Linking protein function to complex traits

Michael William Dorrity

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Joseph Felsenstein, Chair

Christine Queitsch

Stanley Fields

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Linking protein function to complex traits

Michael William Dorrity

Chair of the Supervisory Committee:
Professor Joseph Felsenstein
Biology, Genome Sciences
Adjunct Professor of Computer Science and Statistics

In this dissertation, I aim to establish the utility of linking large-scale protein mutagenesis to complex trait selection in order to better understand how subtle alterations to protein function lead to shifts in trait values. I demonstrate this approach using three different complex traits in yeas and finally present a novel method to generate mutant phenotypes with native polypeptides. In the second chapter of this dissertation, I focus on the interaction between two complex traits in yeast: mating and invasion. Using exhaustive mutagenesis of Ste12, a key transcription factor that contributes to both traits, I uncover an unexpected inverse relationship between mating and invasion, and find that an altered DNA-binding mode underlies this phenomenon. In the third chapter, I investigate the relationship between cell growth rate and the heat-sensitive

trimerization of a conserved transcriptional regulator of the heat-shock response, Hsf1. With a set over 400,000 variants of Hsf1 with mutations in the trimerization domain, I examine growth rates of each variant under basal and heat-shock conditions to find key features that contribute to temperature-sensitive growth regulation. I find patterns of a trimeric helical geometry in variants that reduce function, and a role for two outer helical positions in conferring a fitness benefit under heat-shock. In the fourth chapter, I present a novel method for the identification of dominant negative polypeptides, and demonstrate its utility in yeast. I confirm the sensitivity of the method by identifying hundreds of dominant negative inhibitors derived from the yeast *URA3* and *HSF1* genes. Lastly, I show that the method can scale to large, complex libraries for unbiased selection of novel inhibitors of cell growth. In the fifth chapter, I describe a method for the identification of DNA sequences that act as enhancers of gene expression adapted for use in live plant tissues. The sensitivity of the method is confirmed using a known plant enhancer derived from a virus, and individual subsets of the enhancer are tested to find the smallest sequences with maximum enhancer activity. In the final chapter, I discuss the goals of the field of genomics-enabled protein science, and challenges that remain in translating these studies to better predict effects of genetic variants in the clinic and more broadly. I end the dissertation with a brief prospective analysis of new methods to understand protein function in light of emerging technologies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Many people have contributed to the will required to produce the work in this dissertation. I thank members of both the Genome Sciences and Biology departments for their support and camaraderie.

I thank Christine Queitsch and Stan Fields. I was very lucky to have two advisors, and luckier still that, as a pair, they represented the best of academic cooperation. Christine and Stan are collaborators in the truest sense, equal in their ability to share their excitement and to challenge each other's assumptions. I thank both of them for shaping me into a better scientist.

Both Christine and Stan have built their lab groups around people rather than projects. I benefited greatly from seeing the effectiveness of this approach during my time in the labs: exciting projects were born during light-hearted conversations and matured in the hands of motivated individuals. I thank members of both labs for their constant willingness to hear new ideas and discuss them.

I thank Joe Felsenstein, who lent me a helping hand early in graduate school, and who was constantly available for insightful discussions throughout.

I thank many friends who have made Seattle into a home. I am grateful to Melissa Lacey for an unflinching dedication to friendship and the assembly of good company. I thank Leander D.L. Anderegg for scientific discussions so "interdisciplinary" that another word ought to be used for them.

I thank my parents, who constantly inspire me to work harder.

I thank Katrina, whose encouragement kept me from getting stuck and whose insights constantly inspire me to look closer. I also thank her for keeping me active and engaged with people, art, nature, and everything else that isn't yeast.

The work presented in this thesis has been generously supported by the Washington Research Foundation, the National Institute of Health, and the National Science Foundation Graduate Research Fellowship.

DEDICATION

This work is dedicated to Kathleen Dorrity, a mother of three.

Chapter 1. INTRODUCTION

Complex traits are challenging to study because there are numerous underlying genetic factors, interactions between genes and the environment, and interactions between the traits themselves. From crop yield in plants to autism in humans, variants identified in genome-wide association studies (GWAS) explain only a small fraction of the heritable phenotypic variation, leaving a significant gap in our understanding, the so-called "missing heritability." Further, the multigenic architecture underlying complex traits makes targeted therapeutics a challenging prospect for treatment of disease. I address this challenge by making controlled modifications to the key regulators underlying complex traits and examining their phenotypic output to develop expectations for the translation of genotype to phenotype.

In this chapter, I outline the current experimental methods that exist to link subtle changes in protein sequence to phenotype. I also provide an overview of complex traits that can be studied in the model yeast *Saccharomyces cerevisiae*, and discuss mechanisms by which these traits respond to changes in temperature, a consistent theme throughout my thesis work.

1.1 HIGH-THROUGHPUT ANALYSIS OF PROTEIN FUNCTION

When I started my PhD work, new technologies had begun to expand the field of protein science to include a breadth that complemented the exceptional depth of traditional biochemical techniques. Precise measurements of binding kinetics, specific activity, and structure determination each reveal fundamental properties of protein function that be leveraged for drug discovery (Jafari et al., 2014), industrial applications (Sherif et al., 2013), and evolutionary analyses (Bloom, Labthavikul, Otey, & Arnold, 2006), but these measurements are challenging to make, time-consuming and not efficiently scaled for analysis of mixed protein variants.

Unsurprisingly, experiments that do analyze protein variants are rarely exhaustive in their exploration of mutational space, which leaves significant gaps in our understanding of the link between protein function and genotype. Genomic technologies, generally associated with broad, genome-wide analyses, were applied to address this scaling challenge: could precise measurements of protein function be taken simultaneously in a mixed population of protein variants? The key innovations that merged the tools of genomics with protein science were (1) the ability to generate large, complex mutant libraries of protein variants at a targeted locus; (2) the ability to separate functional and non-functional variants via selection on a mixed population of variants; and (3) the use of a DNA sequencer as a counting tool to measure frequencies of protein variants before and after functional selection.

A singular challenge limiting all experiments that predate the common use of high-throughput DNA sequencing is the necessity for a "positive selection" to identify variants with altered function. If a variant with altered function is rare, it can be easily identified from a large, mixed population so long as the altered function can be selected for. As an example, the yeast *CAN1* gene encodes a membrane permease that can import an amino acid analog, canavanine, that is toxic to yeast cells (Broach, James R., Strathern, Jeffrey N., Hicks, 1979). Spontaneous mutants that disrupt function of the yeast *CAN1* gene can be easily identified by growing cells on plates containing the arginine analog canavanine; colonies with robust growth will presumably represent a clonal population with a non-functional *CAN1* variant genotype. Many mutant screens have taken advantage of positive similar positive selection schemes (Boeke, La Croute, & Fink, 1984; Johnson et al., 1999).

The simplicity of positive selection is attractive and the screens effective; schemes in which altered function manifests in the occurrence of an event (viability, fluorescence, drug

resistance) allow more sensitivity, as neither background signal nor the large numbers of proteins without altered function have a large impact on the ability to detect variants with altered function. Positive selection also retains the desirable quality of facilitating analysis by dramatically reducing complexity in a library of variants: if only three mutants in a pool of one million variants is functional, then a positive selection ought to quickly reduce that population to only functional members. In contrast, negative selection depends on the ability to track altered function in a population that is otherwise functional, i.e. altered function results does not result in the occurrence of an event. Negative selection tends to be more technically challenging, as population complexity does not change greatly through the course of selection, meaning that all types must be individually tracked before and after selection to identify those with altered function. This tracking was not feasible before the common use of high-throughput sequencing, and therefore represents a largely unexplored design for large-scale, pooled experiments analyzing protein function.

1.2 Deep mutational scanning

Many of the initial examples of large-scale analyses of protein function using sequencing of mixed pools of variants were demonstrated in yeast (Araya et al., 2012; Dunham & Fowler, 2013; Fowler et al., 2010; Hietpas, Jensen, & Bolon, 2011; Melamed, Young, Gamble, Miller, & Fields, 2013; Starita et al., 2013). Yeast is an ideal model for development of novel methods to study protein function: populations are large, cells are easily transformed to produce populations of 10^6 unique genotypes, many proteins show functional conservation from human to yeast, and much is known about the biology of budding yeast (Botstein & Fink, 2011). Nearly all mutational scanning experiments follow a similar overall scheme: generating variation at a target

locus, subjecting a population of cells containing those variants to functional selection, and estimating frequencies of each genotype before and after selection using high-throughput DNA sequencing. Though I will focus on application of this method to protein function, this scheme can be applied similarly for analyzing other features encoded in DNA sequence, such as ability to promote gene expression or splicing (Bhagavatula, Rich, Young, Marin, & Fields, 2017; Rich et al., 2016). The initial application of this method, and many subsequent experiments, focused mainly on interrogating known aspects of protein function.

Selection schemes used in many deep mutational scanning experiments depend on reporters for known protein functions. For example, in an analysis of variants of the *BRCA1* gene, Starita and colleagues (Starita et al., 2015) separated functional from non-functional variants in two activities of *BRCA1*, ubiquitination, and protein-protein interaction with *BARD1*. The set of non-functional variants showed overlap in both selections, but several were specific to one selection or another, reflecting an expectation of modular protein function: protein variants unable to perform ubiquitination are not necessarily precluded from binding their protein partner, whereas variants unable to bind an essential partner cannot carry out ubiquitination. This type of experiment allows precise classification of the effect of mutation as it relates to a protein's mechanism of action. This approach is useful because those specific mechanisms have the greatest chance of being targeted through specific therapeutics.

My thesis work has focused primarily on expanding the approaches used in genomics-enabled protein science to address the role of protein variants in the expression of complex traits. Specifically, I have attempted to design selections to capture both known and unknown aspects of protein function, and classify protein variants according to their effect on a whole-organism phenotype, rather than an individual biochemical function. In doing so, I have taken advantage of

the substantial foundation for genomics-enabled protein science and the awesome power of yeast genetics.

1.3 COMPLEX TRAITS OF S. CEREVISIAE

We are more familiar with complex traits in human populations (e.g. height, lifespan, prevalence of disease) than we are with complex traits in yeast (e.g. cell shape, viability, mating). However, the traits most strongly tied with fitness in multi-cellular organisms and single-cell organisms are identical: viability and fecundity. It is not surprising, then, that the cellular pathways underlying these traits are numerous, complex in their organization, and dynamic in their response to the environment. In *S. cerevisiae*, 15-20% of all genes are essential for viability (Mnaimneh et al., 2004) and these essential genes function in myriad cellular pathways of transcription, splicing, translation, DNA replication, membrane biogenesis, nuclear transport and cytoskeleton. Similarly, 345 *S. cerevisiae* genes (6% of total genes) contribute to mating efficiency (M. Dunham, personal communication). Each of these traits has a complex genetic architecture, and therefore they represent excellent models for the study of complex traits generally.

Yeast also exhibit traits unique to their lifestyle and fungal lineage, namely the ability to adhere to and invade into surfaces. Invasion is the primary trait associated with virulence for both animal- and plant-pathogenic fungi (Gauthier & Keller, 2013; Mendgen, Hahn, & Deising, 1996). Fungal species are widely diverged at the genome sequence level, but their ability to invade is conserved, and invasion depends on similar cellular pathways that underlie mating. Further, the molecular components themselves are strongly conserved; a single transcription factor unique to the fungal lineage, Ste12, affects invasion in all tested pathogenic species, in

addition to its well-understood role in both invasion and mating in *S. cerevisiae* (Hoi & Dumas, 2010). A better understanding of the mechanisms underlying increased invasion in yeast has relevance for human health, agriculture, and pathogen evolution.

1.4 Mechanisms of temperature sensing at the protein level

Within minutes, sharp increases in temperature produce a dramatic shift in physiology and core cellular processes of transcription and translation (Gomez-Pastor, Burchfiel, & Thiele, 2018). The cell focuses all its transcription on the small proportion of genes required to guard against protein unfolding and to survive heat-shock. Translation of a specific subset of new protein, as well as specific degradation within the existing pool of proteins, is required for heat-shock survival (Lindquist, 1981; Shalgi et al., 2013). Metabolism of stress-induced sugars serve to affect viscosity of the surrounding cytoplasm, further guarding against protein unfolding (Attfield, 1997). How can a cell produce such a dramatic response in a matter of minutes? In all eukaryotic cells, a conserved response is activated by the central transcriptional regulator of the heat shock response, Heat-shock Factor 1 (Hsf1). Upon sensing heat-shock, or a variety of other proteotoxic stresses, Hsf1 trimerizes and gains an increased DNA-binding affinity along with an increased capacity for transcriptional activation. Though chaperone cofactors contribute to this rapid response, Hsf1 is exceptional in its ability to directly sense temperature and other proteotoxic stresses (Larson, Schuetz, & Kingston, 1995).

Hsf1 is not the sole protein capable of direct temperature sensing. This phenomenon has been observed in other eukaryotic proteins like transient receptor potential channels (Clapham & Miller, 2011), as well as the prokaryotic transcription factor TlpA (Hurme, Berndt, Normark, & Rhen, 1997). That many proteins unfold and become non-functional at high temperature, and that

some proteins gain function in this condition suggests that temperature resistance (or temperature sensitivity) can be encoded at the level of amino acid sequence. Indeed, the identification of temperature sensitive mutants of many essential genes confirms that temperature dependent function is not uncommon, and that it can arise from a single amino acid change (Hartwell, 1967; Hartwell & McLaughlin, 1968b, 1968a).

Many complex traits are known to be influenced by temperature. Elevated temperature alters the time and fidelity of development (Waddington, 1960), *Arabidopsis thaliana* flowers earlier at high temperature (Balasubramanian, Sureshkumar, Lempe, & Weigel, 2006), and pathogenic yeast invade more readily, but mate more poorly at high temperature (Hui Dong & Courchesne, 1998). If the expression of a complex trait is regulated by the action of an individual protein *A*, and most proteins are one or two amino acid changes away from temperature-sensitive function, then it follows that variants of protein *A* could contribute to temperature-sensitive trait expression. I have made a strong effort in my thesis work to pursue this idea by cataloging the functional effects of tens to hundreds of thousands of protein variants in both standard and high-temperature environments. A more thorough understanding of the relationship between temperature-sensitive protein function and trait expression has implications in understanding natural variation in temperature adaption, and the evolution of warm-blooded animal pathogens.

Chapter 2. DISSECTION OF CO-REGULATED TRAITS IN YEAST

Abstract

Few mechanisms are known that explain how transcription factors can adjust phenotypic outputs to accommodate differing environments. In *Saccharomyces cerevisiae*, the decision to mate or invade relies on environmental cues that converge on a shared transcription factor, Ste12. Specificity toward invasion occurs via Ste12 binding cooperatively with the cofactor Tec1. Here, we determine the range of phenotypic outputs (mating vs. invasion) of thousands of DNA-binding domain variants in Ste12 to understand how preference for invasion may arise. We find that single amino acid changes in the DNA-binding domain can shift the preference of yeast toward either mating or invasion. These mutations define two distinct regions of this domain, suggesting alternative modes of DNA binding for each trait. We characterize the DNA-binding specificity of wild-type Ste12 to identify a strong preference for spacing and orientation of both homodimeric and heterodimeric sites. Ste12 mutants that promote hyperinvasion in a Tec1-independent manner fail to bind cooperative sites with Tec1 and bind to unusual dimeric Ste12 sites composed of one near-perfect and one highly degenerate site. We propose a model in which Ste12 alone may have evolved to activate invasion genes, which could explain how preference for invasion arose in the many fungal pathogens that lack Tec1.

2.1 INTRODUCTION

Transcription factors interact with DNA, with cofactors and with signaling proteins to allow cells to respond to changes in their environment. Despite the requirement to manage these multiple levels of regulation, most eukaryotic transcription factors possess a single DNA-binding domain. Distinct responses must therefore be mediated by diversity in cofactors, organizations of binding

sites and conformational changes in the transcription factor itself. For example, human GCM1 gains a novel recognition sequence when paired with the ETS family factor ELK1 (Jolma et al., 2015); auxin-responsive transcription factors regulate expression differentially depending on the arrangement of their binding sites (Boer et al., 2014); and the heat shock factor Hsf1 senses increased temperature by changing its conformation, which allows it to bind a unique recognition sequence (Hentze, Breton, Wiesner, Kempf, & Mayer, 2016).

We sought to investigate the features of a single transcription factor that uses distinct cofactors, binding sites and environmental inputs to mediate a cellular decision. The Ste12 protein of the yeast *Saccharomyces cerevisiae* governs the choice of mating or invasion. Each of these traits contributes to cellular fitness: mating of haploid yeast cells is required for meiotic recombination, and invasion allows a cell to forage for nutrients or to penetrate tissues, a characteristic of pathogenic fungi. Mating is initiated by binding of the appropriate pheromone, which activates an evolutionarily conserved G-protein-coupled mitogen-activated protein kinase (MAPK) pathway (Cook, Bardwell, & Thorner, 1997; Mody, Weiner, & Ramanathan, 2009) (Fig. 1A). By contrast, invasion is initiated in response to increased temperature and limited nutrient availability (Beyhan, Gutierrez, Voorhies, & Sil, 2013; Hiten D. Madhani & Fink, 1998; Perfect, 2006). The shared protein kinase Ste11, a client of the chaperone Hsp90 (Louvion, Abbas-Terki, & Picard, 1998), likely contributes to the environmental sensitivity of both traits.

The two pathways converge on Ste12, which interacts differentially with cofactors to activate either mating or invasion. For mating, Ste12 can bind at pheromone-responsive genes as a homodimer or with the cofactors Mcm1 or Matα1 (Dolan, Kirkman, & Fields, 1989; Errede &

Ammerer, 1989; Yuan, Stroke, & Fields, 1993). The consensus DNA-binding site of Ste12 is TGAAAC, known as the pheromone response element (PRE) (Dolan et al., 1989). For invasion, Ste12 and its cofactor Tec1 are both required to activate genes that mediate filamentation (Chou, Lane, & Liu, 2006; Hiten D. Madhani & Fink, 1998; Zeitlinger et al., 2003). Some of these genes contain a Ste12 binding site near a Tec1 consensus sequence (TCS) of GAATGT, an organization for heterodimeric binding known as a filamentation response element (FRE). An alternative model, however, posits that expression of invasion genes is driven by a complex of Ste12 and Tec1, acting solely through Tec1 binding sites (Chou et al., 2006). An environmental component of trait specificity has been shown in fungi that are animal pathogens; upon recognition of host body temperature (37^{o}C) (Beyhan et al., 2013), *Cryptococcus neoformans* has decreased mating efficiency, but increased ability to invade (Dong & Courchesne, 1998). Although trait preference in *S. cerevisiae* depends on Ste12, the role of this protein in regulating mating and invasion in response to increased temperature is unknown.

Because the highly conserved Ste12 DNA-binding domain ultimately enacts the choice between mating and invasion, here we sought to precisely define its DNA-binding specificity and examine the consequences of mutations and increased temperature on the balance of these traits. We reveal distinct organizational preferences for both homodimeric binding sites and cooperative heterodimeric binding sites with Tec1. Furthermore, single amino acid changes suffice to shift the preference of yeast cells toward one or the other trait; they also suffice to confer dependence on the chaperone Hsp90 and responsiveness to higher temperature. Some hyperinvasive separation-of-function mutations are independent of the cofactor Tec1, thought to be essential for invasion. We show that these Ste12 variants bind to homodimeric Ste12 binding

sites in which one site is highly degenerate. Such a binding preference provides a plausible mechanism to activate the expression of invasion genes in a Tec1-independent manner, and provides a model for the regulation of these genes in the many fungal species that have no copy of *TEC1*.

2.2 RESULTS

STE12 is present in nearly all fungal genomes, but most lack a _TEC1_ orthologue.

As the target of two different MAPK signaling cascades, Ste12 acts to define pathway specificity at the level of transcription (Fig. 2.1A). This role in both mating and invasion appears to be conserved in other fungal species, and presumably derives from properties of its DNA-binding domain, the most conserved segment of the protein (Fig. 2.1B) (Hoi & Dumas, 2010). Although the Ste12 DNA-binding domain has no match outside of the fungal kingdom, several residues as well as the overall predicted secondary structure are conserved in fungi (Fig. 2.1B, 2.1C). We addressed the co-occurrence of Ste12 and Tec1 by examining 1229 fungal species for the presence or absence of these two genes. Nearly every fungal species contains a copy of *STE12* (97.7% of species), while fewer than one third (31.7%) appear to have a copy of *TEC1* (Suppl. Fig. 2.1). Furthermore, even among fungal pathogens with a characterized role for Ste12 in invasion, we found several examples in which a *TEC1* gene is not present (Fig. 2.1D). Therefore, fungal species must have evolved Tec1-independent strategies to regulate mating and invasion.

We tested the extent to which Ste12 and Tec1 binding at adjacent sites within filamentation response elements could explain invasion-specific activation by Ste12 DNA in *S. cerevisiae*.

Cooperative binding at adjacent (within 24 bp) Ste12 and Tec1 binding sites (Filamentation Response Elements, or FREs) has been demonstrated *in vivo* and *in vitro* (H. D. Madhani, 1997), although most invasion genes do not contain FREs (Chou et al., 2006). For all promoter sequences in *S. cerevisiae,* we assessed the frequency and spacing of sites, and called adjacent sites (<24 bp). We found 792 unique promoters with matches ($p < 1e^{-4}$) to a Ste12 binding site, with 220 (27.8%) of these also having a Tec1 binding site (Fig. 2.1E). The median distance between Ste12 and Tec1 sites is 360 base pairs, and only twelve pairs of sites (5%) are within 24 base pairs of each other (Fig. 2.1F). Among these twelve, four have overlapping motifs consisting of a tail-to-tail arrangement, with the Tec1 motif overlapping the 3' end of the Ste12 motif (Fig. 2.1F, inset); this organization is more likely than non-overlapping sites to occur by chance and may not be functional, though overlapping sites have been observed in cooperative binding of mammalian transcription factors (Jolma et al., 2015). The small number of sites organized for cooperative binding with Tec1 contrasts with the hundreds of genes upregulated under invasion conditions (V. M. Boer, De Winde, Pronk, & Piper, 2003; Prinz et al., 2004) as well as with the dozens of genes bound by Ste12 under invasion conditions (57 genes unique to invasion, 100 overall) (Zeitlinger et al., 2003). Thus, the model of Ste12 and Tec1 cooperatively binding to FREs within invasion genes cannot alone account for the broad transcriptional response during invasion observed in *S. cerevisiae.* Specificity may derive from instances of long-range looping interactions, or binding of Ste12 to DNA indirectly through its interaction with Tec1 bound at Tec1 consensus sequences (Chou et al., 2006). However, given the absence of Tec1 in many species, this model is also unlikely to be applicable in fungi more broadly.

Ste12 and Tec1 preferentially bind DNA in defined spacings and orientations.

Because the yeast genome contains few filamentation response elements, we sought to determine at high-resolution the *in vitro* DNA-binding preferences of Ste12 and Tec1. We used high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX), a method that determines a protein's DNA-binding specificity by isolating and sequencing bound fragments present in a large, random pool of sequences (Jolma et al., 2010, 2013, 2015). For the pool of DNA, we used fragments with a random sequence of 36 base pairs, which allowed us to capture instances of Ste12 and Tec1 binding sites in all orientations and at every possible spacing of 24 base pairs or fewer. In the Ste12 sample, we captured monomeric or homodimeric instances of the expected Ste12 binding site of TGAAAC, with dimeric sites found 20-fold more frequently than monomeric sites (Suppl. Fig. 2.2C). Many transcription factors exhibit strong preferences for spacing and orientation of sites, and this preference is conserved within transcription factor families (Jolma et al., 2013, 2015). Ste12 showed a strong preference (33% of 2,229,168 bound output sequences) for tail-to-tail binding sites with a three base pair spacer (Fig. 2.2A, top). While the most enriched sequences contained two perfect TGAAAC sites with this spacing (Fig. 2.2B, top), most of these sequences contained one perfect TGAAAC site paired with a mismatched site. This strong spacing preference had not been identified in previous protein binding microarray studies, which used shorter target sequences (Gordan et al., 2011), but is consistent with the known binding of Ste12 as a homodimer (Y. L. Yuan & Fields, 1991). We found two instances of homodimeric sites with a perfect TGAAAC pair in the *S. cerevisiae* genome: in the promoters of *GPA1*, encoding the α-subunit of the G protein involved in pheromone response, and of *STE12* itself (Fig. 2.2C). Gpa1 and Ste12 are a component of the initial signaling point after pheromone sensing and the ultimate transcription target of the MAPK signaling cascade, respectively, and both are among the most highly induced genes in response to

pheromone (Roberts et al., 2000). Elsewhere in the genome, three base pair-spaced sites in which one site is perfect and one is mismatched are found within several known pheromone-regulated genes. This configuration occurs in 29 genes in the *S. cerevisiae* genome, and 12 of the 50 most pheromone induced genes contain these sites (examples in Fig. 2.2C). To analyze the ability of these sites to drive Ste12-dependent expression, we used a reporter assay in yeast. The *in vitro* preferences for the dimeric sites, as well as for the base flanking the TGAAAC motifs, correlated with *in vivo* activation (Suppl. Fig. 2.2A, 2.2B).

HT-SELEX with Tec1 protein recapitulated the known binding site of (A/G)GAATGT (Fig. 2.2B, middle panel) (Heise et al., 2010). We found that the first base of this motif showed a strong preference for a purine base, and that the final T showed less specificity than the rest of the motif. Unlike with Ste12, we did not observe any spacing and orientation preferences for two Tec1 sites, indicating that bound sequences with multiple Tec1 binding sites are likely the result of independent binding events (Fig. 2.2A, middle panel).

