UNIVERSITY OF CALIFORNIA SAN DIEGO

In Vivo Platform Plasmid Method using *Saccharomyces cerevisiae* for Determining Binding Affinity Profiles of Transcription Factors

A thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Alexis Julia Cugini

Committee in Charge:

Professor Scott Rifkin, Chair
Professor Sergey Kryazhimsky
Professor James Posakony

2020

The thesis of Alexis Julia Cugini is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2020

TABLE OF CONTENTS

Signature Page…… iii

Table of Contents……... iv

List of Figures…… vi

List of Tables……. vii

Acknowledgements…… viii

Abstract of the Thesis…… ix

Chapter 1: Introduction……. 1

1.1 Transcription Factor Introduction……... 1

1.2 *In Vitro* Methods……. 3

1.3 *In Vivo* Methods…….. 14

1.4 *In Silico* Methods and Data Display…… 18

1.5 Platform Plasmid System Introduction……... 19

Chapter 2: System Description……..... 20

2.1 Overview……. 20

2.2 Inducible LexA System……... 20

2.3 Platform Customization……... 25

2.4 Mutagenesis……. 27

Chapter 3: System Construction…… 28

3.1 Bacteria and Yeast Strains…….. 28

3.2 Plasmid Platform Component Isolation…….. 28

3.3 Platform Plasmid Assembly Steps…….. 29

3.4 Positive Control Platform Plasmid Assembly……. 31

3.5 FRP178 Yeast Transformation……... 32

3.6 Platform Yeast Transformation…….. 33

3.7 Positive Control Yeast Transformation…….. 33

Chapter 4: Future Steps…… 34

4.1 Progress Report.…… 34

4.2 Next Milestones…….. 33

4.3 Pilot Study and Flow Cytometry Phenotyping……... 35

4.4 Results Predictions and Discussion…… 36

References…… 38

LIST OF FIGURES

Figure 1: Surface Plasmon Resonance (SPR)…….. 5

Figure 2: Systemic Evolution of Ligands by Exponential Enrichment (SELEX)…… 6

Figure 3: Protein-binding Microarray (PBM)…….. 8

Figure 4: Chromatin Immunoprecipitation Sequencing (ChIP-seq)…… 15

Figure 5: Inducible LexA Transcription Factor Production……. 21

Figure 6: Inducible LexA Transcription Factor Activation…….. 22

Figure 7: Fluorescence vs. β-Estradiol Concentration…….. 23

Figure 8: Target Transcription Factor Production…… 24

Figure 9: Neongreen Protein Production…….. 25

Figure 10: Platform Plasmid CRISPR Features……... 26

Figure 11: Mutant Target Transcription Factor Production……. 27

Figure 12: Platform System Construction Overview Flowchart…….. 28

Figure 13: Platform Plasmid Component PCR Cloned Steps……...…… 29

Figure 14: Platform Plasmid Assembly Steps…….. 31

Figure 15: Positive Control Platform Plasmid Assembly Steps……... 32

Figure 16: LexA Inducible System Yeast Transformation……... 32

Figure 17: Platform Yeast Strain Transformation…… 33

Figure 18: Positive Control Strain Transformation……..... 33

Figure 19: Platform Plasmid Gel Electrophoresis Final Check…… 34

Figure 20: Fluorescence Phenotyping by Flow Cytometry…….. 35

LIST OF TABLES

Table 1: *In Vitro* Methods Overview……. 12

Table 2: *In Vivo* Methods Overview…….. 17

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Scott Rifkin for his guidance and support as my committee chair. Despite many delays and difficulties, his flexibility, patience and advice allowed me to make the progress that I did.

I would also like to acknowledge postdoctoral researcher, Bing Yang, for his invaluable support as my mentor. Through many months in the lab, one project change and one pandemic, he was there to provide hands-on experience and recommendations that changed my research for the better.

Lastly, I would like to acknowledge the entirety of the Rifkin Lab for their feedback and support. Their unique skillsets and research topics contributed to my own research in ways I had not anticipated.

This thesis is coauthored with Cugini, Alexis J., Rifkin, Scott and Yang, Bing. Cugini, Alexis J. was the primary author of this paper.

ABSTRACT OF THE THESIS

In Vivo Platform Plasmid Method using *Saccharomyces cerevisiae* for Determining Binding Affinity Profiles of Transcription Factors

by

Alexis Julia Cugini

Master of Science in Biology

University of California San Diego, 2020

Professor Scott Rifkin, Chair

While transcription factors are a crucial part of gene regulatory networks, their role in evolutionary processes is still not fully understood by researchers. Recent advancements in studying protein-DNA interactions have revealed that transcription factors have a binding affinity profile where there are both primary and secondary binding sites. The existence of these binding profiles hints to the possibility that transcription factors may have more potential for change than previously thought. How binding profiles may be affected by changing environmental conditions or mutations in the binding domains and binding sites

remains to be seen. Existing *in vitro* and *in vivo* methodologies do not have the ability to produce the high throughput, informative data required to answer these questions. In this paper, an inducible platform, integrated in the yeast genome, is described that allows for the study of multiple transcription factors *in vivo*. By measuring an observable phenotype (fluorescence), how introduced mutations and altered environmental conditions affect the binding abilities of a chosen transcription factor can be investigated. The platform plasmid has already been assembled and the inducible system has been transformed into the yeast cells. The next steps are to transform the platform plasmid into the yeast and to insert a target transcription factor system for a pilot study. Once completed, this system will allow for more comprehensive studies of the binding space of various transcription factors.

Chapter 1: Background and Introduction

1.1 Transcription Factor Introduction

Transcription factors (TFs) are proteins that control gene expression by binding to specific sequences in *cis* regulatory DNA elements located on the promoters and enhancers of a target gene. The interaction between a binding domain (BD) on the TF and its corresponding binding site (BS) on the DNA can regulate positive or negative expression of a target gene. In addition to interacting with DNA, TFs have their own BDs that can interact with other proteins to regulate transcription (17). TFs, along with their target promoters and their coregulators, are a significant part of gene regulatory networks (GRNs) which also include intercellular signaling molecules, transcription factor proteins, and *cis*-regulatory module DNA (10, 17). Understanding the workings of GRNs allows evolutionary biologists to predict future and reconstruct past evolutionary pathways. However, understanding GRNs in their entirety is difficult since even their individual components, such as TFs, require extensive research to understand how they are affected by factors such as environmental conditions and mutations (33). Studying the binding properties of TFs is crucial for understanding how components within GRNs interact with each other and understanding physical mechanisms of gene control (10, 34).

Studying transcriptional regulation is difficult due to the large numbers of different TFs, cell types and environmental states involved in gene regulation (19). Additionally, TFs can experience changes in their binding preference due to post-translational BD modifications and exposure to coregulators (10). Binding affinity refers to the strength of binding between a TF and DNA sequence which can refer to how long a TF stays bound or how likely a TF is to bind to a specific BS when other sites are available. Binding

specificity refers to how selective a TF is towards which BSs it interacts with. The TF with the highest affinity for a specific BS may not be the TF with highest specificity for that BS (16). Further complicating TF studies is the fact that knowing where a TF binds does not explain why it binds there (19).

Until relatively recently, it was thought that TFs were highly conserved over evolutionary time due to their importance in expressing genes critical for survival. Most TFs are pleiotropic (able to interact with multiple BSs in different contexts), meaning an adaptive mutation in one context can be deleterious in other potentially crucial interactions (17, 22, 33). However, recent technologies have revealed that TFs can evolve ways to minimize pleiotropy and diversify their roles in GRNs (5, 17). Utilizing more advanced methods, it was revealed that most TFs have both a primary binding motif and secondary binding motif where the primary motif is the most preferred BS for that TF (5, 17). The existence of secondary binding motifs can allow for TF evolution since the primary motif can be conserved, maintaining crucial gene functions, while the secondary motifs can be altered without jeopardizing survival. There have already been multiple studies published that have discovered evidence of this secondary BS divergence occurring (5, 17). Regarding the primary binding motif, it is unknown if it is a single sequence or a group of closely related sequences that can be described using a sequence logo. The exact distinctions between the primary sequences and the secondary sequences need to be investigated.

There are two main evolutionary questions of interest that this paper seeks to investigate. First, how do mutations in the BDs and BSs affect binding affinity profiles? Second, how do binding affinity profiles change with environmental conditions? Regarding studying mutations specially, uncovering the degree of coevolution and epistasis that may

occur between the BS and the BD is of most importance. The ability to investigate these questions has been limited by methodical constraints in the past, but this paper introduces a new methodology that aims to gain a greater understanding on what factors affect TF binding affinities and to what degree these factors alter the binding abilities of TFs and their BSs.

Methods for studying protein, specifically TF, and DNA interactions fall into two general categories: *in vitro* and *in vivo*. Since the method in development described in this paper seeks to combine the benefits of both *in vitro* and *in vivo* experimentation, a review of the advantages and disadvantages of both methods is necessary to fully demonstrate the advantages of the new method. Additionally, this review provides a brief overview of how vast the range of technologies for studying protein-DNA interactions is and how the field has advanced over the years.

1.2 *In Vitro* Methods

In vitro methods refer to studies conducted outside of cells, using just the proteins and DNA, which generally seek to identify BSs preferred by a specific TF and to determine binding energy landscapes (10). Most *in vitro* studies measure the binding affinity of a selected TF by exposing it to a range of DNA sequences while some studies do the opposite by using TFs with manipulated DNA-recognition abilities (10). They are useful at gaining a quantitative understanding of how TFs function and determining the biophysical aspects surrounding specific bindings. However, a critical disadvantage of *in vitro* methods is that they are conducted outside of the biological context in which the TF normally operates. Therefore, results can fail to account for the impact that coregulators and having multiple available target sites can have on the binding performance of specific TFs. To compensate

for this, the accuracy of *in vitro* binding profiles can be checked against *in vivo* measurements or the conditions of *in vitro* experiments can be modified to stimulate the intracellular environment in which most binding reactions occur (10). Additionally, producing and handling proteins *in vitro* tends to be difficult and labor intensive. All of the methods described are useful in their own ways and still in use today.

In the earliest days of TF binding studies, methods were extremely labor intensive, expensive and mostly only good for developing rough binding models due to very low throughput (18). The oldest *in vitro* approach to understanding TF and DNA interactions is the use of electrophoretic mobility shift assays where protein bound DNA travels more slowly across the gel than unbound DNA (10, 34). While this approach does provide some useful characterization, there is low throughput which quantitation makes difficult. A main takeaway from gel assays is distinguishing different binding preferences between a select family of TFs and DNA sequences of interest (20).

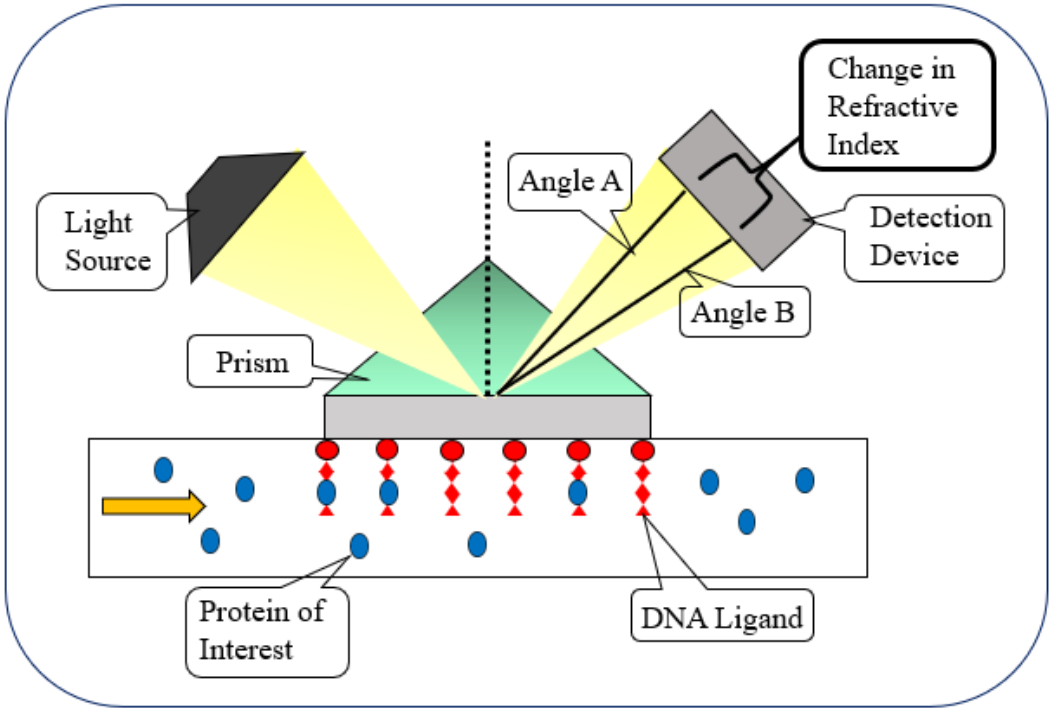

Figure 1: Surface plasmon resonance (SPR) basic principles: a polarized light beam is projected through a prism on top of a thin metal sensor chip with DNA ligands attached to the bottom; resonance occurs between the light photons and surface plasmons of the chip at a specific angle (Angle A); proteins of interest are exposed to the DNA ligands and cause a change in the refractive index when binding occurs (Angle B) (25)

