Quantitatively Assessing the Effects of Mutations on Signaling Pathway Function

by

Purnima Srividya Kompella

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Cell and Systems Biology
University of Toronto

Quantitatively Assessing the Effects of Mutations on Signaling Pathway Function

Purnima Srividya Kompella

Doctor of Philosophy

Department of Cell and Systems Biology
University of Toronto

2019

Abstract

Cells receive diverse stimuli from their surroundings and process them into distinct physiological responses through signal transduction pathways. Mitogen-activated protein kinase pathways are responsible for many cellular functions. The overarching goal of this thesis was to study the effects of mutations on MAPK signaling function using complementary systems biology and genetic engineering methods.

First, I asked how mutations could rescue a decrease in signaling caused by overexpression. I identified several variants of the yeast MAPK, Fus3, that rescued mating function in an overexpression background. In addition to finding Fus3 variants that rescued pathway function by improving kinase activity, I found variants containing premature stop codons (PSC), which are naturally readthrough, that rescued pathway function. The location of the mutated residue in the protein structure and the identity of the residue play a role in determining the impact of the PSC on pathway function. These results suggest that one way in which selection can compensate for protein overexpression is by introducing PSCs at permitted positions.

Next, I quantified the effects of mutations on yeast mating and HOG signaling. I used the information theory measure of mutual information to measure signaling and introduced change

in mutual information as the measure of the effects of mutations on signaling. I showed that information transmission in HOG signaling is more robust to spontaneous mutations than mating signaling. I controlled for some of the possible reasons for this difference, redundancy and mutational target size, and found that although HOG signaling mutational robustness is compromised without the redundancy, it is still more robust than the mating signaling suggesting that other mechanisms play a role in maintaining this mutational robustness. Finally, I did a direct comparison of the effects of the mutations in a shared component on information transmission in the pathways and found that large effect mutations in this protein have large effects on amount of mutual information in both signaling pathways and that some of these effects are opposite.

Overall, in this thesis work, I used various mutagenesis methods to test the effects of mutations on signaling in living cells. In the first part, I used an engineered random mutant library of a protein kinase to assess the functional significance of mutations on a yeast MAPK pathway. In the second part, along with genetic engineering, I used experimental evolution to quantify the mutational robustness of two yeast MAPK pathways. We found both non-intuitive and intuitive results and ways to quantify both. Applying the tools of systems biology and genetic engineering is a comprehensive and complementary approach to answering evolutionary questions.

Acknowledgments

Firstly, I would like to thank my graduate supervisor Dr. Alan Moses for his support and guidance. Dr. Moses is an inspiring advisor and I am grateful for his patience and encouragements. I also thank him for giving me the opportunity to complete my doctoral research in his team after my first supervisor left the department. I feel fortunate to have been part of his team and of the Cell and Systems Biology Department, as I received extensive support to present my research at international conferences and to be exposed to new ideas in those meetings. I am also thankful to Dr. Peter McCourt and Dr. Francis Ouellette for their direction and advice as part of my supervisory committee; their help throughout these years has been an invaluable asset to my research. I thank Dr. David Guttman for his guidance. I would also like to thank Dr. Sergio Peisajovich for giving me the chance to start my research at the University of Toronto. I must also mention my gratitude towards Ian Buglass for his help navigating the paperwork at CSB and for his timely responses to my numerous questions over the years.

To the members of the Moses and Peisajovich labs both past and present, thank you for creating an atmosphere where I could find support not only at the scientific level, but also at the personal level. Reta Aram, Ian Hsu, Alex Lu, Paloma Sato, Benjamin Scott, Bob Strome, Clarissa Wirianto, Shadi Zabad, Taraneh Zarin; you have made an invaluable contribution to my career as a scientist and to my well-being, and I thank you for that.

Pursuing a doctoral degree is a privilege. I thank my parents, Usha and Subbarao, for this opportunity, for the sacrifices they made and for their incredible support and love. I thank my sister, Yashoda, and her family, Prasad, Shashwat and Upasini, for their love, kindness and understanding throughout these years. I love you all so much. I thank my grandmother, Annapurna, and all family and friends for their support. I also thank my late grandparents, Suryanarayanamurthy, Yasoda, and Sriram, for their love and support, and for encouraging me to pursue my dreams. Finally, I thank my husband, Julio, for his love, patience, support, and guidance. I love you more than I could know to write.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BFP	Blue Fluorescent Protein
FACS	Fluorescence Activated Cell Sorting
GDP	Guanosine Diphosphate
GEF	Guanine nucleotide exchange factor
GFP	Green Fluorescent Protein
Gln	Glutamine
GPCR	G-protein coupled receptor
GTP	Guanosine Triphosphate
HOG	High-Osmolarity Glycerol
HOT	Highly Optimized Tolerance
IDR	Intrinsically Disordered Region
KCl	Potassium Chloride
MAPK	Mitogen-Activated Protein Kinase
MAPKK	MAPK Kinase
MAPKKK	MAPKK Kinase
MI	Mutual Information
mRNA	messenger RNA
PDB	Protein Data Base
Pro	Proline

PCR	Polymerase Chain Reaction
PROVEAN	Protein Variation Effect Analyzer
PSC	Premature Stop Codon
RA	Ras-associating domain
RNA	Ribonucleic Acid
SAM	Sterile Alpha Motif
Ser	Serine
Thr	Threonine
Tyr	Tyrosine
VWA	Von Willebrand factor type A domain
WT	Wildtype
YPD	Yeast Extract Peptone Dextrose

1 Chapter 1: General Introduction

1.1 Abstract

Cells receive diverse stimuli from their surroundings and process them into distinct physiological responses through signal transduction pathways. These pathways, and biological systems in general, exhibit robustness to perturbations such as mutations. The effect of mutations on signaling networks is area of current research interest as mutations in signaling genes such as mitogen-activated protein kinases, are implicated in many diseases. Mitogen-activated protein kinase pathways are responsible for many cellular functions and are conserved across eukaryotes. *Saccharomyces cerevisiae, a single-cell eukaryote,* is an excellent model system to study these signaling networks.

1.2 Robustness

1.2.1 Origins of Robustness

Biological organisms have survived and propagated themselves for billions of years despite genetic and environmental changes suggesting that they are incredibly robust to such perturbations. Why aren't complex living things more fragile? There is some debate regarding the origins of mutational robustness - if it is an intrinsic property of the genetic and physiological constraints of the system, evolved independently due to selection against the deleterious effects of mutation ("direct selection" hypothesis) or as a by-product of selection for environmental robustness, the "congruence" hypothesis (Masel and Trotter 2010). Evidence exists both for direction selection and congruence hypotheses and it is as yet unknown whether mutational robustness is a by-product of environmental robustness (Proulx *et al.,* 2007). Regardless of the origin, the consequence of mutational robustness is the accumulation of neutral genetic variants, that is mutations that do not affect the fitness of an organism (Wagner 2005). This cryptic genetic variation allows potential evolutionary innovations - solutions to challenges such as a mutation that is neutral in the current environment being adaptive in a new environment - to be present in a population. In this way, robustness promotes evolvability, the capacity to generate heritable variation.

1.2.2 Evolutionary Consequences of Robustness

Robustness, resisting change, and evolvability, allowing change, might seem contradictory, but we can resolve this contradiction with some clarification. Genotypic robustness and genotypic evolvability share an antagonistic relationship but phenotypic robustness and phenotypic evolvability do not (Wagner 2008). Mechanisms of robustness that block genetic changes from occurring, such as proofreading and repair, hinder evolvability whereas, robustness of a phenotype to genetic change ("mutational robustness") promotes evolvability (Kirschner and Gerhart 1998; Lenski *et al.,* 2006; Masel and Trotter 2010; Wagner 2008). In fact, although the distribution of fitness effects of new mutations is strongly bimodal with one mode representing lethal mutations and the other mode representing small effect mutations that are slightly deleterious, cryptic genetic variation is enriched for potentially adaptive alleles (Eyre-Walker and Keightley 2007). This is because when the effects of mutations is partially masked, small effect mutations would be effectively neutral but lethal mutations would still exhibit enough lethality that selection would remove them (Masel and Siegal 2009). This type of pre-adaptation can occur at different levels - for example, Bloom *et al*. showed that mutations that increase protein stability but are neutral with respect to protein function can allow the protein to take on subsequent, functionally beneficial but destabilizing mutations (Bloom *et al.,* 2006). This idea of partial robustness is also supported by findings by Draghi *et al*. that populations with intermediate levels of robustness adapt faster than those with little or no robustness (Draghi *et al.,* 2010). They show that the relationship between mutational robustness and evolvability is dependent on the population size, mutation rate and the structure of the fitness landscape.

1.2.3 Mechanisms of Robustness

Robust systems are not ones that do not change regardless of perturbations, but are instead ones that maintain specific function(s) in the face of perturbations (Kitano 2004). In order for a system to be robust, it may have to change its mode of operation in a flexible way. Proposed mechanisms for mutational robustness include feedback loops, redundancy, buffering, and modularity (Masel and Siegal 2009). Robust responses are achieved by both positive and

negative feedback loops with the former amplifying the stimuli to allow the system to distinguish the stimulated state from the unstimulated one and the latter allowing to system to adapt (Kitano 2004). Redundancy is achieved either by having multiple identical components or more commonly, by having distinct components with overlapping functions. It aids robustness as in the case of one component failing to perform its function, the other can substitute and rescue the function. However, Wagner argues that distributing robustness throughout the network is more important to the organism than having spare parts (Wagner 2007). The evidence for buffering as a mechanism of mutational robustness is circumstantial - while the deletion of a molecular chaperone reveals cryptic genetic variation, since these variants are not new mutations, it is premature to say that this molecular chaperone aids in mutational robustness (Masel and Siegal 2009). Nevertheless, the requirement of only about 20% of the ~6000 *Saccaharomyces cerevisiae* genes for cell viability suggests the evolution of extensive buffering against genetic perturbations (Dixon *et al.,* 2009). A key design principle of signaling networks that may contribute to their robustness is modularity – the ability to separate the function of an entity (for example, a motif responsible for docking interaction) from other entities and allows a complex system to be broken down into simpler parts (Hartwell *et al.,* 1999; Moses and Landry 2010; Pawson 1988; Pereira-Leal *et al.,* 2006). A specific example of a module contributing to robustness is the three-tiered mitogen-activated protein kinase (MAPK) cascade evolutionarily conserved across eukaryotes to be both robust and adaptable (Tian and Harding 2014). Sato et al., showed that a yeast MAPK pathway is robust to domain rearrangements that change protein interactions as long as catalytic activity is not impaired (Sato *et al.,* 2014). Previous work also shows that domain rearrangements in this MAPK pathway can generate novel phenotypes (Peisajovich *et al.,* 2010).

Another mechanism that aids robustness is the complexity of a network defined by the number of components in a pathway, the connections among these components and the spatial relationship between the components (Weng *et al.,* 1999). Costanzo and colleagues mapped the genetic interactions of ~75% of *S. cerevisiae* genes and found that most genes had very few interactions and that deletions of the few genes that are highly connected, network hubs, resulted in severe fitness defects (Costanzo *et al.,* 2010). This design principle of minimizing the number of connections between genes seems to decrease negative fitness effects of deleting any individual gene. The robustness of a network is closely tied to its complexity with higher complexity

leading to higher robustness (Soyer and Bonhoeffer 2006). Knowing the number of genes affecting a trait is not enough to determine robustness as the other feature to keep in mind is the effect size of each gene on that trait. It is known for example, that the mean effect size of all genes on all traits is reduced as a way to gain mutational robustness and the more important a trait is to the fitness of an organism, the smaller the mean effect size of the genes involved in that trait (Ho and Zhang 2014). Although complexity optimizes the system for specific perturbations, one of its downsides is making the system hypersensitive to design flaws or unexpected perturbations resulting in what Carlson and Doyle call, the "robust, yet fragile" system (Carlson and Doyle 2002). They suggest that this inherent and unavoidable fragility is the trade-off for robustness exemplified in the highly optimized tolerance (HOT) structures of biological systems. A study by Kim *et al*., showed this design characteristic in human signaling networks (Kim *et al.,* 2014). They decomposed the networks into an evolvable core and a robust core, finding that there exists a subgroup of interactions that promote mutational robustness and another distinct subgroup of interactions that promote evolvability.

1.3 Effects of mutations on signaling pathway

Many cellular processes from cell division to cell death require cells to receive diverse extracellular and intracellular cues and process them into distinct and appropriate physiological responses through signal transduction pathways (Berggard *et al.,* 2007). Through versatile, well-regulated and redundant signaling components, cells ensure that these important functions are performed with fidelity (Azeloglu and Iyengar 2015). Understanding how signaling pathways respond to mutations can reveal how cells process information (Laub 2016). For example, if information processing is very robust to mutations, this suggests that cells have mechanisms that adapt or compensate when signaling input is not producing the expected output, or that cellular information processing has many parallel redundant processes, so that even if some of these may be perturbed, signaling function is retained. In any case, robustness to mutations is expected under more complex signaling networks, while classical linear cascades are expected to be more affected by mutations (Soyer and Bonhoeffer 2006; Weng *et al.,* 1999).

Studying how mutations affect the function of signaling networks is also important for understanding the mechanisms by which signaling pathways evolve. Little is known in general about the evolutionary mechanisms underlying signaling pathway evolution, but diversification of signaling molecules is correlated with evolutionary increases in organismal complexity (Mody *et al.,* 2009). Pathways capable of tolerating mutations can allow the accumulation of large neutral variation in populations. Since diverse neutral pools of mutations are reservoirs for potentially adaptive variation, robust pathways may facilitate adaptation (Draghi *et al.,* 2010; Wagner 2008).