We next conducted HT-SELEX in the presence of both Ste12 and Tec1 to identify patterns of cooperative binding events between the two proteins. The DNA-binding domain of Ste12 (1-215) is sufficient for cooperative binding with Tec1 (1-280) *in vitro* (Heise et al., 2010; Olson et al., 2000). We detected bound sequences containing sites for both proteins, with the highest abundance preference being a tail-to-head orientation of Tec1 and Ste12 sites separated by two base pairs (Fig. 2.2A, 2.2B, lower panel; Suppl. Fig. 2.3). Several known invasion genes in the *S. cerevisiae* genome (*GPB1, BST1, PRM6, BUD8,* and *SIT4*) contain this type of heterodimeric site, although other genes with this binding site organization (*CHS7, SRL3, TIR1, SYS1,* and

BCY1) have not been previously associated with invasion (Fig. 2.2D). This arrangement is similar to homodimeric Ste12 sites, except that one Ste12 site is replaced by a Tec1 site, suggesting that a similar protein interaction surface may be used for this heterodimeric binding mode of Ste12 as is used for the homodimeric binding mode. However, even in this sample, the pool of bound sequences was dominated by two Ste12 sites with three base pair spacing, suggesting that Ste12 has higher affinity to two PRE sites than to an FRE requiring binding with Tec1.

Mutations in the DNA-binding domain of Ste12 separate mating and invasion functions. The organization of DNA-binding sites selected by Ste12 alone or in combination with Tec1 suggested that Ste12 balances the expression of mating and invasion genes by its mode of binding to DNA. We sought to determine whether this balance could be shifted by mutations within the Ste12 DNA-binding domain. We conducted deep mutational scanning (Fowler & Fields, 2014) of a segment of this domain by generating ~20,000 protein variants over 33 amino acids, including single, double, and higher order mutants, and subjected yeast cells carrying this variant library to selection for either mating or invasion (Fig. 2.3A, 3B; Suppl. Fig. 2.4; Suppl. Fig. 2.5A). As a control, we employed selection for a third trait, response to osmotic stress, which shares upstream pathway components but does not involve Ste12. Mating selection and osmotic stress selection were carried out in the BY4741 strain background. However, as this strain contains a *flo8* mutation that prevents invasion, we carried out invasion selection in a related strain, Sigma1278b (Winzeler, 1999), which has been used for large-scale invasion phenotyping (Ryan et al., 2012). Although Sigma1278b and BY4741 have many genetic

differences, Ste12 and Tec1 and their essential roles in balancing mating and invasion are conserved between the two strains.

We expected that most *STE12* mutations would affect mating and invasion similarly, because of the conservation of the Ste12 DNA-binding domain and its requirement for both traits. Indeed, we found positions in which almost any amino acid substitution was highly deleterious to both mating and invasion (Fig. 2.3B, 3C; Suppl. Fig. 2.5). We identified sites that were more sensitive or less sensitive, on average, to mutation, by calculating a positional mean score from all mutations tested at that site. Each mutation was tested in triplicate, and the experimental error was calculated for each mutant individually such that standard error of the positional mean represents the variability among different amino acid substitutions, rather than experimental noise (Fig. 2.3C, Suppl. Fig. 2.4E). Conservation only partially explained these results, as some of the most deleterious positions are invariant among fungi (W156, C144), whereas others are not (K149, Q151, K152). Differing expression levels among *STE12* variants were not predictive for either mating or invasion phenotypes (Suppl. Fig. 2.6). Since invasion was measured in a different strain background than mating for each variant, we verified that strain background did not account for the trait differences. We selected five variants that conferred differential mating and invasion phenotypes (K149E; K150I; K150A; K152L; and (see below) the double mutant S177K, Q180R, denoted SKQR), and assayed their mating efficiencies in the Sigma1278b background. The mating efficiencies for each variant correlated (r^2=0.7) between strains, with the average difference for the five variants equal to 21% (Suppl. Fig. 2.6D).

The mutational analysis revealed separation-of-function mutations that primarily reduced either mating or invasion. Substitutions with mostly deleterious effects on mating clustered in residues N-terminal to the conserved W156 (Fisher's exact test p-value = 0.002, designated region I), while substitutions with increased effects on invasion appeared more frequently in this region (p-value = 0.0003). Substitutions that increased mating were rare. By contrast, substitutions at some positions, almost all within region I, both increased invasion and decreased mating. We verified separation-of-function mutants, and further explored the apparent structure of mutational effects, with a much larger pool of 20,000 double mutants. Double mutants involving two positions in region I were mostly defective for mating, while those involving two positions in region II were mostly defective for invasion, confirming the bipartite arrangement of the mutagenized segment, split by the central W156 (p-value = 1.3e-5, Fig. 2.4A). Double mutants between regions I and II showed variability in mutational effects, with some pairwise combinations increasing mating and others decreasing mating; the effects on invasion, however, for these same combinations did not follow the same pattern, suggesting epistatic interactions (examples highlighted in Fig. 2.4A, Suppl. Fig. 2.7). Thus, mutations in the Ste12 DNA-binding domain can impose preference for mating or invasion rather than similarly affecting both traits, suggesting that the DNA-binding domain itself contributes to trait specificity.

To explore the apparent tradeoff between mating and invasion in more detail, we used the Pareto front concept (Shoval, 2012). This concept, rooted in engineering and economics, defines all feasible solutions for optimizing performance in two essential tasks. Here, we consider all feasible genotypes that affect mating or invasion. We plotted the single mutation mean positional scores for both traits, identifying positions close to the Pareto front that show increased invasion

and deleterious effects on mating. (Fig. 2.4B, Suppl. Fig. 2.8). We found that this effect is recapitulated in the effects of individual mutants tested in both trait selections at extreme positions on the front (Suppl. Fig. 2.9). Positions well below the Pareto front, including most prominently W156, had deleterious effects on both traits. These positions are significantly more conserved among fungi, consistent with variation at these sites being disfavored given the constraint on Ste12 to maintain both mating and invasion function in the fungal lineage (Fig. 2.4C). That opposing preference for each trait is better predicted by positional means than every individual mutation suggests that trait preference is encoded in particular functional regions of the Ste12 DNA-binding domain rather than through overall features like protein stability.

Having established that trait preference can be modulated by mutations in Ste12, we asked whether temperature affected specificity toward mating and invasion, and if that specificity could be altered by mutation. Furthermore, since chaperones maintain protein function at increased temperatures and Hsp90 modulates pheromone signaling in *S. cerevisiae* (Louvion et al., 1998), we also asked whether mating or invasion changed in the presence of radicicol, a pharmacological inhibitor of Hsp90. We confirmed that mating is modulated by temperature and by Hsp90 function by mating cells expressing wild-type Ste12 at increased temperature or in the presence of radicicol (Fig. 2.4D). We then subjected cells containing the Ste12 variant library to a mating selection at 37°C or a mating selection in the presence of radicicol. Most Ste12 variants responded to increased temperature or Hsp90 inhibition as did wild-type Ste12 (Suppl. Fig. 2.11). However, there were two positions, K150 and K152, in which mutations resulted in highly temperature-responsive and Hsp90-dependent mating (Fig. 2.4E). This pair of lysines resides within the mutagenized segment that modulates mating and invasion specificity. We validated

the temperature and Hsp90 effect on a variant (K150I) that conferred mating at near wild-type levels in the absence of heat or radicicol treatment (Fig. 2.4F). However, in the presence of heat or radicicol treatment, mating of cells with the K150I variant was severely decreased, indicating that this mutation led to buffering by Hsp90. Thus, Hsp90 could facilitate a mutational path toward the pathogenic lifestyle by minimizing mating costs at 30°C and enhancing invasion at 37°C. To test this idea, we conducted selection for invasion at 37°C, which yielded results comparable to Hsp90 inhibition for mating. Indeed, the mean effect of all variants at K150 was to increase invasion at high temperature with a concomitant decrease in mating, and this was also true of the individually validated variant K150I (Fig. 2.4G), demonstrating that K150 variants were not simply unstable at high temperature but gained a novel function. Thus, mutations in *STE12* can interact with an environmental factor to further bias cellular decision-making toward invasion over mating. Mutation of *STE12* could allow *S. cerevisiae* to mimic the behavior of fungal pathogens like *Cryptococcus neoformans*, for which the sensing of the increased body temperature of their animal hosts facilitates the transition toward an invasive lifestyle (Perfect, 2006).

We examined whether natural variation in fungal Ste12 DNA-binding domains includes amino acid residues found to affect *S. cerevisiae* trait preference, as fungal species have differential capacities for mating and invasion (Hoi & Dumas, 2010; Jones & Bennett, 2011). We examined variation in the DNA-binding domain present in species with and without a *TEC1* gene (Suppl. Fig. 2.12) and detected variation outside of the mutagenized region that might be expected to benefit invasion. For example, the Ste12 DNA-binding domain of the invasive pathogen *Cryptococcus gattii* (Fraser et al., 2005) contains two additional positive residues immediately C-

terminal to region II; such changes are found in many other species that lack Tec1 (Suppl. Fig. 2.12). Introducing these *C. gattii*-specific residues (S177K, Q180R; denoted SKQR) into the *S. cerevisiae* Ste12 DNA-binding domain yielded a dominant invasion phenotype (Fig. 2.5A) with neutral effects on mating (Suppl. Fig. 2.6D) in *S. cerevisiae*. We introduced a negatively charged residue (K175E) into this same region and abolished invasion entirely (Fig. 2.5A). Similar to SKQR, the region I mutation K150A conferred decreased mating and an increase in invasion. The dominant invasion phenotype for each variant suggest that these Ste12 variants engage in altered binding to DNA, altered homodimerization, or altered heterodimerization with Tec1.

Mutations in Ste12 can promote Tec1-independent invasion and alter DNA-binding specificity.

Tec1 activates invasion genes, with Ste12 either binding directly to DNA cooperatively with Tec1 or indirectly as part of a complex with Tec1 (Chou et al., 2006; Zeitlinger et al., 2003). However, in contrast to both established Ste12 binding modes, the invasion phenotype due to the SKQR variant was independent of Tec1, as was the even stronger invasion phenotype due to the K150A variant (Fig. 2.5A). Thus, a single mutation in the Ste12 DNA-binding domain might change the binding preference of this domain such that it can be recruited in the absence of Tec1 to binding sites sufficient for invasion.

To understand the relationship between the phenotypes conferred by Ste12 variants and their DNA-binding specificity, we conducted HT-SELEX experiments. We chose the SKQR and K150A variants, each of which showed an altered invasion phenotype, and asked whether they

could recognize the homodimeric Ste12 sites favored by the wild-type protein. SKQR showed a greatly reduced (by 45%) preference for two sites separated by three base pairs (Fig. 2.5B). Consistent with this reduced preference, SKQR conferred reduced activation *in vivo* from a minimal promoter containing such homodimeric sites (Suppl. Fig. 2.13A). Furthermore, an examination of the most enriched sequences bound *in vitro* by the SKQR variant showed that it lost specificity for the sequence of the second site, which was also reflected in its *in vivo* sequence preference (Fig. 2.5D, Suppl. Fig. 2.13B). In contrast, K150A retained a strong preference for the three base-pair spaced sites, nearly identical to wild-type Ste12 (Fig. 2.5B). However, analysis of its most enriched sequences revealed that, like SKQR, the K150A variant showed reduced specificity for one site of the homodimeric pair (Fig. 2.5D). The reduced specificity in these variants should expand the number of their recognition sites. Because both of these variants recognized at least one perfect, or near perfect, Ste12 site, we conclude that the primary change to DNA-binding specificity in these variants derives not from loss of specific DNA contacts, but rather from their diminished capacity for symmetric dimeric binding.

We also carried out co-binding SELEX experiments with Tec1, as previously conducted with wild-type Ste12 (reproduced in Figure 2.5C, top). Consistent with their ability to promote Tec1-independent invasion, both SKQR and K150A showed altered patterns of cooperative binding with Tec1. K150A showed no preference for the two base pair-spaced heterodimeric sites favored by the wild-type Ste12 in co-binding experiments with Tec1, and SKQR had a greatly reduced preference for these motifs (Figure 2.5C).

2.3 DISCUSSION

Transcription factors maintain their capacity to enact complex regulatory programs over the course of evolution, despite many changes to the genomic locations of their binding sites and to the repertoires of their available cofactors. Nonetheless, by analyzing thousands of mutations in a yeast transcription factor, we found variants that can mediate only one of two alternative programs, and do so in the absence of a requisite cofactor. These variants of the *S. cerevisiae* Ste12 protein have single amino acid changes that lead to DNA binding preferences dramatically different from the wild-type protein, as well to an increased ability of yeast to invade, rather than mate, at high temperature. They thus reveal how natural variation in the DNA-binding domain of a transcription factor might affect an environmentally-sensitive decision to carry out one developmental process rather than another.

In *S. cerevisiae,* Ste12 drives expression of mating genes as a homodimer and of invasion genes as a heterodimer with the invasion cofactor Tec1. The single mutations in Ste12 that shift trait preference towards invasion in a dominant and Tec1-independent manner provide a plausible model (Figure 2.6) that can explain how fungal species, including many pathogens, without Tec1 or similar cofactors accomplish activation of invasion genes. In mating genes, two properly oriented binding sites facilitate the interaction of two Ste12 proteins; one site tends to be perfect and the other slightly mismatched, thereby likely reducing cooperativity. Under invasion conditions, the Ste12 dimer interface instead allows cooperative interaction between Ste12 and Tec1. We posit that mutations in Ste12 that lead to hyperinvasiveness change this interface such that the conformation of the homodimer is altered and interaction with Tec1 is not possible. In this altered conformation, which may be naturally present in species without Tec1, Ste12 loses

its preference for the perfect/near-perfect homodimeric binding sites and instead recognizes a degenerate version of its binding site adjacent to a near-perfect site. The binding mode of these Ste12 mutants exemplifies how transcription factors may access novel binding sites by changes in their recognition of preferred dimeric sites rather than of individual site specificity.

Mutations at positions in Ste12 implicated in shifting trait preference toward invasion also conferred dependence on temperature and the chaperone Hsp90. For some fungal pathogens, increased temperatures associated with warm blooded animals promotes invasive growth (Perfect, 2006). In the non-pathogenic *S. cerevisiae*, increased temperature decreases mating and promotes invasion (Perfect, 2006). The conserved role of Ste12 in invasion and mating, which can be modified by mutations that affect the environmental sensitivity of this trait choice, suggests that variation in Ste12 may potentiate pathogenic transformation in other species. The temperature- and Hsp90-dependent Ste12 variants do not resemble typical temperature-sensitive mutants in simply losing function under non-permissive conditions. In response to high temperature or Hsp90 inhibition, these variants failed to promote mating yet conferred an increased ability to invade. These variants thus represent the unusual phenomenon whereby perturbation of Hsp90 or equivalent environmental stress results in a protein gaining a novel function (Jarosz, Taipale, & Lindquist, 2010; Sangster, Lindquist, & Queitsch, 2004).

As the ultimate mediators of cellular decisions, transcription factors occupy a predominant position in the signaling networks that drive patterns of gene expression. For this reason, they are susceptible to even subtle mutations that may result in consequential phenotypes. Such effects will be less predictable than the full loss-of-function phenotype resulting from, for example, a

failure to bind to DNA. Thus, the type of detailed analysis of missense variants as performed here will be useful in understanding the shifts in phenotype that enable adaptation to novel environments.

Mutations in transcription factors that generate alternative DNA-binding modes have been previously identified. For example, rare coding variants in the homeodomain recognition helices of several transcription factors alter DNA-binding specificity, which likely contributes to associated disease (Barrera et al., 2016). These homeodomain variants bind at sites not dramatically different from the wild-type sites, yet this promiscuity is associated with a deleterious disease phenotype. The demonstration here of a surprising malleability of transcription factor binding stands in stark contrast to the fact that sequences of transcription factors tend to be highly conserved. This study, together with those revealing only weak purifying selection in functional regulatory regions, poses the challenge of reconciling flexible transcription factor activity and often highly variable regulatory regions with the longstanding observation that gene expression patterns are robust to most perturbations and conserved throughout evolution.

2.4 METHODS

HT-SELEX. We expressed and purified fragments of *STE12* (1-215) and *TEC1* (1-250) cloned into pGEX-4T-2 vectors. These protein fragments have been used previously and are sufficient for both individual cooperative binding *in vitro.* Fragments of each protein, as well as protein variants, were purified using a GST tag and subsequently used for HT-SELEX. SELEX reactions with homogenous and mixed protein populations were performed identically to previous work

(Jolma et al., 2010, 2015). Briefly, a 50uL reaction containing purified Ste12 and Tec1 (1:25 molar ratio with DNA), 200ng non-specific competitor double-stranded nucleic acid poly (dI/dC), 100ng selection ligand (36N) was incubated in binding buffer (140 mM KCl, 5 mM NaCl, 1 mM K_2HPO_4, 2 mM $MgSO_4$, 20 mM HEPES [pH 7.05], 100 μM EGTA, 1 μM $ZnSO_4$) for 2 hours. GST Sepharose (GE) beads were then added to each reaction, incubated for 30 minutes, and unbound ligand was removed using seven buffer washes. Output reactions were amplified by PCR after each round, and these products were subsequently used to prep high-throughput sequencing libraries. SELEX motif enrichments were analyzed using Autoseed software (1). Briefly, a pool of binding-selected output sequences was compared against a fully random input sequence to identify sequences, motifs and orientations enriched relative to the unselected oligo pool.

Generation of *STE12* mutant libraries. The *STE12* locus from *Saccharomyces cerevisiae* strain BY4741, including the intergenic regions, was introduced into the yeast vector pRS415 containing a *LEU2* marker (Mumberg, Müller, & Funk, 1995). Degenerate DNA sequence encoding a 33 amino acid (99bp) segment of Ste12's DNA binding domain was generated by 2.5% doped oligonucleotide synthesis (Trilink Biotechnologies, San Diego, CA). Invariant 30bp sequences were designed on either side of the mutagenized fragment; these flanking sequences contained NotI and ApaI cut sites found in the coding sequence of *STE12*, and were unique in the *STE12* plasmid construct. Both fragment and plasmid were double-digested with NotI and ApaI, and the plasmid library was assembled by standard ligation. *STE12* libraries were transformed into electrocompetent *E. coli* (ElectroMAX DH10B, Invitrogen), and amplified overnight in selective media. Efficiency of ligation was verified by Sanger sequencing across the

mutagenized region of 96 transformants (Sanger, Nicklen, & Coulson, 1977); no assembly errors were detected, and mutant proportions reflected those expected for doped oligo synthesis at 2.5%. Plasmid libraries were used to transform yeast (BY4741 MATa or Σ1278b-α) with a deleted endogenous copy of *STE12* by high-efficiency lithium acetate transformation (Gietz & Woods, 2002). The same plasmid library was used to transform yeast (Σ1278b-α) for invasion selection. Individual point mutations were generated in wild-type *STE12* plasmids by site-directed mutagenesis (Q5, New England Biolabs).

Large-scale trait selection. The BY4742 MATα strain was used as the mating partner for the library-transformed BY4741 MATa in all selections and mating assays. Transformed yeast cells were grown to late log-phase in a single 500mL culture, and cells were harvested to determine plasmid variant frequencies in the input population. The same culture was used to seed 36 independent mating selections for each treatment: one million MATa cells with *STE12* variants were mixed with 10-fold excess wild-type MATα cells and allowed 5 hours to mate (Leu & Murray, 2006). Depending on treatment type, cell mixtures were left at 30C with DMSO, 30°C with the Hsp90 inhibitor radicicol, or 37°C with DMSO. A 5uM concentration of the Hsp90 inhibitor radicicol (Sigma-Aldrich, R2146) was chosen due to its measureable effect on mating efficiency and lack of pleiotropic growth defects. Radicicol was chosen over the Hsp90 inhibitor geldanamycin (Sigma-Aldrich, G3381), because five-fold higher concentrations of geldanamycin were required to achieve the same phenotypic effect as with radicicol (data not shown). The temperature of 37°C was chosen for the similarity of effects on mating between temperature and radicicol treatments. After mating was completed, cell mixtures were plated using auxotrophic markers present only in mated diploids. Plasmids containing *STE12* variants were extracted from

this output population, as well as from the pre-mating input, for subsequent deep sequencing. Mating selection was repeated in triplicate. For invasion, Σ1278b-α yeast cells transformed with the Ste12 plasmid library were grown to late log-phase in a single 500mL culture, diluted (10,000 cells per plate), and plated onto 40 plates of synthetic complete medium lacking leucine (2% agar). Plates were incubated at 30°C or 37°C for 72 hours to allow for sufficient invasion, as previously described (Ryan et al., 2012). After three days, cells were washed from the plate surfaces, enriching for cells embedded in the agar. Agar pucks were removed from plates with a razor. Using the "salsa" blender setting (Hamilton Beach, Glen Allen, VA), a coarse slurry was generated and subsequently poured over a vacuum apparatus lined with cheesecloth. The resulting liquid cell suspension was spun down at 5000rpm to collect cells for subsequent deep sequencing. Individual invasion assays (as in Fig. 2.5A) were treated identically, but 10uL aliquots of OD-normalized cultures were plated to image colonies for each strain. For high osmolarity growth, selections were conducted using the BY4741 MATa library-transformed population grown overnight in media containing 1.5M Sorbitol, as described previously (Posas, Witten, & Saito, 1998). Populations were sequenced before and after growth to determine enrichment scores that defined the 95% confidence interval used in Fig. 2.3C. For all trait selections, we chose sample sizes that were at least 10-fold higher than the variant library size, ensuring that each variant would be adequately sampled in each of the three biological replicate selections.

Sequencing and determination of trait scores. Sequencing was completed on Illumina's MiSeq or NextSeq platforms. Sequencing libraries were prepared by extracting plasmids from yeast populations (Yeast Plasmid Miniprep II, Zymo Research, Irvine, CA) before and after

selection. These plasmids were used as template for PCR amplification that added adaptor sequences and 8bp sample indexes to the 99bp mutagenized region for sequencing (all libraries amplified < 15 cycles). Paired-end reads spanning the mutagenized region were filtered to obtain a median of 5 million reads per sample. Using ENRICH software (Rubin et al., 2016), read counts for each variant before and after selection were used to determine mating efficiency and invasion ability of *STE12* variants. Briefly, counts for a particular variant in the input and output libraries were normalized by their respective read totals to determine frequency in each, and a ratio of the output and input frequencies determine a variant's functional score. Finally, enrichment scores are normalized by the enrichment of wild-type Ste12 in each selection experiment. Treatment scores are calculated identically, and in all cases where difference scores are shown, the score represents log2(treated) – log2(untreated).

Calculating intramolecular epistasis scores. Intramolecular epistasis scores were defined as the deviation of double-mutant's functional score (W_{ij}) from the multiplied scores of its constituent single-mutants (w_i* w_j). A negative epistasis score indicates that the deleterious effect of one mutation is increased by the presence of the partner mutation, while a positive epistasis score indicates that the partner mutation decreases the deleterious effects of individual mutations (Araya et al., 2012).

Quantitative mating assay. Individual variants tested for mating efficiency were treated identically to the large-scale mating selection experiments, except genotypes were scored individually on selective plates for either mated diploids (2N) or both diploids and unmated MATa haploids (2N, 1N). The proportion of mated individuals (mating efficiency) is taken as the

ratio of colony counts on diploid to counts on diploid + haploid plates (2N/2N1N). Ste12 variants were always tested alongside wild-type to determine relative changes in response to treatment.

STE12 variant RNA-seq. RNA was extracted from yeast cells harboring the *STE12* mutant library grown under non selective conditions using acid phenol extraction as previously described (Cuperus, Lo, Shumaker, Proctor, & Fields, 2015). *STE12*-specific cDNA was created using a gene-specific cDNA primer and Superscript III (Life Technologies). cDNA was amplified in a manner similar to the plasmids, and prepared for sequencing using Illumina Nextseq.

Large-scale analysis of Ste12 binding sites *in vivo*. A *HIS3* reporter gene was used for testing large populations of binding site variants (Suppl. Fig. 2.13) in media lacking histidine. Although Ste12 does bind single PREs as a monomer, two sites are needed for signal detection in reporter assays (Su, Tamarkina, & Sadowski, 2010). We designed oligonucleotides that maintain the central portion of the native PRE, but randomized six surrounding bases on either side (NNNNNNTTTCAAAATGAAANNNNNN). This library was cloned into a promoter from the same plasmid used in the luciferase assay, at the same position as the native PRE. Strains containing one of four Ste12 protein variants were transformed with the same binding site library reporter population, and grown overnight in synthetic media lacking histidine with 10mM of the His3 competitive inhibitor 3-amino-triazole. Three biological replicate selections were conducted for each Ste12 protein variant tested against the binding site library. Cells were collected and sequenced before and after selections to determine enrichment scores for each binding site

variant. Computational analysis of binding site enrichment scores was identical to pipeline for STE12 protein variants. Enrichment of all binding site variants are shown relative to the enrichments of empty plasmids, which were spiked-in to the binding site library as control. The 250 binding sites with the highest enrichments were grouped and Weblogo (Crooks, Hon, Chandonia, & Brenner, 2004) was used to generate base preference plots.

Evolutionary analysis of Ste12 DNA-binding domain among fungi. Using fungal genomes deposited into NCBI and from the fungal sequencing project at Joint Genome Institute, we created a BLAST database of translated coding sequences and queried with the *S. cerevisiae* Ste12 protein sequence. Full protein sequences from BLAST hits were aligned using MUSCLE (Edgar, 2004) and used to determine conservation at each site in the mutated segment. The secondary structure of *S. cerevisiae*'s Ste12 DNA-binding domain was determined using Psipred (McGuffin, Bryson, & Jones, 2000).