Surface plasmon resonance (SPR) is another early method for studying protein-DNA interactions that detects binding through changes in the refractive index of the experimental apparatus (Figure 1) (34, 25). This method observes the reactions in real time as well as produces the kind of quantitative data needed for meaningful systems biology studies. The main disadvantage of SPR is that experiments are only successful when binding events differ in their kinetic constants. The reason for this drawback is that the results of SPR experiments are gained from response curves that are the sum of all the binding events between the proteins and DNA (25). An SPR technology known as the BIAcore was made available for commercial use in 1990 and is used to study binding constants, stoichiometry and DNA-protein interaction thermodynamics (25). However, the quality of results gained from BIAcore is dependent on the experimental design working within the mass transport

limitations of the technology. Mass transport limitations refer to when the binding of the protein analyte to the DNA ligand is limited by the diffusion of the analyte to the surface of the chip (25). To address this issue, workarounds include lowing the amount of immobilized DNA, increasing the flow rate of the analyte, upgrading to newer BIAcore models and removing glycerol or sucrose from the sample solution (25). Studying the effects of different mutations on binding affinity using SPR would be difficult since each mutant will require their own separate round of experimentation. Likewise, there would be limitations in how much environmental conditions could change and what kinds of environmental conditions can be changed.

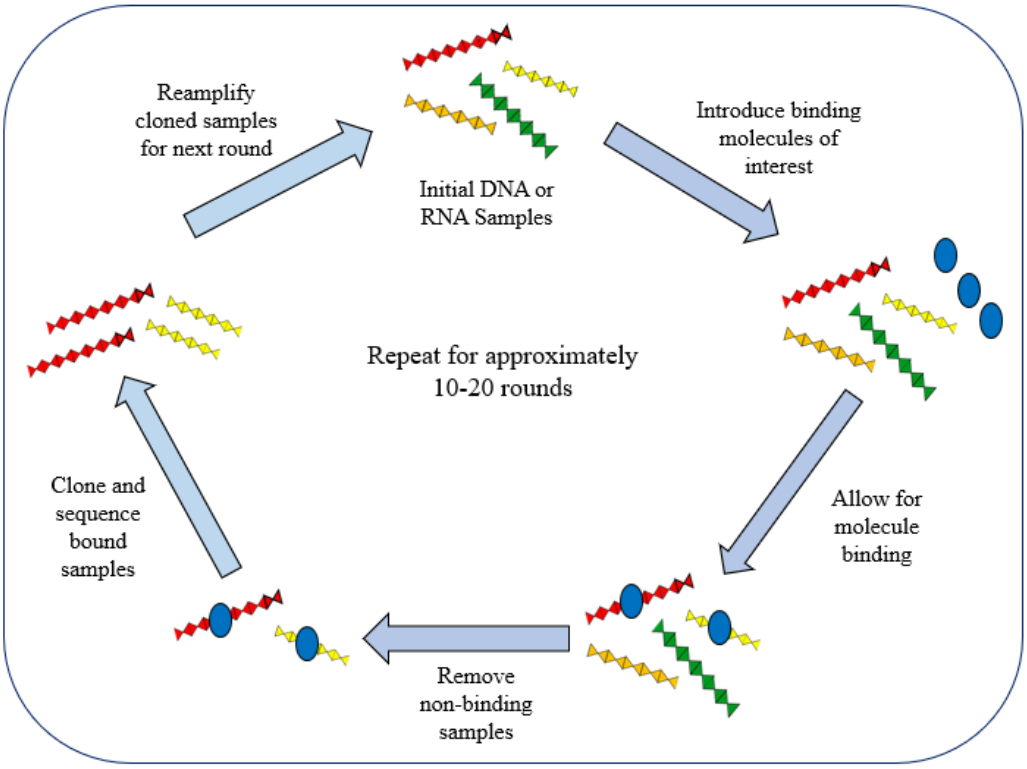

Figure 2: Systemic evolution of ligands by exponential enrichment (SELEX) (10)

Systemic evolution of ligands by exponential enrichment (SELEX) is a technique that can determine the BS consensus of a TF without any prior information (10, 31). SELEX is a technique where a sample of DNA or RNA is exposed to a binding molecule of interest

such as a TF (Figure 2). Unbound DNA or RNA fragments are washed away while the bound fragments are eluted to be cloned and sequenced. The copied DNA or RNA is reamplified for use in the next SELEX round and the process is repeated for ten to twenty rounds to build a binding profile. One disadvantage of this method is that having too few molecules in the analysis can lead to bottleneck effects in the data. Additionally, the PCR amplification steps in the process can introduce bias into the data while the extensive protein preparation steps result in relatively low throughput (18). However, the main disadvantage of SELEX is that only a select group of high affinity sites are selected for and amplified (10). This drawback results in an incomplete binding profile since low affinity sites are overlooked. Using SELEX is not ideal for this study since all BSs, even low affinity sites, need to be accounted for in order to get a complete binding affinity profile. Like SPR, SELEX would require a separate experimental set up for each separate mutation and environmental condition studied.

There have been several improvements to SELEX since its beginnings that have overcome some of its initial setbacks. When SELEX is coupled with massively parallel sequencing, the required cloning steps are eliminated, and the number of individual sequencing reads increases (18). As a result, this particular type of SELEX's binding profiles are 100 to 1000 times greater than original SELEX with the increased sequencing reads leading to better quality control by decreasing the statistical error in the profiles (10). Furthermore, Jolma and company developed a computational analysis that checks for lack of enrichment, cross contaminated samples, failed SELEX rounds and TFs binding to constant sequences (18). Despite these improvements, SELEX is still not suitable for this

study since the issues described in the previous paragraph have not been fully addressed and SELEX remains a poor candidate to collect the necessary comprehensive binding data.

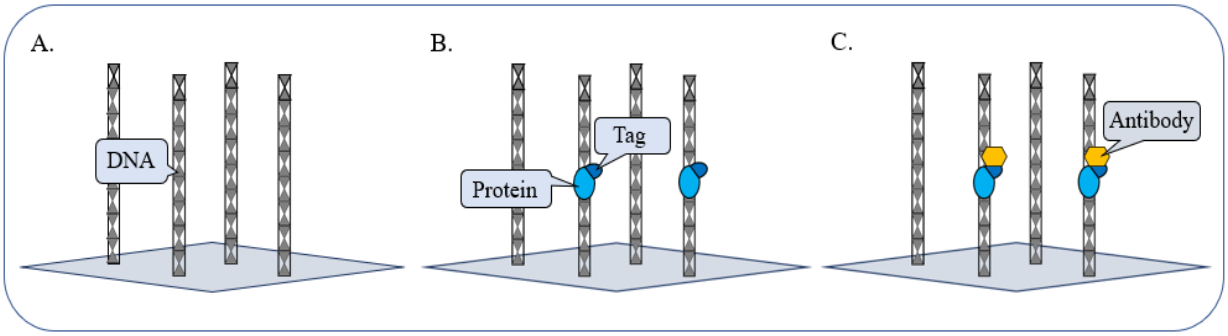

Figure 3: Protein-binding microarray diagram: **A.** single stranded DNA fragments are made double stranded by primer extension on the surface of the microarrays **B.** GST-tagged binding proteins of interests bind their target sites on the DNA strands **C.** fluorophore-labeled anti-GST antibodies respond to GST tag to reveal DNA-protein binding (3)

Protein-binding microarrays (PBMs) are a commonly used *in vitro* method for studying DNA-protein interactions. PBMs work by tagging a DNA-binding protein with an epitope tag, purifying the protein and sticking it to a dsDNA microarray (Figure 4) (3, 35). A fluorophore attached antibody made to match the tag is then applied to the microarray allowing for the BS motif to be identified based on what parts of the DNA experience binding. The microarrays are very high throughput with increased BS detection (10, 18). Despite being able to examine a wide range of TFs, the microarrays are usually limited by high costs, positional effects and the number of sequences that can fit on the array (2, 38). PBMs can have difficulty profiling certain TF families with longer BSs since the microarrays usually accommodate BSs less than ten base pairs long (19). Since the procedure requires a vast amount of purified proteins, proteins that are difficult to obtain, such as those with low expression or that have post-transcriptional modifications, can be difficult to analyze (18). Most of the datasets obtained from PBMs are presented as position weight matrices which can show how certain TF binding reactions are interconnected (10).

Studies involving PBMs uncovered the existence of secondary binding sites due to their capability to create comprehensive binding affinity profiles (17). Despite their role in uncovering the complexities of a TF's binding profile, PBMs are a not viable way to answer the questions this paper wants to investigate. The reasons why PBMs are not viable are less to do with the shortcomings of PBMs but rather the shortcomings of *in vitro* methods as a whole. This point is further elaborated on after *in vivo* methods have been explained.

Fluorescence anisotropy-based assays work by suspending nucleic acids and proteins freely in a solution with a part of the DNA complex being labelled with a fluorophore (16, 23). The anisotropy of the labeled species is measured to create a baseline before the binding protein is added to the solution. When the protein binds its labeled target, the whole complex should tumble more slowly, increasing anisotropy and allowing for the protein's DNA-binding properties to be measured. The main advantage of using fluorescence anisotropy is that, as a solution-based equilibrium technique, this method allows for measurements to be taken without potentially disrupting the reaction (16, 23). Other advantages include being a real-time assay and producing high precision data that makes it easier to differentiate between reactions of varying affinities. However, preparing properly labeled DNA probes is difficult and this technique does not allow for simultaneous observation of both bound and unbound species. Another disadvantage is that the concentration of fluorophores can fall below the detection level of the equipment (23). This technique does not allow for the customization this study requires.

Single molecule fluorescence microscopy (SMRM) is a method in which a DNA strand of interest is tagged with a single fluorophore and glued onto a surface (34). Proteins binding to the immobilized DNA can be detected and visualized. This method is capable of

imaging the same molecule for long periods of time, so it is useful for studying binding profiles and dynamics on a molecule-by-molecule basis (34). Another feature is being able to filter multiple molecule interactions to look at the single protein-DNA interactions. Observing different mutants and environmental conditions would require different experimental setups.

MITOMI devices are microfluidic devices with a high throughput that use the "mechanically induced trapping of molecular interactions" to detect binding interactions (10). This technique can measure thousands of interactions in a single experiment and can measure binding affinities in the nanomolar and micromolar range (10). As a result of intense protein and DNA washing, even low affinity interactions can be detected. However, designing and fabricating microfluidic devices can be a costly and time-consuming process. The intense preparation steps make this method unsuitable for this study since a separate device would be needed to study different mutations and environmental conditions.

Traditional Immuno-PCR (IPCR) is a process in which a primary antibody is attached to a surface to capture a protein of interest. After the protein is captured, secondary antibodies and oligonucleotide primers are added to the system and attach to the captured protein. The sample is then amplified and quantified to create a binding profile. IPCR combined with nanoparticle-based bio-barcode techniques allowed for highly sensitive results but depended on the use of the specific antibodies (15). An updated technique developed by Hou and company combined an exonuclease III (ExoIII) foot printing assay with PCR to gain high throughput data in a way that does not require specialized antibodies (15). They claimed that their technique has less time-consuming preparation steps and is easily customizable for different TFs. However, this universal real-time PCR assay is

dependent on the complete digestion of ExoIII for full detection sensitivity (15). This aspect of the PCR assay makes it unsuitable for this study since the factors in question may affect the performance of ExoIII.

High throughput sequencing with fluorescent ligand interaction profiles (HiTS-FLIP) is another technique used to collect quantitative protein-DNA binding affinity data. The procedure involves building and sequencing millions of clusters of genomic or random synthetic DNA, denaturing the DNA and washing away the second strand. The double stranded DNA is then rebuilt and exposed to a fluorescently tagged protein. Protein-bound DNA is tested for florescence so that the bound clusters can be mapped to corresponding sequences to create a quantitative binding affinity landscape (27). Without requiring multiple washing and drying steps, this technique allows for the analysis of long or complex binding motifs and observation of interdependencies between different genomic positions (27). HiTS-FLIP allows for changing conditions and reimaging binding affinities without having to sequence the DNA again. Utilizing HiTS-FLIP for this study is not possible since access to the specialized hardware this process is severely limited. Acquiring the necessary equipment would be a substantial investment that is not guaranteed to yield results.