Finally, improper function of signaling pathways is implicated in various human diseases including Alzheimer's disease, Parkinson's disease and many cancers (Kim and Choi 2010). In many of these disorders, disease causing mutations produce quantitative effects on signaling pathways (Creixell *et al.,* 2015). Of interest also are the pleiotropic effects of mutations since mutations in many disease-associated genes affect multiple traits (Solovieff *et al.,* 2013).

1.4 *Saccharomyces cerevisiae*

Yeast is an ideal model in which to study eukaryotic cellular processes and the genome sequences of many yeast species are publicly available (Dujon 2010). Yeast-based model systems have been used to study human pathologies indicating their relevance to human disease (Khurana and Lindquist 2010; McGary *et al.,* 2010; Treusch *et al.,* 2011). In recent years, various studies have also used yeast to study evolutionary processes such as adaptive evolution (Gresham *et al.,* 2008; Venkataram *et al.,* 2016), speciation (Leducq *et al.,* 2017), effects of rare mutations (Zhu *et al.,* 2017), evolutionary dynamics of mutations (Levy *et al.,* 2015), regulatory dynamics of gene expression (Metzger *et al.,* 2017), hybrid genome stability (Gibson *et al.,* 2017), genetic suppression interactions (van Leeuwen *et al.,* 2016), selection on quantitative traits (Zarin *et al.,* 2017), and effects of domain rearrangements on signaling (Sato *et al.,* 2014).

The model organism, *Saccharomyces cerevisiae*, commonly known as baker's or brewer's yeast, is a microscopic, single cell, and genetically tractable eukaryote. It is relatively easy to manipulate the *S. cerevisiae* genome, introducing genes either on plasmids or through genomic integration, or deleting genes from yeast chromosomes through homologous recombination

(Duina *et al.,* 2014). *S. cerevisiae* is easy to propagate and divides rapidly (once every ~90 minutes) through a process of budding whereby the smaller daughter cell buds off the mother cell. This budding yeast is also the first eukaryote to have its genome sequenced over twenty years ago (Goffeau *et al.,* 1996). The 12.1 Mb genome is divided into 16 chromosomes carrying 5138 verified open reading frames (as of July 22, 2018 on the *Saccharomyces* Genome Database, http://www.yeastgenome.org) in the S288C reference strain. Studies in yeast have expanded our understanding of many basic cellular processes such as regulated cell division, cell differentiation, vesicle trafficking, allowed identification of the structure of the components necessary for transcription, and the mechanism of chaperones (Duina *et al.,* 2014). Furthermore, the core eukaryotic cellular processes were established long before the last common ancestor of yeast and human, making work in budding yeast relevant to the study of human disorders (Hoffman *et al.,* 2015).

1.5 Mitogen-Activated Protein Kinase Signaling Pathways

1.5.1 The three-tiered Mitogen Activated Protein Kinase Cascade

A frequently used protein kinase cascade in eliciting various types of cellular responses such as cell-cycle progression, stress responses and cell differentiation is the mitogen-activated protein kinase (MAPK) cascades. MAPK pathways are highly conserved across all eukaryotes (Bardwell, 2005). Many of the components of these pathways were first identified in *S. cerevisiae* (Chen and Thorner 2007). The canonical MAPK cascade contains three sequentially acting protein kinases – a MAPK kinase kinase (MAPKKK) activates a MAPK kinase (MAPKK) which activates a MAPK (Chen and Thorner 2007). MAPKKKs are serine/threonine-specific kinases that contain an N-terminal regulatory domain and a C-terminal serine/threonine protein kinase domain. They phosphorylate MAPKKs at two conserved serine or threonine residues in the activation loop. MAPKKs are dual-specificity serine/threonine or tyrosine kinase that phosphorylate the conserved Thr-X-Tyr motif in the activation loop of MAPKs. MAPKs are serine/threonine kinases that phosphorylate their targets at Ser/Thr-Pro motifs.

In *S. cerevisiae*, there are at least five mitogen-activated protein kinase (MAPK) pathways that allow cells to respond to various extracellular stimuli (Chen and Thorner 2007). In my thesis work, I have studied the mating and high osmolarity glycerol (HOG) MAPK pathways (Figure 1-1).

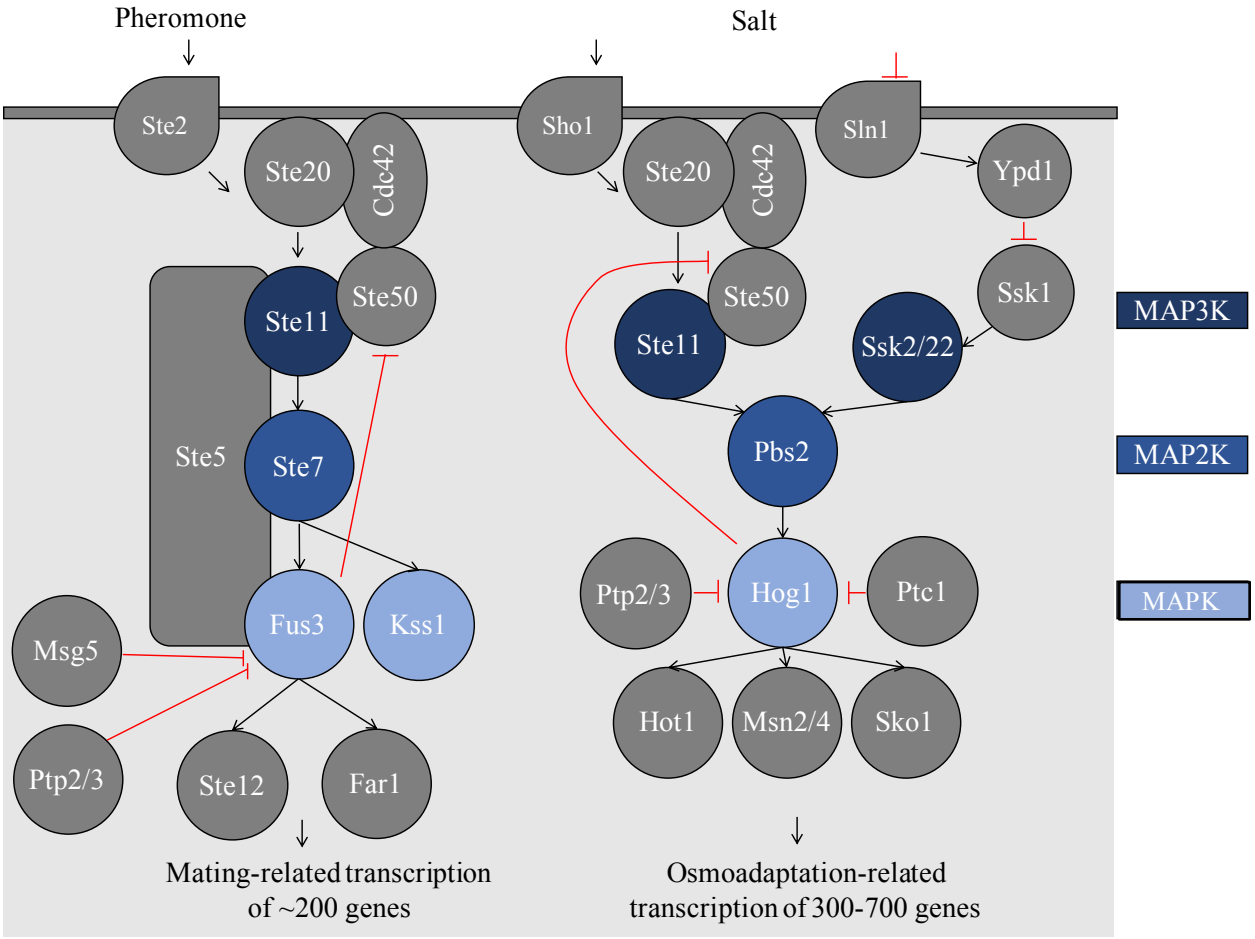


Figure 1-1: Schematic of Mating and HOG MAPK pathways. The yeast mating pathway (left) responds to pheromone and turns on mating related genes including FUS1 and the HOG pathway (right) responds to hyperosmotic stress and turns on osmoregulation related genes, including STL1. Negative regulatory interactions are indicated with red lines and positive regulatory interactions are indicated with black arrows.

1.5.2 Mating Pathway

During mating, *S. cerevisiae* haploid cells of opposite mating types, MATa (a-cell) and MATα (α-cell), communicate with each other through the secretion of peptide mating pheromones (Chen and Thorner 2007). These cells mate by cellular and nuclear fusion generating a third, diploid, cell type, MATa/MATα. The mating pheromones, a/α-factor, bind to the pheromone receptor, Ste2/3, a G-protein coupled receptor (GPCR). The α-factor that α-cells secrete bind to Ste2 on a-cells and the a-factor that a-cells secrete bind to Ste3 on α-cells. The GPCR couples to a heterotrimeric G-protein with Gα and Gβγ subunits. The Gα subunit is known as Gpa1, Gβ as Ste4 and Gγ as Ste18. The binding of the mating pheromone allows the receptor to serve as a guanine nucleotide exchange factor (GEF), facilitating the release of GDP and the binding of GTP by the Gα subunit. This frees up the Gβγ complex to recruit the scaffold protein, Ste5, to the membrane. Gα and Gβγ remain tethered to the membrane allowing the recruitment of downstream MAPKs to the membrane site with the highest ligand-occupied pheromone receptors. Another membrane-tethered protein, Bem1, brings Ste20, a p21-activated protein kinase, close to Cdc42, a small monomeric Ras-related GTPase. GTP loading activates Cdc42, which in turn allows Cdc42 to bind to the autoinhibitory CRIB motif on Ste20, relieving autoinhibition and activating Ste20. Ste20 also binds to the Gβγ complex. Ste11, the mating MAPKKK, is recruited to the membrane by Ste50, a small adapter protein that also plays a role in sustaining the mating pheromone-induced signal (Xu *et al.,* 1996). Activated Ste20 phosphorylates Ste11 and triggers the phosphorylation cascade. Ste11 phosphorylates the MAPKK, Ste7 which phosphorylates the MAPK, Fus3. This phosphorylation is dependent on the catalytic unlocking of Fus3 by the VWA domain of the scaffold protein, Ste5, to which all three MAPKs are bound (Good *et al.,* 2009). Activated Fus3 localizes to the nucleus and turns on mating related transcription factor, Ste12, resulting in the activation of about 200 genes responsible for cell cycle arrest (Far1), the formation of mating projections, called “shmoos”, toward the mating partner, and ultimate result in cellular and nuclear fusion (Fus1) (Aymoz *et al.,* 2018; Roberts *et al.,* 2000).

Exposure to pheromone results in sustained Fus3 activation and prolonged cell cycle arrest as the cells undergo the morphological changes involved in the mating response (Baltanas *et al.,* 2013). Negative regulators of Ste12, Dig1/2, prevent aberrant activation of Ste12 target genes (Bardwell 2005). To promote desensitization and recovery, post pheromone exposure, other mechanisms of

negative feedback include degradation of the pheromone by Bar1/Sst1, phosphorylation (by Fus3) and endocytosis of the pheromone-bound receptor, Sst2 phosphorylation by Fus3 enhancing its stability, Fus3 autoregulation limiting the magnitude and duration of its own phosphorylation, dephosphorylation of the various pathway kinases by phosphatases and their degradation (Bardwell 2005).

Expression of key mating proteins such as the Ste2/3 receptor, important for sensing the mating pheromone, is ensured by the basal pathway activity (Thomson *et al.,* 2011). Since mating triggers cell cycle arrest, the basal mating activity must not be high enough to trigger inappropriate growth arrest. Similarly, upon exposure to pheromone, the pathway output must be high enough to trigger the cell cycle arrest and other physiological changes necessary for the mating response. This difference between the basal and induced system output, known as the dynamic range, is large enough that the system can respond distinguishably to different concentrations of pheromone (Yu *et al.,* 2008). The stoichiometric balance of the components of the mating pathway is important for optimal pathway response. Specifically, the absolute system output and the dynamic range of the mating pathway is sensitive to the amounts of Fus3, the MAPK and Ste5, the scaffold, respectively (Thomson *et al.,* 2011). This sensitivity may also be thought of as a point of fragility in the mating pathway as mutations that increase the concentrations of these proteins would not be tolerated. Hence, the cellular concentrations of these proteins are tightly regulated to avoid negative effects on signaling function. In addition to triggering cell cycle arrest, once activated, Fus3 limits activation of Kss1, the MAPK responsible for triggering the filamentous growth response under conditions of nutrient limitation (Winters and Pryciak 2018). Therefore, improper activation of Fus3 would both trigger cell cycle arrest, and restrict cell growth under starvation.

1.5.3 High Osmolarity Glycerol Pathway

When the external osmolarity range is higher than physiological levels, yeast cells trigger a complex adaptation response that includes cell cycle arrest and synthesis and retention of the osmolyte glycerol (Saito and Posas 2012). These responses are mainly governed by the HOG pathway. There are two functionally partially redundant but mechanistically distinct HOG pathway branches, the Sln1 branch and the Sho1 branch. Although both branches respond to high

osmolarity, the Sln1 branch has a much more prominent role in responding to moderate increases in osmolarity than the Sho1 branch (O'Rourke and Herskowitz 2004). The two branches also respond more robustly and faster than one branch alone (Schaber *et al.,* 2012). Granados and colleagues also show that mutants lacking either branch of the HOG pathway pay a cost either in terms of the speed of the response or the accuracy of the response (Granados *et al.,* 2017).