Analysis of Ste12 ChIP data. Genes bound by Ste12 only in mating conditions were compared to those bound by Ste12 only in invasion conditions (Zeitlinger et al., 2003). FIMO(Grant, Bailey, & Noble, 2011) was used to extract all matches to the Ste12 binding site in each of these gene sets. The frequency of the each possible base at position 6 of the core STE12 motif was determined relative to the genomic background (STE12 binding sites in all upstream sequences, no separation by mating or invasion gene function). The core 6 base frequencies relative to genomic background was then compared for genes bound in mating or invasion conditions. We used the same method to look at core 6 base frequencies at genes bound by both Ste12 and Tec1 or those bound by Ste12 alone during invasion.

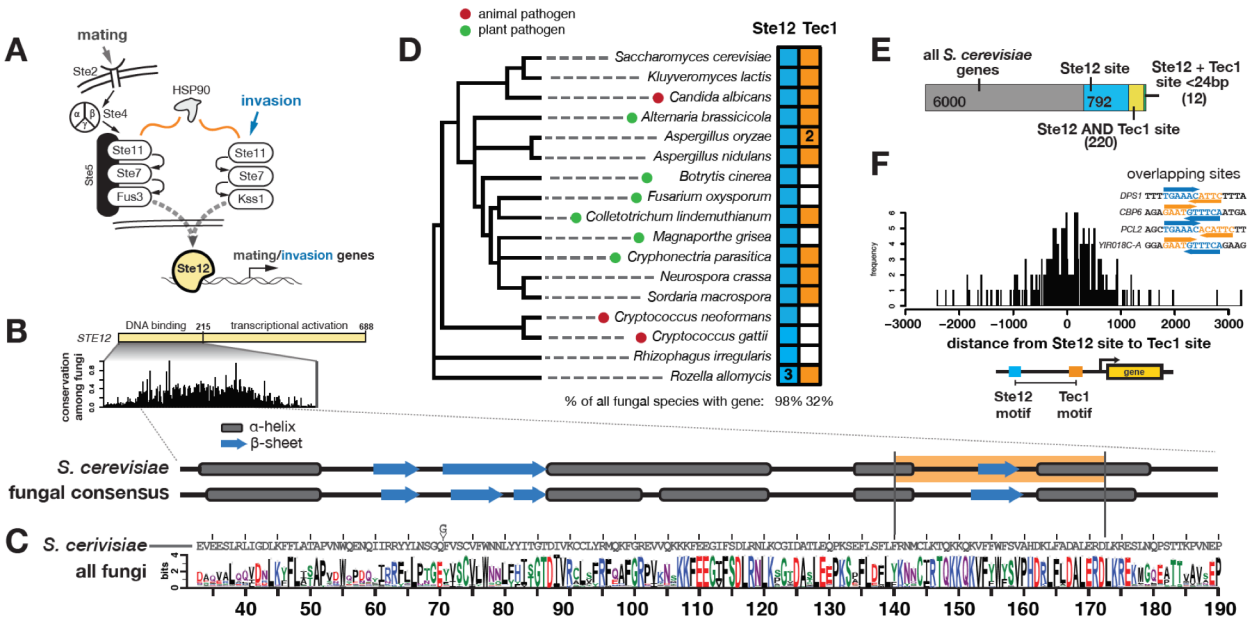


Figure 2.1

Conservation of Ste12 and Tec1 and their DNA-binding sites. (A) The yeast mating and invasion pathways contain shared signaling components, and both depend on Ste12 for activation of distinct regulatory programs. (B*) S. cerevisiae* Ste12 protein has a non-conserved transcriptional activation domain, as well as a highly conserved DNA-binding domain. The secondary structure of this domain is predicted to contain a pattern of alpha-helices (grey boxes) interspersed with β-sheets (blue arrows) that are conserved among all fungal species. The region of the DNA-binding domain chosen for mutagenesis is shaded in orange. (C) A logo plot of the conserved portion of Ste12's DNA-binding domain generated from 1229 fungal species. (D) A phylogeny of fungal species selected for those with functionally tested *STE12* genes (except for the basal fungi *R. irregularis* and *R. allomycis* shown as outgroups). Pathogenic species are indicated with circles, and colored according to plant (green) or animal (red) hosts. All fungal species were queried for presence of a *STE12* (blue) or *TEC1* (orange) gene, and filled squares indicate presence of either gene, numbers inside boxes indicate species with multiple gene copies. Results for all species are shown below each gene's column. (E) Proportional bar chart showing the number of S. cerevisiae genes that contain Ste12 sites (blue), Tec1 and Ste12 sites (yellow), and adjacent sites less than 24 bases from each other (green). (F) Histogram showing

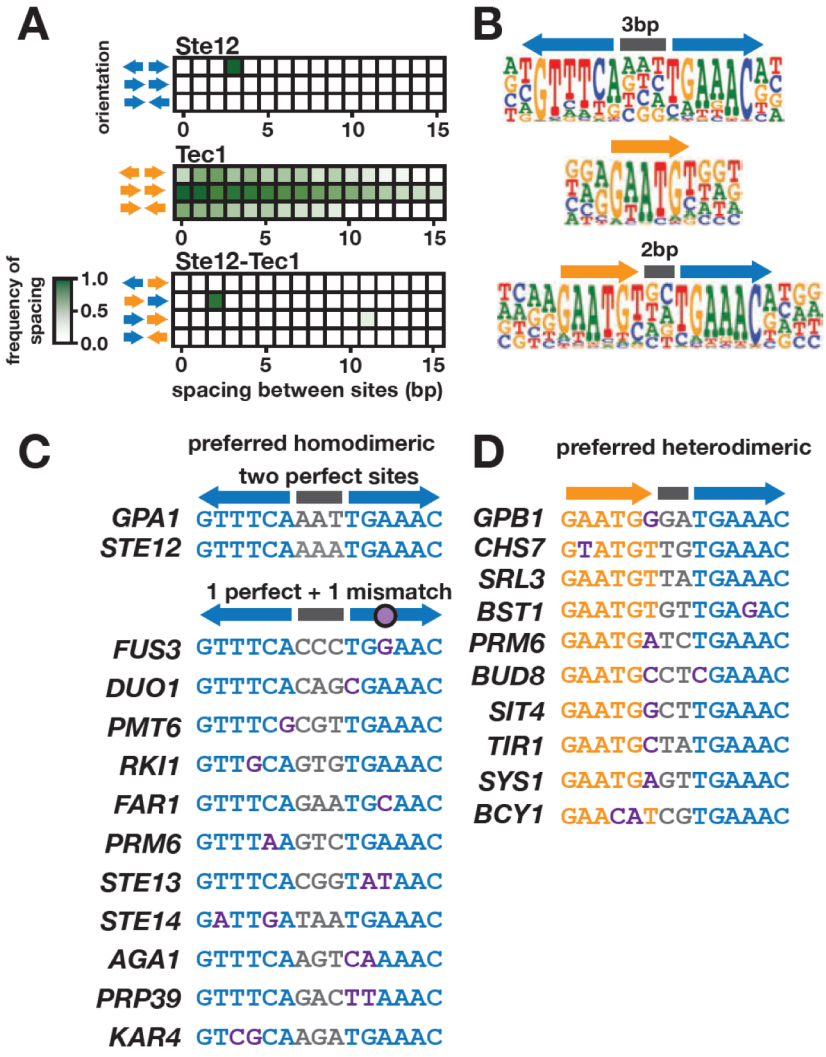


Figure 2.2

Identification of the DNA-binding preferences of Ste12 and Tec1 by HT-SELEX. (A) Heatmap showing relative frequencies of the possible orientations and spacings of the primary 6-mer selected in Ste12 (TGAAAC, upper) and Tec1 (GAATGT, middle) binding reactions. The bottom heatmap shows the frequency of each respective 6-mer in the co-binding sample containing both proteins. In the co-binding sample, we excluded sequences with homodimeric Ste12 sites, which were present at a 30-fold higher level than heterodimeric sites. The single dark green box in the upper and bottom heatmaps indicates the most frequent orientation and spacing of sites; white boxes are at most 30% as frequent as the maximum. No frequent dimeric site organizations were observed for Tec1 (middle heat map). (B) Full motifs identified by Autoseed software (1). (C) Instances of the Ste12 homodimeric tail-to-tail, 3 base-pair spaced motif were used to query native yeast promoters. Two pheromone-induced genes, *GPA1* and *STE12*, contain two perfect sites, while most other genes contain a perfect site paired with a site containing one or two mismatches, with the 3 base-pair spacing intact. (D) A similar search of yeast promoters using the preferred Ste12 and Tec1 heterodimeric site identified a set of invasion-associated genes, as well as those not previously linked to invasion.

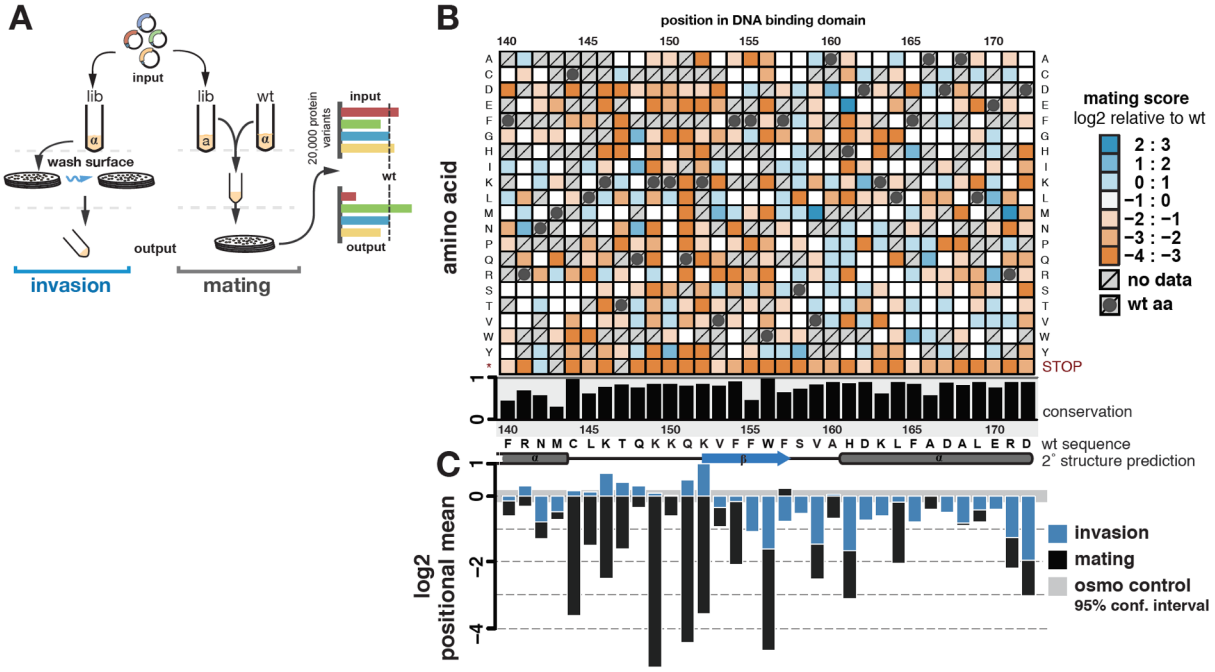


Figure 2.3

Deep mutational scanning identifies Ste12 DNA-binding domain mutants with altered mating and invasion function. (A) Two yeast populations were transformed with the same *STE12* variant library. For assaying invasion, SIGMA1278b *MAT*α cells (n = 400,000 per replicate, 3 biological replicates) were plated on selective media, and grown. Selection was performed by washing colonies from plate surfaces and collecting cells embedded in the agar for sequencing. For assaying mating, BY4741 *MATa ste12Δ* cells were mated to *MAT*α cells, and diploids (n = 500,000 per replicate, 3 biological replicates) were selected, scraped from the agar, and sequenced. In both cases, input variant frequencies were defined prior to selection. (B) The effects of single amino acid substitutions within Ste12's DNA-binding domain on mating ability are shown. On the x-axis, the wild-type Ste12 sequence is shown, along with its predicted secondary structure (helices shown as tubes) and conservation. Conservation was determined as fraction of identity among 1229 fungal species. On the y-axis, amino acid substitutions are shown. Variants increasing mating efficiency are in shades of blue, and variants decreasing mating efficiency are in shades of orange based on log2 enrichment scores relative to wild-type. Dark grey circles indicate the wild-type Ste12 residue, and crosses indicate missing data. Ste12 variants showed comparable expression levels (Suppl. Fig. 2.6). (C) Positional mean scores for single amino acid substitutions are shown for mating (black) and invasion (blue), excluding stop codons. Grey horizontal bar indicates confidence interval for experimental noise determined from selection for Ste12-independent high osmolarity growth.

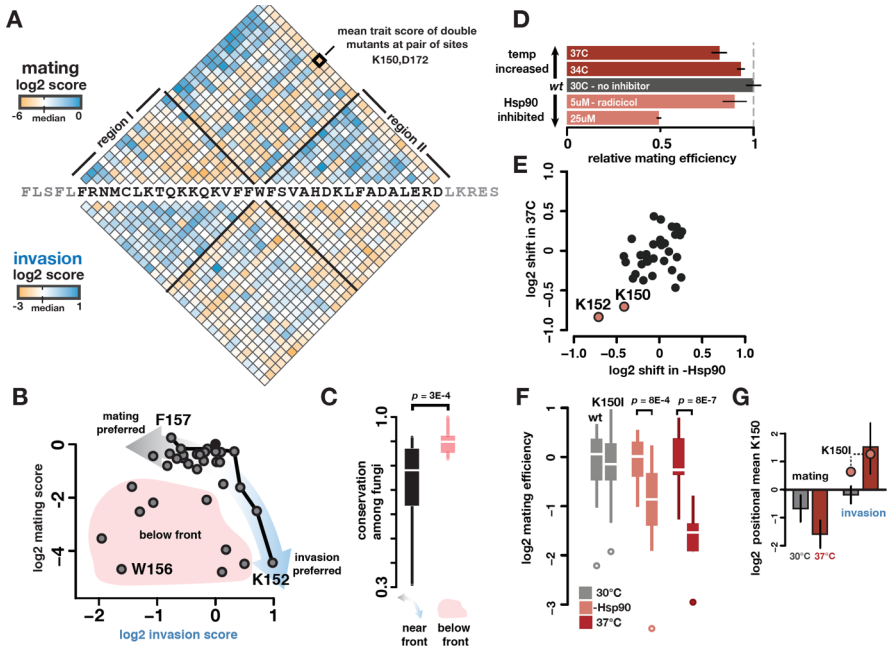


Figure 2.4

Mutations residing in a region of the Ste12 DNA-binding domain increase invasion at the cost of mating, with exceptional mutations doing so in a temperature- and Hsp90-dependent manner. (A) Mean effect of double mutations between all combinations of positions for mating (above wild-type sequence) and invasion (below wild-type sequence). A specific pair of positions is shown in bold. Effects are shown as log2 fold change relative to wild-type and color-coded as in Figure 3, functional variants in shades blue, nonfunctional variants are shown in shades of orange. Note difference in scale ranges between mating and invasion. Black lines emanating from W156 represent boundaries of region I, in which mutations primarily reduced mating, and region II, in which mutations primarily reduced invasion. (B) Scatterplot of positional mean scores for both mating and invasion showed inverse relationship, indicating a tradeoff between both traits. A point is shown for each of the 33 mutagenized positions of *STE12*. The tradeoff is visualized as a Pareto front (black line) determined empirically from the positional data; positions near the front maintain high values for one trait and minimize costs to the other. Arrows indicate preference for either mating (grey) or invasion (blue). Positions near the front were distinguished from those below the front (shaded in red) by calculating Euclidian distances. (C) Boxplots show conservation for positions near and below the front; positions below the front are significantly more conserved among fungi than those near the front (two-sided t-test). (D) Mating efficiency of yeast cells with wild type Ste12 at increased temperature (dark red) or in the presence of Hsp90 inhibitor radicicol (pale red) is reduced relative to an untreated control (black), error bars represent standard error of the mean (s.e.m). (E) The mating efficiency of Ste12 variants at high temperature or with Hsp90 inhibition is shown as the shift in mean effect at each mutated position. Two positions, K150 and K152 (pale red), showed greatest sensitivity to both treatments. (F) K150I was tested in a quantitative assay to validate its temperature-sensitive and Hsp90-dependent mating activity, shown relative to wild type (left boxplot in each pair) in each treatment (n = 20 for each sample). (G) Ste12 variants at K150 (mean effect) decreased mating (left panel) and invasion (right panel) at standard temperature (grey bars). At high temperature (red bars), mating further decreased; however, invasion increased. Error bars represent s.e.m. K150I variant is individually highlighted for comparison.

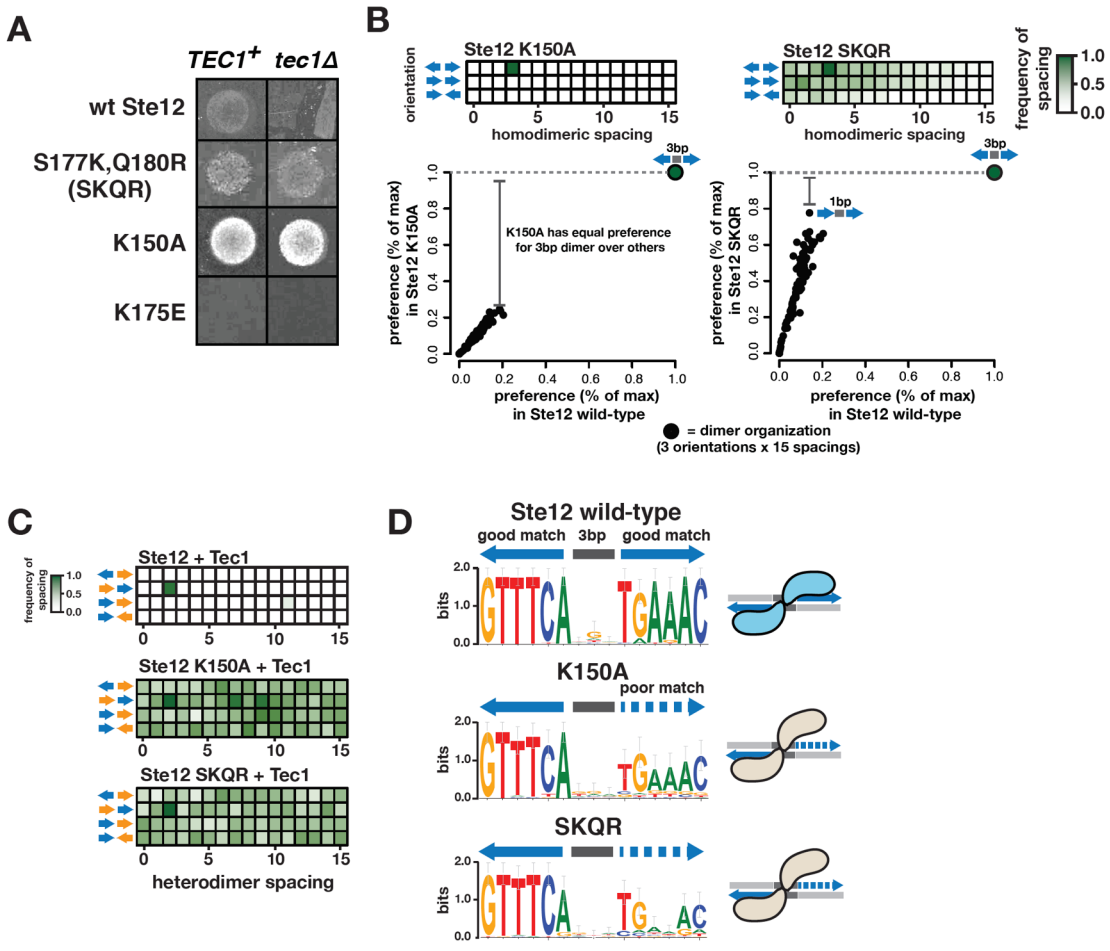


Figure 2.5

Ste12 DNA-binding domain mutants that confer invasion independent of the cofactor Tec1 have altered binding specificity. (A) Wild-type Ste12 requires the cofactor Tec1 for invasion; a tec1Δ strain fails to invade. The region I mutation K150A showed a hyperinvasive phenotype with or without Tec1. Introduction of the two positively charged residues found in *C. gattii* (SKQR) also led to Tec1-independent invasion, whereas introduction of a negative charge (K174E) eliminated invasion. The hyperinvasive Ste12 mutants showed a dominant invasion phenotype, as they were tested in the presence of wild-type Ste12, in the SIGMA1278b background. (B) Spacing heatmaps for Ste12's binding site are shown as in Fig. 2A. The K150A retains the same preference for tail-to-tail 3 base pair-spaced sites as the wild-type Ste12, but SKQR's preference is reduced. To visualize this difference, the scatterplots below show frequencies of site organizations relative to the most preferred site (% of max) for each variant (y-axis) compared to wild-type Ste12 (x-axis), where each point represents one orientation and spacing combination (example of SKQR's reduced preference is shown: its preference for head-to-tail, 1bp is ~78% that of the tail-to-tail 3bp site). (C) Spacing heatmaps are shown for the K150A and SKQR variants in co-binding reactions with Tec1; unlike wild-type Ste12 (repeated from Fig. 2A lower panel for clarity), neither variant shows strong preference for any heterodimeric site organization. (D) Logo plots generated from the 50 most highly enriched sequences in HT-SELEX represent ideal sites for each variant alongside wild-type Ste12.

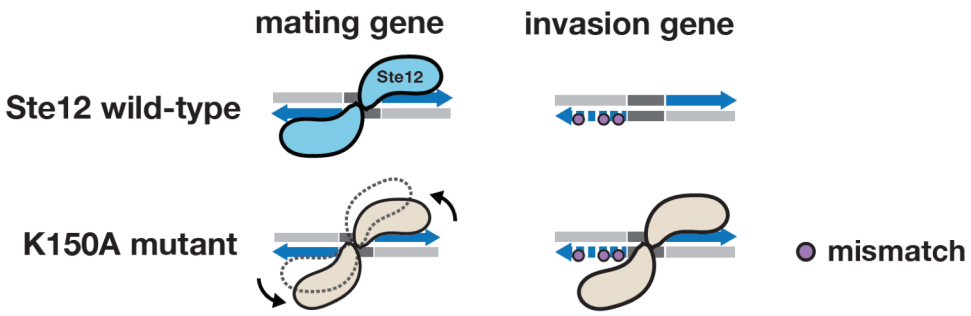
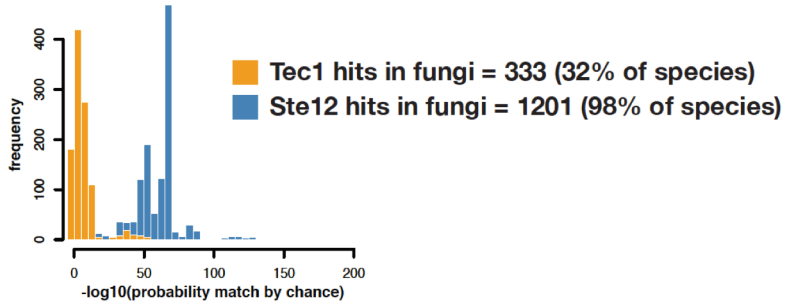


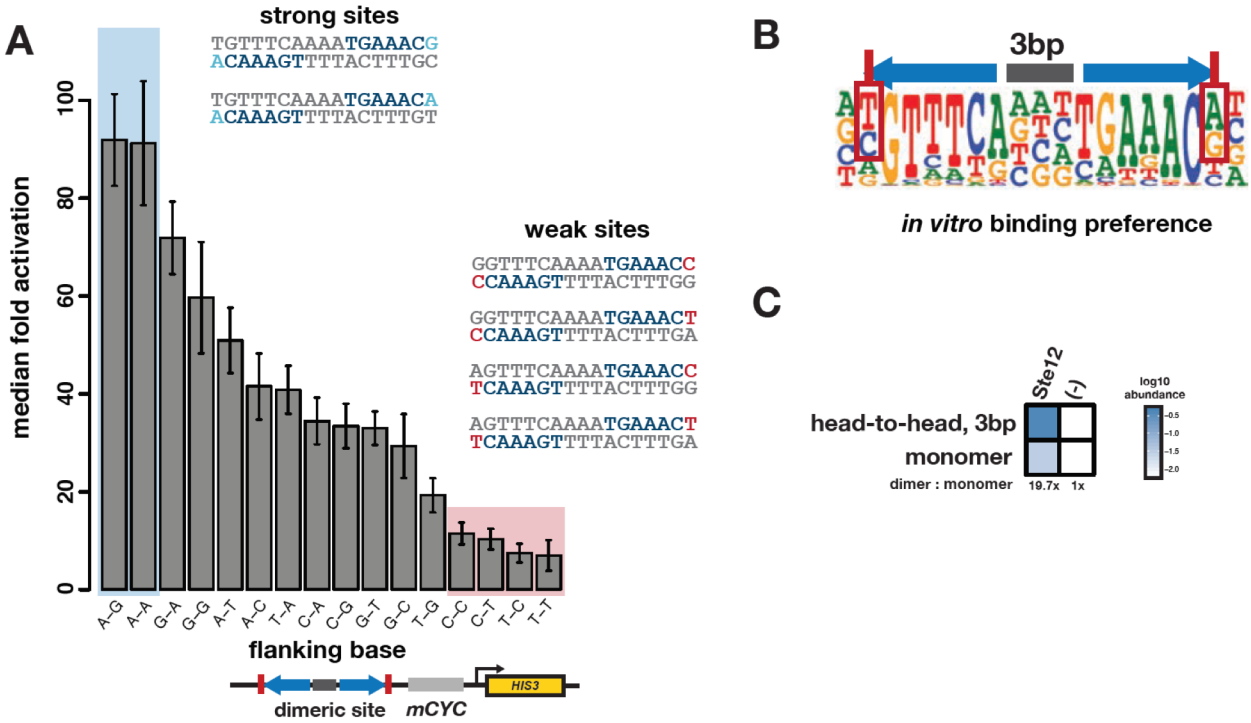
Figure 2.6

A model for novel site recognition by Tec1-independent Ste12 variants. Wild-type Ste12 protein (blue) is compared to the K150A mutant (tan) for binding at a mating gene with two perfect recognition sites (tail-to-tail with a 3 base pair spacing, shown in blue) and an invasion gene with a single perfect site along with a degenerate site. Due to its altered dimer interface, the "kinked" mutant K150A is less capable of binding symmetrically at two perfect sites, consistent with its phenotype of reduced mating efficiency. However, the conformation of K150A allows the variant protein to occupy novel sites within invasion genes that contain one perfect or nearly perfect site paired with a degenerate site; wild-type Ste12 is unable to occupy these sites.