Bind-n-Seq is a high throughput method that works by binding proteins to randomized oligonucleotide DNA targets, sequencing the bound oligonucleotides with massively parallel technology and then finding motifs among the sequences (38). This method can analyze multiple binding reactions at once with the use of barcoded oligonucleotides while does not requiring multiple rounds of binding and amplification (9, 38). Additionally, this method is not confined to testing 10 base pair long BSs like most microarrays, resulting in more complete binding profiles. Bind-n-seq is not an appropriate

method because it would be time consuming to run a separate experiment for each mutant. While each mutant protein would need its own separate purification and binding reaction, the sequencing of all the mutants can be conducted simultaneously using specialized oligonucleotide tags that identity which experiment the mutant originated. However, there are still other methods that can obtain similar results with both simultaneous sequencing and experimentation.

Table 1: *In Vitro* Methods Overview

Method	Advantages	Disadvantages	Reasons for exclusion in this study
Electrophoretic Mobility Shift Assays (EMSA)	Provides some useful characterization for most TFs	Low throughput, low resolution	Low throughput and lack of quantifiable data (method still used for checking PCR products)
Surface Plasmon Resonance (SPR)	Real time observations, quantitative data	Binding constants need to differ significantly, mass transport limitations	Each mutant/environmental condition would require separate experiments
Systemic Evolution of Ligands by Exponential Enrichment (SELEX)	No prior information needed for BS consensus	Bottleneck effects, PCR amplification bias, low throughput, only selects high affinity BS	Does not provide complete binding affinity profiles and separate experimental setups would be required
Protein-binding Microarrays (PBMs)	High throughput, increased BS detection, examines wide range of TFs	Limited by high costs, positional effects, limited number of sequences fit on array, high amounts of purified proteins required	*In vivo* methods are more suitable for this study

Fluorescence Anisotropy-based Assays	Solution-based equilibrium technique, real-time assay, high precision data	DNA probes are difficult to prepare, no simultaneous observation of bound vs. unbound species, detection limit on equipment	Does not allow required level of customization
Single Molecule Fluorescence Microscopy (SMFM)	Image same molecule for long periods of time, filter out multiple molecule interaction	Must be able to label samples with fluorophores	Would require different experimental setups for each mutant
MITOMI Devices	High throughput, precise binding affinity measurements, detects low affinity BSs	Costly and time consuming to design and fabricate	Time consuming preparation steps
Immuno-PCR (IPCR)	Sensitive results	Requires specialized antibodies	Would require specialized antibodies
Exonuclease III (ExoIII) Foot Printing Assay with PCR	Does not require specialized antibodies, high throughput, less time consuming than IPCR	Full detection sensitivity depends on complete digestion of ExoIII	The mutations and environmental conditions studies may affect the behavior of ExoIII
High Throughput Sequencing with Fluorescent Ligand Interaction Profiles (HiTS-FLIP)	Delivers quantitative protein-DNA binding affinity data, does not require multiple washing and drying steps, can analyze	Required hardware is not widespread	Obtaining the necessary equipment may not be a worthwhile investment

	long or complex binding motifs		
Bind-n-Seq	High throughput, does not require multiple binding and amplification rounds, can analyze multiple binding reactions at once	Requires separate purification processes and binding reactions for each protein	Would require multiple experimental runs

The most prevalent issues regarding most existing *in vitro* methods is that each unique mutant and environmental conditions would require their own separate experimental setup. Current *in vitro* methods are not equipped for studying possible coevolution between the BS and BD or detecting epistasis as it requires a large sample size of BS and BD variants. The evolutionary questions would be better answered using *in vivo* methods. However, most current *in vivo* methods have their own disadvantages that render them unsuitable for this study.

1.3 *In Vivo* Methods

In vivo methods refer to experiments conducted in living cells or whole organisms as opposed to *in vitro* methods which use isolated proteins and DNA (10). Like *in vitro* methods, *in vivo* methods can provide information on preferred BSs of a specific TF. *In vivo* methods have the advantage of preserving the biological context of sequence-specific interactions that is lost *in vitro* (10). While occurring in a biological context has its advantages, the results are limited to that specific biological context. *In vivo* experiments

often suffer from low resolution in BS identification and any data gained is usually qualitative or semiquantitative which makes numerical analysis difficult (10). However, our study requires *in vivo* methods to acquire the data needed its evolutionary questions.

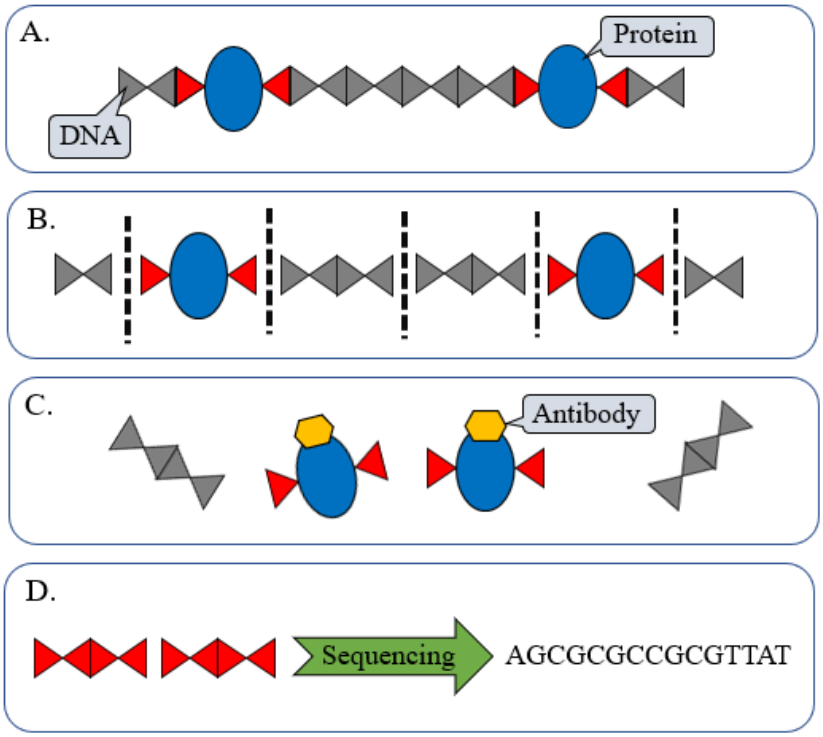

Figure 4: Chromatin immunoprecipitation sequencing (ChIP-seq): **A.** binding proteins of interest are cross-linked to their target sites on a genomic DNA sequence, **B.** DNA fragments are sheared apart, **C.** antibodies immunoprecipitate the protein of interest and unbound fragments are removed, **D.** proteins are unlinked and the target DNA is sequenced in ordered to be mapped on the genomes (10)

Chromatin immunoprecipitation (ChIP) is a commonly used *in vivo* method for searching for genome-wide TF binding (10, 30). This method is usually coupled with DNA microarray technology (ChIP-chip) or massive parallel sequencing (ChIP-seq) to allow for high throughput in determining protein bindings in the genome (10, 19). Of the two, ChIP-seq performs better with higher resolution, lower noise levels and greater range (Figure 4) (10). However, both ChIP-chip and ChIP-seq have limited by how abundant the protein of interest is and cross-linkage efficiency (38). For both ChIP methods, there is a high number of required sequencing reads which results in knowing where the binding occurs but not

much else (18). Furthermore, the main limitation of these methods is finding a specific antibody for experimentation (18, 38). Since our study involves investigating different TF mutants and environments, that specific antibody may become ineffective and not respond to a bound TF. However, gene editing methods, such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), can allow researchers to attach a Guanosine-5'-triphosphate (GTP) or an amino acid tag to the TF species that the antibody can attach to it. With this workaround, you can guarantee that the antibody will still be effective, but the tag must not interfere with the target protein's function. Nonetheless, even with the gene engineering, ChIP would not be an efficient method for creating a large enough sample size of BS and BD variants for this study.

Escherichia coli DNA adenine methyltransferase (DamID) requires the protein of interest to be bound to *E. coli* DNA (10). After the binding, the nucleotides that are close to the BSs are methylated, immunoprecipitated, analyzed with microarrays and then sequenced. Since methylation only occurs with the adenine in the GATC sites, the distance between two consecutive sites limits the resolution of this method. This method was limited to proteins that can bind *E. coli* but has expanded and been exploited in other systems. This method was excluded from use due to lack of flexibility.

DNaseI hypersensitivity assays use microarrays and massive parallel sequencing to obtain an unbiased, genome-wide mapping of protein binding (10). An advantage is that this approach distinguishes between nucleosome bound and unbound genomic loci. Like PBMs, each new mutant and environment will require separate experimental runs to collect the data needed to answer the questions of interest.

Yeast and bacterial two- and one- hybrid systems are *in vivo* methods that allow for experimentally manipulated stringency in protein-DNA interaction analysis (38). However, the hybrid systems suffer from low sequencing throughput and a limited library of target sites. The throughput can be improved by decreasing the library, but since the library is already small, there is not difference overall (38). Since this study needs high throughput to draw meaningful conclusions about binding behaviors, this method is not a good candidate to create comprehensive binding profiles unless some modifications are made.

Table 2: *In Vivo* Methods Overview

Method	Advantages	Disadvantages	Reasons for exclusion in this study
Chromatin immunoprecipitation (ChIP) coupled with DNA microarray technology (ChIP-chip) or massive parallel sequencing (ChIP-seq)	high throughput, ChIP-seq boasts higher resolution, lower noise levels and greater range	limited by protein abundance and cross-linkage efficiency, high number of required sequencing reads, specific antibodies required	Requires specific antibodies (despite gene engineering techniques, there are more efficient methods available)
Escherichia coli DNA adenine methyltransferase (DamID)	Maps protein interaction sites	Resolution limited by the distance between two consecutive sites of interest	Would require separate experimental runs
DNaseI hypersensitivity assays use microarrays and massive parallel sequencing	Provides an unbiased, genome-wide mapping of protein binding, distinguishes between nucleosome bound and unbound genomic loci	Relies on proper DNaseI digestion	Would require separate experimental runs

Yeast and bacterial two- and one-hybrid systems	Allows for experimentally manipulated stringency	Low throughput, limited target site library	Has low throughput (but could be remedied)

In vivo methods are less varied than *in vitro* methods. They are biased toward the specific type of TF and DNA sequence used in the study and rarely produce quantitative data. However, *in vivo* binding observations are more likely to resemble natural binding than *in vitro* observations due to being conducted in the TF's environmental context. While using *in vivo* methods for this study is likely to deliver a more contextually accurate binding profile, preexisting *in vivo* methods are not adequate.

To this day, *in vitro* and *in vivo* experiments often produce significantly different results despite improvements made to both methodologies. As such, many researchers combine both methods in hopes of observing results that are closer to actual binding behaviors. Most *in vitro* analyses use isolated TF-DNA binding domains while *in vivo* methods generally use full length TFs (19). It is unknown if isolated BDs and full-length TFs have different binding profiles in general, but this uncertainty could be the cause for the differences between the methods. However, an analysis by Jolma and company found that the primary binding specificity of TFs is determined by the binding domain (19).

1.4 *In Silico* Methods and Data Display

In silico methods refer to techniques that use computers to analyze and predict DNA and TF interactions (10). Most computer models are based on experimental data and then extrapolate beyond that data. Computers can oversimplify their results is ignoring how biological processes can affect the efficiency with which transcription factors bind their target sites (10). *In silico* predictions still tend to differ from *in vitro* and *in vivo* results.

Most models still need to account for how TFs interact with one another in addition to how TFs interact with DNA.

There are multiple ways to display TF binding affinity data. Cognate site identifier (CSI) analysis works by examining the sequence preferences of DNA binding molecules and comparing their affinity for about half a million similar sequences (5). Position weight matrices (PWMs) model the binding specificities of a given TF and can determine the presence of DNA sequences that are significantly more like the PBM than the background (4). A drawback of PBMs is that they often compress more specific binding specificities in larger data sets. Sequence specificity landscapes (SSLs) display multiple binding motifs with the best match in the center and display the relative affinities of a specific transcription factor to each motif at once (4).