The Sln1 branch is an example of a complex two-component system characterized by the autophosphorylation of Sln1 at a Histidine residue, and the following conserved phosphotransfer to an Aspartate residue also on Sln1. This phosphate is then transferred to a Histidine residue on Ypd1 and then to an Aspartate on Ssk1 (Ssk1-P). Under normal osmotic conditions, Ssk1 is constitutively phosphorylated by Ypd1 which in turn is constitutively phosphorylated by Sln1. When external osmolarity increases and cells shrink, the resulting decrease in turgor pressure inactivates Sln1. This leads to Ypd1 inactivation and unphosphorylated Ssk1 (Ssk1-OH) accumulates. This Ssk1-OH activates two functionally redundant, homologous MAPKKKs, Ssk2 and Ssk22 which turn on the phosphorylation cascade. Since Ssk1-OH triggers pathway activation, there are several ways in which the pathway minimizes aberrant activation, including degrading Ssk1-OH under normal osmotic conditions. Active Ssk2/22 phosphorylates Pbs2, the MAPKK, which goes on to phosphorylate Hog1, the MAPK.

The Sho1 branch responds to high osmolarity through interaction between various osmosensors including Sho1 which activates Ste20 and Cla4, brought to the membrane by Cdc42. These PAK-like proteins then phosphorylate the MAPKKK Ste11 (recruited to the membrane by Ste50 through association with the membrane anchor Opy2). Ste11 phosphorylates Pbs2. Sho1 and Pbs2 act as co-scaffolds in this pathway. The two branches converge on Pbs2 so downstream signal transduction proceed similarly for the two pathways.

Following activation Hog1 translocates to the nucleus and regulates several transcription factors responsible for osmoresponsive genes, including GPD1 and GPP2, which are involved in glycerol biosynthesis and STL1, encoding a glycerol/proton symporter. Genomic expression profiling studies revealed that about 300-600 genes are regulated in response to stress (Gasch *et al.,* 2000). Hog1 is activated transiently ensuring that once the osmostress is relieved, addition glycerol is not produced. This adaptation response is managed by several negative feedback mechanisms in the HOG pathway including accumulation of glycerol removing the osmostress

signal, dephosphorylation of Hog1 by protein phosphatases (resulting in the return to cytoplasm), Hog1 phosphorylation of Sho1 (perhaps causing a disruption on Sho1 oligomerization) and Ste50 (decreasing its affinity to the membrane anchor, Opy2).

1.5.4 Signal specificity

Despite the MAPK pathways sharing many components, the pathways are well insulated from another and there are several mechanisms by which leakage of signal or crosstalk between the various MAPK pathways is prevented (Bardwell 2006). One is the presence of scaffold proteins. For example, activation of the mating pathway requires the scaffold protein Ste5 and activation of the HOG pathway requires the scaffold Sho1. There are also specific docking interactions that determine pathway specificity (Mody *et al.,* 2009). MAPK pathways show examples of insulation to limit crosstalk. As evidence for the role of Hog1 in limiting a mating response to osmostress, in *hog1Δ* cells, the mating MAPK Fus3 is inappropriately activated by osmostress (O'Rourke and Herskowitz 1998). In addition to activating Fus3, Ste7 also activates the filamentation/invasive growth MAPK, Kss1. However, signal fidelity is ensured by Fus3 and Ste5 - Fus3 limits the magnitude and duration of Kss1 activation, and unlike its activation of Fus3, Ste7 activation of Kss1 does not require Ste5 (Good *et al.,* 2009; Sabbagh *et al.,* 2001). Kss1 (but not Fus3) is activated by osmotic stress however Hog1 phosphorylation of Ste50 is known to limit the duration of Kss1 activation thus preventing an inappropriate filamentation response under high osmolarity conditions (Hao *et al.,* 2008).

In addition to Fus3 and Hog1, another protein kinase, a casein kinase I, phosphorylates Ste50 at a specific threonine residue in the SAM domain (Chen and Thorner 2007). This phosphorylation is required for the mating response but not HOG response indicating the importance of Ste50 interactions in ensuring Ste11 activates the appropriate MAPK. Mutations in the SAM domain of Ste50 decrease osmoresistance and increase crosstalk with the mating and filamentation pathways under hyperosmotic conditions (Jansen *et al.,* 2001). Ste50 interacts with Ste11 via its Sterile Alpha Motif (SAM) domain and with membrane proteins Opy2, Sho1 and Cdc42 via its Ras-associating (RA) domain (Hao *et al.,* 2008; Yamamoto *et al.,* 2010). There is also a cluster of MAPK consensus phosphorylation sites in the intrinsically disordered region (IDR) that are known to modulate MAPK signaling (Zarin *et al.,* 2017).

While there is no HOG response to pheromone under normal conditions, cells that are pre-adapted to an osmolyte increase HOG related transcription in response to pheromone although this is not due to a failure to insulate the response, but due to a mating related glycerol release triggering the response to decreased osmolarity (Baltanas *et al.,* 2013).

MAPK pathways have grown and changed topology during evolution and although the MAPK cascade is highly conserved, these proteins play an important role in allowing their pathways to acquire new specificities and interaction partners (Mody *et al.,* 2009). By exploiting the modular nature of MAPK interactions and changing a few residues in the mating and HOG pathway MAPKs, Fus3 and Hog1, the specificity of these pathways can change whereby the mating pathway can be induced by osmotic stress and the HOG pathway by pheromone (Mody *et al.,* 2009). Since signaling specificity is maintained in part by docking interactions, changes to these can rewire the pathways to create new stimulus-output relationships. Systematically studying the effects of mutations in shared components on signaling specificity can help reveal the mechanisms controlling this important feature.

1.6 Flow Cytometry

Flow cytometry is an invaluable tool routinely used in the high-throughput analysis of gene expression (Duveau *et al.,* 2017; Kompella *et al.,* 2017; Metzger *et al.,* 2017; Peisajovich *et al.,* 2010; Sato *et al.,* 2014; Strome *et al.,* 2018; Zarin *et al.,* 2017). Gene expression can be monitored using a fluorescent protein expressed from the promoter of the gene of interest and using the fluorescence intensity as a readout for the promoter activity (Taher 2017).

Three main systems, fluidics, optics and electronics, make up a flow cytometer (Givan 1992). The fluidics system is responsible for transporting particles in a single stream to the laser beam whereby the optical system consisting of the excitation sources (lasers) illuminate the particles and direct the resulting light and fluorescence scatter signals to the appropriate detectors through the use of optical mirrors, and filters. These light signals are converted electronic signals, voltages, by the detectors in the electronics system. Here, the electronic pulses are converted to channel numbers corresponding to the various parameters analyzed, such as fluorescence.

One of the major strengths of flow cytometry analysis of reporter gene expression is the throughput whereby, hundreds of thousands to millions of cells can be analyzed within second to minutes (Ducrest *et al.,* 2002). These measurements generate a distribution of reporter expression levels representative of the variability in gene expression that isogenic populations display (Flores 2013). In flow cytometry, numerous fluorescent reporters can be detected simultaneously, allowing the quantification of multiple gene expression profiles (Ragan *et al.,* 2004). By using a flow cytometer equipped with Fluorescence Assisted Cell Sorting capabilities, live cell populations can be separated ("sorted") based on quantitative differences in many parameters such as expression level of multiple fluorescent proteins and cultured for further experimentation. Some of the disadvantages of flow cytometry include not being able to monitor individual cells over time and lack of spatial resolution preventing the analysis of sub-cellular features both of which can be overcome by the complementary use of the relatively lower-throughput technique of fluorescence microscopy (O'Connor 1996).

1.7 Research Objectives and thesis overview

This thesis presents a study of the effects of mutations on yeast signaling pathways. The specific aims of this thesis are:

1) Characterize variants of the yeast mating MAPK that are able to rescue over-expression driven decrease in pathway response.

2) Quantify the relative robustness of yeast mating and HOG pathways by experimental evolution and targeted mutagenesis.

In Chapter 2, we identify the variants of the yeast mating MAPK, Fus3, that rescue an over-expression driven decrease in pathway response (Kompella *et al.,* 2017). Specifically, we find that variants carrying premature stop codons are able to compensate for the over-expression. Through western blot analysis of protein levels, we show that the premature stop codons are readthrough at low levels resulting in a decrease in Fus3 levels. In Chapter 3 we quantify the robustness of the mating and HOG pathways to whole genome mutations introduced by mutation

accumulation and to mutations in a shared component introduced by error-prone PCR. We find that mutations have a stronger effect on the mating pathway.

In Chapter 4, I present a discussion of future directions.

2 Chapter 2: Introduction of Premature Stop Codons as an Evolutionary Strategy to Rescue Signaling Network Function

Previously published as: Purnima S. Kompella, Alan M. Moses, and Sergio G. Peisajovich. (2017). Introduction of Premature Stop Codons as an Evolutionary Strategy to Rescue Signaling Network Function. ACS Synthetic Biology. 6(3):446-454.

2.1 Abstract

The cellular concentrations of key components of signaling networks are tightly regulated, as deviations from their optimal ranges can have negative effects on signaling function. For example, overexpression of the yeast mating pathway mitogen-activated protein kinase (MAPK) Fus3 decreases pathway output, in part by sequestering individual components away from functional multiprotein complexes. Using a synthetic biology approach, we investigated potential mechanisms by which selection could compensate for a decrease in signaling activity caused by overexpression of Fus3. We overexpressed a library of random mutants of Fus3 and used cell sorting to select variants that rescued mating pathway activity. Our results uncovered that one remarkable way in which selection can compensate for protein overexpression is by introducing premature stop codons at permitted positions. Because of the low efficiency with which premature stop codons are read through, the resulting cellular concentration of active Fus3 returns to values within the range required for proper signaling. Our results underscore the importance of interpreting genotypic variation at the systems rather than at the individual gene level, as mutations can have opposite effects on protein and network function.

2.2 Introduction

Cells receive stimuli from their surroundings and process them into physiological responses through signal transduction pathways (Berggard *et al.,* 2007). As protein–protein interactions are crucial for the proper function of signaling pathways, it is not surprising that expression levels of individual components within signaling complexes are tightly regulated (Bekaert and Conant 2011). In fact, synthetic attempts at rewiring signaling networks need to consider how changes in the level of individual components might disrupt the precise interactions needed for function (Bashor *et al.,* 2008). Deleterious consequences of changes in the expression levels of one component of a complex were first observed in studies of transcriptional regulators (Gill and Ptashne 1988). An increase in the cellular concentration of a transcription regulator can result in transcriptional inhibition, due to the sequestration of limiting components, a phenomenon known as squelching (Cahill *et al.,* 1994; Gill and Ptashne 1988). Though one could draw easy parallels

between the effects of changes in the concentration of individual components in gene regulatory complexes and signaling complexes, still little is known about how overexpression of signaling pathway components may affect signaling interactions and consequently function. Furthermore, as mutations could affect steady state levels of signaling components, it is also interesting to explore how natural selection might compensate for such a perturbation.

Saccharomyces cerevisiae, the budding yeast, is an excellent model system to study signaling networks. Budding yeast has five pathways mediated by the canonical mitogen activated protein kinase (MAPK) cascade (Bardwell 2005). The three sequentially activated kinases in these pathways are conserved across diverse species. In *S. cerevisiae*, the mating specific MAPK, Fus3, acts in a complex with the scaffold protein, Ste5, the MAPK Kinase (MAPKK) Ste7, and MAPK Kinase Kinase (MAPKKK), Ste11. In addition, Fus3 interacts with several other proteins to propagate the signal, as well as to regulate it (Figure 2-1A).