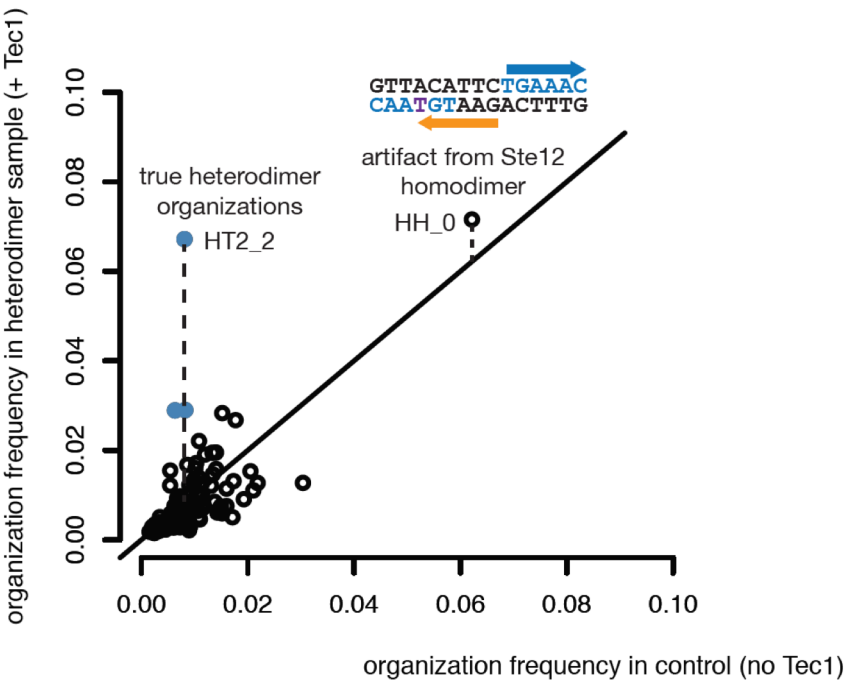
Supplemental Figure 2.1

***TEC1* is frequently absent in fungi whereas *STE12* is maintained.** We queried a protein database of 1229 fungal genomes and identified matches to either *S. cerevisiae* Tec1 (orange) or Ste12 (blue) sequence.



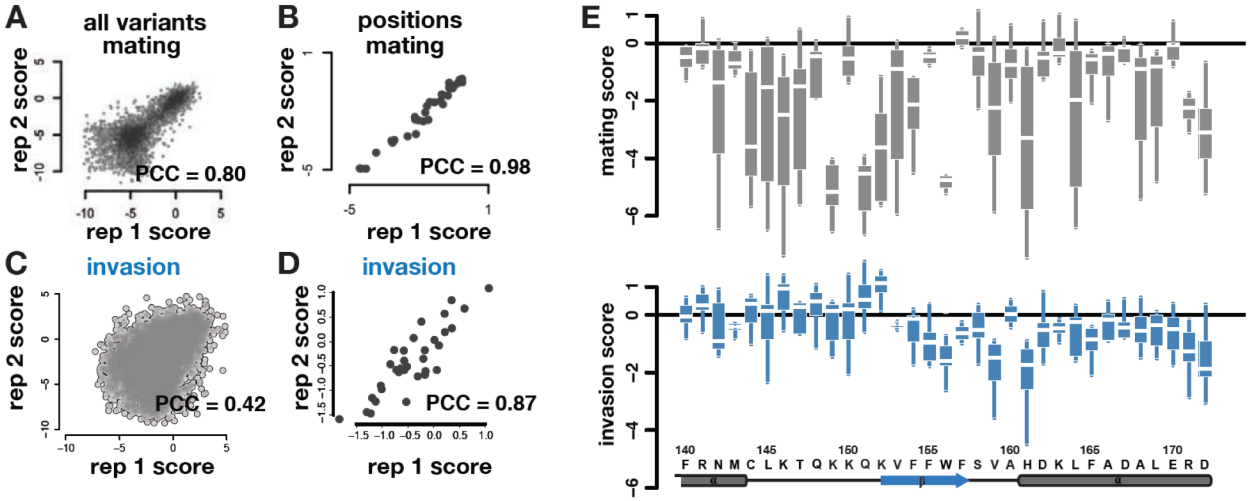
Supplemental Figure 2.2

Wild-type flanking base preferences of Ste12 binding *in vitro* are correlated with basal activity in vivo. (A) We collected activity scores among all sites containing perfect dimer sites GTTTCANNNTGAAAC, separated by the combination of flanking bases following the final base of the core motif (n > 50 for all flank combinations), and plotted the median activation of these sets of sites. Nearly an order of magnitude of Ste12's activation can be explained by the combination of flanking bases on a dimeric site. (B) Ste12's flanking base preferences identified *in vitro* by HT-SELEX (from Fig. 2B). (C) Ste12 was unable to activate a monomeric site *in vivo*, and this preference is reflected in its in vitro preference for dimeric binding.



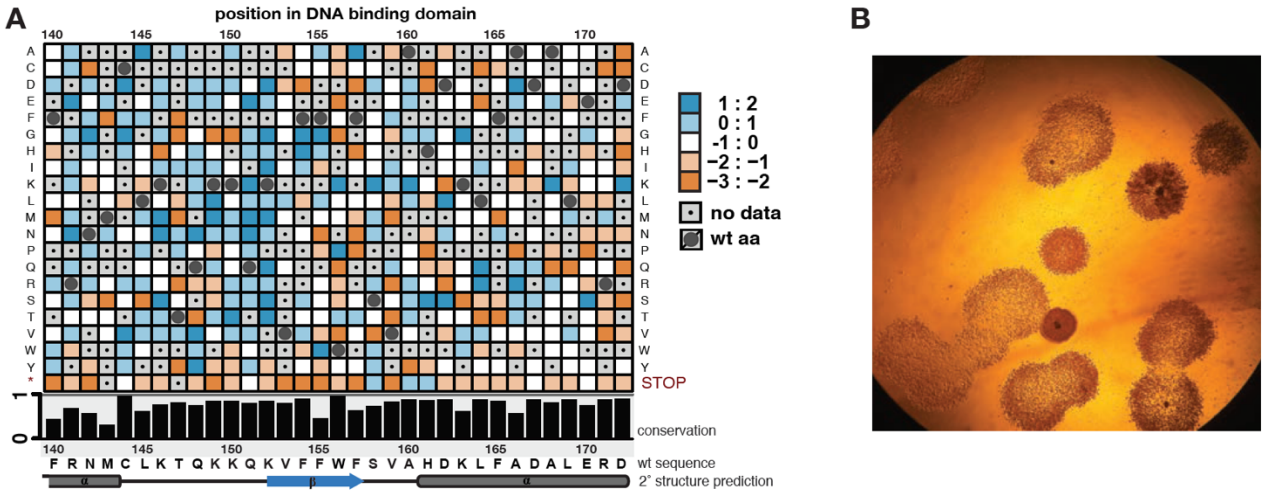
Supplemental Figure 2.3

Comparison of co-binding sample with Ste12 alone sample allows removal of false Tec1 sites. Each point represents a particular spacing and orientation combination. The frequency of each combination found in output SELEX samples is shown for Ste12 alone (x-axis) and for Ste12 + Tec1 (y-axis). The most frequent organization in the Ste12 + Tec1 sample is head-to-head 0 bp, but this site arises from fortuitous spacer sequences between Ste12 homodimer sites that resemble a Tec1 site (example indicated above head-to-head 0bp point). Normalizing each organization by its representation in the Ste12 alone sample reveals the true heterodimeric sites (colored in blue). (HH_0 = head-to-head, 0 bp spacer, HT2_2 = head-to-tail Tec1 before Ste12, 2 bp spacer)



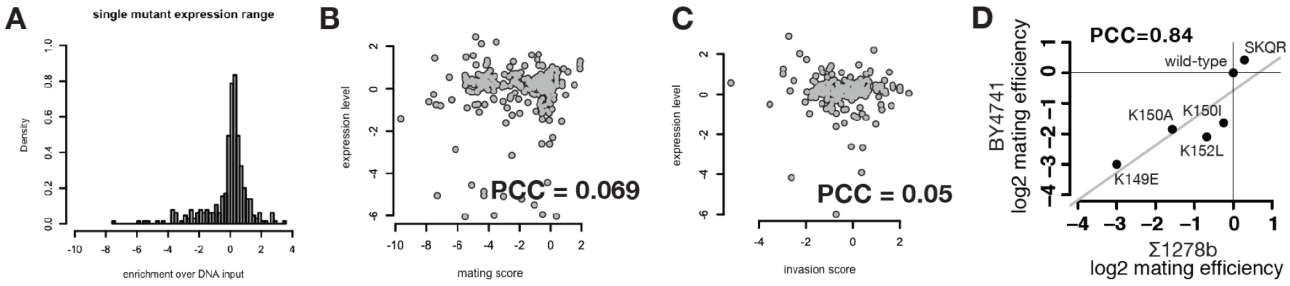
Supplemental Figure 2.4

Mating and invasion enrichment scores are reproducible across replicate experiments. Selection for each trait was done in three biological replicates. We computed Pearson correlation coefficients (PCC) between replicates for all variants (A,C) as well as positional means (B,D). Shown are correlations for replicate 1 and 2, results were similar for other replicate pairs. Invasion selections were less well correlated. Nevertheless, positional mean scores were strongly correlated for both traits across replicates. (E) Variability in per-position effect of mutation is shown as boxplots of all mutations tested at each site for mating (grey) and invasion (blue).



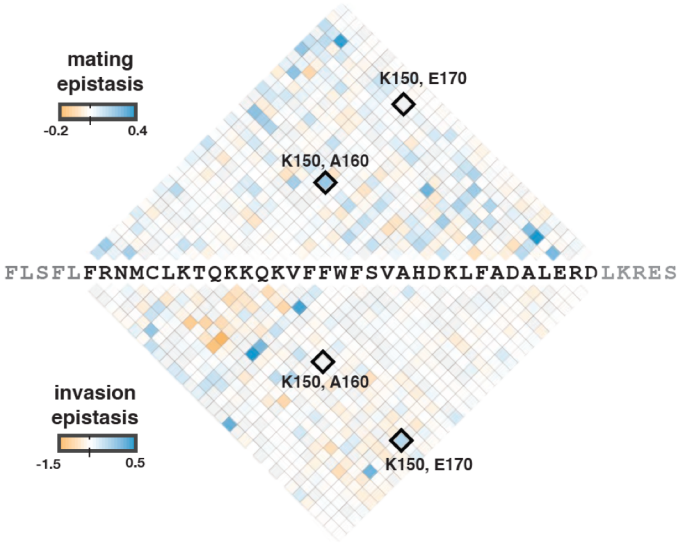
Supplemental Figure 2.5

Variation in the highly conserved DNA-binding domain of Ste12 generates a range of invasion phenotypes. (A) A heatmap of invasion scores for all single mutants is displayed as in Fig. 3B for mating scores. (B) A 10x magnified image of a washed plate containing invaded colonies with different Ste12 variants shows phenotypic variation in invasion.



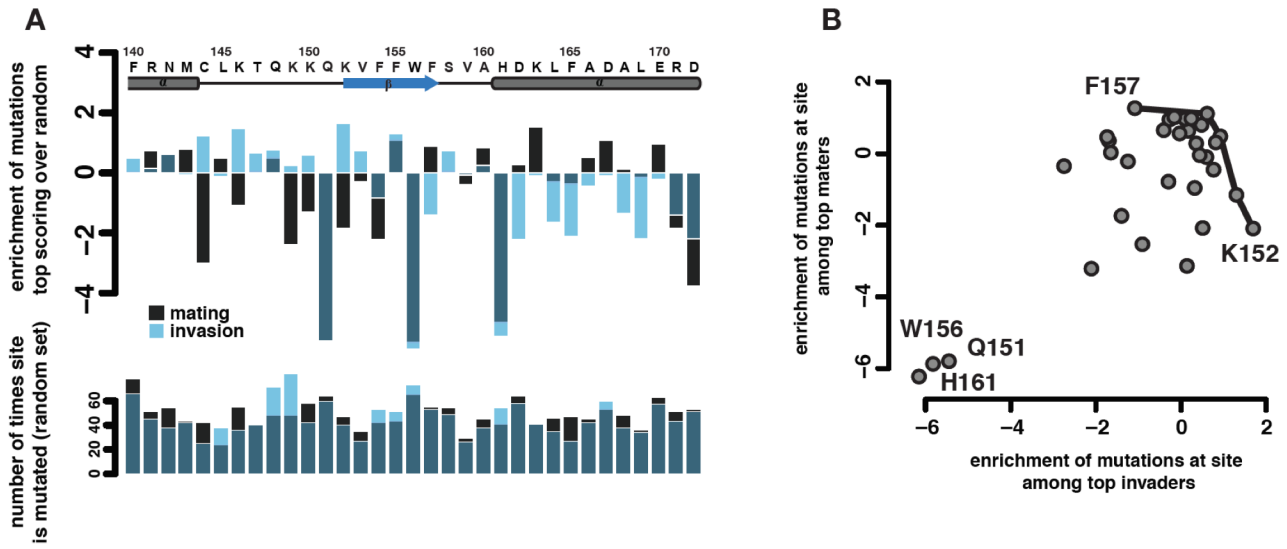
Supplemental Figure 2.6

Phenotypic effects of Ste12 variants are not due to differential expression levels or differences in strain background. (A) The distribution of log2 enrichment of transcripts relative to plasmid counts for each Ste12 variant rarely deviates from zero, indicating low variation in expression level among variants. Each variant's expression level is plotted against each variant's trait enrichment score, for (B) mating and (C) invasion, demonstrating that variant expression levels did not affect phenotype. (D) Mating defects inferred from large-scale experiments in BY4741 strain were confirmed in the strain used for invasion selection, Σ1278b. Mating efficiencies of *STE12* variants tested in each strain background were correlated (r^2=0.7), suggesting the strain background did not contribute to differences in measured phenotypic effects. The mating efficiency of the SKQR variant in both Σ1278b and BY4741 is shown, as this variant was not present in the initial library. The sterile phenotype of K149E is represented as a log2-fold relative to wild-type of -3, the median effect of all stop-containing variants in the large-scale mating selection.



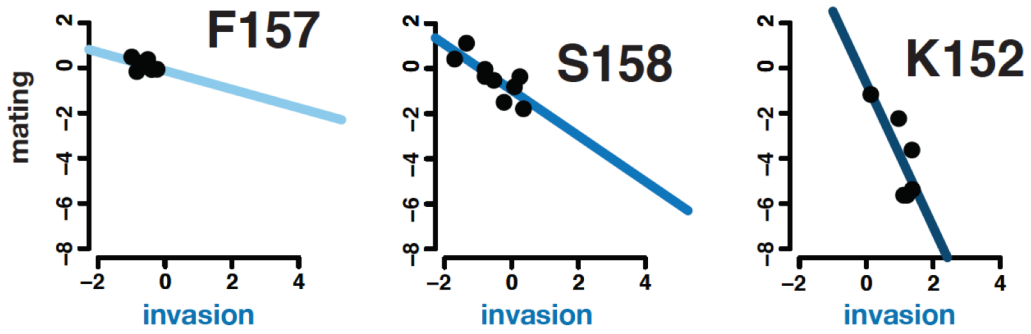
Supplemental Figure 2.7

Double mutant analysis identifies pairs of positions with strong epistasis, and such pairs differ between mating and invasion. For each double mutant, we calculated epistasis as the deviation of that mutant's effect from the multiplied effect of its constituent single mutants (see Methods). Scores are displayed as in Fig. 2A; epistasis scores across all combinations of mutations tested at each pair of sites was used to calculate the displayed mean epistasis score. Sites with strong positive (shades of blue) or negative epitasis scores (shades of orange) differ between traits. Strong intramolecular epistasis occurs more frequently between residues in close proximity, suggesting that Ste12 conformation differs between both traits. Two pairs with trait-specific epistatic interactions are indicated in boxes.



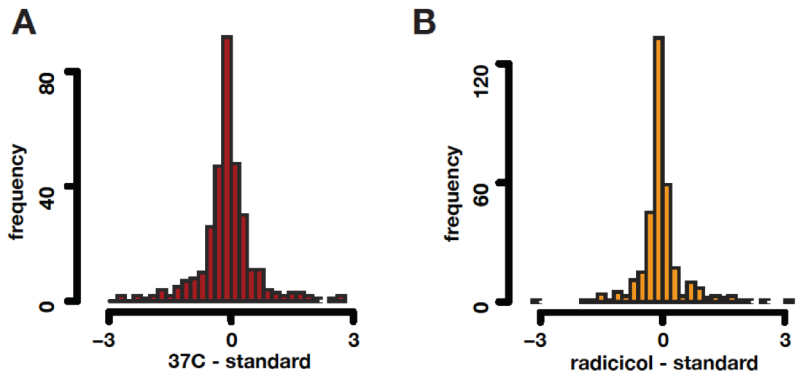
Supplemental Figure 2.8

Orthogonal analysis reveals positions contributing to 'separation-of-function' between mating and invasion. Per-site mutation frequencies were calculated for the top 750 mating variants, top 750 invasion variants, and a random set of 750 variants from each dataset. (A) Enrichment of per-site mutations for top performing variants in each trait recapitulates bipartite arrangement of effects ('separation-of-function'), as well as sites increasing one trait at the cost of the other (e.g. see K152 and F157, tradeoff indicated with blue/black bars). (B) Per-site enrichment scores identify similar sites on the empirical Pareto front (Fig. 4B), as well as the same sites in which mutations are deleterious in both traits.



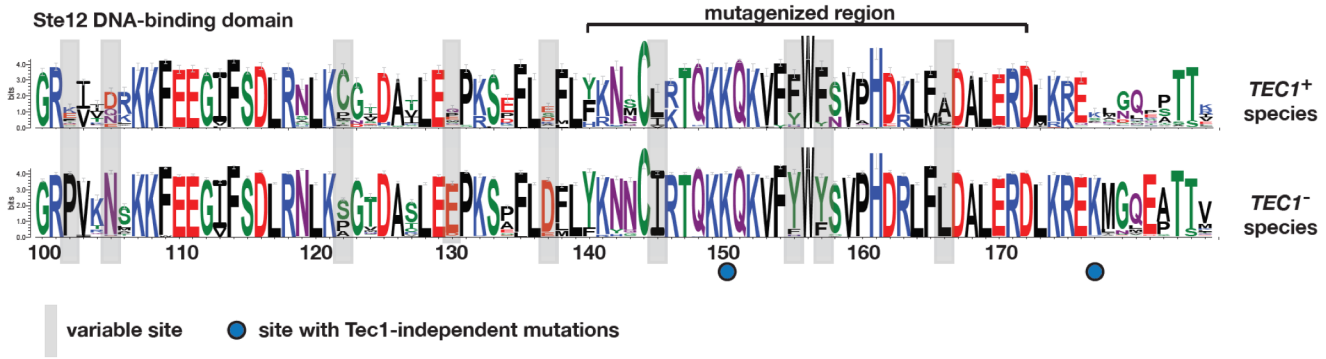
Supplemental Figure 2.9

Individual mutations tested at sites critical for separation of function between mating and invasion. Mating and invasion scores for all single amino acid change mutations at each indicated position were plotted, and a simple linear model was fitted. Steeper slopes indicate that mutations at that position are more likely to shift trait preference toward invasion.



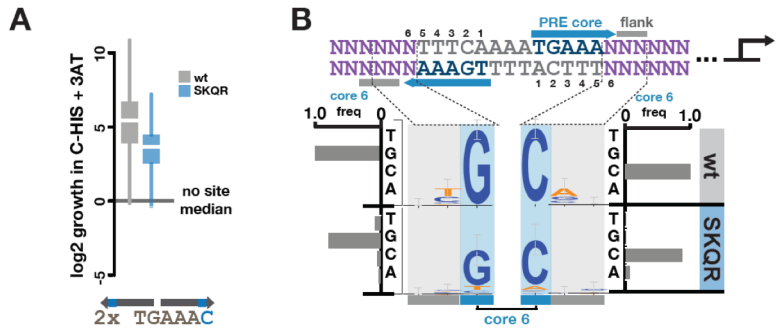
Supplemental Figure 2.10

The majority of Ste12 variants respond like wild-type Ste12 to increased temperature or Hsp90 inhibition. The distribution of differences of log2 mating scores for all single mutants (treated - untreated) are plotted for high-temperature (red, A) and Hsp90-reduced (orange, B) conditions.



Supplemental Figure 2.11

Variation in the DNA-binding domain of Ste12 is associated with the presence of a *TEC1* gene. A weblogo of theDNA-binding domain of Ste12 was generated as in Figure 1D , except that the fungal species were split into two groups: those whose genome contains a copy of *TEC1* (n=333) and those whose do not (n=715).



Supplemental Figure 2.12

SKQR shows reduced activation at perfect match dimeric sites compared to wild-type Ste12, but can activate degenerate sites. A library of binding sites was cloned into a *HIS3* reporter gene and used in growth selections in the BY4741 *ste12Δ* background. (A) The activity of wild-type Ste12 and SKQR variant with all binding sites containing two canonical PREs is shown as a boxplot (n = 725, 747). (B) The *in vivo* sequence preferences for wild-type Ste12 and SKQR are shown as logo plots of three positions (core position 6 and two flanking bases, no preferences were seen at further flanking bases) of each PRE. Grey error bars on logo plots represent 95% confidence interval at each base. Bar charts show the frequency of base preference at core position 6.

Chapter 3. TEMPERATURE-SENSITIVE CONTROL OF GROWTH IN YEAST

Abstract

The activity of Heat shock transcription factor 1 (Hsf1) in regulating the heat shock response is conserved among all eukaryotes. Hsf1 plays a role in such biological processes as longevity, climate adaptation, and oncogenic transformation. Upon sensing an increase in temperature, Hsf1 trimerizes and binds DNA to activate heat shock genes. Here, we use high-throughput fitness measurements of >500,000 variants in the trimerization domain of the *Saccharomyces cerevisiae* Hsf1 to investigate mechanisms of temperature-specific Hsf1 function. We find that the rate of cell division under heat-shock can be modulated by mutations in the coiled-coil trimerization domain. Trimerization domains derived from the Hsf1 in other yeast species also confer temperature-specific phenotypes when substituted into *S. cerevisiae* Hsf1. Furthermore, exceptional Hsf1 trimerization variants increase the fitness of yeast cells under heat-shock conditions beyond that of wild-type Hsf1. Hsf1 trimerization variants have altered DNA-binding patterns *in vivo*, and altered transcriptional programs, suggesting that trimerization modulates target specificity. We suggest that Hsf1 variants with altered trimerization can modulate cell division in response to changing temperatures.

3.1 INTRODUCTION

The heat shock transcription factor (Hsf1) is the central regulator of heat shock-inducible gene expression and the cytoplasmic unfolded protein response in eukaryotes. The regulatory function of Hsf1 depends on its ability to sense heat stress and activate genes necessary for survival. Hsf1 contains two highly conserved domains (Peteranderl et al., 1999; Sakurai &

Enoki, 2010): the characteristic Hsf1 DNA binding domain, and a hydrophobic repeat oligomerization domain. Standard models of Hsf1 regulation present an inactive monomer as the predominant form of Hsf1 under non-stress conditions, with stress exposure inducing post-translational modification and trimerization of the molecules into an active state, resulting in upregulation of heat shock target genes.

In addition to its well-studied role in inducing rapid gene expression in response to stress, Hsf1 contributes to non-stress-induced cellular processes: promoting development and fertility in mice (X. Z. Xiao et al., 1999) and serving as an essential gene in many model organisms, including *C. elegans*, *D. melanogaster*, *S. pombe*, and *S. cerevisiae* (Gallo, Prentice, & Kingston, 1993; Hajdu-Cronin, Chen, & Sternberg, 2004; Jedlicka, Mortin, & Wu, 1997; Morton & Lamitina, 2013; Phosphorylation, Sorger, & Pelham, 1988). Hsf1 activity contributes to development (Hajdu-Cronin et al., 2004; Jedlicka et al., 1997; Morton & Lamitina, 2013; X. Z. Xiao et al., 1999), aging (Hsu, Murphy, & Kenyon, 2003), cancer (Dai & Sampson, 2016; Mendillo et al., 2012), innate immunity (Singh & Aballay, 2006), and metabolism (Ma et al., 2015).

We sought to better understand the unique properties of Hsf1 activity in basal and stress conditions. We used deep mutational scanning of the *Saccharomyces cerevisiae* Hsf1 oligomerization domain because of its role in stress-induced activation. Mutation of its DNA-binding domain would likely alter the composition of its native transcriptional programs. The oligomerization domain influences not only the ability of Hsf1 to trimerize (Sorger & Nelson, 1989), but also intramolecular interactions (Hentze et al., 2016), post-translation modification (Hashikawa, Yamamoto, & Sakurai, 2007; He et al., 2003), and protein-protein interaction. The high conservation of this domain across eukaryotes emphasizes its functional importance and

increases the likelihood that we can translate our observations in yeast to other organisms, including humans.