1.5 Platform Plasmid System Introduction

In this paper, a custom platform plasmid system is transformed into yeast (*Saccharomyces cerevisiae*) in order to explore the binding space of a TF of interest. The platform plasmid was developed especially to study BS and BD coevolution and to investigate possible evolutionary pathways involving the target TF. This method is a flexible system, capable of providing a broad range of experimental conditions. Furthermore, the platform plasmid can test multiple protein variants at the same time instead of having to set up multiple experiments for the different variants. The next section will describe all the aspects of the system that make it ideal for studying TF binding affinities with multiple *cis* regulatory elements.

Chapter 2: System Description

2.1 Overview

In summary, the goal of the platform plasmid system is to use the expression of a reporter gene as a measure of the binding affinity of a TF to a BS (*in vivo*). The platform plasmid system is transformed into the yeast (*Saccharomyces cerevisiae*) genome to create a system that produces all necessary TFs and other proteins *in vivo*. The "platform" is called that because it is a framework that we can build experimental yeast strains upon. Since all necessary proteins are produced inside of the cells, the difficult process of cultivating the proteins *in vitr*o is unnecessary. One advantage of using yeast is that the organism has a short life cycle which allows for quick experimental preparation with low wait times. Since yeast is so widely used as a model system, there are a variety of genetic tools and tests capable of analyzing yeast systems. Yeast cells are especially easy to phenotype with fluorescence which is beneficial since the reporter gene is a fluorescent protein..

2.2 Inducible LexA System

This system utilizes two TFs: a constant LexA TF and an interchangeable target TF of interest. The LexA TF contributes to the production of the target TF while the target TF regulates expression of a fluorescence reporter gene. The systems that produce the target TFs and the LexA TFs are on separate plasmids and can be integrated into different locations in the yeast genome

One of the benefits of this system is that it is inducible as the target TF will not be produced without the presence of the inducer. This feature allows for the establishment of a baseline fluorescence level for the yeast system without any target TFs present. This baseline will include the naturally occurring TFs produced in an unaltered yeast system.

These preexisting yeast TFs will allow for the study of competitive binding since the target TFs may have to compete with them for access to the BSs. This aspect of the system stimulates natural conditions where multiple TFs interact and interfere with each other.

The inducible system utilized is the LexA complex developed by Ottoz and company to create a regulated system that controls the expression of a target gene through a specific, well-known input (28). The advantages of their systems include: a broad expression range for the target gene, no system activity without the input, being unaffected by and not affecting metabolism of the cell and lastly, having no toxic effects on the cell.

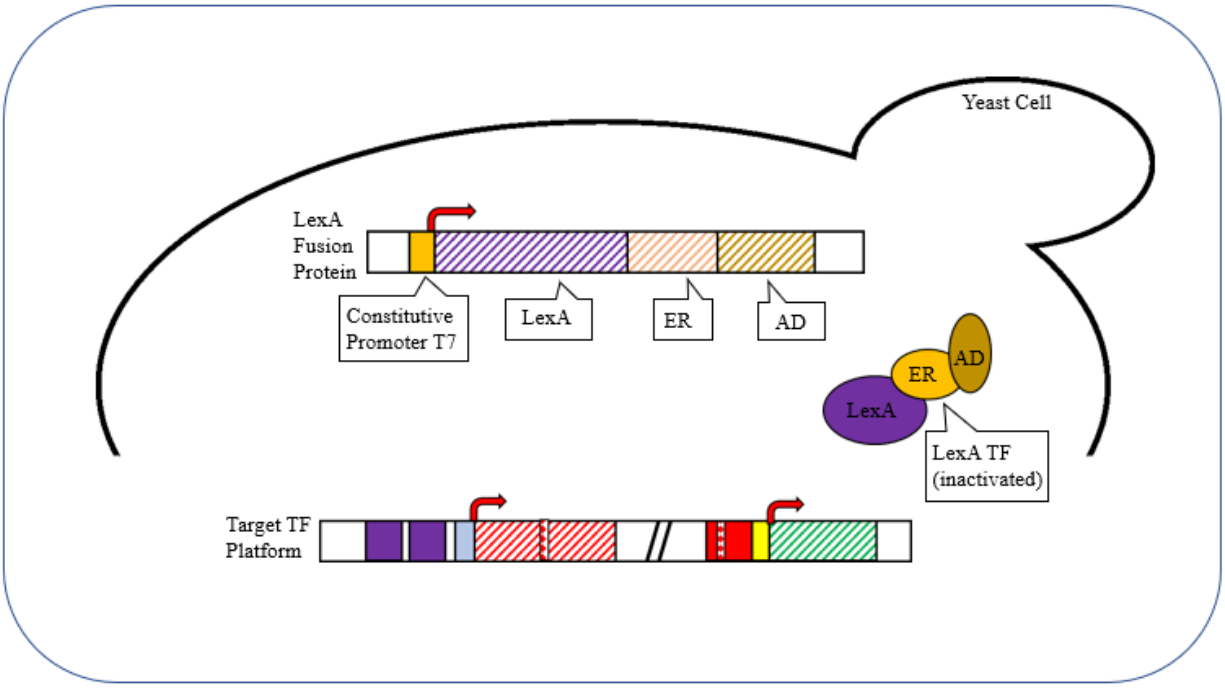

Figure 5: LexA TF is constitutively produced by the LexA fusion protein, but is inactive without the presence of its inducer

To establish an inducible system, a LexA fusion protein is inserted into the yeast (Figure 5) (11). The chimera TF produced by this system contains three parts: the bacterial LexA DNA-binding protein (LexA), the human estrogen receptor (ER) and the activation domain (AD) (28). These three components are combined to create a LexA TF that is inactive without the input of the inducer. Constitutive promoter T7 regulates this LexA

system and constantly produces the LexA TF product into the intracellular space. However, without the inducer, the LexA TF will not bind to its BS and will not have any effect on the yeast cell as a result.

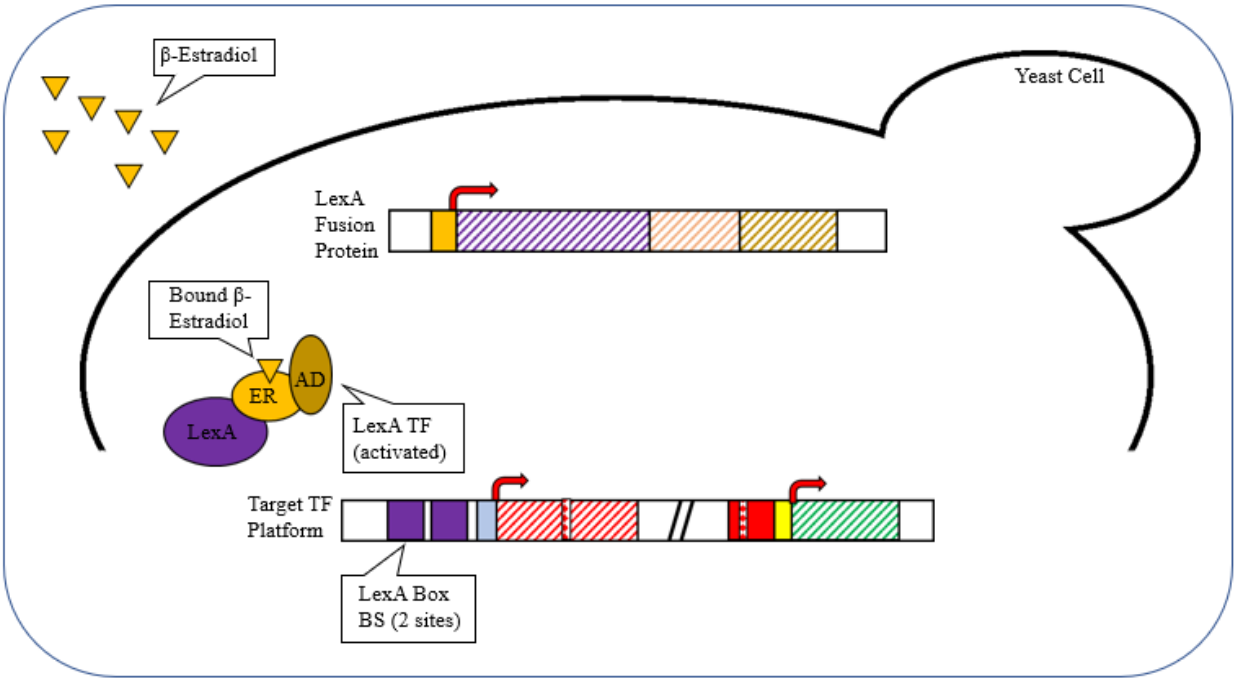

Figure 6: LexA TF is activated after the inducer, β-Estradiol, binds the human estrogen receptor (ER), allowing for the LexA TF to bind its target site, the LexA box.

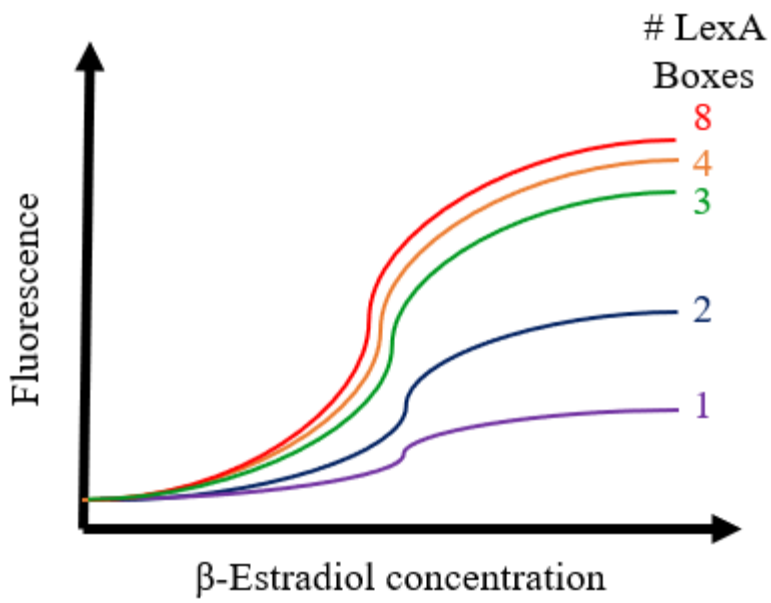

Figure 7: Changing the number of LexA box BSs alters the scale of β-Estradiol concentration vs fluorescence expression levels; 2 LexA boxes is ideal for measuring differences in fluorescence when exposed to the inducer, β-Estradiol (28)

The required inducer is β-Estradiol, which binds to the human estrogen receptor (ER) present on the LexA TF (Figure 6). After the ER is bound by the β-Estradiol, the TF is activated and capable of binding to the LexA box BS. There are two LexA box BSs present on the platform plasmid because the expression levels of the target genes can be scaled based on the number of LexA box binding sites present in the system (Figure 7). Two LexA boxes deliver an optimal tradeoff between how gradually the fluorescence changed with inducer concentration and the maximal fluorescence achieved. While three or more LexA boxes achieve higher fluorescence levels, there is a sharp transition from low to high fluorescence over a short range of β-Estradiol concentration.

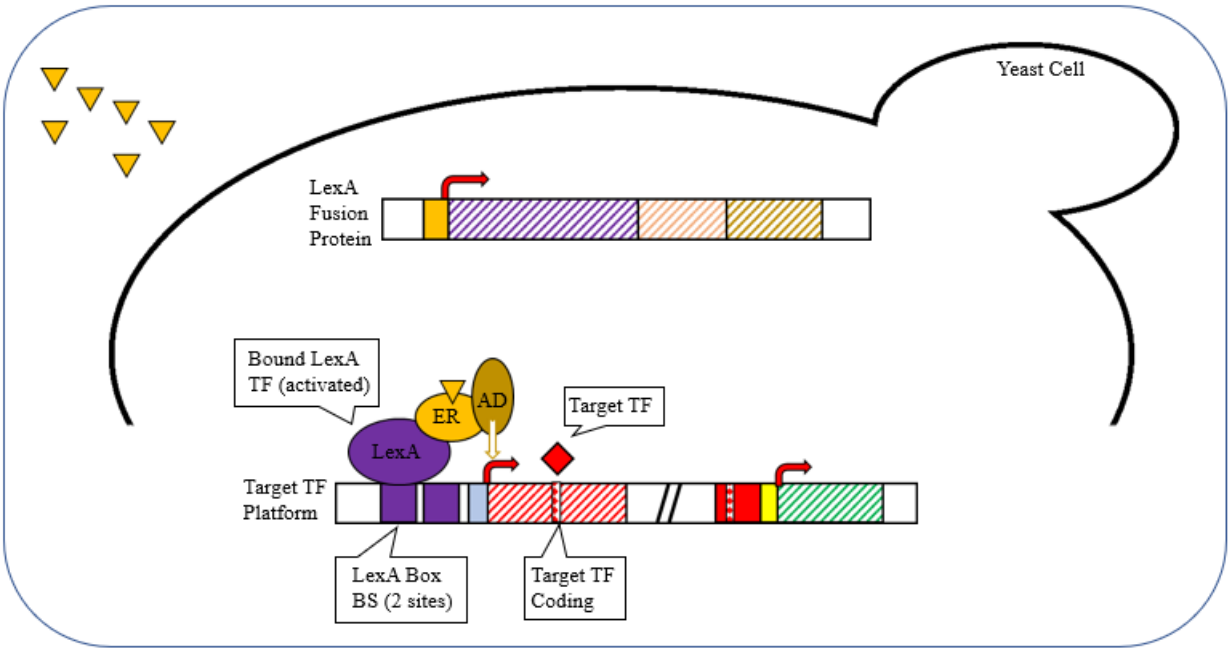

Figure 8: The activation domain (AD) on the LexA TF activates the production of the target TFs

The LexA DNA-binding protein binds to its target LexA BS on the platform, allowing the AD to regulate the target TF system promoter. After the LexA TF binds to the LexA boxes, the AD regulates the promoter of the target TF production system and allows for the transcription of the target TFs of interest.