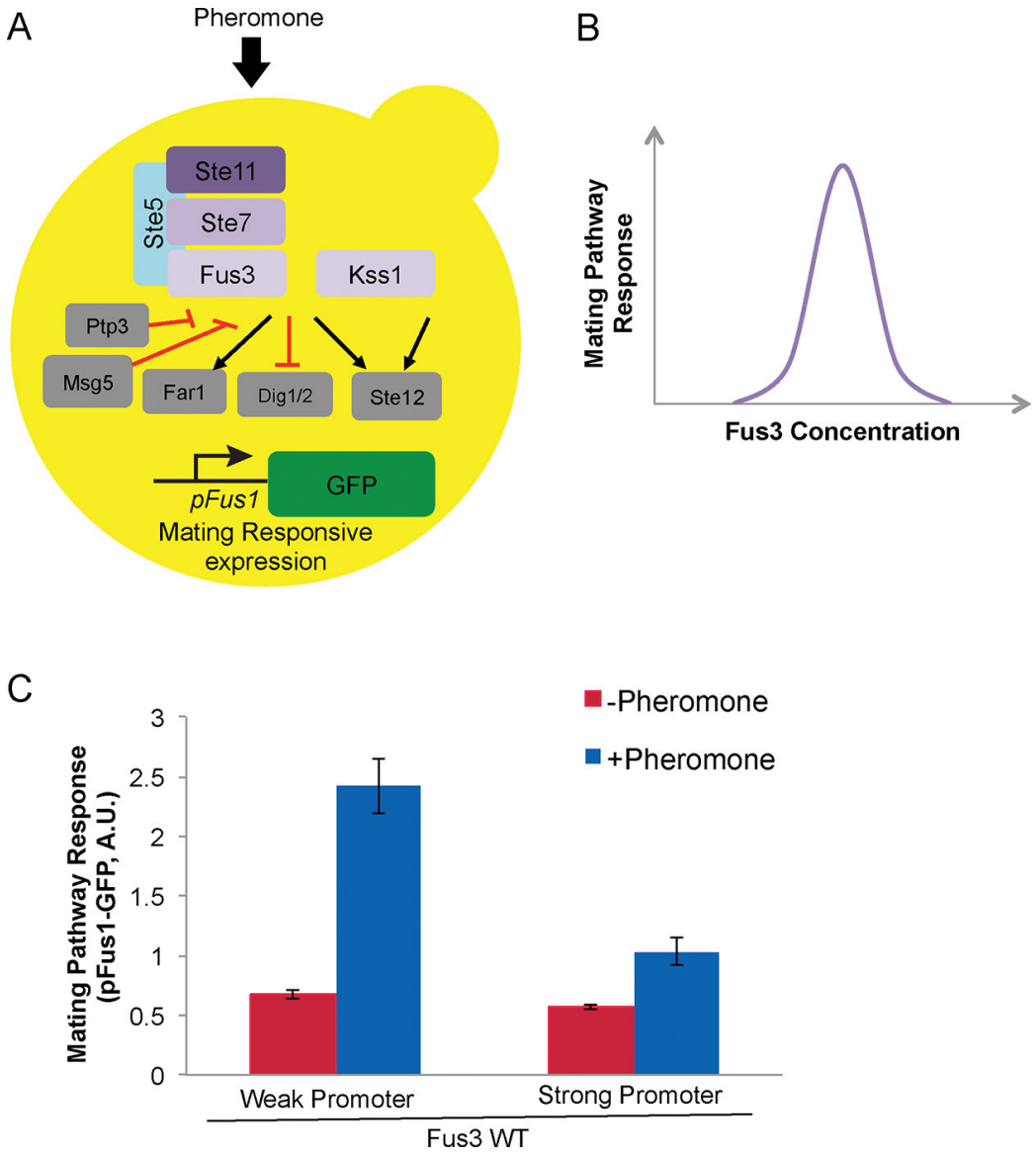


Figure 2-1: Yeast mating pathway activity depends on Fus3 levels. (A) Yeast mating pathway activity can be measured by flow cytometry using a pathway responsive promoter fused to a fluorescent reporter. Fus3 interacts with many proteins. (B) Optimal amount of Fus3 is required for mating pathway activity. Deviations from optimal Fus3 expression decrease mating pathway activity. (C) Fus3 overexpression-driven decrease in mating pathway activity can be triggered by expressing Fus3 from the strong TEF1 promoter.

Fus3 overexpression has been shown to result in signal dampening by displacing the stoichiometric balance of the complex (Levchenko *et al.,* 2000; Suderman and Deeds 2013; Thomson *et al.,* 2011) (Figure 2-1B). Excess Fus3 can bind to its partners individually, sequestering them away and therefore decreasing the levels of the fully functional signaling complex. Overexpression of Fus3 could lead to the accumulation of both inactive Fus3 and monophosphorylated Fus3 (Fus3-P), both of which dampen mating pathway signal transduction. Inactive Fus3 inhibits mating pathway activity by binding to Gpa1, the α subunit of the heterotrimeric G-protein (Errede *et al.,* 2015). Unlike dually phosphorylated Fus3 (Fus3-PP), Fus3-P dampens mating pathway activity (Nagiec *et al.,* 2015).

Here, using the yeast mating pathway as a model system, we use a synthetic biology approach to explore potential evolutionary mechanisms that could overcome the negative effects that overexpression of one component of a protein complex has on signaling pathway function. We find that mutations that create stop codons and are subsequently read through *in vivo* restore wild-type-like expression levels of the MAPK leading to normal signaling function. We therefore propose that stop-codon readthrough is an unexpected compensatory mechanism to correct for overexpression of signaling components.

2.3 Results

2.3.1 Overexpression of Fus3 Reduces Pathway Output

To confirm the deleterious effect of Fus3 overexpression on mating pathway activity, we expressed Fus3 from promoters of varying strengths, in a strain where a fluorescent reporter is fused to a mating pathway responsive promoter (pFUS1-GFP) (Peisajovich *et al.,* 2010). As shown in Figure 2-1C, pathway output, as represented by GFP fluorescence, is strongly affected by Fus3 expression levels: GFP levels are higher when Fus3 is expressed from a weak promoter (pCYC1) than when Fus3 is expressed from a strong promoter (pTEF1), confirming that overexpression of Fus3 decreases mating pathway activity.

While a trivial way to restore pathway activity would be to lower the strength of the promoter, we are interested in identifying mutations in the coding region that increase pathway activity. These mutations would represent potential mechanisms that evolution could employ to overcome squelching. Thus, we focus our attention in mutational changes exclusively affecting the Fus3 coding sequence. To identify Fus3 variants that rescue pathway activity, we overexpressed a library of random mutants of Fus3 in a pFUS1-GFP reporter strain, in which wild type (WT) Fus3 and its homologue Kss1 (which can also activate the mating pathway) had been deleted (Figure 2-2A). We then used fluorescence activated cell sorting (FACS) to isolate yeast cells expressing GFP levels higher than the WT strain (i.e., capable of rescuing mating pathway activity). We recovered the plasmids from the sorted cells and retransformed them into the Fus3ΔKss1Δ strain to confirm activity. Next, we sequenced these variants to determine the mutations responsible for the mating pathway rescue phenotype.

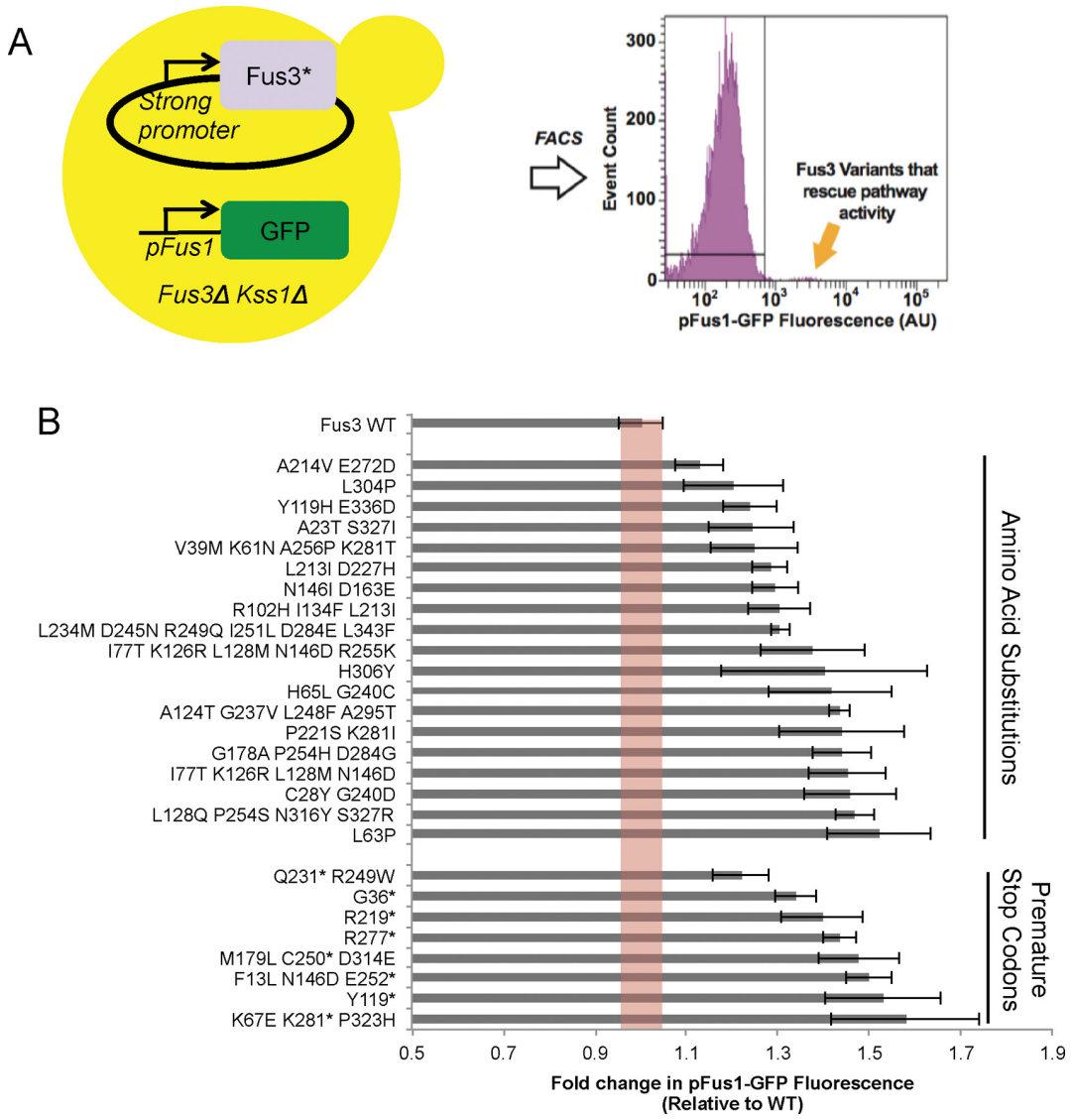


Figure 2-2: Identification of pathway activating Fus3 variants. (A) Fus3 is overexpressed from a strong promoter in a Fus3ΔKss1Δ pFus1-GFP background. Pathway activating variants are isolated by FACS. (B) Fus3 variants with higher than WT mating pathway induction ratio (fold change in pFus1-GFP fluorescence over WT).

2.3.2 Overexpressed Fus3 Mutants Can Rescue Pathway Activity in Different Ways

We grouped the sorted mutants into two categories based on the type of mutations they carried: variants encoding substitutions of one or more amino acids, and, to our surprise, variants encoding premature stop codons (Figure 2-2B). Among the variants encoding amino acid substitutions, three carried the previously characterized Fus3 activating mutations L63P, D227, and C28Y (Brill *et al.,* 1994), indicating that our selection is capable of isolating variants with pathway activating mutations. These mutants demonstrate that increasing activity of the kinase is a potential mechanism to overcome the effects of squelching.

Strikingly, eight of the selected Fus3 variants able to rescue mating pathway activity contain premature stop codons (PSCs). If translated correctly, many of these variants would result in significant truncations eliminating large protein fragments that often include many conserved residues vital for function (e.g., all of the active site). An extreme example would be the Fus3 variant truncated at residue G36, as a stop codon at that position would eliminate ~90% of the protein, including the ATP binding site (K42) (Brill *et al.,* 1994), the activation loop (135–138 HRD) (Endicott *et al.,* 2012), the active site (155–158 DFGL) (Endicott *et al.,* 2012), and the common docking site into which upstream activators and downstream substrates bind (D314, D317) (Remenyi *et al.,* 2005) (Figure 2-3A). Given the intriguing role that PSC may play in restoring pathway function, we focused our attention exclusively on the PSC-containing Fus3 variants. The rest of the sorted variants will be analyzed in an independent study.

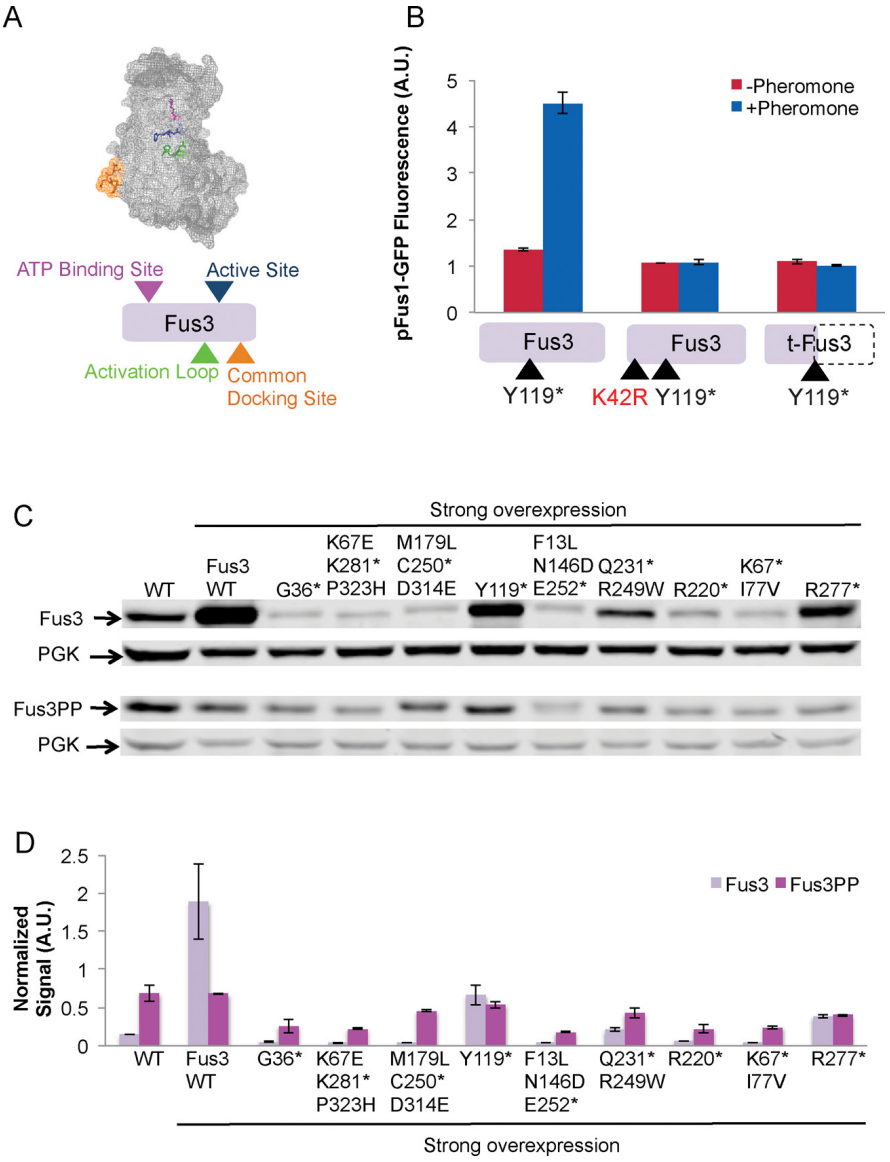


Figure 2-3: Genetic analysis of Fus3 variants carrying PSCs. (A) Functionally important, conserved residues are shown on Fus3 structure (PDB: 2B9F) (B) Fus3 carrying PSC at residue 119 is able to activate the mating pathway. Fus3 with a loss-of-function mutation (K42R) upstream of the PSC cannot activate the mating pathway. Fus3 truncated at residue 119 cannot activate the mating pathway. (C) Western blot analysis of Fus3 mutants with PSCs. All Fus3 mutants with PSC express full-length, active Fus3. (D) Fus3 variants with PSCs express less Fus3 than WT, even when expressed from the same strong promoter, as indicated by the normalized signal intensity.