The high conservation of Hsf1 among eukaryotes is echoed by the high conservation of its target DNA sequence motif, the heat shock element (HSE). A canonical HSE consists of three tandem inverted repeats of the sequence nGAAn, but variations on this theme exist. The number and arrangement of the repeats is known to influence the strength of Hsf1 binding (Bonner, Ballou, & Fackenthal, 1994; H. Xiao, Perisic, & Lis, 1991), and mutations that enhance human Hsf1 oligomerization enhance binding to variant HSEs (Takemori et al., 2009). Hsf1 target genes, and reportedly whole transcriptional programs, use different variations on the HSE triplet theme (Hahn, Hu, Thiele, & Iyer, 2004; Hashikawa et al., 2007; Mendillo et al., 2012), making oligomerization state a candidate mechanism for regulating transcriptional specificity. The unusual nature of Hsf1 target binding offers an exceptional opportunity to modulate target specificity without directly altering the DNA-binding domain.

Analysis of the Hsf1 oligomerization domain may furthermore inform our understanding of the mechanism by which Hsf1 senses and responds to stress. There are many levels to Hsf1 regulation, both dependent on and independent from the cellular context. Hsf1 is regulated through interaction with chaperones and negative regulators as well as post-translational modification and intramolecular interactions (Morimoto, 1998). The importance of the cellular environment is illustrated by the fact that human Hsf1, for which 37°C is its natural basal temperature, responds to 37°C as a heat stress when expressed in *Drosophila* cells or *Xenopus* (Baler, Dahl, & Voellmy, 1993; Clos, Rabindran, Wisniewski, & Wu, 1993). However, some level of temperature sensing regulation clearly occurs within the Hsf1 molecule itself: Hsf1 protein *in vitro* is capable of induced DNA binding with heat shock (Goodson, Michael L.,

Sarge, 1995; Zhong, Orosz, & Wu, 1998) and purified monomeric Hsf1 trimerizes upon heat shock (Hentze et al., 2016). In our analysis of the functional consequences of mutation of the Hsf1 oligomerization domain, we sought to better understand the role of this domain in temperature discrimination.

We present here a deep mutational scan of the oligomerization domain of *S. cerevisiae* Hsf1 and the effects of these variants on cellular fitness at both basal and heat stress growth temperatures. The results highlight the integral nature of the oligomerization domain to Hsf1 function in yeast at all temperatures, and demonstrate that point mutations in this domain can confer temperature-specific phenotypes. We identify critical residues in the trimerization domain, and find that patterns of reduced function derive from its helical structure. We uncover a set of rare variants in which changes to a pair of hydrophobic residues outside of the canonical helical phase of Hsf1's oligomerization domain increase growth rate under heat stress. Our results show the usefulness of large mutational datasets for understanding the molecular mechanisms regulating cell division in response to changing temperature.

3.2 RESULTS

Natural variation in the trimerization domain affects fitness in a temperature-specific fashion.

Hsf1 function (Fig. 3.1A) depends on two domains present among all eukaryotes: a DNA-binding domain and a helical oligomerization domain (Fig. 3.1B). Because variation within the Hsf1 oligomerization domain affects activation of the heat-shock response (Hashikawa et al. 2007), we and others have speculated about a role for oligomerization in Hsf1 functional specificity. In *S. cerevisiae*, Hsf1 function is essential in both basal (30°C) and heat shock (37°C)

conditions, with increasing temperature accompanied by shifts in oligomeric state and changes in target specificity. Because a transcription factor's target specificity derives from its DNA-binding domain, we would predict coevolution between the DNA-binding and trimerization domains if both contribute to specificity. We analyzed coevolution of these domains across 704 species in the fungal kingdom, the largest available dataset of Hsf1 sequence variation. Consistent with our prediction, we find that mutations in the trimerization domain co-vary with mutations in the DNA-binding domain, suggesting that altered trimerization has a role in Hsf1 target specificity (Figure 3.1C).

We observed substantial variation in the trimerization domain among fungi, even in the periodic hydrophobic residues that define the core interactions of the oligomeric state (Figure 3.1B). Since variation in coiled-coil structures has been associated with altered oligomerization states, we asked if any naturally-occurring oligomerization domain variants would be predicted to confer an oligomer state other than trimer. Using a model trained on all known coiled-coil structures, we predicted the highest likelihood oligomeric state (parallel dimer, antiparallel dimer, trimer, or tetramer) for all fungal Hsf1 oligomerization domains (Figure 3.1D). The majority of these (79.5%) are predicted to act as trimers, but we find a small subset (16/704, 2.3%) where a parallel dimer (or even numbered symmetry) is predicted to be favored over the trimer. Present in this subset are several species that thrive at low-temperatures, physiologically consistent with poor trimerization capacity by these heat shock factors. Conversely, fungal species with the strongest predicted trimerization propensity – well above that of *S. cerevisiae* and possibly indicating a constitutive trimer state – contained several thermophiles. To test whether the trimerization domains of these species could affect temperature response, we replaced the *S. cerevisiae* oligomeriation domain (residues 342-403) with the oligomerization

domain of six different fungi of varying predicted trimerization propensity (Fig. 3.1E). These Hsf1 variants were tested for their ability to influence *S. cerevisiae* growth rate in a background of tet-repressed endogenous Hsf1, at 30°C and 37°C growth temperatures. At 30°C, all natural trimerization variants rescued growth compared to a strain lacking Hsf1, and affected growth only slightly, suggesting that these variants do not strongly affect the basal function of Hsf1 (Fig. 3.1F). At 37°C, we observe a greater dependence on the trimerization function; two of three trimerization domains from cold-adapted species (*M. irregularis, M. sympodialis*) with low predicted trimerization were unable to rescue heat-shock growth. Conversely, thermophilic species with strong trimerization predictions were able to rescue growth, though these trimerization domains did not provide a fitness benefit beyond that of wild-type Hsf1 (Fig. 3.1F).

High-throughput characterization of temperature-dependent fitness effects of trimerization domain variants.

Given that natural variation in the oligomerization domain can influence the effect of temperature on fitness, we sought to explore this relationship more exhaustively by testing synthetic variants of the Hsf1 oligomerization domain in *S. cerevisiae*. We first confirmed the specific role of the trimerization domain in both basal and induced function in *S. cerevisiae* (Fig. 3.2A). Cells containing no Hsf1 or an Hsf1 with no oligomerization domain grew at less than half the rate of wild-type cells after conditional knockdown of the endogenous *HSF1* gene. We used deep mutational scanning to exhaustively interrogate amino acid substitutions at each position within a region of the trimerization domain. We targeted these 36 amino acids (position 365 – 400) because of the potential for mutations at these positions to have functional consequences (Fig. 3.2D, top). Human Hsf1 cannot substitute for the essential function of Hsf1

in yeast unless it carries point mutations, with a high density of such mutations at residues residues 161-189 (Neef et al. 2013); these residues correspond to 365-396 in the yeast protein. The conservation of the hydrophobic heptad repeats in this region also allows predictions to be made about critical residues involved in oligomerization; Sorger and Nelson (1989) visually annotated residues likely involved in the hydrophobic coiled-coil interaction. Consistent with the presence of periodic hydrophobic residues, this domain is predicted to be a single, large alpha helix (Drozdetskiy, Cole, Procter, & Barton, 2015) (Fig. 3.2D, top).

HSF1 is an essential gene in yeast (Sorger and Nelson 1988, Wiederrecht et al. 1988). We therefore conducted our screen in a background in which the endogenous Hsf1 can be repressed by treatment with tetracycline (Fig. 3.2B). A library of Hsf1 variants was created by doped oligo synthesis of the 108 base-pairs encoding amino acids 365-400, with a 2.5% chance of mutant base incorporation at each DNA position. The resulting library was transformed in bulk into the tet-off Hsf1 strain. Endogenous Hsf1 expression was permitted during transformation and expansion of the library in order to reduce selection pressure before starting the experiment. Samples of the library before yeast transformation as well as immediately post-transformation were collected and processed for sequencing. The yeast library contained over 400,000 variants in total, including 620 single amino acid mutations out of the 720 possible single amino acid substitutions (Table 1).

We silenced the endogenous *HSF1* gene and carried out a growth selection at both basal and heat shock temperatures on yeast cells carrying the variant library. The representation of each variant in the population was compared before and after extended growth. Variants that confer a fitness defect decreased in frequency over the growth interval, while those with a fitness benefit increased in frequency, compared to wild type. After the initial 4 hours of 30°C growth

with anhydrotetracycline (ATc) to commence repression of endogenous Hsf1, the culture was split into three and growth was continued at 30°C, 35°C, or 37°C. Three replicates were growth at 30°C or 35°C for 14 hours, and at 37°C for 17 hours because of the slower growth under stress. Variant frequencies were determined by high-throughput sequencing of the plasmids following each growth selection and were compared to the initial plasmid library to generate a fitness score. Correlation in variant scores was high between the triple replicates at each condition (Table 2).

Variants with nonsense mutations displayed severe growth defects, as expected, based on the essential nature of Hsf1 under all conditions (Fig. 3.2D). Mutations with synonymous codon changes formed a distribution around zero (indicating wild-type fitness) (Fig. 3.2D). Mutations to proline were as detrimental as stop codons at all temperatures (Fig. 3.2D). This results suggests that the integrity of an alpha-helical structure in this domain is necessary for function at both basal and high temperatures. We found rare instances of epistatic double mutants wherein the effect of a proline mutation could be rescued by a mutation at another site within the trimerization domain. This phenomenon was extremely rare, but a pair of mutations (N372I, A392P) that show this rescue effect was validated (data not shown). Variants effects determined in high-throughput were validated by individual growth assays, which confirmed our ability to detect a range of mutational effect sizes at each temperature (r^2 = 0.93 for 37°C, r^2 = 0.84 for 30°C) (Fig. 3.2C).

The role of trimeric structure in basal and induced function of Hsf1.

The Hsf1 oligomerization domain is a trimeric coiled coil, which imposes a constraint on the types of amino acid residues, as well as their spacing, within this domain. We tested these

constraints by analyzing the mean effects of mutation at the level of individual position in the domain (365 – 400), as well as the expected positions in an alpha helix (*a,b,c,d,e,f,g*), in both basal and induced conditions. Under basal conditions, the most deleterious positions, on average, tended to be those that participate in the hydrophobic core of the trimeric Hsf1 (Fig. 3.3A). This effect was consistent under heat-shock conditions, but the most deleterious position differed at each temperature. At 37°C, L375 at position *d* in the helix was most sensitive to mutation, but was much less so at 30°C (Fig. 3.3A). The most sensitive position at 30°C was the adjacent W376, in helix position *e*, suggesting that these positions may play a role in the separation of basal and induced function of Hsf1 (Fig. 3.3A). Both of these positions are highly invariant among fungi (Fig. 3.1B), suggesting that both basal and induced functions of Hsf1 influence patterns of conservation.

We next examined the effects of mutation at the level of helix position. Because different hydrophobic residues can substitute at 1 and 4 positions in this domain, we looked at specifically at those small or charged residues (Gly, Ser, Lys, Arg, Glu, Asp) that cannot support coiled-coil interactions. At 37°C, small or charged residues were more deleterious on average in helix positions *a* and *d* (Fig. 3.3B). We found a similar pattern of helix position effects at 30°C, with sensitivity to mutation for helix position *c* as well. These results suggest a role for trimerization at 30°C, and the importance of helical structure in separating basal and induced growth.

Rare variants show increased fitness relative to wild-type.

Because variants with decreased fitness showed striking specificity in the types of residues, and in the helix position, of their mutations, we wondered whether these patterns also manifested in Hsf1 variants of cells with increased fitness under heat-shock. We used the full distribution of

mutational effects to identify variants with significant deviations from the mean fitness relative to cells containing wild-type Hsf1 (Fig. 3.4A). Most variants conferring this phenotype contained multiple mutations; only one variant had a single amino acid change (A382C). We validated the fitness effects of six variants (N369Y, E373K; A382C; R398I, N372I; M381K, R387K; M380K, M381I; and M381T, F399Y) with increased fitness using individual measurements of growth rate, and found growth rates up to 20% higher than wild type at 37°C. Although less benefit was observed at 30°C, there was no fitness trade-off at the two temperatures (Fig. 3.4C).

We examined the patterns of mutation in the variants that conferred an increase in fitness over the wild-type Hsf1. These variants were depleted for mutations at position *d* in the helix, confirming the deleterious nature of these mutations. On the outer face of the helix (positions *b* and *c*), we found a striking enrichment of specific mutations (Fig. 3.4B). We classified each variant by its mutation at helix positions *b* and *c*. Increased fitness was driven by heptad repeat 6, whose wild-type residues at the *b* and *c* positions are MM. The most enriched mutations to these two residues change one of each of these methionines to a hydrophobic leucine or isoleucine (Fig. 3.4D). These results demonstrate that modification of the trimerization domain outside the core hydrophobic residues can affect fitness under heat-shock.

3.3 DISCUSSION

Hsf1 enacts a temperature-specific regulatory program whose sensitivity depends on trimerization. Heat-shock and other proteotoxic stresses activate Hsf1 by triggering the shift to a predominantly trimerized state. Despite the importance of this process in determining cellular physiology and fitness, the relationship between protein sequence identity of the trimerization domain and cellular fitness has been largely unexplored. We exhaustively determined the relationship between sequence identity in the trimerization domain and the essential role of Hsf1

in growth in basal and heat-shock conditions. Our approach revealed patterns of Hsf1 function that could not be predicted from natural variation of this domain. Further, we identified exceptional mutants that increased heat-shock fitness, and this phenotype could not be conferred by simply swapping in oligomerization domains from thermophilic species. We found some species whose Hsf1 contains mutations we found to be associated with increased growth under heat-shock, and though these species are not annotated as thermophiles, they include human pathogens like *Candida parapsilosis.* Thus, the exhaustive genotype-phenotype map we have generated for Hsf1 trimerization domain may be useful in predicting temperature-sensitive growth responses in other species.

We found that mutation of the regular helical structure predicted from the geometry of the Hsf1 trimer disrupted the wild-type function of Hsf1 under basal and heat-shock conditions. We also confirmed a role for trimerized Hsf1 for growth under basal conditions in *S. cerevisiae*; residues making up the hydrophobic core of the Hsf1 trimer showed equal contributions to basal and heat-shock fitness. This result confirms the earliest predicted models of Hsf1 trimerization (Sorger & Nelson, 1989).

Surprisingly, we also found that positions on the outer face of the Hsf1 trimer are associated either with reduced basal function or a stress-resistant phenotype. The mechanism underlying these changes in native Hsf1 function is unclear. For example, the Q389K variant, whose function is reduced under basal conditions, but not under heat-shock could be explained by: (1) altered trimerization propensity, such that Hsf1 can no longer trimerize under basal conditions, but readily trimerizes at high temperature; (2) altered interactions with cofactors that specifically contribute to basal function; or (3) overall stability of that Hsf1 variant.

A second class of variants that are affected by positions on the outer face of the Hsf1 trimer conferred increased growth rate under stress. These variants are significant in their similarity to disease phenotypes associated with Hsf1 function. For example, Hsf1 is over-active in many tumors, despite lack of heat-shock. The subsequent upregulation of chaperone proteins potentiates oncogenesis by providing a protein folding buffer for other mutant proteins that increase rate of cell division (Mendillo et al., 2012). Further study of the proteomic and transcriptomic effects of expressing Hsf1 variants with increased growth could be used to better understand the similarities of these effects to those seen in cancer.

3.4 METHODS

Deep mutational scan of S. cerevisiae HSF1 trimerization domain. *S. cerevisiae HSF1* was cloned under its native promoter (664bp) into pRS415. In order to facilitate large scale library cloning, silent restriction enzyme sites were introduced on either end of the target region - AvrII (CCTAGG) at base pairs 1041-1046 and Bpu10I (GCTTAGG) at base pairs 1244-1250 – to generate plasmid pEM74. The region targeted for mutagenesis was base pairs 1093-1200 of the *HSF1* coding sequence. A doped oligo library of 168 bases was ordered (TriLink) containing 30 bases of invariant flanking sequence on either side of the 108 bases of the doped target region (2.5% base misincorporation).

Sequencing and determination of trait scores. Sequencing was completed on Illumina's MiSeq or NextSeq platforms. Sequencing libraries were prepared by extracting plasmids from yeast populations (Yeast Plasmid Miniprep II, Zymo Research, Irvine, CA) before and after selection. These plasmids were used as template for PCR amplification that added adaptor

sequences and 8bp sample indexes to the 108bp mutagenized region for sequencing (all libraries amplified < 15 cycles). Paired-end reads spanning the mutagenized region were filtered to obtain a median of 5 million reads per sample. Using ENRICH software (Rubin et al., 2017), read counts for each variant before and after selection were used to determine growth enrichments for *HSF1* variants. Briefly, counts for a particular variant in the input and output libraries were normalized by their respective read totals to determine frequency in each, and a ratio of the output and input frequencies determine a variant's functional score. Finally, enrichment scores are normalized by the enrichment of cells containing wild-type Hsf1 in each selection experiment. Treatment scores are calculated identically, and in all cases where difference scores are shown, the score represents log2(treated) – log2(untreated).

Growth validation assays of HSF1 variants. Point mutations identified from the DMS were introduced by inverse PCR of pEM74, followed by Gibson assembly with a 60-base oligo containing the mutation. *HSF1* variant clones were transformed into a tet-off *HSF1* strain(Mnaimneh et al., 2004). The Δtri mutant was created by ligation of an inverse PCR that omitted the HR-A/B domain (amino acids 342-403). *S. cerevisiae* HR-A/B domain (amino acids 342-403) were replaced with HR-A/B domains of other fungal species by inverse PCR and Gibson assembly with synthesized gBlocks of the domains.

RNA-seq of HSF1 variants. *HSF1* variant clones transformed were into tet-off *HSF1* yeast. Independent transformants were isolated for each of three biological replicates. Clones were grown overnight at 30°C in C-Leu media and back-diluted the following morning to OD 0.45. These back-dilutions were grown for 2hr at 30°C and then ATc was added to a final

concentration of 0.1 μg/mL and the cultures were given 4hr more at 30°C. Cultures were diluted to OD 0.1 in C-Leu + 0.1 μg/mL ATc in two sets of tubes, one of which was placed at 30°C and the other at 37°C. After 16hr of growth, 5mL of culture were spun down and yeast pellets were frozen in liquid nitrogen. RNA was isolated via acid-phenol extraction, and polyA-specific cDNA was synthesized. cDNA was subsequently prepared for high-throughput sequencing using Illumina Nextera transposition. Sample indexes were added in an 8-cycle PCR, and each sample was sequenced at a minimum of 5 million reads. Reads were aligned using STAR, and per-gene counts were analyzed for differences using DEseq.

Table 1

# MUTATIONS	OBSERVED	POSSIBLE	%
0	1	1	100.0
1	620	720	86.1
2	35694	252000	14.2
3	243501	57120000	0.4

Table 2

GROWTH TEMPERATURE	ALL VARIANTS CORRELATION (PEARSON'S R)	SINGLES CORRELATION (PEARSON'S R)
30°C	0.86	0.92
35°C	0.73	0.91
37°C	0.86	0.93

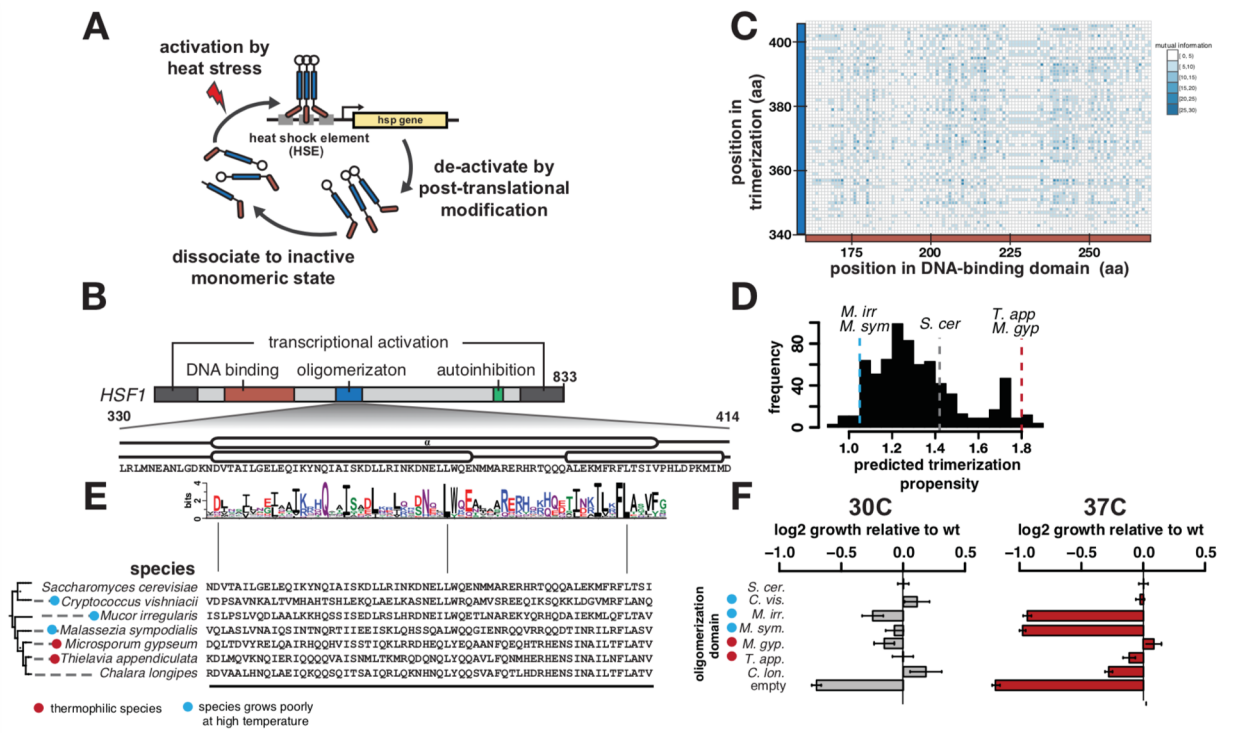


Figure 3.1

Natural variation in the Hsf1 trimerization domain is associated with heat-shock specific function. (A) The standard model of Hsf1 activation by trimerization under heat-shock. (B) A schematic of known domains in the Hsf1 transcription factor: DNA-binding (red), trimerization (blue), and autoregulation (green). For the trimerization domain, a sequence logo is shown below that represents the sum of natural variation amount >1000 fungal species. Note periodicity in conserved positions. (C) A summary of co-variation in the DNA-binding and trimerization domains among all fungal species. Darker shades of blue indicate more co-variation, as measured by mutual information. (D) Histogram of model predictions of trimerization propensity for >1000 fungal sequences, where higher values indicate a higher probability for trimerization. *S. cerevisiae*, and several species with extreme prediction values are indicated. (E) Sequences of trimerization domains from fungal species with very low (*C. vis., M. irr., M. sym.)*, or very high predicted trimerization values (*M. gyp., T. app.)*. Temperature-related growth annotations are added for the species if available. (F) Individual growth rates for *S. cerevisiae* strains containing trimerization domain swaps from all indicated species in (E).

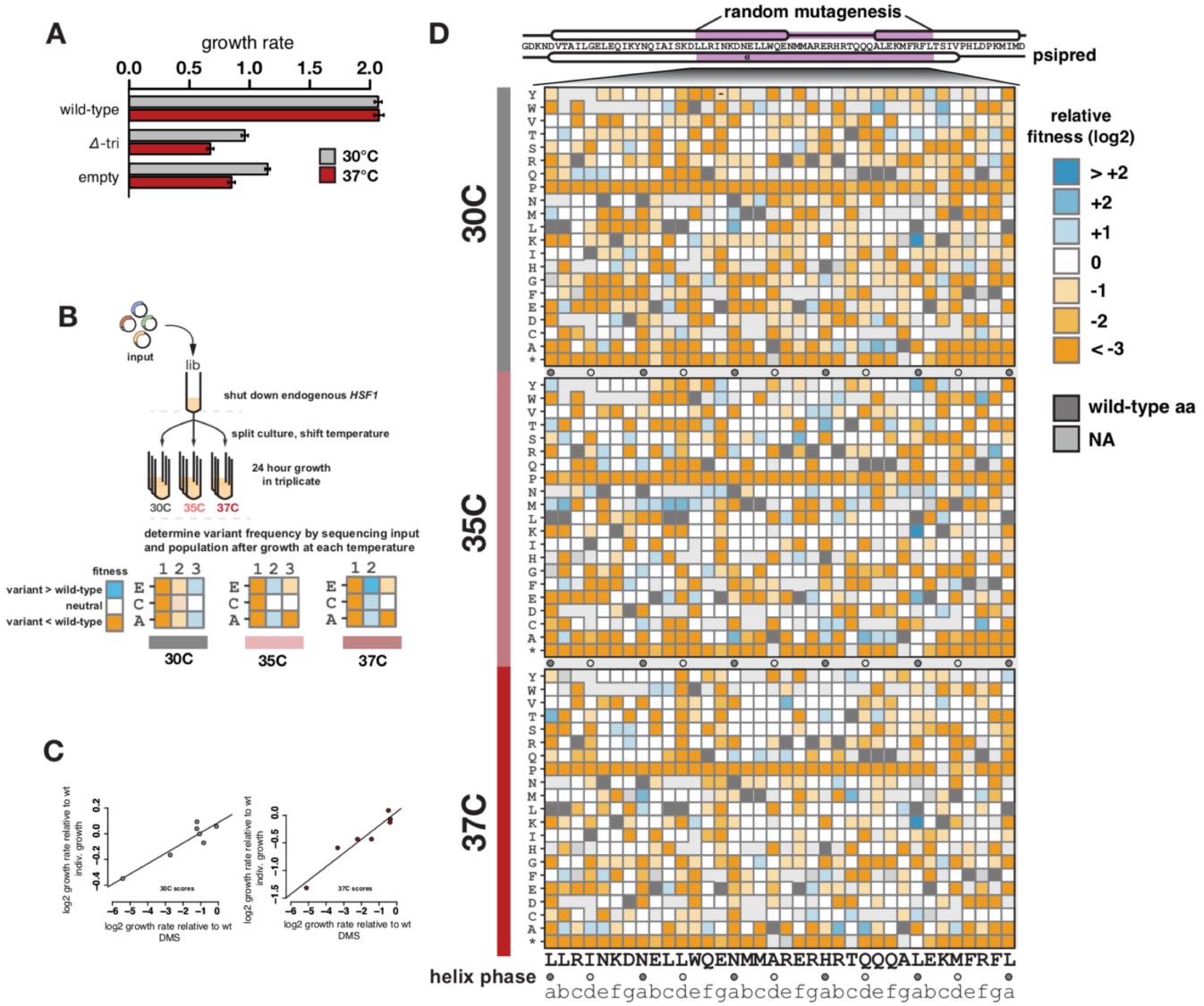


Figure 3.2

Deep mutational scanning of the Hsf1 trimerization domain reveals temperature-sensitive fitness effects. (A) Individual growth rates in S. cerevisiae strains containing either wild-type Hsf1, Hsf1 with the trimerization domain removed, or with no trimerization domain. (B) Experimental schema for temperature-specific growth selections of the Hsf1 variant library. (C) Effects of Hsf1 variants on growth rate determined by sequencing are confirmed by individual validation of growth rates. In each case, deep mutational scanning (DMS) score is shown on the x-axis and individual growth rates are shown on the y-axis for basal (left) and heat-shock growth temperatures (right). (D) Heatmap depicting mutational effects all nearly all single mutations in the Hsf1 trimerization domain. Deleterious variants are shown in shades of orange, neutral variants are white, and fitness benefits are shown in shades of blue. Variants not present in the library are shown in grey, and the wild-type amino acid identity at each site is shown in dark gray. A secondary structure prediction for the trimerization domain is shown above the heatmap.