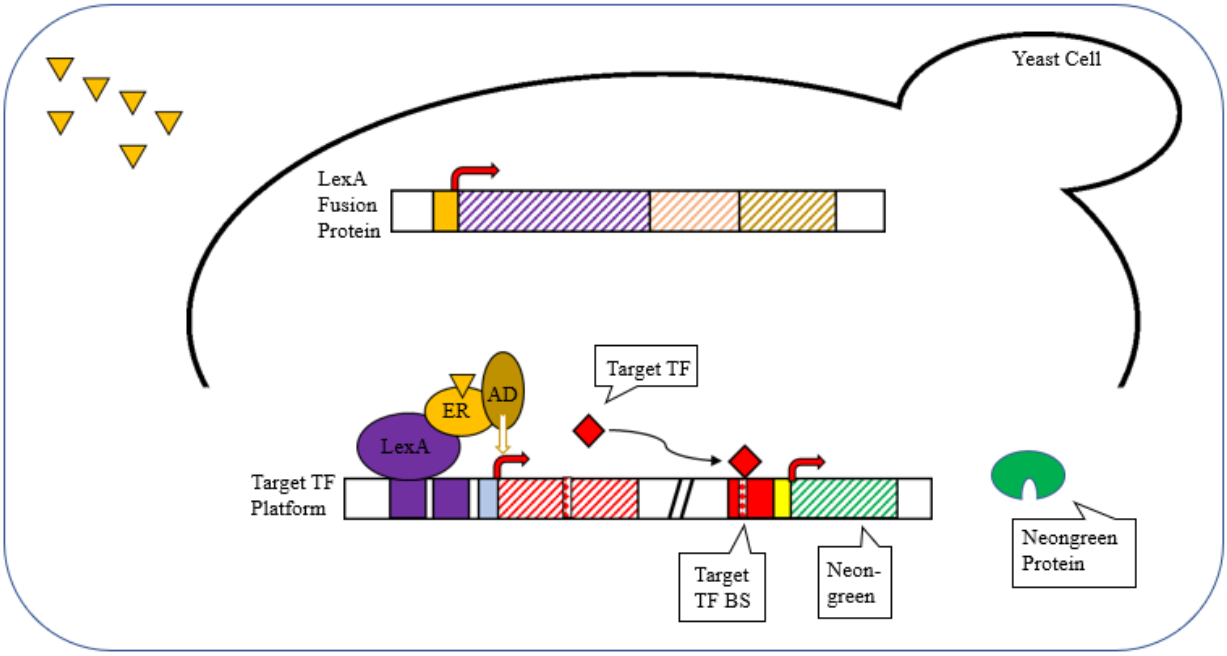

Figure 9: The target TF binds to its target BS and activates the production of Neongreen proteins

The target TFs regulate the expression of a gene that codes to produce synthesized Neongreen fluorescent proteins (21, 32). These Neongreen proteins reflect how strongly the TF interacted with the BS.

2.3 Platform Customization

One distinction to make is that there is the platform (containing a URA3 gene placeholder) and then there are the actual sequences for the target TFs that are used in experiments. Both the platform and the experimental sequences are constructed using bacterial plasmids since plasmid engineering is a well-established method. When it is time to conduct experiments, the bacterial plasmid is cut and inserted into a yeast genome that already has the LexA component integrated. No longer a plasmid, the platform exists as a platform strain of yeast that has both the LexA component and the DNA from the cut platform plasmid integrated into its genome (Figure 10A).

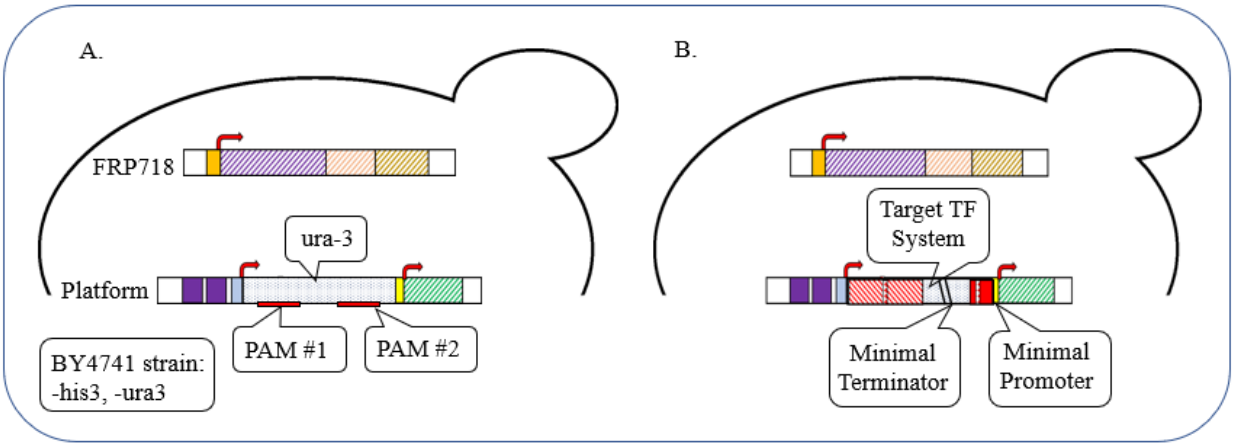

Figure 10: CRISPR features of the platform plasmid; the LexA components exists on FRP178 while the target TF system replaces the URA3 component on the platform

The chosen yeast strain needs to have a -URA3 deletion to check if the platform was properly integrated into the yeast genome (Figure 10A). The yeast should be able to survive on -URA3 plates if the integration was successful since the platform will provide the URA3 gene functions that the yeast lacked prior to the transformation.

To add in the experimental sequences, the URA3 fragment on the platform has the protospacer adjacent motifs (PAM) needed for the gene editing technique CRISPR (Figure 10A). The experimental plasmid can be cut with the appropriate restriction enzymes while the platform can be cut with CRISPR-Cas9. The experimental sequences can then replace some of the platform sequences to insert the experimental target TF and BS into the genome.

Due to limitations on length of the Illumina paired end sequencing system, the variant BD and variant BS need to be close enough together (<500 base pairs apart) to be on a single DNA sequence (Figure 10B). To accomplish this, a short terminator is used as the end of the target TF coding system (7). Additionally, a short promoter (about 60 base pairs) is used in front of the Neongreen coding sequence (29).

2.5 Mutagenesis

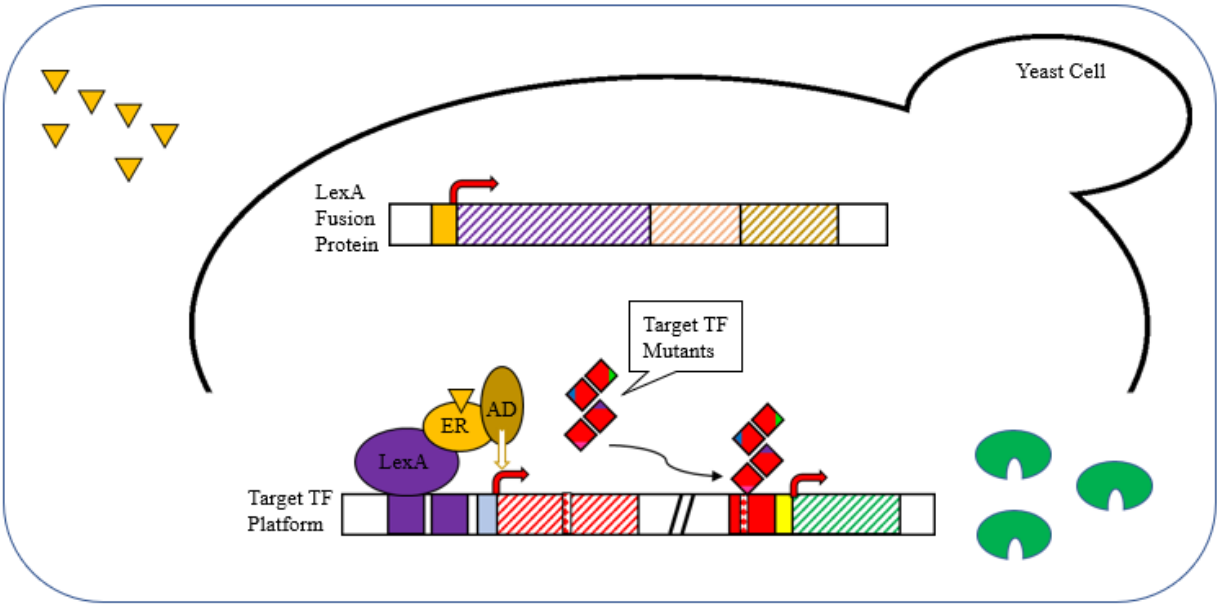

Figure 11: Production of mutant variants of the target TFs

There are two ways to introduce mutations into the platform system: random synthesis with mixed nucleotides or synthesized specific variants (8, 26, 36, 37). The former will be used to study binding of the variance of *cis* regulatory elements while the latter will be used to study a specific TF and its matching *cis* regulatory elements. Another option for studying mutations is to introduce random amino acids into the target TF and observing how its binding behavior changes (1).

Chapter 3: System Construction

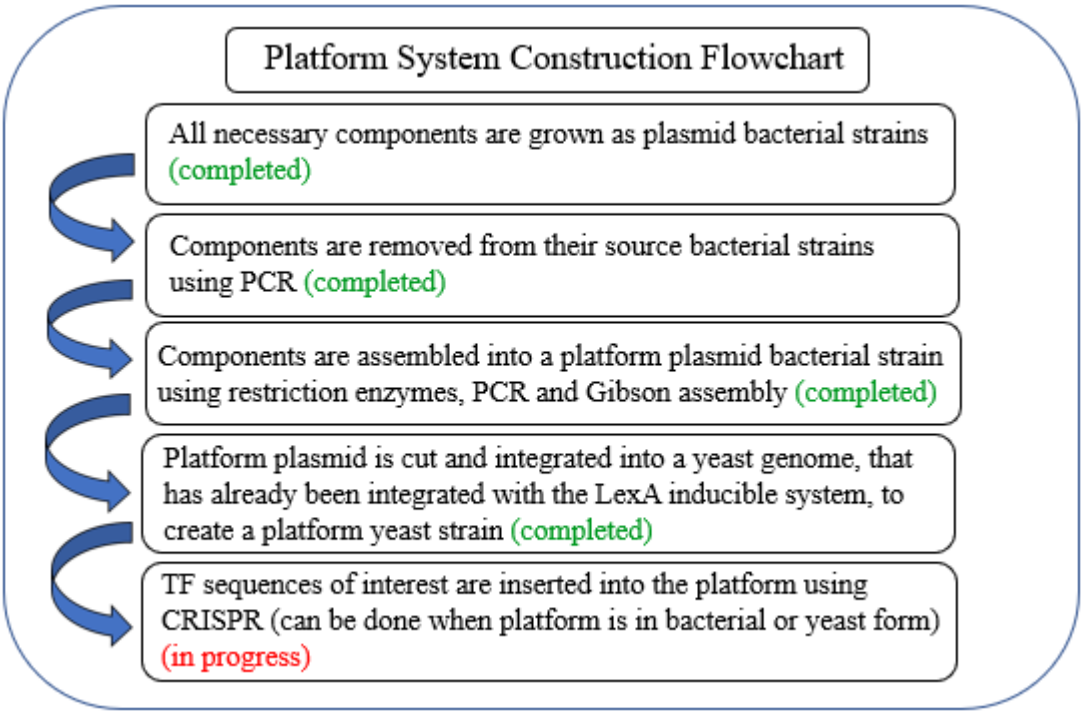

Figure 12: Platform system construction overview flowchart

3.1 Bacteria and Yeast Strains

For the bacterial strains, *Escherichia coli* DH5α was used. For the yeast strains, *Saccharomyces cerevisiae* BY4741 was used. All strains were cultivated under normal growing conditions.