2.3.3 Further Analysis of PSC Containing Variants Suggests Readthrough

To ensure that the PSC containing variants were active and that the assignment of the PSC was not an error in sequencing analysis, we introduced a PSC in Fus3, by site-directed mutagenesis, at one of the residues found to have a PSC among the sorted variants, Y119. Analysis of this mutant by flow cytometry confirmed that this PSC containing variant is capable of rescuing mating pathway activity (Figure 2-3B), suggesting the PSC might be being read through. To further explore this hypothesis, we introduced a mutation, K42R, known to abolish Fus3 activity upstream of the PSC. As shown in Figure 2-3B, this mutant, Fus3 K42R Y119*, is unable to rescue mating pathway activity in a Fus3ΔKss1Δ strain, confirming that the Y119* mutation is necessary and sufficient for rescue of mating pathway activity. To further determine that the selected PSC containing Fus3 variants were effectively being readthrough, we deleted the region of Fus3 downstream of the PSC at position 119, therefore mimicking the protein product that would result from the accurate translation of the stop codon at that position. As before, we overexpressed this mutant, truncated Fus3, in the Fus3ΔKss1Δ pFus1-GFP strain and measured mating pathway activity by flow cytometry. As shown in Figure 2-3B, Fus3 truncated at residue 119 cannot activate the mating pathway. Our integrated results so far suggest that translation is not stopping at the PSC, but rather may be reading through the PSC, resulting in the correct mRNA transcript and ultimately in an active kinase.

2.3.4 Readthrough of PSC Leads to Full-Length Active Kinases

To confirm that the PSCs are being readthrough, resulting in full-length active kinases, we measured Fus3 protein levels for all eight variants containing PSC in the Fus3Δ Kss1Δ strain by Western blots. As shown in Figure 2-3C, all PSC containing variants express full length Fus3, thus demonstrating that the PSCs are read through leading to full-length kinases. Moreover, to determine if the expressed kinases are properly folded substrates of their upstream MAPKK activator Ste7 and also able to interact with the mating scaffold Ste5, as Ste7-mediated activation occurs only when Fus3 is bound to Ste5, we measured the levels of phosphorylated Fus3 using an antiphospho-Fus3 specific antibody. As shown in Figure 2-3C (and Figure 2-4, showing the full picture of the Western blot), all eight PSC-containing variants are phosphorylated by Ste7 *in vivo*, indicating that they are properly folded and capable of interacting with Ste7 and Ste5.

Together with the flow cytometry experiments that demonstrate pFus1-GFP expression mediated by the PSC-containing Fus3 variants, our results confirm that readthrough of PSC leads to Fus3 variants fully capable of mediating the activation of the mating pathway.

Readthrough of PSCs has been shown to occur naturally in yeast, though at a low rate, (Baudin-Baillieu *et al.,* 2014) as readthrough occurs as a result of a decrease in the efficiency of translation termination (Namy *et al.,* 2003). In agreement, our results show that, while different PSC-containing Fus3 variants are present in the cell at different levels, all of them are present at levels lower than that of WT Fus3 (Figure 2-3D). Indeed, our results suggest that the intrinsically low efficiency of readthrough compensates for the high levels of expression of the pTEF1 promoter, and, as a result, restores pathway activity. As shown in Figure 2-5 and Figure 2-6, similar PSC readthrough is observed when Fus3 variants are expressed in different *S. cerevisiae* strains, indicating that the observed effects are not an anomaly of a peculiar strain.

2.3.5 PSC Containing Variants Can Mediate a Full Mating Response

To determine if a PSC containing variant is capable of mediating a full mating response, we assessed the ability of an "a" mating type strain carrying Fus3 Y119* to mate with an "α" mating strain encoding WT Fus3. In our experiment, each of the two mating strains was auxotrophic for a distinct marker, so that only the diploids resulting from the mating of the two strains can grow on minimal media. As shown in Figure 2-7A, the Fus3 Y119* mutant is able to mediate a full mating response in a Fus3ΔKss1Δ strain. Furthermore, as mating can occur because of cell-to-cell proximity even in the absence of pheromone-induced polarized growth (also called "shmooing"), we also measured the ability of the PSC containing mutant to shmoo. As shown in Figure 2-8B, cells expressing Fus3-Y119* in a Fus3ΔKss1Δ strain can develop shmoos that are indistinguishable from those seen in WT cells. Together, these results confirm that the PSC-containing variants can mediate a full mating response.

2.3.6 PSCs are Preferentially Tolerated at Surface Residues

In yeast, readthrough of PSCs results in amino acid replacements that depend on the specific identity of the premature stop codon (Blanchet *et al.,* 2014; Roy *et al.,* 2015). Blanchet *et al*., and Roy *et al*., show that, in yeast, the UAA and UAG premature stop codons are most often replaced by glutamine, tyrosine, or lysine, while the premature stop codon UGA is most often

replaced by tryptophan, cysteine, and arginine, all with distinct frequencies. Roy *et al*. also showed that the specific causes of readthrough (e.g., genetic background or pharmacological agents) can affect these frequencies. Since we are not using pharmacological agents to induce readthrough, and we see strain-independent results confirming readthrough (Figure 2-5 and Figurer 2-6), we assume that our results are due to unprogrammed readthrough inherent to wild-type cells. On the basis of these measured replacement frequencies, it is possible to hypothesize what the most likely amino acid replacements would be for the PSC present in the sorted Fus3 variants. For instance, for the variant with a PSC at residue 119, the most likely replacement would be a Tyr/Gln, which makes it possible for the substitution to result in the original Tyr present at position 119 in WT Fus3 (Table 2-1). In contrast, for the mutant with a PSC at residue 36, the wild-type Glycine is not one of the possible replacement amino acids–the most likely replacement would be a tryptophan. Thus, while it is expected that multiple proteins with different substitutions at the same PSC might coexist within an individual cell, it is likely that, given the narrow range of possible replacements available, PSC should be tolerated only at defined locations. In particular, we hypothesize that the tolerated locations would be those in which the most likely replacement reverts back to the wild-type amino acid, as in position 119 in Fus3, or those in which the mutated residue would not have detrimental consequences in protein function.

To investigate this hypothesis, we visualized the location of all PSC on the 3D structure of Fus3. Not surprisingly, we found that all the PSC containing mutants occurred at surface exposed residues (Figure 2-8A), supporting the idea that the resulting amino acid replacements (which most likely lead to Gln, Tyr, Lys, Trp, Cys, or Arg substitutions) may be more easily tolerated at locations that do not seriously affect folding and function.

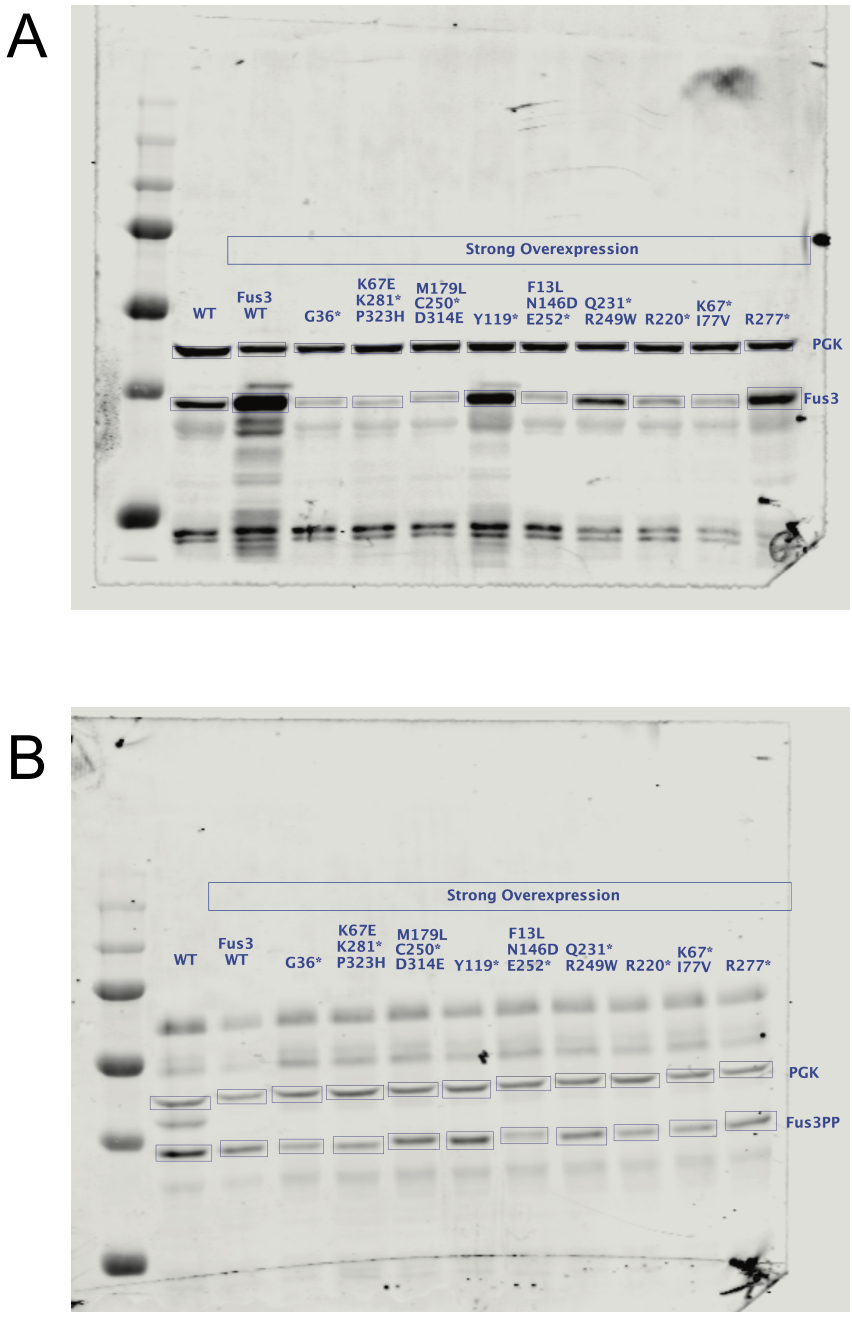


Figure 2-4: Genetic analysis of Fus3 variants carrying PSCs. Western blot analysis of Fus3 mutants with PSCs. All Fus3 mutants with PSC express full-length (A), and active Fus3 (B).