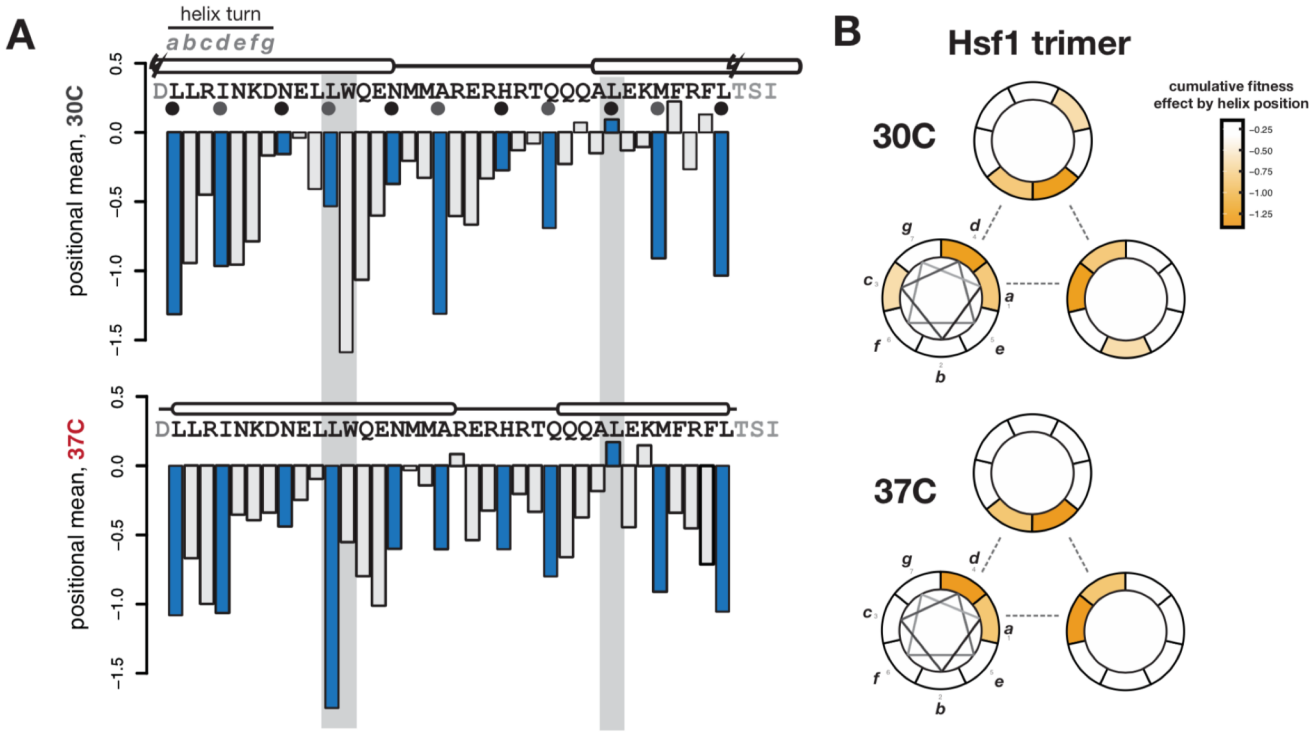


Figure 3.3

Trimeric helical geometry explains positional effects of mutation in the trimerization domain. (A) The mean effects of all single mutants tested at each position for basal (top) and heat-shock (bottom) is shown. Positions at a or d positions in the helix phase are indicated in blue. A pair of positions with temperature-specific effects is highlighted in grey. (B) The effects of all mutations were summarized by helix position in basal (top) and heat-shock (bottom) conditions are shown. More deleterious effects are shown in shades of orange. Helix positions are depicted by their expected theoretical geometry, and each helix is repeated to represent the positioning expected for the Hsf1 trimer.

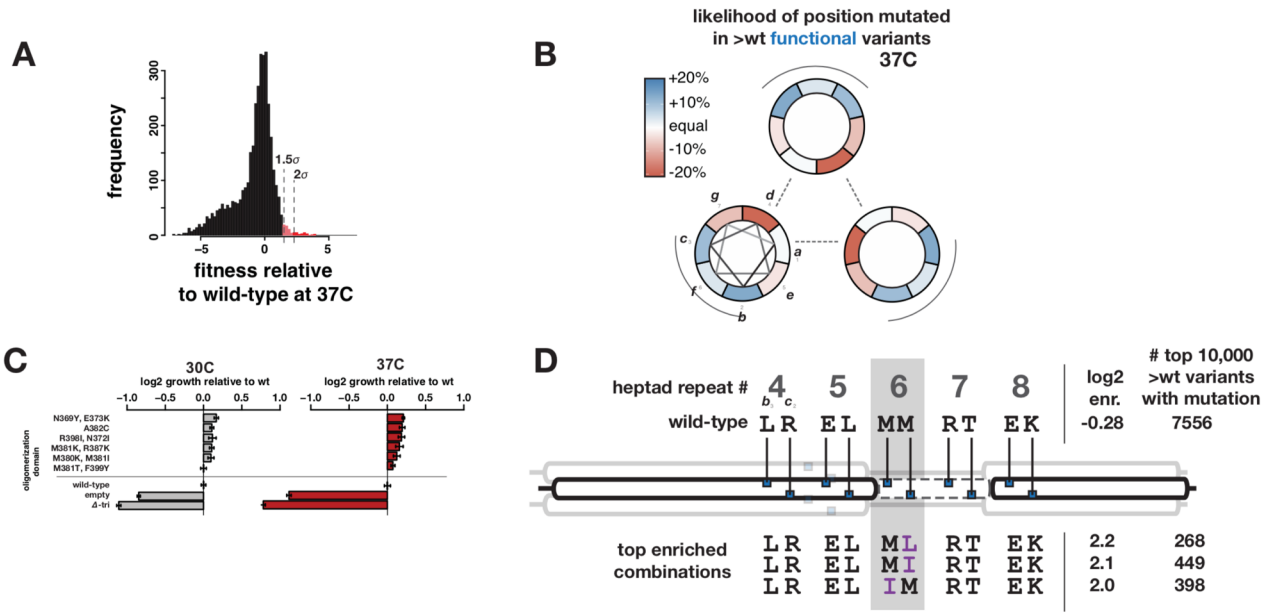


Figure 3.4

Hsf1 variants with fitness benefits contain mutations on the outer helical face of heptad repeat six. (A) A histogram showing growth of all variants under heat-shock; a population of variants with growth effects 1.5 and 2 standard deviations from the mean are shown in red. (B) Variants with fitness benefits are enriched at outer helix face positions. Helix position plots are shown as in 3.3 (B), except that shading now represents the frequency of mutations that increase growth rate at particular helix positions; blue represents an enrichment, and red represents a depletion. (C) Variants with growth rate greater than wild-type were individually validated at basal (left) and heat-shock (right) temperatures. (D) The population of variants showing a fitness benefit were analyzed for enrichment of any particular mutations at the b and c helix positions (wild-type sequence of helix b, c positions indicated) that showed enrichment in (B). Many of these variants contained mutations specifically in heptad repeat 6, which is specified by two methionine residues in the wild-type Hsf1 trimerization domain.

Chapter 4. HIGH-THROUGHPUT IDENTIFICATION OF DOMINANT NEGATIVE POLYPEPTIDES

Abstract

Dominant negative polypeptides act as inhibitors by binding to the wild type protein or by titrating a ligand essential for activity. Here, we use DNA sequencing to track the frequencies of large numbers of protein fragments as yeast cells are subjected to a growth selection. We demonstrate that this method can be applied to a cDNA library to identify dominant negative polypeptides at large scale and at exquisite domain resolution.

4.1 INTRODUCTION

Dominant negative mutants disrupt the activity of a wild-type protein present in the same cell(Herskowitz, 1987), allowing the detailed study of cellular processes. In contrast to the effect of more common recessive loss-of-function mutations, dominant negative inhibition can occur without a genetic alteration to the wild type gene. The predominant mechanism of action is that a dominant negative mutant retains interaction with another molecule while lacking a critical activity. In the case of proteins that function as oligomers – estimated to represent 35% of the proteins in a cell (Goodsell & Olson, 2000)– a dominant negative polypeptide might bind to a wild type subunit to form an inactive mixed dimer or higher order structure. In other cases, a

dominant negative polypeptide may titrate out an interacting protein, a DNA site or a small molecule. For example, a dominant negative fragment of the oligomerization domain of the *E. coli lac* repressor forms nonfunctional mixed tetramers (Betz & Fall, 1988); a dominant negative Rag GTPase binds to the mTORC1 complex and inhibits the complex's ability to sense amino acid levels (Sancak et al., 2008); and a truncated version of the human transcription factor Stat5 that retains the DNA-binding domain but lacks the transcriptional activation domain saturates Stat5 DNA-binding sites (Moriggl et al., 1996). Dominant negative alleles have been linked to human disease, particularly splice variants that result in truncated proteins (Holbrook, Neu-Yilik, Hentze, & Kulozik, 2004); a variant human β-globin causes β-thalessemia due to its ability to bind heme (Thein et al., 1990).

Dominant negative mutants have multiple advantageous properties. Unlike RNAi approaches which require the targeted protein to be degraded before an effect is observed, dominant negative mutants can inhibit protein activity almost immediately. Unlike CRISPR/guide RNA approaches that permanently alter a gene (Shi et al., 2015; Wang et al., 2015), dominant negative mutants can be conditionally induced and expressed at variable levels, features particularly useful for disrupting the activity of essential genes. Furthermore, because dominant negative inhibitors bind their targets based on protein structural features, rather than nucleic acid sequence specificity, they have the potential to inhibit individual functions of a protein. In addition, binding to structural features allows a single dominant negative polypeptide to inhibit multiple proteins that possess a common domain. Despite their utility, dominant negative alleles are typically discovered only at small scale, usually in the context of a broader mutant screens (Reddy & Hahn, 1991). Larger-scale attempts have been made in the yeast *Saccharomyces*

cerevisiae to identify polypeptide fragments that, when overexpressed, produce dominant negative effects, but these studies were limited in their sensitivity and their capacity to track potential inhibitory fragments (Akada, Yamamoto, & Yamashita, 1997; Boyer et al., 2004; Ramer, Elledge, & Davis, 1992; Stevenson, Kennedy, & Harlow, 2001).

Here, we use high-throughput sequencing of DNA libraries encoding yeast polypeptide fragments to identify dominant negative inhibitors based on their depletion during cell growth. Unlike naturally occurring dominant negative mutants, our experimental strains are diploid at only a single locus. In addition to endogenous gene copy that exists in the haploid yeast genome, we express a fragment of that gene in the same genetic background to identify dominant negative activity of that fragment.

4.2 RESULTS

We initially tested the method with individual proteins known to oligomerize. Yeast cells require orotidine-5'-phosphate decarboxylase (Ura3 protein) to synthesize pyrimidines and thus to grow in media lacking uracil. We fragmented the *URA3* gene to a median length of 200 bp, and cloned >7,000 fragments into a galactose-inducible, low-copy expression vector (~18.2% of fragments in-frame). Triplicate cultures transformed with this library were induced with galactose and grown 24 generations in media lacking uracil (Fig. 4.1A). Cells were collected before induction and at the 48 hour timepoint. Sequencing was used to track the frequencies of each fragment before and after selection, allowing us to calculate a depletion value for each fragment based on its change in frequency. Depletion values ranged from highly depleted (2 - 6% of starting frequency) to neutral (Fig. 4.1A, lower panel).

While most (981/1318, 74.4%) in-frame fragments had little to no effect, 337 fragments (25.6%) showed dominant negative activity, with depletion below 1/16 of their starting frequency. We aligned these to the *URA3* gene, controlling for coverage biases by also aligning equivalent numbers of in-frame fragments with neutral depletion scores and of out-of-frame fragments to establish a per-position depletion score. Depleted fragments and per-position depletion scores were highly correlated (r^2=0.94, r^2=0.9) across replicates. The greatest magnitude of depletion encompassed amino acids 95 to 135 (Fig. 4.1B). This region corresponds to the Ura3 dimer interface (homodimeric contacts at R92, K93, A95, D96, N99, T100, L103, Q104, H122, G123, V124, Q135, L153, Y164, R235), and it also contains key residues in the active site required to convert orotidine-5'-phosphate into uridine monophosphate (Fig. 4.1B, below). A local minimum within the dominant negative region occurs at amino acid 100, a threonine residue that makes a critical contact to the substrate (Fig. 4.1B). The smallest and most depleted dominant negative inhibitor was a 26-amino acid segment (S131-K156), and the most frequent region covered by small inhibitors (<30 amino acids) was near the active site (fragments: 97-123, 101-127, 102-128, and 105-131). The inhibitory region (Fig. 4.1C, red) contains four a-helices and two b-strands. Two smaller fragments from this region (T100-K142, K142-I183) reduced growth in media lacking uracil by 25%, while an out-of-frame control fragment had no effect (Fig. 4.1D).

We carried out a similar experiment with the transcription factor Hsf1, which contains three conserved domains: a DNA-binding domain, a coiled-coil helix required for homo-trimerization, and a repressive autoregulatory helix. Hsf1 activates genes required for both basal growth at

30°C and the heat-shock response at 37°C, and is essential at both temperatures. We generated a library of >12,000 fragments of *HSF1,* transformed yeast, induced expression and grew cells at either 30°C or 37°C. Most fragments showed no change in depletion across the selection at either temperature, but 377 (19.8%) in-frame fragments were significantly depleted (>2 sd) under basal condition, and 436 (23%) under heat shock. These fragment sets largely overlapped (319/494, 64.6%), suggesting that most dominant negative fragments of Hsf1 act independent of temperature. Alignments of depleted and control fragment sets reveal three regions with dominant negative effects (Fig. 4.2A), corresponding to the known domains. Dominant negative fragments of the trimerization domain were more depleted under heat shock, as were fragments of the DNA-binding domain, albeit to a lesser extent (Fig. 4.2A, lower panel). We also observed a subtle depletion in fragments containing the autoregulatory domain (Fig. 4.2B), despite the observation that this domain was nonfunctional in *S. cerevisiae* (Liu, Liu, Santoro, & Thiele, 1997).

We measured both basal and heat shock growth rates of cells expressing defined dominant negative fragments. The DNA-binding domain had the largest deleterious effect, which was similar at 30°C and 37°C (Fig. 4.2B). In contrast, the trimerization domain reduced growth at 30°C by only 4%, but under heat-shock by 56% (Fig. 4.2B). This effect is consistent with previous dominant negative Hsf1 mutants, and suggests that the trimerization domain alone produces nonfunctional trimers with wild-type Hsf1 and inhibits its heat shock function. The autoregulatory domain had no effect at 30°C (Fig. 4.2B, left panel), but reduced growth at 37°C by 15%, suggesting similar activity as the autoregulatory domain of the human protein. The minima of the dominant negative domains (Fig. 4.2C) were mapped onto the crystal structure. Of

two minima in the DNA-binding domain, one (DBD_A, in red, Fig. 4.2C) centered around a metal-coordinating residue necessary for the winged-helix fold, while the other (DBD_B) centered on the recognition helix that contacts the major groove of DNA (Fig. 4.2D). Three minima adjacent to and in the trimerization domain included two regions in the linker sequence that precedes the predicted trimerization helix, and a region spanning the first two heptad repeats of the predicted trimerization helix (Fig. 4.2C).

To test this approach across many genes at once, we transformed yeast cells with a library of >172,000 fragments derived from cDNA and subjected them to a growth selection over 48 hours to obtain depletion scores (Fig. 4.3A). We expected that the use of cDNA would enrich for fragments from abundantly expressed proteins. Individual fragment enrichments correlated (r^2 = 0.95) across replicates. From sequencing of the plasmid library, 132,267 fragments had adequate coverage (at least 5 reads per fragment). Of the 6,713 annotated ORFs in the yeast genome, 3,311 had at least one in-frame fragment. Of these 3,311, 1,866 had at least one strongly (1/250 of starting frequency) depleted in-frame fragment, representing possible dominant negative inhibitors. The 5,000 strongest dominant negative polypeptides were analyzed for unique biochemical properties compared to neutral polypeptides. Dominant negative fragments showed similar hydrophobicity and charge relative to the neutral set ($p > 0.01$), but significantly higher propensity for secondary structure and significantly lower flexibility (p = 2e-8, p = 1e-7).

We next sought to identify gene regions enriched for dominant negative fragments. We identified the 20,000 most depleted in-frame fragments and corresponding control fragment sets and mapped them back to individual genes. We filtered this set to those genes with at least 5 in-frame

fragments, identifying 949 genes (corresponding to 25,340 fragments) with sufficient coverage. We next filtered for genes with at least 4 overlapping dominant negative fragments, and enrichment over control sets, leaving a set of 157 genes for which a dominant negative region could be identified.

The regions of most of these 157 genes were enriched for dominant negative fragments of size 40-100 amino acids. The dominant negative hits showed significant functional enrichment for components of the ribosome and for translation overall, consistent with their essential nature and their mRNA abundance (Fig. 4.3B). Fourteen ribosomal proteins contained an enriched region of overlapping dominant negative fragments, eleven derived from the cytoplasmic ribosome and three from the mitochondrial ribosome. Seven of the cytoplasmic ribosomal proteins could be mapped onto the cryo-EM structure of the eukaryotic translating ribosome (Fig. 4.3A). We also observed several hits in other essential processes, including tRNA synthetase genes, enriched in their domains for anticodon recognition; this domain has been shown to act as a dominant negative for a tRNA synthetase(Michaels, Schimmel, Shiba, & Miller, 1996) (Fig. 4.3B).

To identify condition-specific inhibitors, we applied a heat stress during the growth of cells expressing the cDNA library. The large majority (97.8%) of in-frame fragments depleted in basal conditions were also depleted under heat shock (Fig. 4.3D). We found 508 in-frame polypeptide fragments with strong depletion (> 4-fold) only in heat shock, with the most extreme 26 of these fully dropped out in the heat-shock selection (Fig. 4.3D). The most strongly depleted fragment corresponded to the linker region preceding the trimerization domain of Hsf1, which we had earlier identified as having temperature-specific dominant negative activity. We also found

evidence for temperature-specific inhibitors of translation, identifying dominant negative fragments specific to heat shock from two translation initiation factors (*TIF3* encoding eIF3i, and *TIF34* encoding eIF-4B), and an alanyl-tRNA synthetase (*ALA1*); each displays heat-triggered aggregation (Wallace et al., 2015). That fragments of genes expressed in both basal and heat-shock conditions inhibited growth only under heat stress suggests that they target a stress-specific function. We found about twice as many fragments (1248) that showed the opposite trend, with strong dominant negative activity (depleted > 4-fold, 126 fully dropped out) in basal growth conditions but not under heat shock. The high temperature may affect the accessibility of the target or the binding interaction itself.

4.3 Discussion

The method developed here can be applied broadly to specific genes or entire libraries from any organism with suitable genetic tools. It can be readily adapted to screens and selections besides viability, including ones that can be performed under multiple conditions to obtain condition-specific inhibitors. The resulting data should allow the definition of functional domains corresponding to interaction surfaces, particularly domains that mediate homo-oligomerization or protein-protein interaction. Some of these inhibitory polypeptides may disrupt protein complexes by displacing native components with truncated versions.

Remarkably, the method is capable of pinpointing residues corresponding to local minima that may be of special relevance, such as their acting as contact sites for DNA or small molecules. These exceptional residues can serve as anchor points when selecting the smallest possible polypeptide fragment with dominant negative activity. Such small fragments may be useful in

the design of therapeutics(Arkin & Wells, 2004; Chevalier et al., 2017). Although the dominant negatives we identified are from yeast, the homology of many of them to proteins from human cells suggests that similar small fragments of mammalian proteins can be found.

4.4 METHODS

Fragment library construction and transformation. For individual gene libraries, the *URA3* and *HSF1* genes were amplified via PCR, and fragmented to an average length of 200bp using Illumina Nextera transposition. To generate a cDNA-derived fragment library, RNA was isolated from yeast cells via acid-phenol extraction, polyA-specific cDNA was synthesized, and subsequently transposed to obtain gene fragments. Transposed fragments were amplified via PCR and cloned into a low copy, galactose-inducible expression vector. Because each fragment contained a 5' adaptor sequence (AGATGTGTATAAGAGACAG) and 3' adaptor sequence (CTGTCTCTTATACACATCT) resulting from the transposition, the vector was designed to translate through these sequences such that each perfectly in-frame fragment would be translated as **MKDVYKRQ**-NNN-**LSLIHILTD****. Electromax E. coli cells were transformed with the library assemblies, and used to amplify a large pool of plasmid for subsequent yeast transformation. Yeast cells (W303 strain) were transformed with fragment libraries such that each cell received a single fragment on a low-copy plasmid with a *TRP1* gene for plasmid selection.

Cell growth selections. For *URA3* experiment, yeast cells containing the *URA3* fragment library were grown up without selection for Ura3 function in synthetic complete medium lacking tryptophan (SC -TRP, for plasmid selection only). After 2 hours of pre-induction with galactose,

these cells were then washed to remove excess media, and transferred to induction media lacking uracil (SC -TRP, -URA, +galactose) for the selective growth condition. After reaching log phase, the cells were back-diluted and grown again in fresh SC -TRP -URA, +galactose for another round of growth selection. The *HSF1* experiment was conducted in a similar fashion, except that after the 2 hour pre-induction, the culture was split such that one population continued growth at 30°C (SC -TRP,+galactose), and the second population was shifted to 37°C (SC -TRP, +galactose). Each culture was back-diluted after reaching log phase to undergo additional growth selection. Growth selection for cDNA experiment were conducted identically to the *HSF1* experiment. In all growth selections, care was taken to transfer enough cells for at least a 1000-fold coverage of the initial fragment library size.

High-throughput sequencing and analysis of polypeptide fragments. Sequencing was completed on Illumina's NextSeq platform. DNA sequencing libraries were prepared from plasmids extracted from yeast populations (Yeast Plasmid Miniprep II, Zymo Research, Irvine, CA) before and after selection, as well as the initial plasmid input. These plasmids were used as template for PCR amplification (all libraries < 15 cycles) that added sequencing adaptors, as well as 8bp sample indexes. Paired-end reads spanning the fragment library were generated at a median of 1 million reads per sample for smaller individual gene fragment libraries, and between 15-50 million reads per sample for the larger cDNA fragment libraries. The position and orientation of each fragment was determined by aligning each set of reads to the respective gene of origin, or to the full set of coding ORFs (SGD). Reads were aligned using Bowtie2. The frequency of each fragment was determined by counting unique fragments with identical gene of

origin, DNA start site, and DNA stop site, and orientation (ex. HSF1_202_431_+). Each fragment was assigned a reading frame based on its start position.

Analysis of biochemical properties of dominant negative inhibitors. Hydrophobicity, net, charge, secondary structure propensity, and flexibility were calculated using the "Peptides" R package. Each metric represents a principal component calculated from over 600 known biochemical features of amino acids.