3.2 Platform Plasmid Component Isolation

The first step of the platform plasmid assembly process was to PCR the various components of the system from their sources into bacterial strains. Each PCR run was checked for the correct product size using gel electrophoresis. Additionally, the DNA concentration of each PCR product was checked with NanoDrop spectrophotometer.

To begin, the URA3 component was removed from PML104 via PCR, creating overhanging ends (Figure 13A). Next, a fragment from FRP1640 was made to overlap with the FRP1640 backbone used in the final product and the URA3 fragment for PCR stitching via PCR

(Figure 13B). The URA3 fragment and the FRP1640 fragment were then combined into one fragment with PCR stitching (Figure 13C). The 3' MET15 homology arm and 5' MET15 homology arm were removed from S288c in separate PCR runs (Figure 13D). The *end-1* gene from *C. elegans* N2 was removed in another PCR run.

Synthetically produced Neongreen was ordered. These fragments were synonymously mutated to remove certain restriction sites, specifically BstEII sites.

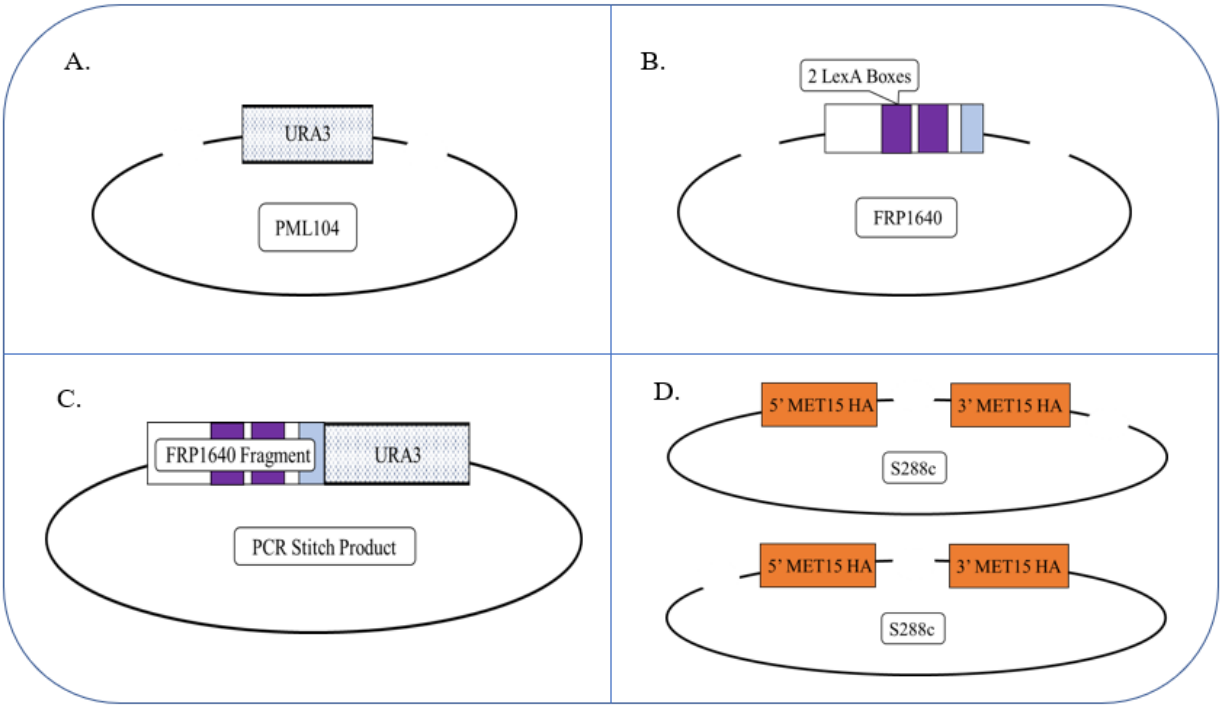

Figure 13: Platform plasmid component PCR cloned steps; **A.** URA3 fragment is removed from PML104, **B.** Fragment containing the 2 LexA boxes and a terminator is removed from FRP1640, **C.** Products from A. and B. are PCR stitched into a single fragment, **D.** 3' MET15 homology arm and 5' MET15 homology arm are removed from S288c

3.3 Platform Plasmid Assembly Steps

The next step of the process to assemble all the PCR components into a completed platform plasmid bacterial strain (13). Each intermediate plasmid was checked via gel electrophoresis for correct size and checked for adequate DNA concentration with NanoDrop

spectrophotometer. If the intermediate plasmid passed both checks, it was stored as a glycerol stock.

A FRP1640 plasmid was digested with KpnI (NEB #R3142) and SalI (NEB #R3138) restriction enzymes. The digested fragment was then attached to the 3’ MET15 homology arm using Gibson assembly to create intermediate plasmid A (Figure 14A).

Intermediate plasmid A was then digested with the restriction enzyme, BstEII (NEB #R0554S), in preparation for a Gibson assembly that attached the digested intermediate plasmid A to the 5’ MET15 arm. The product of this Gibson assembly was intermediate plasmid B (Figure 14B).

Intermediate plasmid B was digested with restriction enzymes, HindIII (NEB #R3104) and BamHI (NEB #R3136). The digested intermediate plasmid B was attached to the PCR stitched URA3/FRP1640 fragment and the custom Neongreen (Figure 14C). The completed platform plasmid was stored as a glycerol stock after being checked via gel electrophoresis and NanoDrop spectrophotometer.

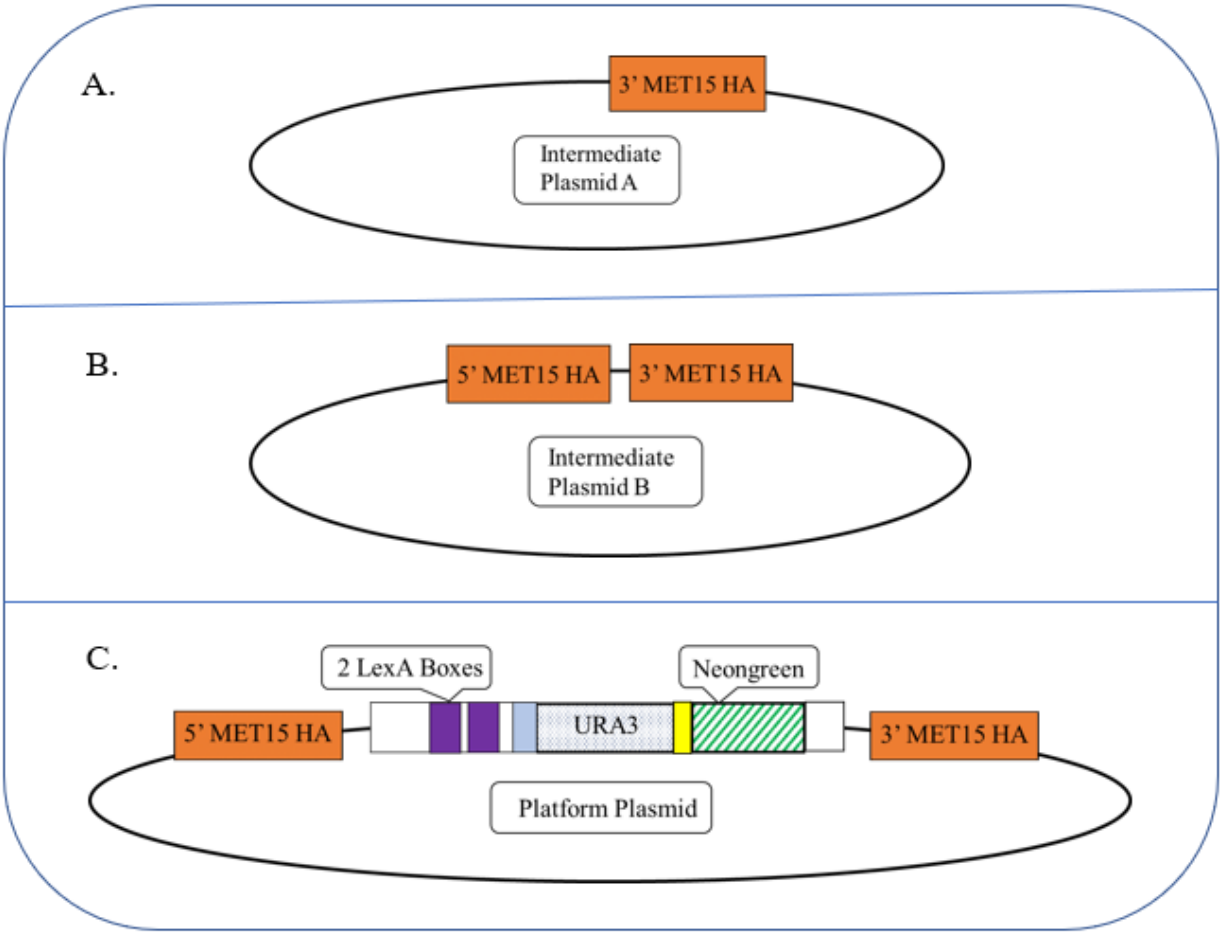

Figure 14: Platform plasmid assembly steps; **A.** intermediate plasmid A, **B.** intermediate plasmid B, **C.** completed platform plasmid, the pale blue box between the LexA boxes and URA3represent the terminator that separate the LexA complex and the URA3, the yellow box between the URA3 and Neongreen represent the promoter for the Neongreen

3.4 Positive Control Platform Plasmid Assembly

The platform plasmid was digested with restriction enzymes, BamHI (NEB #R3136) and PacI (NEB # R0547S), to remove the URA3 fragment. The digested platform plasmid, the *end-1* fragment and the synthetic Neongreen fragment were combined with Gibson assembly to create a positive control platform plasmid bacterial strain (Figure 15). A glycerol stock was created after being checked via gel electrophoresis and NanoDrop spectrophotometer.

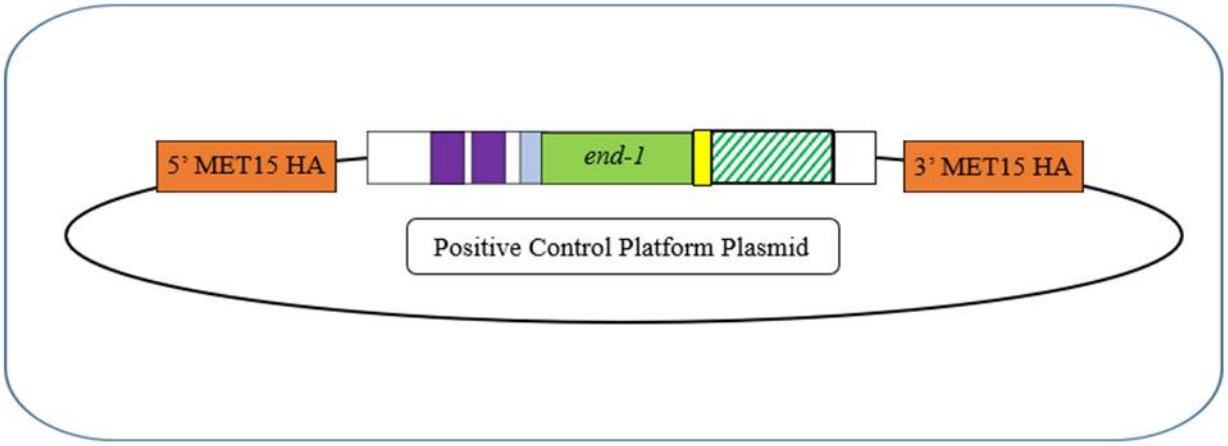

Figure 15: URA3 is cut from platform plasmid and replaced with the *end-1* gene that was removed from *C. elegans* N2

3.5 FRP178 Yeast Transformation

FRP178 was cut using restriction enzyme, NsiI (NEB #R3127). The cut plasmid is integrated into BY4741 yeast using a standard yeast transformation protocol (Figure 16) (6, 12, 14).