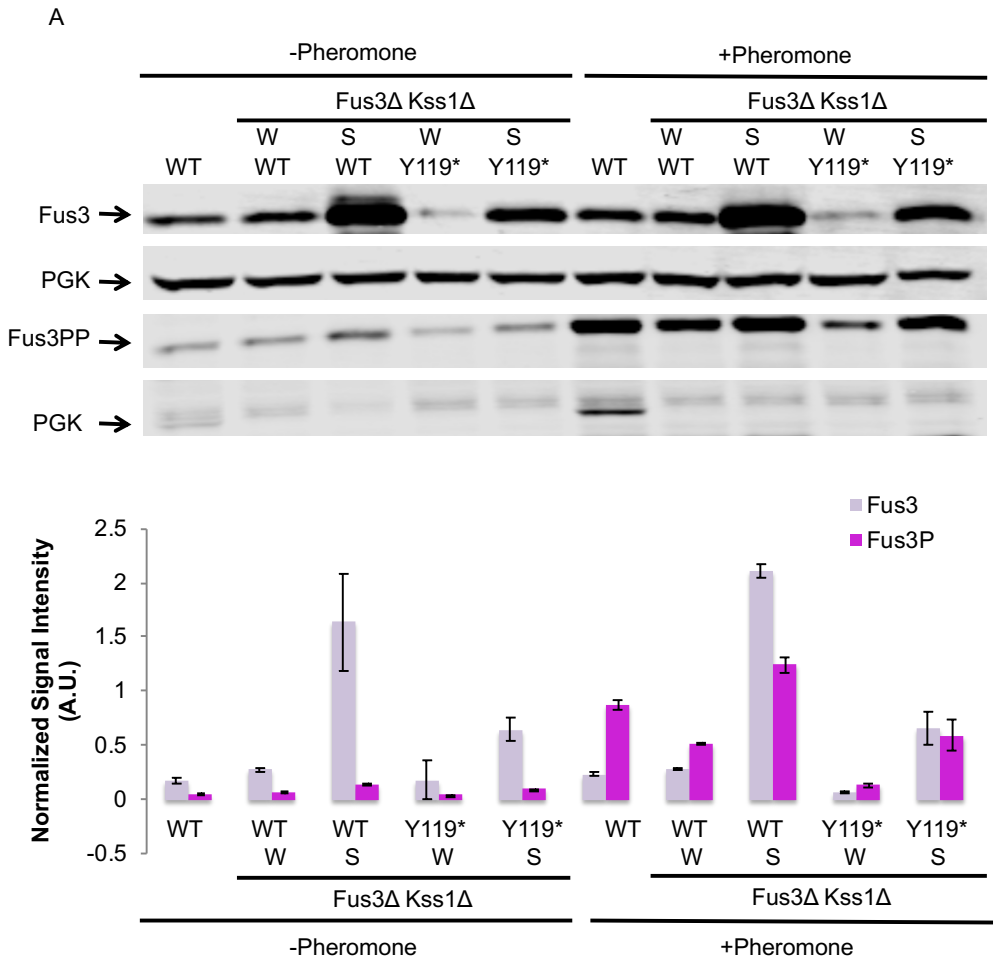


Figure 2-5: Genetic analysis of Fus3 variants carrying PSCs in an S288c strain. Western blot analysis of Fus3 expressing PSC at residue 119 expressed from pCYC1 (Weak, W, overexpression) and pTEF1 (Strong, S, overexpression) in Fus3 and Kss1 deletion strain with the following genotypes: S288C MATa his3, met15, leu2, ura3.

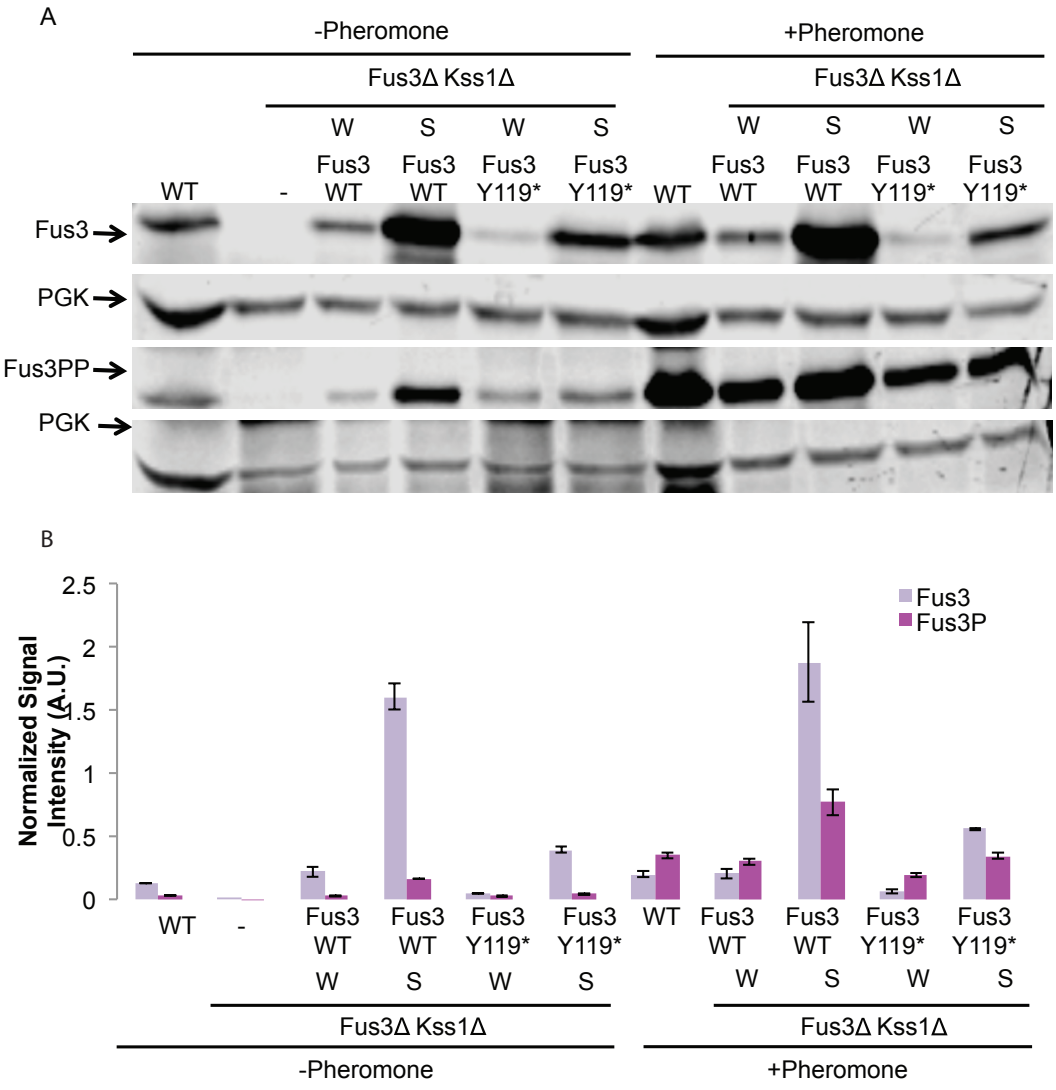


Figure 2-6: Genetic analysis of Fus3 variants carrying PSCs in another W303 strain. Western blot analysis of Fus3 expressing PSC at residue 119 expressed from pCYC1 (Weak, W, overexpression) and pTEF1 (Strong, S, overexpression) in Fus3 and Kss1 deletion strain with the following genotypes: W303 MATa, his3, trp1, leu2, ura3.

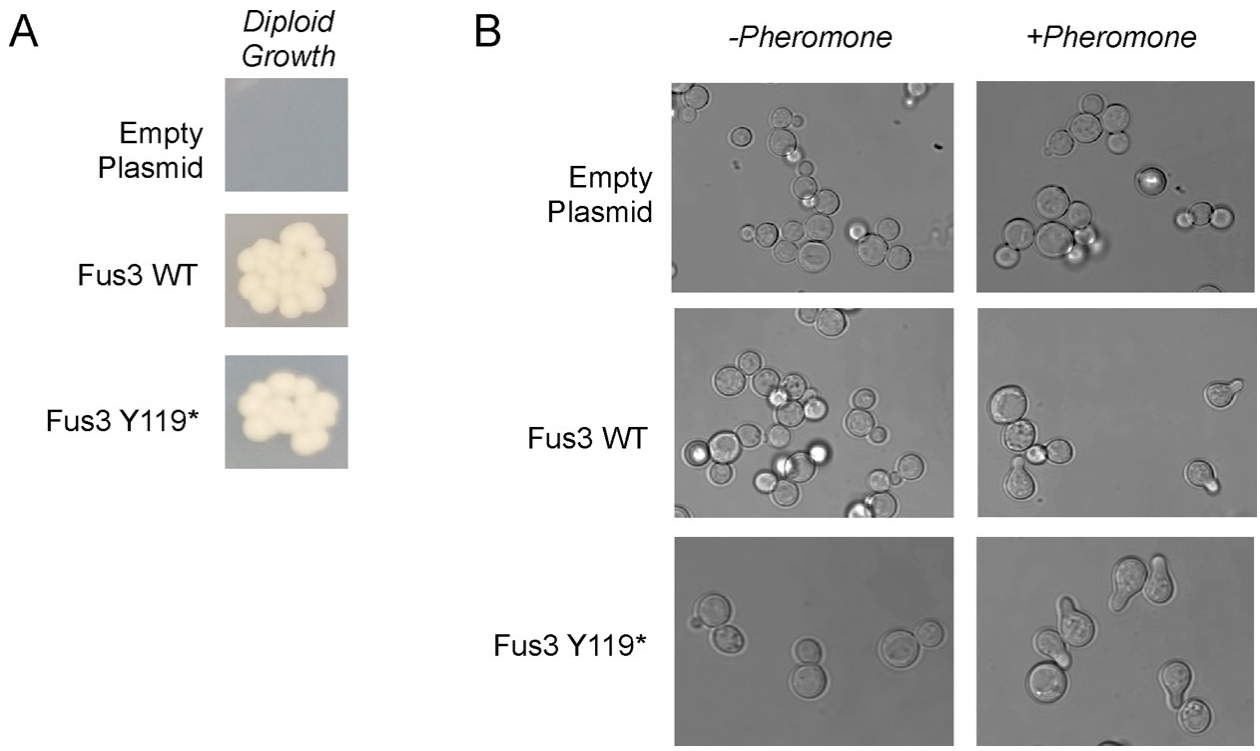


Figure 2-7: Functional analysis of Fus3 variants carrying PSCs. (A) Fus3ΔKss1Δ yeast cells expressing Fus3 WT or Y119* mutant from pTEF1 have normal mating responses, as indicated by growth on minimum media after mating with a tester strain. (B) Yeast cells expressing Fus3 Y119* from pTEF1 have mating-induced morphological responses similar to WT.

Table 2-1: Premature Stop Codon Containing Mutants and Their Likely Amino Acid Substitutions [a].

Premature Stop Codon Mutants	Location	Wild Type Codon	Premature Stop Codon	Wild Type Amino Acid	Possible Replacements	Possible Change to WT
F13L N146D E252*		GAG	TAG	E	Q, Y, K	no
K67E K281* P323H		AAA	TAA	K	Q, Y, K	yes
Y119*		TAC	TAG	Y	Q, Y, K	yes
Q231* R249W		CAA	TAA	Q	Q, Y, K	yes
G36*		GGA	TGA	G	W, C, R	no
M179L C250* D314E		TGT	TGA	C	W, C, R	yes
R219*		CGG	TGA	R	W, C, R	yes
R277*		CGA	TGA	R	W, C, R	yes
A30*	buried	GCA	TAA	A	Y, Q, K	no
C209*	buried	TGC	TGA	C	W, C, R	yes
G237*	surface exposed	GGT	TGA	G	W, C, R	no
Y228*	surface exposed	TAT	TAG	Y	Y, Q, K	yes

[a] The table contains the list of Fus3 variants with premature stop codons tested in this study. Possible replacements based on premature stop codon are identified through sequencing and likely substitutions based on previous works (Blanchet *et al.,* 2014; Roy *et al.,* 2015).

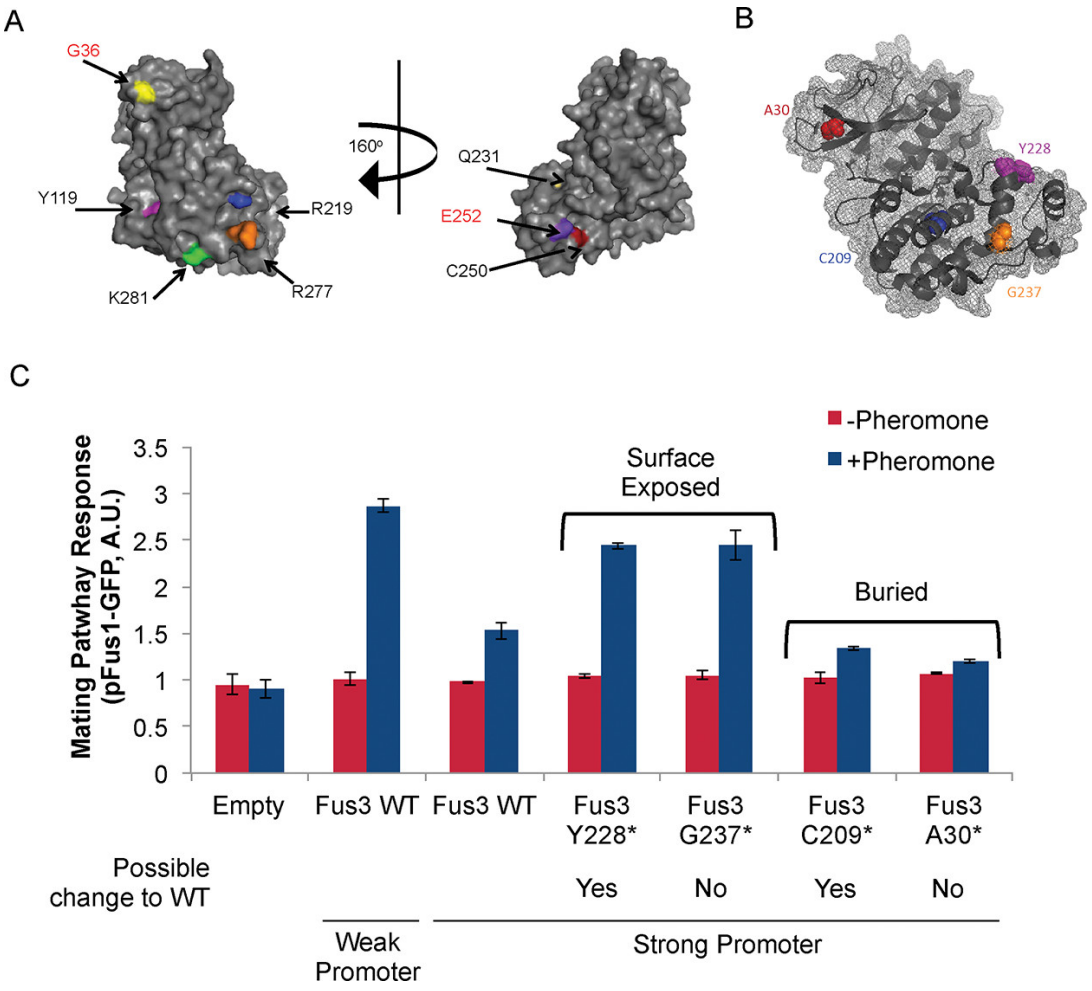


Figure 2-8: Structural analysis of residues that tolerate PSC. (A) All PSCs map to surface residues. On the basis of Blanchet *et al*., we predict the possibility of each of the residues, when mutated to a stop codon to be substituted by the WT residue during readthrough (Blanchet *et al.,* 2014). Residues that cannot revert to WT (red) and residues that could possibly revert to WT (black) are shown. (B) Buried residues, A30 (red) and C209 (blue); and exposed residues Y228 (magenta) and G237 (orange) are shown on the Fus3 3D structure (2B9F). (C) Fus3 with a PSCs at buried residues cannot mediate mating response.