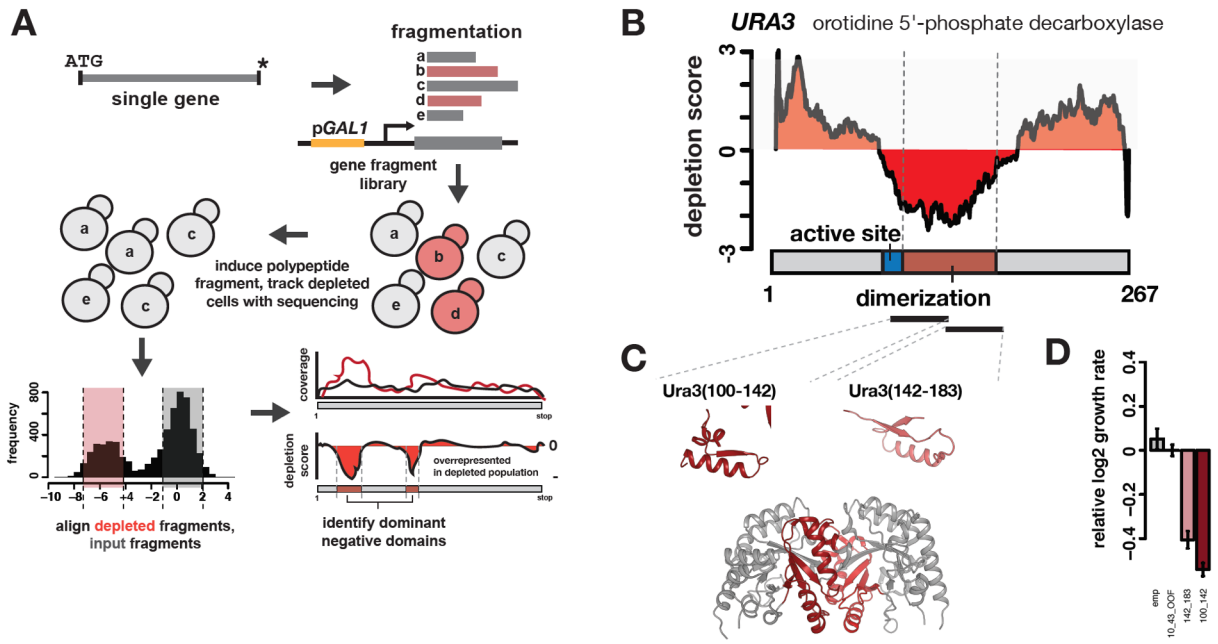


Figure 4.1

Dominant negative polypeptides can be identified in high-throughput and correspond to known protein domains. (A) A schematic of the experimental and computational pipeline to identify dominant negative polypeptides in high-throughput. (B) Dominant negative polypeptides are enriched, relative to neutral polypeptides, in a central region of the *URA3* gene. Negative depletion scores indicate regions with excess dominant negative fragments. A model of the Ura3 protein is shown below, with known domains highlighted. (C) The dominant negative region identified in (B) was mapped onto the crystal structure of the Ura3 homodimer; these region covers nearly all residues in the homodimer interface. Two individual fragments (amino acids 100-142 and amino acids142-183) selected for individual validation are shown. (D) Barplots showing individually validated growth rates for yeast cells containing a wild-type copy of the *URA3* gene and one of three fragments (aa 10-43 out of frame, aa 100-142, aa 142 – 183).

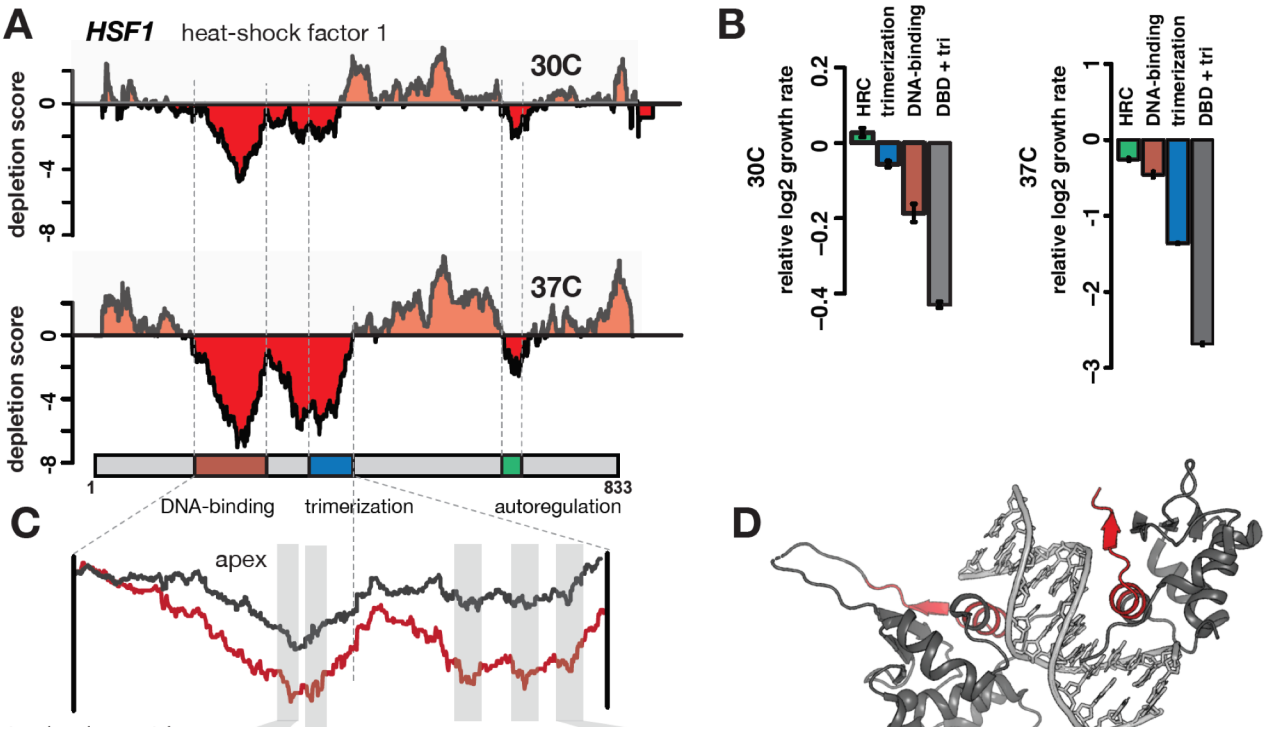


Figure 4.2

Identification of conditional dominant negative polypeptides in the essential heat-shock transcription factor Hsf1. (A) Plots showing depletion scores as in 4.1 (B) for dominant negative identification in the *HSF1* gene. Depletion scores for cells growth in basal conditions are shown in the top panel, and scores for cells grown under heat-shock are shown in the lower panel. All dominant negative regions overlapped one of the known domains of the *HSF1* gene (indicated with dashed lines). (B) Individually validated growth rates for the regions identified in (A) reveal increased dominant negative activity of a trimerization domain fragment under heat-shock conditions, as well as a role for the short autoregulatory domain. (C) A zoomed-in view of depletion scores from (A) focusing on the DNA-binding and trimerization domains. Basal score is shown in black, and heat-shock is shown in red. Local minima in the depletion scores are indicated with gray boxes. (D) Local minima from (C) are mapped onto the known structure of the *Kluveromyces lactis* Hsf1 DNA-binding domain.

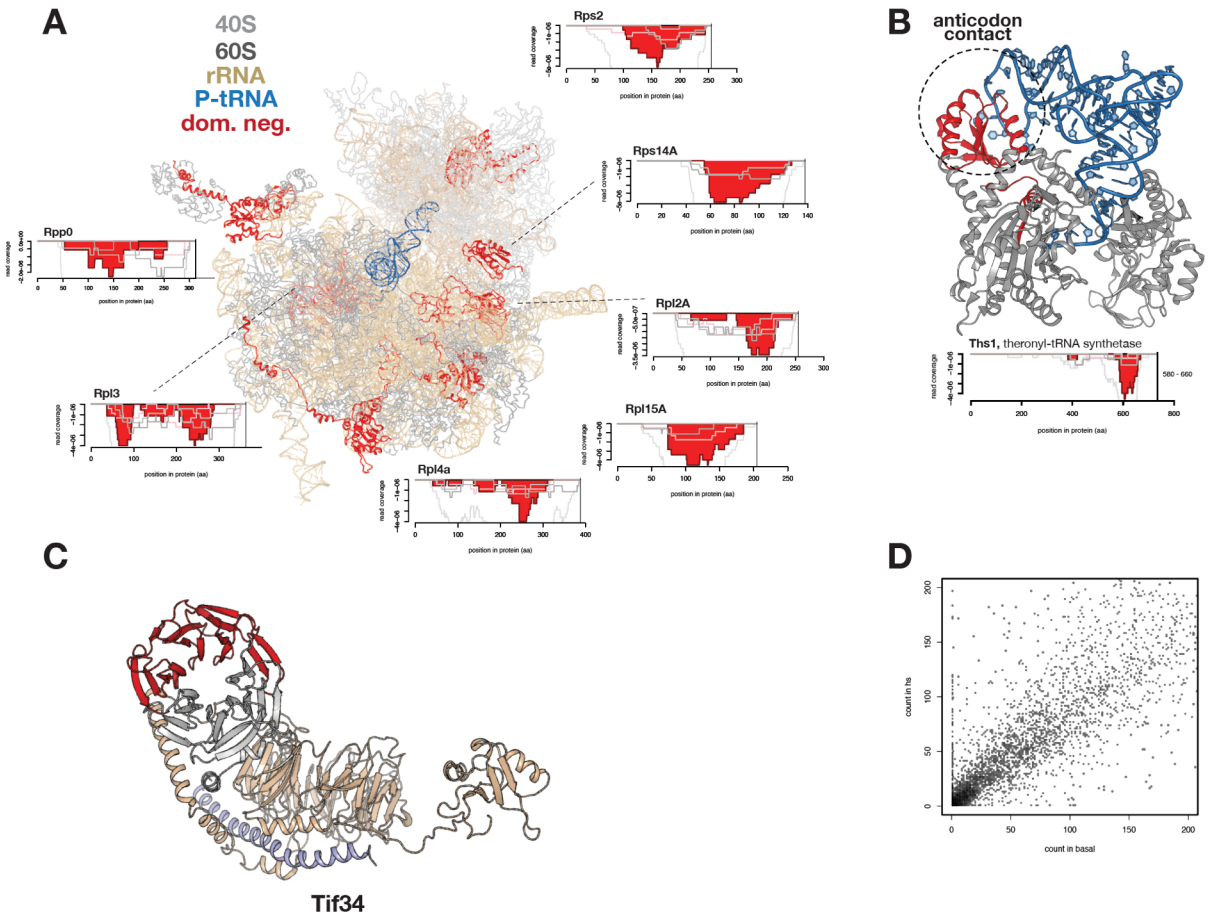


Figure 4.3

Novel inhibitors of translation identified by a full-genome identification of dominant negatives. (A) Several genes with an enriched number of dominant negative polypeptides were found to be components of the ribosome. These genes were mapped onto the cryo-EM structure of the translating ribosome, where the small subunit (40S) is shown in grey, the large subunit (60S) is shown in darker grey, the 18S and 25S rRNAs are shown in gold, the P-site tRNA is shown in blue, and protein components with dominant negative polypeptides are shown in red. Around the structure are depletion plots that show particular domains in these genes (names above) with enrichments for dominant negative polypeptides. (B) The dominant negative region identified (below) was mapped onto the crystal structure of the theonyl-tRNA synthetase (grey) in complex with its cognate tRNA (blue). The dominant negative domain (red) corresponded to the anticodon binding region of the tRNA synthetase. (C) A dominant negative fragment from the translation initiation factor subunit 3 (yeast gene name *TIF34*) whose activity was specific to the heat-shock condition was mapped onto the crystal structure of this protein. (D) The frequency of each fragment in basal (x-axis) and heat-shock (y-axis) conditions shows that many fragments lose activity under heat-shock, and a subset gain dominant negative activity in under heat-shock.

Chapter 5. IDENTIFICATION OF ENHANCER SEQUENCES IN PLANT GENOMES

5.1 INTRODUCTION

Plants respond to environmental stimuli by tightly controlled changes in gene expression, requiring a dynamic regulatory network. Despite this need, a key mode of dynamic regulation, long-range activation or repression, appears to be absent from chromatin interaction datasets (3C, Hi-C) in *Arabidopsis thaliana.* The apparent lack of distal physical interactions among chromosomes suggests that *Arabidopsis* may not use typical enhancers, though they are frequent among other higher eukaryotes. We describe the use of an alternative method, STARR-seq, to identify and functionally characterize enhancers. This method relies on simultaneously measuring the activity of a library of gene constructs, each containing a fragment of genomic DNA inserted into a transcribed sequence. If a fragment does not possess an enhancer element, the construct is expressed only at low levels due to the minimal promoter driving transcription. If a fragment is able to confer distal activation, the gene is expressed at higher level and the functional enhancer sequence, codified in the transcript, increases in abundance. We show the efficacy of the plant-optimized method by testing a known enhancer and demonstrate the scale to which the method can leverage next-generation sequencing to pinpoint functional enhancers in *Arabidopsis* and maize. We foresee this method being useful for

addressing biological questions of gene regulation, as well as for application to emerging challenges in crop design.

5.2 RESULTS

A library of >8000 fragments was generated from a pUC19 vector containing a control enhancer derived from the 35S cauliflower mosaic virus. These fragments were cloned into the STARR expression vector downstream of the minimal plant promoter, and within the coding sequence of a GUS reporter gene. Most sequence fragments in this library are not expected to have enhancer activity, as the 35S enhancer sequence comprises only 300 base pairs of the total 3000 base pairs of the plasmid used to generate the library (Figure 5.1B). We transformed *Agrobacterium* cells with this library, and infiltrated *Nicotiana benthamiana* leaves with the *Agrobacterium* population. We allowed 48 hours for expression of putative sequence fragments in leaves, and extracted RNA to quantify plant-expressed transcripts. We generated sequencing libraries for both the input *Agrobacterium* population, and the output cDNA synthesized from the plant-expressed RNA. High-throughput sequencing was used to determine the frequency of each fragment in the input population and after leaf expression. Any fragments enriched in the plant-expressed fraction presumably represent sequences with enhancer activity in *N. benthamiana* leaves. Fragments from the plant-expressed fraction were mapped back to the 35S plasmid. We found an enrichment of fragments in the plant-expressed fraction covering the 35S enhancer region (Figure 5.1C). This suggests that this method can distinguish enhancer sequences from those without enhancer activity. The same set of individual fragments with enhancer activity were identified in two replicates of the experiment (Figure 5.1D).

5.3 METHODS

Sequencing analysis. Mapping and analysis methods were identical to those used in previous applications of STARR-seq in *Drosophila* S2 cells.

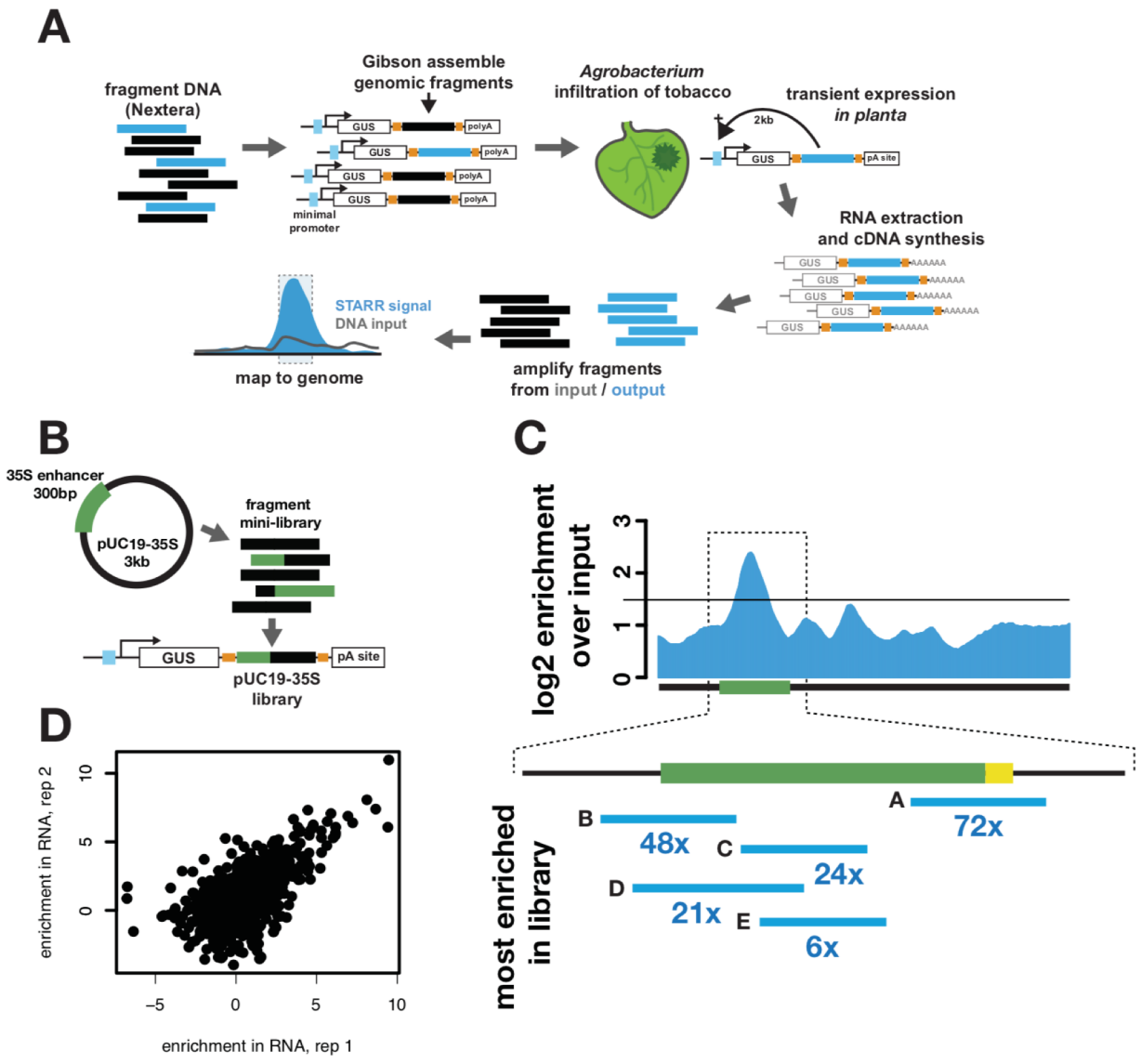


Figure 5.1

Identification of enhancer sequences in live plant tissues. (A) A schematic detailing the method for high-throughput identification of enhancer sequences in plants using STARR-seq. (B) A library of fragments was generated using a pUC19 vector (black) with an insert of the 35S enhancer (green). These fragments were cloned into the STARR-seq expression vector containing a minimal promoter (light blue), the GUS coding sequence, and the fragment library preceding the polyA signal. (C) Plot showing coverage of plant-expressed fragments compared to input fragments. Increasing values on the y-axis indicate a base was more likely to be covered in the plant-expressed population. A linear representation of the input pUC19-35S plasmid is shown below, and the 35S enhancer region contains a peak in the plant-expressed fragment coverage enrichment. A zoom-in of the enhancer region is shown below to highlight the most enriched individual fragments (blue). Each fragments enrichment relative to the input is indicated. (D) A scatterplot showing the enrichment of each fragment in two replicates of leaf tissue expression. The most enriched fragments are found at high levels in both plant-expressed output populations.

Chapter 6. CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I detailed my work to analyze the link between protein function and complex traits. In studies of two key transcriptional regulators in yeast, I showed patterns of mutation that have drastic deleterious effects on trait expression. However, in both cases I found exceptional variants that confer phenotypes not easily predicted by known functions of either protein. I also developed a novel method for the identification of dominant negative mutants derived from libraries of natural polypeptide fragments, and showed the broad applicability of this method in growth selections. In these studies, I also examined the set of mutants in the presence of a high-temperature stress, and, in all cases, found temperature-sensitive effects.

6.1 THE FUTURE OF HIGH-THROUGHPUT PROTEIN SCIENCE

Population scale selection for function paired to individual genotype tracking is an invaluable tool for facile assignment of function to large pools of genotypes. The use of this approach in a population whose genotypes comprise all possible point mutants in a gene is one of many applications, but the utility of the general scheme for a high-throughput sequencing-based analysis of protein (or other function) has not yet been pushed to its limit. DNA-sequencing based readouts allow us to take advantage of layered information encoded in DNA. This information includes amino acid sequence, reading frame, locus-of-origin, splicing signals, and translation effects through synonymous base changes. Critically, all of these features can

simultaneously be read from a sequence which also acts as a unique identifier, providing fertile ground for sequencing-based method development to interrogate protein function.

There are limits to the general scheme of a DNA-sequencing-based analysis of protein function. One of these limits derives from a property often taken for granted in cell-based selections: DNA variant and the expression of its unique protein variant are confined to a single cell. Therefore, the frequency of the DNA sequence encoding any variant tied to its effect on cellular fitness. Unfortunately, this connection is lost if a similar experiment is performed outside a cellular context. For example, a sensitive selection could be performed on a very large pool of DNA-encoded transcription factor variants, transcribed and translated *in vitro,* for their ability to bind a DNA site, but sequencing could not be used to track the association between a variant's identity and its DNA-binding ability. There are promising technologies that may be able to support an experiment like the one described, if unique DNA-sequences can be covalently linked to protein variants (Chan, McGregor, Jain, & Liu, 2017).

For *in vitro* applications in particular, for which the numbers of tested variants can potentially far exceed *in vivo* methods, I suspect that many useful technologies will be based on proteomic methods. Proteomic measurements have some practical advantages over DNA-sequencing-based interrogation of protein function in that direct measurement of protein variants obviates the link between a protein and its encoding DNA. A modified version of the previous example of an *in vitro* DNA-binding experiment, could consist instead of two populations variants, those of transcription factor *A*, and those of transcription factor *B*. The goal would be to determine which variants cooperatively bind as heterodimers to DNA. Collecting heterodimeric, DNA-bound complexes, cross-linking them, and measuring the frequency of cross-linked variant pairs on a mass spectrometer would allow the identification of variants that are capable of

cooperative binding. Completing this experiment in a cell-based method like those used throughout my thesis work is far more challenging, because of the lack of a physical link between the two DNA variants in a cell, even though the expressed protein variants themselves may come into physical contact.

6.2 ENGINEERING TEMPERATURE SENSING

It seems likely that all proteins are capable of temperature sensing, given the right variant of the protein. This contention is supported by the discovery of temperature-sensitive alleles of many essential genes in *S. cerevisiae* and other model organisms. The alleles were initially used as tools – a conditional means to test the function of genes whose full knock-out caused embryonic lethality, or failure to grow. These temperature-sensitive alleles allowed researchers to assign function to essential genes. The availability of many temperature-sensitive alleles in deep mutational scanning experiments (S. Zimmerman, personal communication), as well as the presence of temperature-sensitive alleles in most genes, betray a simple fact: most proteins are only one or two amino acid changes away from temperature-sensitivity.

However, the types of mutations we have observed to confer temperature sensitivity do not appear to have a discernable pattern. Thus, predicting temperature-sensitive variants in other proteins would be challenging or impossible with only our current datasets. The variability in the types of mutations that confer temperature-sensitivity probably derives from the fact that mutations that change local or global protein stability are context-dependent. We found temperature sensitive mutants generally occurred at low frequencies, but these mutants were found in several different proteins of varying structure and molecular function (e.g. Ste12, Ste7, Ste5). From this result, I would predict that naturally occurring proteins have few mutational

paths to temperature-sensitivity, but these paths will be challenging to predict since specific mutations that destabilize the protein will constantly change depending on the context of an individual protein species.

BIBLIOGRAPHY

Akada, R., Yamamoto, J., & Yamashita, I. (1997). Screening and identification of yeast sequences that cause growth inhibition when overexpressed. *Molecular and General Genetics*, *254*(3), 267–274. https://doi.org/10.1007/s004380050415

Araya, C. L., Fowler, D. M., Chen, W., Muniez, I., Kelly, J. W., & Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences*, *109*(42), 16858–16863. https://doi.org/10.1073/pnas.1209751109

Arkin, M. M. R., & Wells, J. A. (2004). Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. *Nature Reviews Drug Discovery*, *3*(4), 301–317. https://doi.org/10.1038/nrd1343

Attfield, P. V. (1997). Stress tolerance: The key to effective strains of industrial baker's yeast. *Nature Biotechnology*, *15*(13), 1351–1357. https://doi.org/10.1038/nbt1297-1351

Balasubramanian, S., Sureshkumar, S., Lempe, J., & Weigel, D. (2006). Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. *PLoS Genetics*, *2*(7), 0980–0989. https://doi.org/10.1371/journal.pgen.0020106

Baler, R., Dahl, G., & Voellmy, R. (1993). Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. *Molecular and Cellular Biology*, *13*(4), 2486–2496. https://doi.org/10.1128/MCB.13.4.2486.Updated

Barrera, L. A., Vedenko, A., Kurland, J. V., Rogers, J. M., Gisselbrecht, S. S., Rossin, E. J., … Bulyk, M. L. (2016). Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science*, *351*(6280), 1450–1454. https://doi.org/10.1126/science.aad2257

Betz, J. L., & Fall, M. Z. (1988). Effects of dominant-negative lac repressor mutations on operator specificity and protein stability. *Gene*, *67*(2), 147–158. https://doi.org/10.1016/0378-1119(88)90392-7

Beyhan, S., Gutierrez, M., Voorhies, M., & Sil, A. (2013). A Temperature-Responsive Network Links Cell Shape and Virulence Traits in a Primary Fungal Pathogen. *PLoS Biology*, *11*(7). https://doi.org/10.1371/journal.pbio.1001614

Bhagavatula, G., Rich, M. S., Young, D. L., Marin, M., & Fields, S. (2017). A Massively Parallel Fluorescence Assay to Characterize the Effects of Synonymous Mutations on *TP53*

Expression. *Molecular Cancer Research*, (9), molcanres.0245.2017. https://doi.org/10.1158/1541-7786.MCR-17-0245

Bloom, J. D., Labthavikul, S. T., Otey, C. R., & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences*, *103*(15), 5869–5874. https://doi.org/10.1073/pnas.0510098103

Boeke, J. D., La Croute, F., & Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular and General Genetics MGG*, *197*(2), 345–346.