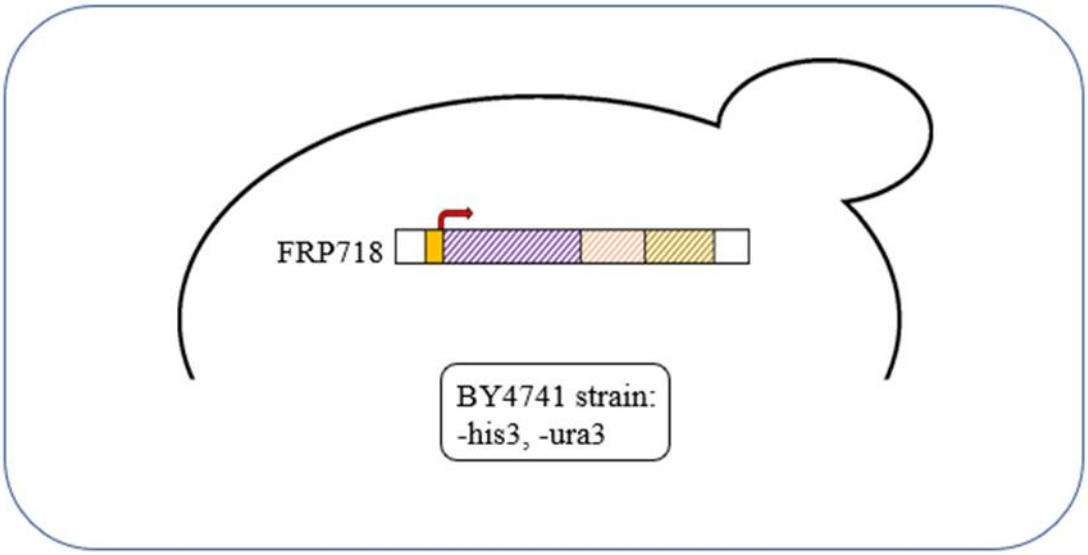

Figure 16: FRP178 containing the LexA inducible system is integrated into BY4741 yeast

3.6 Platform Yeast Transformation

The platform plasmid was cut using restriction enzyme, XhoI (NEB # R0146S). The cut platform plasmid is integrated into the BY4741 yeast strain that already contains the FRP178 integration using a standard yeast transformation protocol (Figure 17).

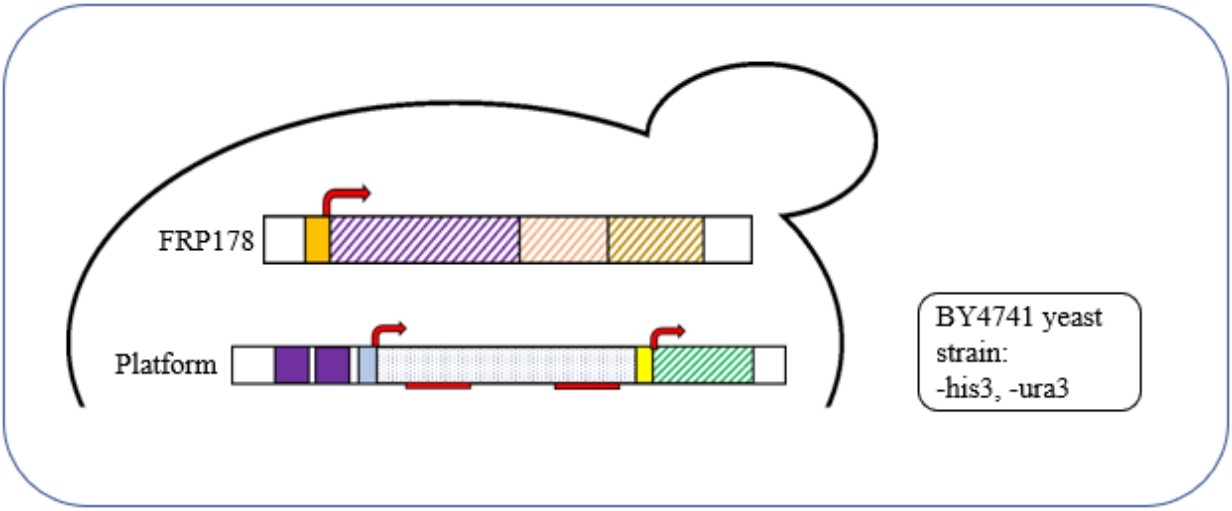

Figure 17: The platform is integrated into the BY4741 yeast containing FRP178

3.7 Positive Control Yeast Transformation

The positive control platform plasmid was cut using restriction enzyme, XhoI (NEB # R0146S). The cut positive control plasmid is integrated into the BY4741 yeast strain that already contains the FRP178 integration using a standard yeast transformation protocol (Figure 18).

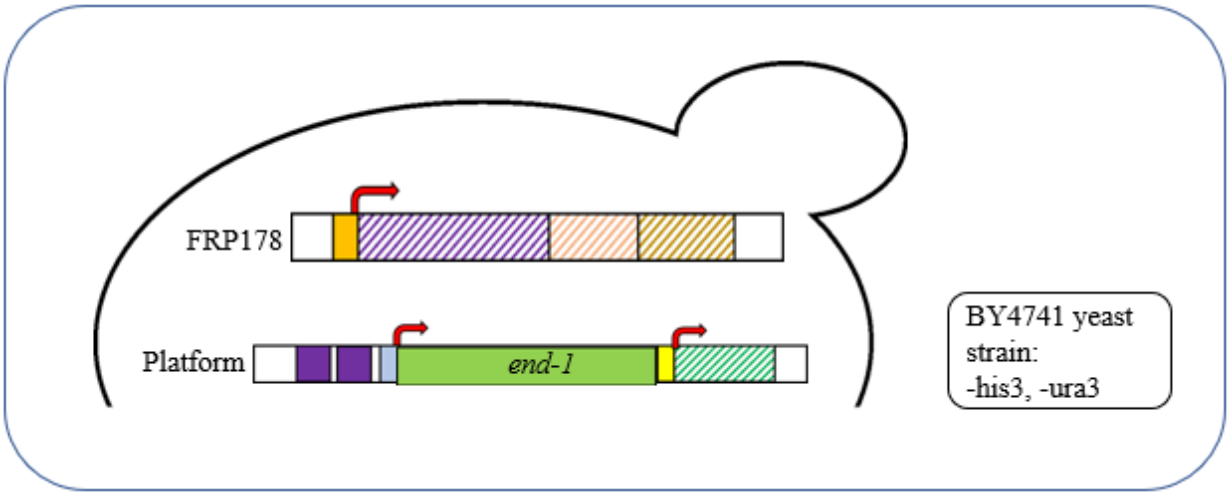

Figure 18: The positive control platform is integrated into the BY4741 yeast containing FRP17

Chapter 4: Future Steps

4.1 Progress Report

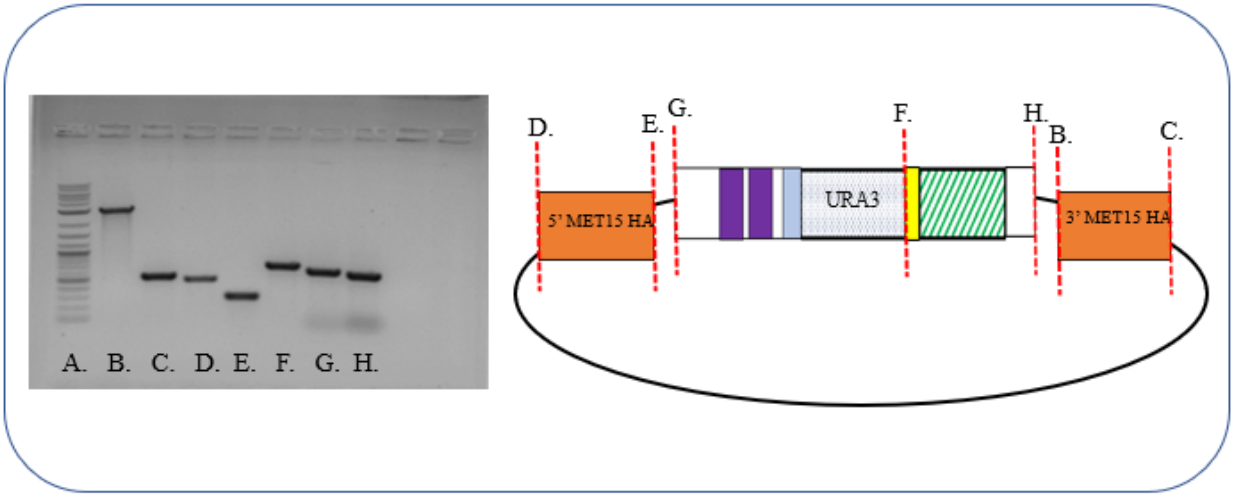

Figure 19: Gel electrophoresis final check of the completed platform plasmid by utilizing each set of primers to cut the platform plasmid into its various components: A. ladder, B. IPA 3'-5'-F and IPA 3'-5'-R (3000 bp), C. IPA 3'-3'-F and IPA 3'-3'-R (530 bp), D. IPB 5'-5'-F and IPB 5'-5'-R (490 bp), E. IPB 5'-3'-F and IPB 5'-3'-R (290 bp), F. neongreen F and neongreen R (700 bp), G. platform 5' F and platform 5: R (660 bp) H. platform 3' F and platform 3' R (660 bp)

So far, the platform plasmid has been fully assembled with all components present and accounted for by gel electrophoresis (Figure 14C, 19) (14). The platform has been successfully integrated into the BY4741 yeast that already contains a FRP178 integration to create the platform yeast strain (Figure 17). The positive control platform has been successfully integrated into the BY4741 yeast that already contains a FRP178 integration to create the positive control yeast strain (Figure 18).

4.2 Next Milestones

Although the URA3 fragment has already been replaced with *end-1* to create the positive control plasmid, different TF variants need to be inserted into the platform to check if the TF insertion feature of the system is consistent (Figure 10).

Eventually, to create a negative control, the *end-1* promoter will be removed via PCR from the positive control plasmid and transformed back into the yeast system. If the *end-1* gene

does not perform sufficiently as a positive control, the human *HoxD13* is a possible candidate to test out (1).

4.3 Pilot Study and Flow Cytometry Phenotyping

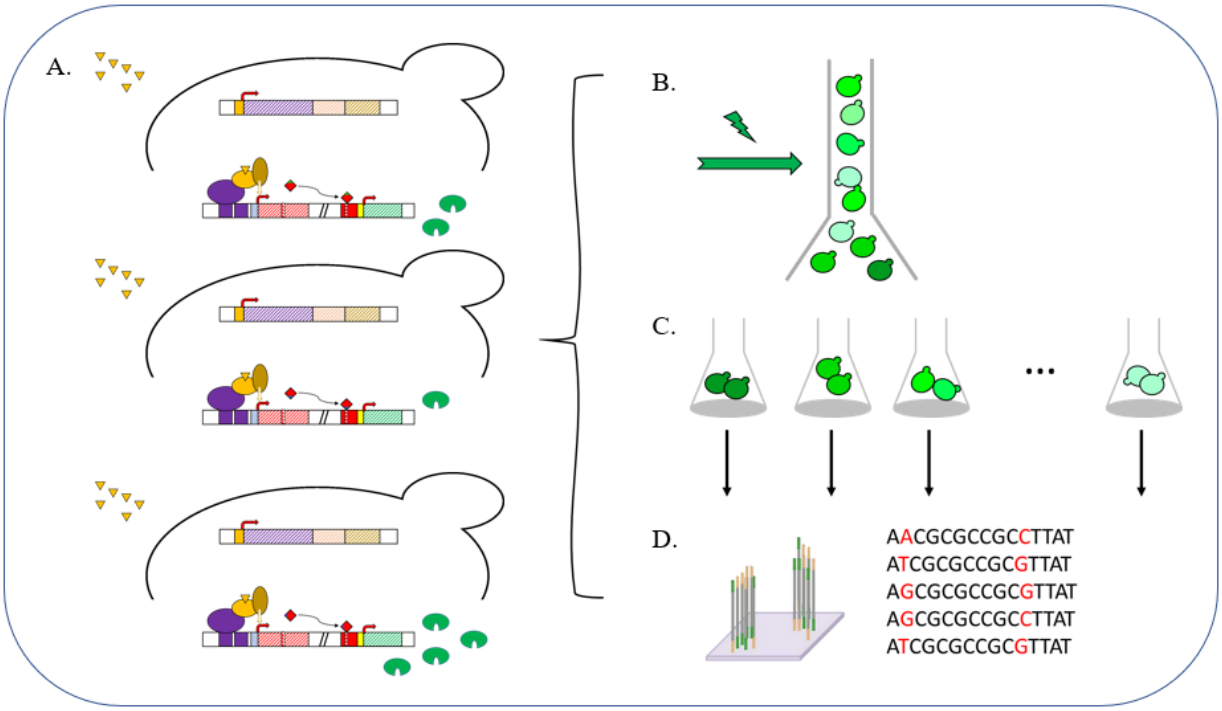

Figure 20: Phenotyping by flow cytometry, **A.** mutant TFs are produced in the yeast system through synthetically mutated DNA sequences leading to altered fluorescence levels due to fluctuations in the amount of neongreen proteins produced, **B.** phenotyping by measuring levels of fluorescence with flow cytometry, **C.** yeast cells are grouped based on their levels of fluorescence, **D.** each group of cells sequenced using a prepared library of variants in both TF and cis regulatory elements

To conduct a trial test, *end-1* from *C. elegans* will be inserted into the platform plasmid. The TFs themselves will be unmutated while there will be random mutations in the BSs (a space of eight nucleotides). For phenotyping, the yeast cells will be processed using flow cytometry (likely AriaII) to measure their fluorescence and grouped according to their fluorescence levels (Figure 20). These groups will be sequenced and the mutations will then be matched to specific levels of fluorescence (24).