To further understand readthrough tolerance, we introduced PSCs at two residues buried in the Fus3 protein core (A30 and C209) or at two residues exposed on the Fus3 surface (Y228 and G237) (Figure 2-8B). Moreover, we introduced specific PSCs into these positions, so that they could either be replaced by the wild-type amino acid during readthrough (C209 and Y228) or by nonsynonymous substitutions (A30 and G237) (Table 2-1). As before, we overexpressed these mutants from a TEF1 promoter in a Fus3ΔKss1Δ pFus1-GFP strain, and measured mating pathway activity by flow cytometry. Yeast cells overexpressing PSCs at the surface-exposed residues Y228 and G237 are able to mediate mating pathway response better than cells overexpressing Fus3 WT, regardless of whether or not the wild-type amino acid is likely substituted for the PSC during readthrough (Figure 2-8C). PSCs at these residues restore mating pathway activation to levels that are close to that mediated by WT Fus3 expressed from a weak promoter. In contrast, cells overexpressing PSCs at buried residues (A30 and C209) are unable to even mediate a response similar to that of overexpressed WT Fus3, confirming that amino acid substitutions are less tolerated at buried positions. Interestingly, overexpression of Fus3-C209* results in significantly higher pathway activity than overexpression of Fus3-A30* ($P = 0.016$, Student's *t* test). This could be explained by the fact that, while the PSC at C209 can be substituted by the wild-type amino acid (albeit at a much lower level than Y228*), the PSC at A30 cannot be substituted by an alanine during readthrough. Thus, we conclude that PSCs are most likely to be tolerated when the resulting amino acid replacements occur at exposed locations with lax structural or functional constraints, while at buried residues PSCs are more likely to be tolerated when they restore, even if at a low frequency, the wild-type residue. Furthermore, our Western blot analysis found that these Fus3 with PSCs express full-length, active kinases (Figure 2-9). In addition, we found that coexpressing Fus3 carrying PSCs with wild-type Fus3 does not affect mating pathway response, indicating that if any truncated species are being expressed, they do not interfere with the function of full-length species (Figure 2-10).

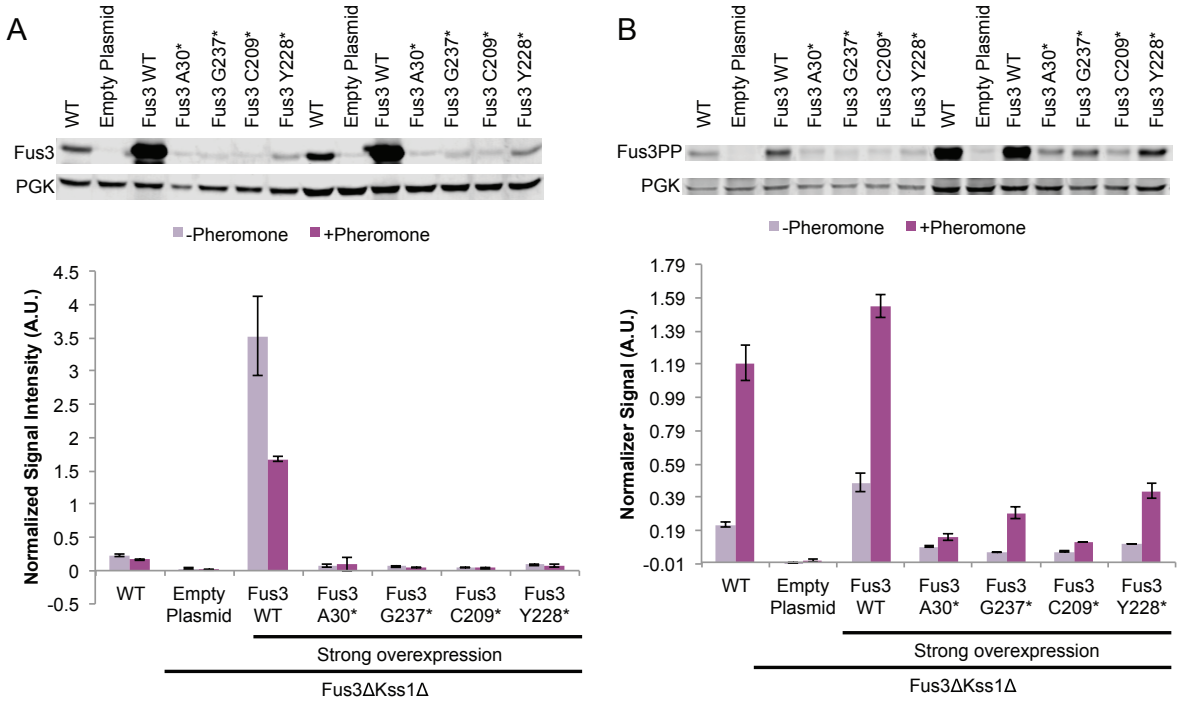


Figure 2-9: Genetic analysis of Fus3 variants carrying PSCs at buried and surface-exposed residues. Western blot analysis of Fus3 expressing PSCs at residues A30, G237, C209, or Y228 expressed from the strong promoter pTEF1 assaying expression full-length (A), and active Fus3 (B).

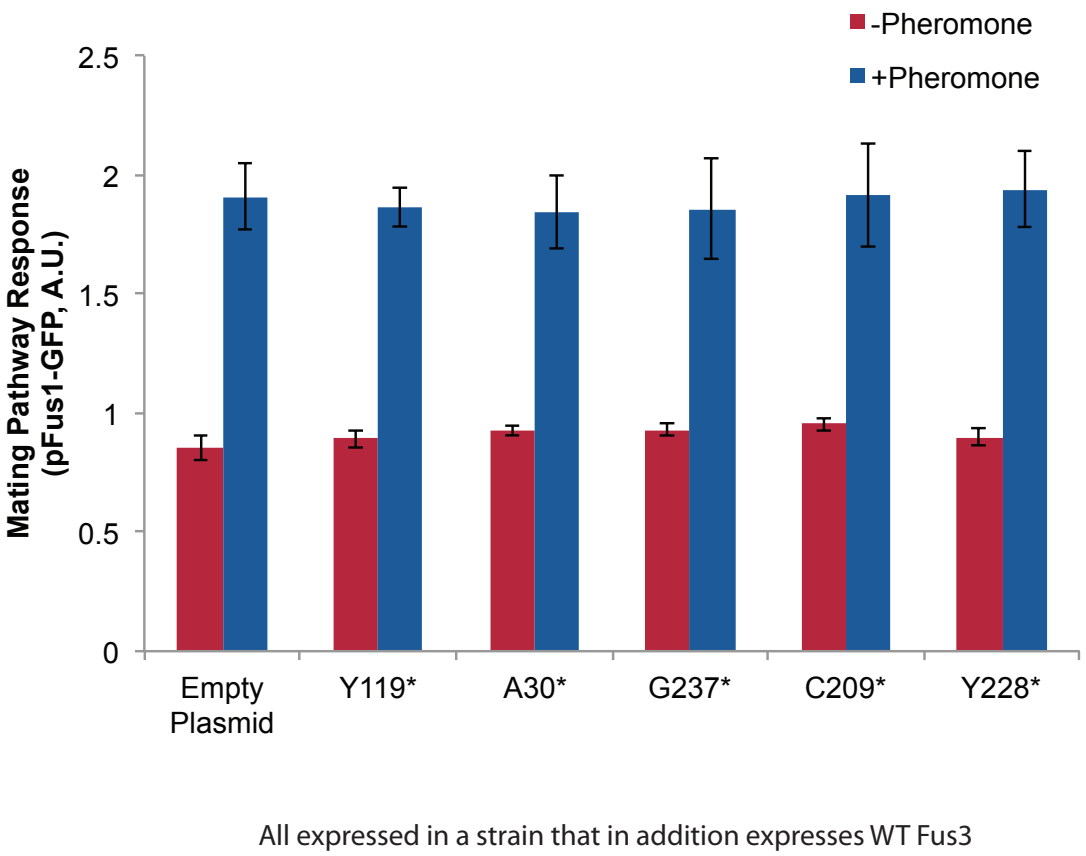


Figure 2-10: Co-expression of Fus3 carrying PSCs with wild-type Fus3 does not affect mating pathway response. When co-expressed in a strain with wild-type Fus3, Fus3 carrying PSCs at residue Y119, A30, G237, C209, or Y228 do not affect mating pathway activity in comparison to empty plasmid.

2.4 Discussion

In this work, we used a synthetic biology approach to investigate potential mechanisms by which selection could compensate for the decrease in signaling activity caused by overexpression of a tightly regulated signaling protein—the yeast mating pathway MAPK Fus3. Specifically, we used cell sorting to select a library of random mutants of Fus3 overexpressed from the strong promoter, pTEF1, in a strain where the WT Fus3 (and its paralog Kss1) had been deleted. In this manner, we isolated several Fus3 variants that rescued mating pathway activity. Our selection revealed that, in addition to the predictable way of improving kinase activity, as previously reported for the selected Fus3 variants L63P, D227, and C28Y (Brill *et al.,* 1994), one unexpected way by which selection could rescue pathway function was through the introduction of PSCs, which are then naturally readthrough.

Readthrough has been shown to occur not just as an error in translation termination, but also as a regulated mechanism of stop codon bypass (von der Haar and Tuite 2007). Several studies have revealed that readthrough can result in protein extensions. For example, using baker's yeast as a model organism, Namy *et al*. found eight adjacent open reading frames separated only by a unique stop codon to have higher than average stop codon bypass levels (Namy *et al.,* 2003). Also, in yeast, Artieri and Fraser reported 19 proteins with conserved C-terminal peptide extensions originating from stop codon read-throughs (Artieri and Fraser 2014). Similarly, functionally important protein extensions through readthrough of stop codons were also identified in viral genes (Brown *et al.,* 1996; Cimino *et al.,* 2011; Firth *et al.,* 2011; Napthine *et al.,* 2012; Pelham 1978), human genes (Jungreis *et al.,* 2011; Loughran *et al.,* 2014; Schueren *et al.,* 2014), and several other species (Dunn *et al.,* 2013; Ivanova *et al.,* 2014; Jungreis *et al.,* 2011; Robinson and Cooley 1997). In addition, programmed readthrough of stop codons is used to change cellular localization of some enzymes, from yeast to humans (Stiebler *et al.,* 2014; Yanagida *et al.,* 2015). In yeast, stop codon readthrough has been linked to the unveiling of otherwise cryptic genetic variation (True and Lindquist 2000). Taken together, these results suggest that stop codon readthrough is a mechanism by which evolution can regulate gene expression.

Our results indicate that the location of the mutated residue in the protein structure, its effect on function, as well as the specific identity of the PSC are all factors that affect the impact that

PSCs have on protein (and pathway) function. Furthermore, our work demonstrates that, in addition to C-terminal extensions, readthrough can occur at internal PSCs, resulting in proteins with WT length, though with possible amino acid replacements. Thus, readthrough can exert its effects not only by changing the length of a protein, but also by lowering protein levels due to the intrinsically low efficiency of the readthrough process. In other words, our results demonstrate that, at least in the laboratory, selection can compensate for protein overexpression not only by changes in promoter strength, as it would have been easily expected, but also by introducing PSCs at permitted positions.

Interestingly, our results suggest that mutations that one would predict to have a detrimental effect on protein function (e.g., introduction of a PSC) could actually have the opposite effect on network function. This apparent contradiction highlights the importance of understanding genotypic variation at the systems level, rather than at the individual gene level. Studies attempting to predict phenotype from genotype computationally frequently categorize both frameshifting indels and premature stop codons as loss-of-function (Bergstrom *et al.,* 2014; MacArthur *et al.,* 2012). However, recent work by Tawfik and co-workers showed that the effects of frameshifting insertion and deletions (indels) on the function of a bacterial enzyme can be phenotypically rescued by subsequent slippage of the ribosome (Rockah-Shmuel *et al.,* 2013). Similarly, our work indicates that PSCs cannot be always considered to detrimentally affect cellular functions.

2.5 Methods

2.5.1 Yeast Strains

Fus3 and Kss1 were targeted for deletion by homologous recombination using Trp and Leu as selectable markers, respectively, in strains with the following genotypes:

W303 MATa, bar1::NatR, far1Δ, mfa2::pFUS1- GFP, his3, trp1, leu2, ura3.

W303 MATa, his3, trp1, leu2, ura3

Deletions were confirmed by PCR, flow cytometry, and Western blots.