Boer, D. R., Freire-Rios, A., Van Den Berg, W. A. M., Saaki, T., Manfield, I. W., Kepinski, S., … Coll, M. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell*, *156*(3), 577–589. https://doi.org/10.1016/j.cell.201312.027

Boer, V. M., De Winde, J. H., Pronk, J. T., & Piper, M. D. W. (2003). The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *Journal of Biological Chemistry*, *278*(5), 3265–3274. https://doi.org/10.1074/jbc.M209759200

Bonner, J. J., Ballou, C., & Fackenthal, D. L. (1994). Interactions between DNA-bound trimers of the yeast heat shock factor. *Molecular and Cellular Biology*, *14*(1), 501–508. https://doi.org/10.1128/MCB.14.1.501.Updated

Botstein, D., & Fink, G. R. (2011). Yeast: An experimental organism for 21st century biology. *Genetics*, *189*(3), 695–704. https://doi.org/10.1534/genetics.111.130765

Boyer, J., Badis, G., Fairhead, C., Talla, E., Hantraye, F., Fabre, E., … Dujon, B. (2004). Large-scale exploration of growth inhibition caused by overexpression of genomic fragments in Saccharomyces cerevisiae. *Genome Biology*, *5*(9), R72. https://doi.org/10.1186/gb-2004-5-9-r72

Broach, James R., Strathern, Jeffrey N., Hicks, J. B. (1979). Transformation in yeast: development of a hybrid cloning vector and isolation of the. *Gene*, *8*, 121–133.

Chan, A. I., McGregor, L. M., Jain, T., & Liu, D. R. (2017). Discovery of a Covalent Kinase Inhibitor from a DNA-Encoded Small-Molecule Library × Protein Library Selection. *Journal of the American Chemical Society*, *139*(30), 10192–10195. https://doi.org/10.1021/jacs.7b04880

Chevalier, A., Silva, D.-A., Rocklin, G. J., Hicks, D. R., Vergara, R., Murapa, P., … Baker, D.

(2017). Massively parallel de novo protein design for targeted therapeutics. *Nature*, *550*(7674), 74–79. https://doi.org/10.1038/nature23912

Chou, S., Lane, S., & Liu, H. (2006). Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *26*(13), 4794–4805. https://doi.org/10.1128/MCB.02053-05

Clapham, D. E., & Miller, C. (2011). A thermodynamic framework for understanding temperature sensing by transient receptor potential (TRP) channels. *Proceedings of the National Academy of Sciences*, *108*(49), 19492–19497. https://doi.org/10.1073/pnas.1117485108

Clos, J., Rabindran, S., Wisniewski, J., & Wu, C. (1993). Induction temperature of human heat shock factor is reprogrammed in a Drosophila cell environment. *Nature*, *364*(6434), 252–255. https://doi.org/10.1038/364252a0

Cook, J. G., Bardwell, L., & Thorner, J. (1997). Inhibitory and activating functions for MAPK Kss1 in the S. cerevisiae filamentous-growth signalling pathway. *Nature*, *390*(6655), 85–88. https://doi.org/10.1038/36355

Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res*, *14*, 1188–1190. https://doi.org/10.1101/gr.849004.1

Cuperus, J. T., Lo, R. S., Shumaker, L., Proctor, J., & Fields, S. (2015). A tetO Toolkit to Alter Expression of Genes in Saccharomyces cerevisiae. *ACS Synthetic Biology*, *4*(7), 842–852. https://doi.org/10.1021/sb500363y

Dai, C., & Sampson, S. B. (2016). HSF1: Guardian of Proteostasis in Cancer. *Trends in Cell Biology*, *26*(1), 17–28. https://doi.org/10.1016/j.tcb.2015.10.011

Dolan, J. W., Kirkman, C., & Fields, S. (1989). The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5703–5707.

Dong, H., & Courchesne, W. (1998). A novel quantitative mating assay for the fungal pathogen Cryptococcus neoformans provides insight into signalling pathways responding to nutrients and temperature. *Microbiology (Reading, England)*, *144 (Pt 6*, 1691–1697.

Dong, H., & Courchesne, W. (1998). A novel quantitative mating assay for the fungal pathogen Cryptococcus neoformans provides insight into signalling pathways responding to nutrients

and temperature. *Microbiology*, *144*(6), 1691–1697. https://doi.org/10.1099/00221287-144-6-1691

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: A protein secondary structure prediction server. *Nucleic Acids Research*, *43*(W1), W389–W394. https://doi.org/10.1093/nar/gkv332

Dunham, M. J., & Fowler, D. M. (2013). Contemporary, yeast-based approaches to understanding human genetic variation. *Current Opinion in Genetics and Development*, *23*(6), 658–664. https://doi.org/10.1016/j.gde.2013.10.001

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Errede, B., & Ammerer, G. (1989). STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. *Genes & Development*, *3*(9), 1349–1361. https://doi.org/10.1101/gad.3.9.1349

Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D., & Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. *Nature Methods*, *7*(9), 741–746. https://doi.org/10.1038/nmeth.1492

Fowler, D. M., & Fields, S. (2014). Deep mutational scanning: A new style of protein science. *Nature Methods*, *11*(8), 801–807. https://doi.org/10.1038/nmeth.3027

Fraser, J. A., Giles, S. S., Wenink, E. C., Geunes-Boyer, S. G., Wright, J. R., Diezmann, S., … Heitman, J. (2005). Same-sex mating and the origin of the Vancouver Island Cryptococcus gattii outbreak. *Nature*, *437*(7063), 1360–1364. https://doi.org/10.1038/nature04220

Gallo, G. J., Prentice, H., & Kingston, R. E. (1993). Heat Shock Factor Is Required for Growth at Normal Temperatures in the Fission Yeast Schizosaccharomyces pombe, *13*(2), 749–761.

Gauthier, G. M., & Keller, N. P. (2013). Crossover fungal pathogens: The biology and pathogenesis of fungi capable of crossing kingdoms to infect plants and humans. *Fungal Genetics and Biology*, *61*, 146–157. https://doi.org/10.1016/j.fgb.2013.08.016

Gietz, R. D., & Woods, R. a. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology*, *350*(2001), 87–96.

Gomez-Pastor, R., Burchfiel, E. T., & Thiele, D. J. (2018). Regulation of heat shock transcription factors and their roles in physiology and disease. *Nature Reviews Molecular Cell Biology*, *19*(1), 4–19. https://doi.org/10.1038/nrm.2017.73

Goodsell, D. S., & Olson, A. J. (2000). S s p f.

Goodson, Michael L., Sarge, K. D. (1995). Heat-inducible DNA binding of purified heat shock transcription factor 1. *J. Biol. Chem.*, *2447*(50).

Gordan, R., Murphy, K., McCord, R., Zhu, C., Vedenko, A., & Bulyk, M. (2011). Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biology*, *12*(12), R125. https://doi.org/10.1186/gb-2011-12-12-r125

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Hahn, J., Hu, Z., Thiele, D. J., & Iyer, V. R. (2004). Genome-Wide Analysis of the Biology of Stress Responses through Heat Shock Transcription Factor. *Society*, *24*(12), 5249–5256. https://doi.org/10.1128/MCB.24.12.5249

Hajdu-Cronin, Y. M., Chen, W. J., & Sternberg, P. W. (2004). The L-type cyclin CYL-1 and the heat-shock-factor HSF-1 are required for heat-shock-induced protein expression in Caenorhabditis elegans. *Genetics*, *168*(4), 1937–1949. https://doi.org/10.1534/genetics.104.028423

Hartwell, L. H. (1967). Macromolecular synthesis in temperature-sensitive mutants of yeast. *J. Bacteriol*, *93*(5), 1662–1670.

Hartwell, L. H., & McLaughlin, C. S. (1968a). Mutants of yeast with temperature-sensitive isoleucyl-tRNA synthetases. *Proc Natl Acad Sci U S A*, *59*(2), 422–428. https://doi.org/10.1073/pnas.59.2.422

Hartwell, L. H., & McLaughlin, C. S. (1968b). Temperature-sensitive mutants of yeast exhibiting a rapid inhibition of protein synthesis. *J Bacteriol*, *96*(5), 1664–1671. Retrieved from http://jb.asm.org/cgi/reprint/96/5/1664?view=long&pmid=5726307

Hashikawa, N., Yamamoto, N., & Sakurai, H. (2007). Different mechanisms are involved in the transcriptional activation by yeast heat shock transcription factor through two different types of heat shock elements. *Journal of Biological Chemistry*, *282*(14), 10333–10340. https://doi.org/10.1074/jbc.M609708200

He, H., Soncin, F., Grammatikakis, N., Li, Y., Siganou, A., Gong, J., … Calderwood, S. K. (2003). Elevated expression of heat shock factor (HSF) 2A stimulates HSF1-induced transcription during stress. *Journal of Biological Chemistry*, *278*(37), 35465–35475.

https://doi.org/10.1074/jbc.M304663200

Heise, B., van der Felden, J., Kern, S., Malcher, M., Brückner, S., & Mösch, H. U. (2010). The TEA transcription factor Tec1 confers promoter-specific gene regulation by Ste12-dependent and -independent mechanisms. *Eukaryotic Cell*, *9*(4), 514–531. https://doi.org/10.1128/EC.00251-09

Hentze, N., Breton, L. Le, Wiesner, J., Kempf, G., & Mayer, M. P. (2016). Molecular mechanism of thermosensory function of human heat shock transcription factor Hsf1, (January), 1–24. https://doi.org/10.7554/eLife.11576

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature*, *329*(6136), 219–222. https://doi.org/10.1038/329219a0

Hietpas, R. T., Jensen, J. D., & Bolon, D. N. A. (2011). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences*, *108*(19), 7896–7901. https://doi.org/10.1073/pnas.1016024108

Hoi, J. W. S., & Dumas, B. (2010). Ste12 and Ste12-like proteins, fungal transcription factors regulating development and pathogenicity. *Eukaryotic Cell*, *9*(4), 480–485. https://doi.org/10.1128/EC.00333-09

Holbrook, J. A., Neu-Yilik, G., Hentze, M. W., & Kulozik, A. E. (2004). Nonsense-mediated decay approaches the clinic. *Nature Genetics*, *36*(8), 801–808. https://doi.org/10.1038/ng1403

Hsu, A.-L., Murphy, C. T., & Kenyon, C. (2003). Regulation of aging and age- related disease by DAF- 16 and heat- shock factor. *Science (New York, N.Y.)*, *300*(5622), 1142.

Hurme, R., Berndt, K. D., Normark, S. J., & Rhen, M. (1997). A proteinaceous gene regulatory thermometer in Salmonella. *Cell*, *90*(1), 55–64. https://doi.org/10.1016/S0092-8674(00)80313-X

Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundbäck, T., Nordlund, P., & Molina, D. M. (2014). The cellular thermal shift assay for evaluating drug target interactions in cells. *Nature Protocols*, *9*(9), 2100–2122. https://doi.org/10.1038/nprot.2014.138

Jarosz, D. F., Taipale, M., & Lindquist, S. (2010). Protein Homeostasis and the Phenotypic Manifestation of Genetic Diversity: Principles and Mechanisms. *Annual Review of Genetics*, *44*(1), 189–216. https://doi.org/10.1146/annurev.genet.40.110405.090412

Jedlicka, P., Mortin, M. A., & Wu, C. (1997). Multiple functions of Drosophila heat shock transcription factor in vivo, *16*(9), 2452–2462.

Johnson, M. A., Waterham, H. R., Ksheminska, G. P., Fayura, L. R., Cereghino, J. L., Stasyk, O. V., … Cregg, J. M. (1999). Positive selection of novel peroxisome biogenesis-defective mutants of the yeast Pichia pastoris. *Genetics*, *151*(4), 1379–1391.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., … Taipale, J. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities, 861–873. https://doi.org/10.1101/gr.100552.109.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., … Taipale, J. (2013). DNA-binding specificities of human transcription factors. *Cell*, *152*(1–2), 327–339. https://doi.org/10.1016/j.cell.2012.12.009

Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., … Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, *527*(7578), 384–388. https://doi.org/10.1038/nature15518

Jones, S. K., & Bennett, R. J. (2011). Fungal mating pheromones: Choreographing the dating game. *Fungal Genetics and Biology*, *48*(7), 668–676. https://doi.org/10.1016/j.fgb.2011.04.001

Larson, J. S., Schuetz, T. J., & Kingston, R. E. (1995). In Vitro Activation of Purified Human Heat Shock Factor by Heat. *Biochemistry*, *34*(6), 1902–1911. https://doi.org/10.1021/bi00006a011

Leu, J.-Y., & Murray, A. W. (2006). Experimental evolution of mating discrimination in budding yeast. *Current Biology : CB*, *16*(3), 280–286. https://doi.org/10.1016/j.cub.2005.12.028

Lindquist, S. (1981). Regulation of protein synthesis during heat shock. *Nature*, *293*(24).

Liu, X. D., Liu, P. C., Santoro, N., & Thiele, D. J. (1997). Conservation of a stress response: human heat shock transcription factors functionally substitute for yeast HSF. *The EMBO Journal*, *16*(21), 6466–6477. https://doi.org/10.1093/emboj/16.21.6466

Louvion, J. F., Abbas-Terki, T., & Picard, D. (1998). Hsp90 is required for pheromone signaling in yeast. *Molecular Biology of the Cell*, *9*(11), 3071–3083. https://doi.org/10.1091/mbc.9.11.3071

Ma, X., Xu, L., Alberobello, A. T., Gavrilova, O., Bagattin, A., Skarulis, M., … Mueller, E. (2015). Celastrol protects against obesity and metabolic dysfunction through activation of a

HSF1-PGC1α transcriptional axis. *Cell Metabolism*, *22*(4), 695–708. https://doi.org/10.1016/j.cmet.2015.08.005

Madhani, H. D. (1997). Combinatorial Control Required for the Specificity of Yeast MAPK Signaling. *Science*, *275*(5304), 1314–1317. https://doi.org/10.1126/science.275.5304.1314

Madhani, H. D., & Fink, G. R. (1998). The control of filamentous differentiation and virulence in fungi. *Trends in Cell Biology*, *8*(9), 348–353. https://doi.org/10.1016/S0962-8924(98)01298-7

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404–405. https://doi.org/10.1093/bioinformatics/16.4.404

Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R., & Fields, S. (2013). Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *Rna*, *19*(11), 1537–1551. https://doi.org/10.1261/rna.040709.113

Mendgen, K., Hahn, M., & Deising, H. (1996). Morphogenesis and Mechanisms of Penetration By Plant Pathogenic Fungi. *Annual Review of Phytopathology*, *34*(1), 367–386. https://doi.org/10.1146/annurev.phyto.34.1.367

Mendillo, M. L., Santagata, S., Koeva, M., Bell, G. W., Hu, R., Tamimi, R. M., … Lindquist, S. (2012). HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell*, *150*(3), 549–562. https://doi.org/10.1016/j.cell.2012.06.031

Michaels, J. E. A., Schimmel, P., Shiba, K., & Miller, W. T. (1996). Dominant negative inhibition by fragments of a monomeric enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(25), 14452. https://doi.org/10.1073/pnas.93.25.14452

Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W. T., Zhang, W., … Hughes, T. R. (2004). Exploration of essential gene functions via titratable promoter alleles. *Cell*, *118*(1), 31–44. https://doi.org/10.1016/j.cell.2004.06.013

Mody, A., Weiner, J., & Ramanathan, S. (2009). Modularity of MAP kinases allows deformation of their signalling pathways. *Nature Cell Biology*, *11*(4), 484–491. https://doi.org/10.1038/ncb1856

Moriggl, R., Gouilleux-Gruart, V., Jähne, R., Berchtold, S., Gartmann, C., Liu, X., … Gouilleux, F. (1996). Deletion of the carboxyl-terminal transactivation domain of MGF-Stat5 results in sustained DNA binding and a dominant negative phenotype. *Molecular and Cellular*

Biology, *16*(10), 5691–5700. https://doi.org/10.1128/MCB.16.10.5691

Morimoto, R. I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes & Development*, *12*(24), 3788–3796. https://doi.org/10.1101/gad.12.24.3788

Morton, E. A., & Lamitina, T. (2013). Caenorhabditis elegans HSF-1 is an essential nuclear protein that forms stress granule-like structures following heat shock. *Aging Cell*, *12*(1), 112–120. https://doi.org/10.1111/acel.12024

Mumberg, D., Müller, R., & Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, *156*(1), 119–122.

Olson, K. A., Nelson, C., Tai, G., Hung, W., Yong, C., Astell, C., … Olson, K. A. M. Y. (2000). Two Regulators of Ste12p Inhibit Pheromone-Responsive Transcription by Separate Mechanisms Two Regulators of Ste12p Inhibit Pheromone-Responsive Transcription by Separate Mechanisms. *Molecular and Cellular Biology*. https://doi.org/10.1128/MCB.20.12.4199-4209.2000.Updated

Perfect, J. R. (2006). Cryptococcus neoformans: The yeast that likes it hot. *FEMS Yeast Research*, *6*(4), 463–468. https://doi.org/10.1111/j.1567-1364.2006.00051.x

Peteranderl, R., Rabenstein, M., Shin, Y. K., Liu, C. W., Wemmer, D. E., King, D. S., & Nelson, H. C. (1999). Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. *Biochemistry*, *38*(12), 3559–3569. https://doi.org/10.1021/bi981774j

Phosphorylation, T., Sorger, P. K., & Pelham, H. R. B. (1988). Yeast Heat Shock Factor Is an Essential DNA-Binding Protein That Exhibits, *54*, 855–864.

Posas, F., Witten, E. a, & Saito, H. (1998). Requirement of STE50 for osmostress-induced activation of the STE11 mitogen-activated protein kinase kinase kinase in the high-osmolarity glycerol response pathway. *Molecular and Cellular Biology*, *18*(10), 5788–5796. https://doi.org/10.1128/MCB.18.10.5788

Prinz, S., Avila-Campillo, I., Aldridge, C., Srinivasan, A., Dimitrov, K., Siegel, A. F., & Galitski, T. (2004). Control of yeast filamentous-form growth by modules in an integrated molecular network. *Genome Research*, *14*, 380–390. https://doi.org/10.1101/gr.2020604

Ramer, S. W., Elledge, S. J., & Davis, R. W. (1992). Dominant genetics using a yeast genomic library under the control of a strong inducible promoter. *Proceedings of the National*

Academy of Sciences of the United States of America, *89*(23), 11589–11593. https://doi.org/10.1073/pnas.89.23.11589

Reddy, P., & Hahn, S. (1991). Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region. *Cell*, *65*(2), 349–357. https://doi.org/10.1016/0092-8674(91)90168-X

Rich, M. S., Payen, C., Rubin, A. F., Ong, G. T., Sanchez, M. R., Yachie, N., … Fields, S. (2016). Comprehensive analysis of the sul1 promoter of saccharomyces cerevisiae. *Genetics*, *203*(1), 191–202. https://doi.org/10.1534/genetics.116.188037

Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., … Hughes, T. R. (2000). Signaling and Circuitry of Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles, *287*(February), 873–881.

Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2017). A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1272-5

Rubin, A. F., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2016). Enrich2: a statistical framework for analyzing deep mutational scanning data. https://doi.org/10.1101/075150

Ryan, O., Shapiro, R. S., Kurat, C. F., Mayhew, D., Baryshnikova, A., Chin, B., … Boone, C. (2012). Global gene deletion analysis exploring yeast filamentous growth. *Science*, *337*(6100), 1352–1356. https://doi.org/10.1126/science.1224339

Sakurai, H., & Enoki, Y. (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. *FEBS Journal*, *277*(20), 4140–4149. https://doi.org/10.1111/j.1742-4658.2010.07829.x

Sancak, Y., Peterson, T. R., Shaul, Y. D., Lindquist, R. A., Thoreen, C. C., Bar-peled, L., & Sabatini, D. M. (2008). The Rag GTPases Bind Raptor and Mediate Amino Acid Signaling to mTORC1, (June), 1496–1502. https://doi.org/10.1126/science.1158042

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sangster, T. A., Lindquist, S., & Queitsch, C. (2004). Under cover: Causes, effects and implications of Hsp90-mediated genetic capacitance. *BioEssays*, *26*(4), 348–362. https://doi.org/10.1002/bies.20020

Shalgi, R., Hurt, J. A., Krykbaeva, I., Taipale, M., Lindquist, S., & Burge, C. B. (2013). Widespread Regulation of Translation by Elongation Pausing in Heat Shock. *Molecular Cell*, *49*(3), 439–452. https://doi.org/10.1016/j.molcel.2012.11.028

Sherif, M., Waung, D., Korbeci, B., Mavisakalyan, V., Flick, R., Brown, G., … Master, E. R. (2013). Biochemical studies of the multicopper oxidase (small laccase) from Streptomyces coelicolor using bioactive phytochemicals and site-directed mutagenesis. *Microbial Biotechnology*, *6*(5), 588–597. https://doi.org/10.1111/1751-7915.12068

Shi, J., Wang, E., Milazzo, J. P., Wang, Z., Kinney, J. B., & Vakoc, C. R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nature Biotechnology*, *33*(6), 661–667. https://doi.org/10.1038/nbt.3235

Shoval, O. (2012). Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. *Science*, *1157*. https://doi.org/10.1126/science.1217405

Singh, V., & Aballay, A. (2006). Heat-shock transcription factor (HSF)-1 pathway required for Caenorhabditis elegans immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(35), 13092–13097. https://doi.org/10.1073/pnas.0604050103

Sorger, P. K., & Nelson, H. C. M. (1989). Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell*, *59*(5), 807–813. https://doi.org/10.1016/0092-8674(89)90604-1

Starita, L. M., Pruneda, J. N., Lo, R. S., Fowler, D. M., Kim, H. J., Hiatt, J. B., … Klevit, R. E. (2013). Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proceedings of the National Academy of Sciences*, *110*(14), E1263–E1272. https://doi.org/10.1073/pnas.1303309110

Starita, L. M., Young, D. L., Islam, M., Kitzman, J. O., Gullingsrud, J., Hause, R. J., … Fields, S. (2015). Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics*, *200*(2), 413–422. https://doi.org/10.1534/genetics.115.175802

Stevenson, L. F., Kennedy, B. K., & Harlow, E. (2001). A large-scale overexpression screen in Saccharomyces cerevisiae identifies previously uncharacterized cell cycle genes. *Proceedings of the National Academy of Sciences*, *98*(7), 3946–3951. https://doi.org/10.1073/pnas.051013498

Su, T. C., Tamarkina, E., & Sadowski, I. (2010). Organizational constraints on Ste12 cis-elements for a pheromone response in Saccharomyces cerevisiae. *FEBS Journal*, *277*(15), 3235–3248. https://doi.org/10.1111/j.1742-4658.2010.07728.x

Takemori, Y., Enoki, Y., Yamamoto, N., Fukai, Y., Adachi, K., & Sakurai, H. (2009). Mutational analysis of human heat-shock transcription factor 1 reveals a regulatory role for oligomerization in DNA-binding specificity. *Biochemical Journal*, *424*(2), 253–261. https://doi.org/10.1042/BJ20090922

Thein, S. L., Hesketh, C., Taylor, P., Temperley, I. J., Hutchinson, R. M., Old, J. M., … Weatherall, D. J. (1990). Molecular basis for dominantly inherited inclusion body beta-thalassemia. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(10), 3924–3928. https://doi.org/10.1073/pnas.87.10.3924

Waddington, C. H. (1960). Experiments on canalizing selection. *Genetics Research*, *1*(1), 140–150.

Wallace, E. W. J., Kear-Scott, J. L., Pilipenko, E. V., Schwartz, M. H., Laskowski, P. R., Rojek, A. E., … Drummond, D. A. (2015). Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell*, *162*(6), 1286–1298. https://doi.org/10.1016/j.cell.2015.08.041

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., … Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science*, *350*(6264), 1096–1101. https://doi.org/10.1126/science.aac7041

Winzeler, E. a. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, *285*(5429), 901–906. https://doi.org/10.1126/science.285.5429.901

Xiao, H., Perisic, O., & Lis, J. T. (1991). Cooperative binding of drosophila heat shock factor to arrays of a conserved 5 bp unit. *Cell*, *64*(3), 585–593. https://doi.org/10.1016/0092-8674(91)90242-Q

Xiao, X. Z., Zuo, X. X., Davis, A. A., McMillan, D. R., Curry, B. B., Richardson, J. A., & Benjamin, I. J. (1999). HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice. *EMBO Journal*, *18*(21), 5943–5952. https://doi.org/10.1093/emboj/18.21.5943

Yuan, Y. L., & Fields, S. (1991). Properties of the DNA-binding domain of the Saccharomyces cerevisiae STE12 protein. *Molecular and Cellular Biology*, *11*(12), 5910–5918.

Yuan, Y. L. O., Stroke, I. L., & Fields, S. (1993). Coupling of cell identity to signal response in yeast: Interaction between the alphal and STE12 proteins. *Genes and Development*, *7*(8),

1584–1597. https://doi.org/10.1101/gad.7.8.1584

Zeitlinger, J., Simon, I., Harbison, C. T., Hannett, N. M., Volkert, T. L., Fink, G. R., & Young, R. A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell*, *113*(3), 395–404. https://doi.org/10.1016/S0092-8674(03)00301-5

Zhong, M., Orosz, A., & Wu, C. (1998). Direct sensing of heat and oxidation by Drosophila heat shock transcription factor. *Molecular Cell*, *2*(1), 101–108. https://doi.org/10.1016/S1097-2765(00)80118-5

VITA

Michael Dorrity earned a Bachelor of Science (BS) in genetics from the University of California, Davis. While at UC Davis, he worked on quantitative genetics of root development traits in the domestic tomato *Solanum lycopersicum.* Before enrolling at the University of Washington, he worked for a year in the lab of Siobhan Brady at the University of California, Davis to develop high-throughput sequencing methods in tomato.

Outside of the lab, Michael performs with the Seattle ensemble Gamelan Pacifica, frequents Seattle art museums, and rides his bike out of the city.