It would be inefficient to sequence the whole yeast genome when specific sequences of interests are needed. The DNA surrounding the variable BS and any variable TF sequences is

fixed so it can be used as a marker for finding which part of the genome is needed for sequencing. Using paired end sequencing these variable regions are captured with the TF paired with the BS from a single DNA strand.

4.4 Results Predictions and Discussion

The proposed pilot study will not investigate how environmental conditions affect binding affinities. So far, this study has been more focused on how mutations alter TF binding affinities and specificities. The platform plasmid system's default environment is that of a yeast cell under standard conditions so altering the exterior environment of the yeast cells should alter the environmental conditions for TFs.

To answer the evolutionary question about how mutations in the BD and BS affect binding affinity and specificity, the group sequencing data can be used to determine which mutations affect binding affinity and specificity. Computer analysis can sort through the sequencing data to determine what randomized mutations result in a level of fluorescence.

To study binding specificity with the platform system, all possible 8-base pair and 10-base pair sequence binding sites need to be accounted. With a list of all possible BSs, the specificity of a chosen TF can be determined by observing the fluorescence that each BS variant produces. The TF has high binding specificity if only a few BSs give high fluorescence. If many BSs give high fluorescence, then that TF has low binding specificity. The exact distinction between TFs with high or low binding specificity needs to be quantified according to how many BSs result in high fluorescence.

Ideally, the fluorescence level of the yeast cell should correspond to the binding affinity of the target TF it carries. Higher binding affinity between the target TF and the BS should result

in higher fluorescence. What fluorescence levels qualify as resulting from high affinity binding reactions needs to be quantified.

Although the platform plasmid system is technically categorized as an *in vivo* method since experiments are carried out in yeast cells, there are still concerns regarding how closely the system stimulates natural conditions. This method puts TFs in an artificial context that differs from their native system albeit one that is more realistic than most *in vitro* methods. Therefore, results regarding a specific TF should be verified by studying that TF in its native system with other methods.

This thesis is coauthored with Cugini, Alexis J., Rifkin, Scott and Yang, Bing. Cugini, Alexis J. was the primary author of this paper.

References

1. L. A. Barrera, A. Vedenko, J. V. Kurland, J. M. Rogers, S. S. Gisselbrecht, E. J. Rossin, J. Woodard, L. Mariani, K. H. Kock, S. Inukai, T. Siggers, L. Shokri, R. Gordân, N. Sahni, C. Cotsapas, T. Hao, S. Yi, M. Kellis, M. J. Daly, M. Vidal, D. E. Hill, M. L. Bulyk, "Survey of variation in human transcription factors reveals prevalent DNA binding changes," Science, 351:6280, 1450-1454 (2016).

2. M. F. Berger, A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep III, M. L. Bulyk, "Compact, Universal DNA Microarrays to Comprehensively Determine Transcription-Factor Binding Site Specificities," Nature Biotechnology, 24:11, 1429–1435 (2006).

3. M. F. Berger, M. L. Bulyk, "Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors," Nature Protocols, 4:3, 393-411 (2009).

4. C. D. Carlson, C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina, A. Z. Ansari, "Specificity Landscapes of DNA Binding Molecules Elucidate Biological Function," Proceedings of the National Academy of Sciences, 107:10, 4544–4549 (2010).

5. G. A. Cary, A. M. C. Jarvela, R. D. Francolini, V. F. Hinman, "Genome-Wide Use of High- and Low-Affinity Tbrain Transcription Factor Binding Sites during Echinoderm Development." Proceedings of the National Academy of Sciences, 114:23, 5854–5861 (2017).

6. 6. R. Cohen, D. Engelberg, "Commonly Used *Saccharomyces Cerevisiae* strains (E.g. BY4741, W303) Are Growth Sensitive on Synthetic Complete Medium Due to Poor Leucine Uptake," FEMS Microbiology Letters, 273:2, 239–243 (2007).

7. K. A. Curran, A. S. Karim, A. Gupta, H. S. Alper, "Use of Expression-Enhancing Terminators in *Saccharomyces Cerevisiae* to Increase MRNA Half-Life and Improve Gene Expression Control for Metabolic Engineering Applications," Metabolic Engineering, 19, 88–97 (2013).

8. E. Firnberg, M. Ostermeier, P. Funkel, "Efficient, Expansive, User-Defined Mutagenesis," PLOS ONE, 7:12, (2012).

9. P. M. Fordyce, D. Gerber, D. Tran, J. Zheng, H. Li, J. L. DeRisi, S. R. Quake, "*De Novo* Identification and Biophysical Characterization of Transcription-Factor Binding Sites with Microfluidic Affinity Analysis," Nature Biotechnology, 28:9, 970–975 (2010).

10. M. Geertz, S. J. Maerkl, "Experimental Strategies for Studying Transcription Factor-DNA Binding Specificities," Briefings in Functional Genomics, 9:5-6, 362–373 (2010).

11. R. Gnügge, T. Liphardt, F. Rudolf, “A Shuttle Vector Series for Precise Genetic Engineering of *Saccharomyces Cerevisiae*,” Yeast, 33:3, 83–98 (2016).

12. W. J. Gong, K. G. Golic, “Ends-out, or Replacement, Gene Targeting in Drosophila,” Proceedings of the National Academy of Sciences, 100:5, 2556–2561 (2003).

13. Y. Goulev, A. Matifas, V. Heyer, B. Reina-San-Martin, G. Charvin, “COSPLAY: An Expandable Toolbox for Combinatorial and Swift Generation of Expression Plasmids in Yeast,” PLOS ONE, 14:8, (2019).

14. P. J. Hastings, C. McGill, B. Shafert, J. N. Strathem, “Ends-In vs. Ends-Out Recombination in Yeast,” Genetics, 135, 973-980 (1993).

15. P. Hou, Z. Chen, M. Ji, N. He, Z. Lu, “Real-Time PCR Assay for Ultrasensitive Quantification of DNA-Binding Proteins,” Clinical Chemistry, 53:4, 581–586 (2007).

16. D. Jantz, J. M. Berg, “Probing the DNA-Binding Affinity and Specificity of Designed Zinc Finger Proteins,” Biophysical Journal, 98:5, 852–860 (2010).

17. A. M. C. Jarvela, V. F. Hinman, “Evolution of Transcription Factor Function as a Mechanism for Changing Metazoan Developmental Gene Regulatory Networks.” EvoDevo, 6:1, (2015).

18. A. Jolma, T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, M. Taipale, J. M. Vaquerizas, J. Yan, M. J. Sillanpaa, M. Bonke, K. Palin, S. Talukder, T. R. Hughes, N. M. Luscombe, E. Ukkonen, J. Taipale, “Multiplexed Massively Parallel SELEX for Characterization of Human Transcription Factor Binding Specificities,” Genome Research, 20:6, 861–873 (2010).

19. A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, “DNA-Binding Specificities of Human Transcription Factors.” Cell, 152:1-2, 327–339 (2013).

20. L. J. Ko, J. D. Engel, “DNA-Binding Specificities of the GATA Transcription Factor Family,” Molecular and Cellular Biology, 13:7, 4011-4022 (1993).

21. J. Li, Q. Liang, W. Song, M. A. Marchisio, “Nucleotides Upstream of the Kozak Sequence Strongly Influence Gene Expression in the Yeast *S. Cerevisiae*.” Journal of Biological Engineering, 11:1, (2017).

22. X. Li, J. Lalić, P. Baeza-Centurion, R. Dhar, B. Lehner, “Changes in Gene Expression Predictably Shift and Switch Genetic Interaction,” Nature Communications, 10:1, (2019).

23. V. J. Licata, A. J. Wowor, “Applications of Fluorescence Anisotropy to the Study of Protein–DNA Interactions,” Methods in Cell Biology, 84, 243–262 (2008).

24. M. Ligr, R. Siddharthan, F. R. Cross, E. D. Siggia, “Gene Expression from Random Libraries of Yeast Promoters.” Genetics, 172:4, 2113–2122 (2006).

25. J. Majka, C. Speck, “Analysis of Protein–DNA Interactions Using Surface Plasmon Resonance,” Adv Biochem Engin/Biotechnol, 104, 13–36 (2007).

26. K. A. Matreyek, L. M. Starita, J. J. Stephany, B. Martin, M. A. Chiasson, V. E. Gray, M. Kircher, A. Khechaduri, J. N. Dines, R. J. Hause, S. Bhatia, W. E. Evans, M. V. Relling, W. Yang, J. Shendure, D. M. Fowler, “Multiplex Assessment of Protein Variant Abundance by Massively Parallel Sequencing,” Nature Genetics, 50:6, 874-882, (2018).

27. R. Nutiu, R. C. Friedman, S. Luo, I. Khrebtukova, D. Silva, R. Li, L. Zhang, G. P. Schroth, C. B. Burge, “Direct Measurement of DNA Affinity Landscapes on a High-Throughput Sequencing Instrument,” Nature Biotechnology, 29:7, 659–664 (2011).

28. D. S. M. Ottoz, F. Rudolf, J. Stelling, “Inducible, Tightly Regulated and Growth Condition-Independent Transcription Factor in *Saccharomyces Cerevisiae*,” Nucleic Acids Research, 42:17, (2014).

29. H. Redden, H. S. Alper, “The Development and Characterization of Synthetic Minimal Yeast Promoters,” Nature Communications, 6:1, (2015).

30. H. S. Rhee, B. F. Pugh, “Comprehensive Genome-Wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution,” Cell, 147:6, 1408–1419 (2011).

31. T. R. Riley, M. Slattery, N. Abe, C. Rastogi, D. Liu, R. S. Mann, H. J. Bussemaker, “SELEX-seq: A Method for Characterizing the Complete Repertoire of Binding Site Preferences for Transcription Factor Complexes,” Methods Mol Biol, 1196, 255-278 (2014).

32. W. Song, J. Li, Q. Liang, M. A. Marchisio, “Can Terminators Be Used as Insulators into Yeast Synthetic Gene Circuits?” Journal of Biological Engineering, 10:1, (2016).

33. S. Venkataram, J. C. Fay, “Is Transcription Factor Binding Site Turnover a Sufficient Explanation for Cis-Regulatory Sequence Divergence?” Genome Biol. Evol., 2, 851–858 (2010).

34. Y. Wang, L. Guo, I. Golding, E. C. Cox, N. P. Ong, “Quantitative Transcription Factor Binding Kinetics at the Single-Molecule Level,” Biophysical Journal, 96:2, 609–620 (2009).

35. J. Wang, J. Lu, Y. Liu, “*In vitro* DNA-binding profile of transcription factors: method and new insights,” Journal of Endocrinology, 210, 15-27 (2011).

36. J. Weile , S. Sun, A. G. Cote, J. Knapp, M. Verby, J. C. Mellor, Y. Wu, C. Pons, C. Wong, N. van Lieshout, F. Yang, M. Tasan, G. Tan, S. Yang, D. M. Fowler, R.

Nussbaum, J. D. Bloom, M. Vidal, D. E. Hill , P. Aloy, F. P. Roth, “A Framework for Exhaustively Mapping Functional Missense Variants,” Molecular Systems Biology, 13:12, 957, (2017).

37. E. E. Wrenbeck, J. R. Klesmith, J. A. Stapleton, A. Adeniran, K. E. J. Tyo, T. A. Whitehead, “Plasmid-Based One-Pot Saturation Mutagenesis,” Nature Methods, 13:11, 928–930, (2016).

38. A. Zykovich, I. Korf, D. J. Segal, “Bind-n-Seq: High-Throughput Analysis of *In Vitro* Protein–DNA Interactions Using Massively Parallel Sequencing,” Nucleic Acids Research, 37:22 (2009).