2 zinc fingers	Y [50]	
Rox1	Heme-dependent repressor of hypoxic genes	Y [50]	Mot3

Table C.3: Functional analysis of SGA library hits: GO annotations. GO slim process terms for the top 16 activators and top 3 repressors identified in the SGA. Searched and exported directly from SGD using the GO SLIM Mapper [15].

GOID	GO term	Gene(s)
6366	transcription from RNA polymerase II promoter	INO2,SPT3,ADA2,GCN4, MOT2,AFT1,SPT10,IXR1 ,SPT8,ACE2,SFP1,GAL80, MOT3,ROX1
6325	chromatin organization	SPT3,ADA2,SPT10, SPT8,BDF1
6974	response to DNA damage stimulus	CKB1,SPT10,IXR1, BDF1
16570	histone modification	SPT3,ADA2,SPT10, SPT8
43543	protein acylation	SPT3,ADA2,SPT10, SPT8
18193	peptidyl-amino acid modification	SPT3,ADA2,SPT10,SPT8
6281	DNA repair	SPT10,IXR1,BDF1
42221	response to chemical stimulus	GCN4,IXR1
33043	regulation of organelle organization	ADA2,BDF1
6629	lipid metabolic process	INO2,MOT3
910	cytokinesis	ACE2
6360	transcription from RNA polymerase I promoter	CKB1
5975	carbohydrate metabolic process	GAL80
6383	transcription from RNA polymerase III promoter	CKB1
51726	regulation of cell cycle	ACE2
51052	regulation of DNA metabolic process	BDF1
7059	chromosome segregation	AFT1
6352	" DNA-dependent transcription, initiation"	GCN4
43934	sporulation	SPT3
7005	mitochondrion organization	EXO5
746	conjugation	SPT3
31399	regulation of protein modification process	ADA2
278	mitotic cell cycle	ACE2
6811	ion transport	AFT1
7124	pseudohyphal growth	SPT3
6354	" DNA-dependent transcription, elongation"	MOT2
51049	regulation of transport	AFT1
6468	protein phosphorylation	CKB1
6401	RNA catabolic process	MOT2
70271	protein complex biogenesis	GCN4
1403	invasive growth in response to glucose limitation	SPT3
48285	organelle fission	AFT1
70647	protein modification by small protein conjugation or removal	MOT2
6970	response to osmotic stress	MOT3
42594	response to starvation	AFT1
8150	biological process unknown	TDA9
other	other	MET18