2.5.2 Library Construction

Random mutagenesis was carried out PCR-based methods by using Fus3 as a template and the Agilent GeneMorph II Random Mutagenesis Kit. The random mutants were constructed using previously described cloning strategy (Bardwell 2005) and expressed from centromeric plasmids with His selection, under the control of a constitutive high expression promoter, Tef1 (pTEF1), and an AdhI transcription terminator. Mutation rate was confirmed by sequencing 18 independent clones from the unselected library.

2.5.3 Site Directed Mutagenesis

Point mutations were introduced into Fus3 by PCR using either Agilent QuikChange II Site-Directed Mutagenesis Kit or by using Pfu Ultra, following the manufacturer's protocols. The PCR products were cloned using the method described under Library Construction under the control of pTEF1 or a constitutive low expression promoter, Cyc1 (pCYC1). Mutations were confirmed by sequencing.

2.5.4 Transformation

Yeast strains were transformed by standard lithium–acetate method except libraries which were transformed by the specific lithium–acetate-based high efficiency, large scale transformation method described by Gietz *et al*. with minor modifications (Gietz and Schiestl 2007).

2.5.5 Fluorescence-Activated Cell Sorting

The Fus3 overexpression random mutant library was sorted by fluorescence-activated cell sorting (FACS) to isolate variants able to elicit a mating pathway response in Fus3 and Kss1 double deletion strains. Specifically, to ensure high library diversity, about 68 000 colonies were

harvested after transformation into *Escherichia coli*. PK001 was transformed with a library of Fus3 mutants replacing the deleted Fus3 with a random variant. To maintain diversity, about 4000 transformants were harvested in selective medium and grown overnight at 30 °C. Overnight cultures were diluted to OD_{600} between 0.1 and 0.2 and grown to log phase. Samples were then induced with 1 μM α-factor and incubated at 30 °C for 2 h. After sonication, cells were gated by side and forward scatter and sorted by GFP expression using a BD FACSAria I cell sorter into selective medium. Sorted cells were plated on selective media and cultured.

2.5.6 Flow Cytometry

Double deletion yeast strains were transformed with plasmids expressing Fus3 variants and grown on selective media. Transformants were grown in triplicate selective medium overnight at 30 °C. Wild-type strain without the double deletion was grown in complete synthetic dropout medium as positive control and either an empty double deletion strain (grown in complete synthetic dropout medium) or one expressing an empty plasmid (grown in selective medium) was used as negative controls. Overnight cultures were diluted to OD_{600} between 0.1 and 0.2 and grown to early log phase. Samples were then induced with 1 μM α-factor and incubated at 30 °C for 2 h. Cells were then treated with the protein synthesis inhibitor cycloheximide for approximately 30 min. The GFP signal of 10 000 cells was measured for each sample with a Miltenyi Biotec MACSQuant VYB. The mean GFP fluorescence and standard deviation of triplicates were calculated using FlowJo.

2.5.7 Western Blots

Western blots were performed as described previously (Zalatan *et al.,* 2012) with minor modifications. Overnight cultures were diluted to OD_{600} between 0.1 and 0.2 and grown to mid-log phase. Samples were then induced with 2 μM α-factor and incubated at 30 °C for 15 min. After the pellet was frozen in liquid nitrogen, the cells were then lysed and run in 10% Tris-Gly, SDS polyacrylamide gels at 200 V for 50 min with the Odyssey Protein Molecular Weight Marker (928-40000) using the Mini-PROTEAN Tetra Cell. The Bio-Rad Trans-Blot Turbo Transfer System was used to transfer the proteins to Bio-Rad low fluorescence PVDF

membranes. Membranes were blocked in Odyssey Blocking Buffer and TBS overnight at +4 °C. The membranes were then incubated with primary antibodies (Fus3: (yC-19) #sc6773, goat polyclonal IgG, Santa Cruz Biotechnology, dilution 1:5000; Fus3-PP, rabbit monoclonal phospho-p44/42 MAPK antibody, Cell Signaling Technology, #4370, dilution, 1:2000; PGK (loading control), Invitrogen 459250 mouse monoclonal, dilution, 1:5000) for 2 h. After washing the primary antibodies with TBS + 0.05% Tween20, membranes were incubated with secondary antibodies (Fus3, donkey antigoat Licor IRDYE 800 antibody, 926-32214; Fus3-PP, goat antirabbit Licor IRDYE 800 antibody, 926-32211; PGK, goat antimouse Licor IRDYE 680LT antibody, 926-68020; all at dilution of 1:10000) for 1 h. Membranes were then washed and visualized on a Licor Odyssey CLx Infrared Imaging System.

2.5.8 Mating Assay

Mating assays were performed as described previously (Peisajovich *et al.,* 2010) with minor modifications. Equal amounts of yeast cells of mating type "a" (W303 *MATa, his3, trp1, leu2, ura3*, Fus3::kanMX4+KIURA3, Kss1::leu) were transformed with either empty vector, Fus3 expressed from pTEF1, or Fus3 Y119* expressed from pTEF1, and yeast cells of mating type "α" were mixed and added to a polycarbonate filter. After incubating at 30 °C for 3 h, cells were extracted by vortexing and aliquots were plated on minimum synthetic media.

2.5.9 Microscopy

Yeast cells expressing WT Fus3 or Fus3 Y119* from pTEF1 were grown to log phase and induced with 1 μM α-factor. Cells were imaged using an automated inverted Leica TCS SP8 confocal microscope.

3 Chapter 3: Mutational Robustness of Mating and HOG Signaling Pathways

3.1 Abstract

Cells receive stimuli from their surroundings and process them into physiological responses through signal transduction pathways. Mutations are an important source of genetic variation and understanding how pathways respond to mutations can reveal how cells process information. Pathway robustness to mutations is thought to facilitate adaptation to future environmental or genetic changes. Here, we quantified the effects of whole genome mutations and pathway specific mutations on *Saccharomyces cerevisiae* mating and high osmolarity glycerol response pathways. Based on the dynamics of the pathway responses, we hypothesized that the HOG pathway would be more robust to mutations. We used an information theory measure, mutual information, to quantify the effects of mutations mating and HOG signaling. First, we find that mutations decrease information in mating signaling more than that in HOG signaling. Second, mutations lead to more variable effects on information transmission in the mating pathway than in the HOG pathway. Third, mutations in a shared pathway component can have opposite effects on the amount of information transmitted by a stimulus to the incorrect pathway ("crosstalk"). Our analysis quantifies the differences in the mutational robustness of different signaling pathways and improves our understanding of the phenotypic significance of mutations in evolutionarily conserved signaling pathways.

3.2 Introduction

Cell signaling networks underlie cellular responses to the environment and are critical for cell fate specification and patterning during development (Basson 2012). Signaling networks allow cells to 'make decisions' and thus provide the information processing capacity of the cell (Azeloglu and Iyengar 2015). Mutations in signaling genes are critical in many human diseases (Li 2012) and changes in these genes are thought to underlie key evolutionary changes (Martin

and Courtier-Orgogozo 2017). Thus, understanding how mutations affect signaling networks is an area of current research interest.

Signaling networks are expected to vary in their tolerance to mutations (robustness). Complex network structures with feedback loops and redundant branches are expected to be more robust to mutations than simple linear cascades (Soyer and Bonhoeffer 2006). Here we set out to test this hypothesis by comparing the robustness of two MAPK signaling pathways in budding yeast, the mating pathway (a largely linear pathway) and the HOG pathway (a branched pathway with many feedback loops) (Chen and Thorner 2007).

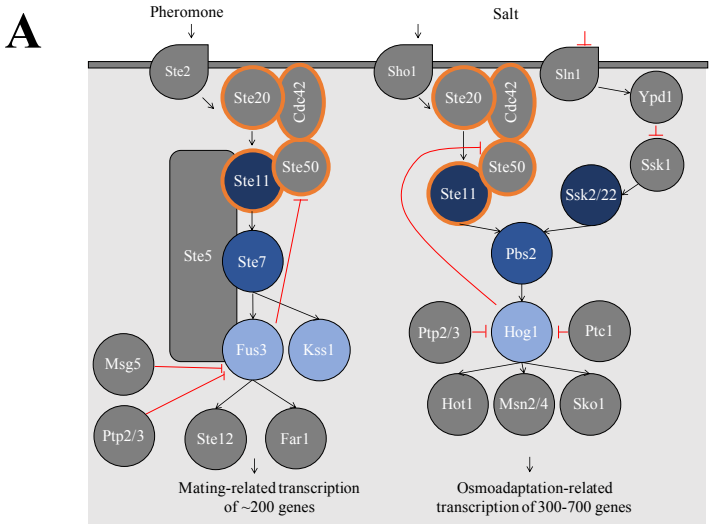
Because the information processing properties of signaling networks depend on the interactions of many proteins, it is difficult to understand the effects on information processing by studying the effects on the biochemical activities of individual network components (Chen and Wu 2012) or through statistical analysis of genetic variation in natural populations (Lage 2014). Here we take a direct experimental approach to test the effects of mutations on signaling pathways in cells.

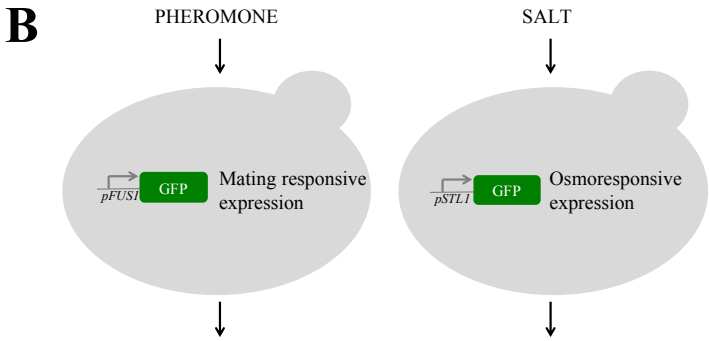
Experimental evolution methods such as mutation accumulation (MA) allow us to study the effects of the entire spectrum of mutations (except strongly deleterious ones). Mutations identified in extant populations are only the ones that are retained after natural selection has acted to remove deleterious mutations and increase the frequency of beneficial mutations (Desai and Fisher 2007). We can limit the effects of selection, and therefore obtain a nearly unbiased view of the effects of mutations, by repeated introductions of population bottlenecks (Landry *et al.,* 2007). While single nucleotide mutations are the most common mutations found in MA studies, we are able to interrogate the effects of all types of spontaneous mutations including indels, larger scale mutations such as duplications and translocations, as well as gene expression changes (Landry *et al.,* 2007; Lynch *et al.,* 2008; Wittkopp *et al.,* 2004; Zhu *et al.,* 2017).

We use well-characterized MAPK pathways in *Saccharomyces cerevisiae*, as a model to compare the effects of mutations on signaling pathway robustness. In *S. cerevisiae*, there are at least five mitogen-activated protein kinase (MAPK) pathways that allow cells to respond to various extracellular stimuli (Chen and Thorner 2007). The mating and high-osmolarity glycerol (HOG) pathways respond to pheromone and salt, respectively (Figure 3-1A). The mating pathway is required for cells to sense pheromone, transmit information about this stimulus, and

under sufficient concentrations, activate the MAPK cascade and downstream effectors leading to cell cycle arrest, polarized growth in the direction of the pheromone and other cellular processes (Shao *et al.,* 2006). The HOG pathway is required for re-establishing optimal cellular osmolarity upon introduction to hyperosmotic shock by triggering the MAPK cascades and initiating glycerol production (Patterson *et al.,* 2010). Unlike the mating pathway, which may be compared to a cell-fate decision system leading to sustained activation of the mating MAPK, the HOG pathway is an adaptation response where the HOG MAPK is activated only transiently, and such adaptation responses may be more tightly regulated (Gasch *et al.,* 2000; Pelet *et al.,* 2011). These differences in dynamics pose a challenge for comparing pathway output directly and motivate our use of mutual information (MI) to assess the effects of mutations on pathways (see Discussion). To study the effects of spontaneous mutations on mating and HOG signaling, we employed a mutation accumulation approach to evolve two *S. cerevisiae* strains, one with a mating pathway reporter and the other with a HOG pathway reporter for ~2000 generations.

Although MA studies can reveal how spontaneous mutations across the whole genome affect a trait of interest, we cannot control the so-called 'mutational target size' or easily map the genotype to phenotype relationship. The probability that a mutation affects a gene depends on the cis- and trans-mutational target sizes and the mutation effect size (Houle 1998; Landry *et al.,* 2007). Furthermore, one pathway may simply have more critical genes and therefore be more sensitive to mutations. Therefore, differences in the effects of mutations in the two pathways could be because of different mutational target sizes. To overcome this limitation, we can introduce mutations through random mutagenesis by error-prone PCR in a specific gene. While the mutations may not reflect those that are naturally occurring, targeting a specific region in the genome will also allow us to quickly map the genotype-phenotype relationship. By taking advantage of signaling pathways with shared components and introducing mutations into these components, we can interrogate how different pathways respond to the same mutations.


A
Pheromone
Salt
Ste2
Ste20
Cdc42
Ste11
Ste50
Ste5
Ste7
Fus3
Kss1
Msg5
Ptp2/3
Ste12
Far1
Sho1
Sln1
Ypd1
Ssk1
Ssk2/22
Pbs2
Hog1
Ptp2/3
Ptc1
Hot1
Msn2/4
Sko1
Mating-related transcription of ~200 genes
Osmoadaptation-related transcription of 300-700 genes


B
PHEROMONE
SALT
pFUS1
GFP
Mating responsive expression
pSTL1
GFP
Osmoresponsive expression
Measure GFP Fluorescence by Flow Cytometry

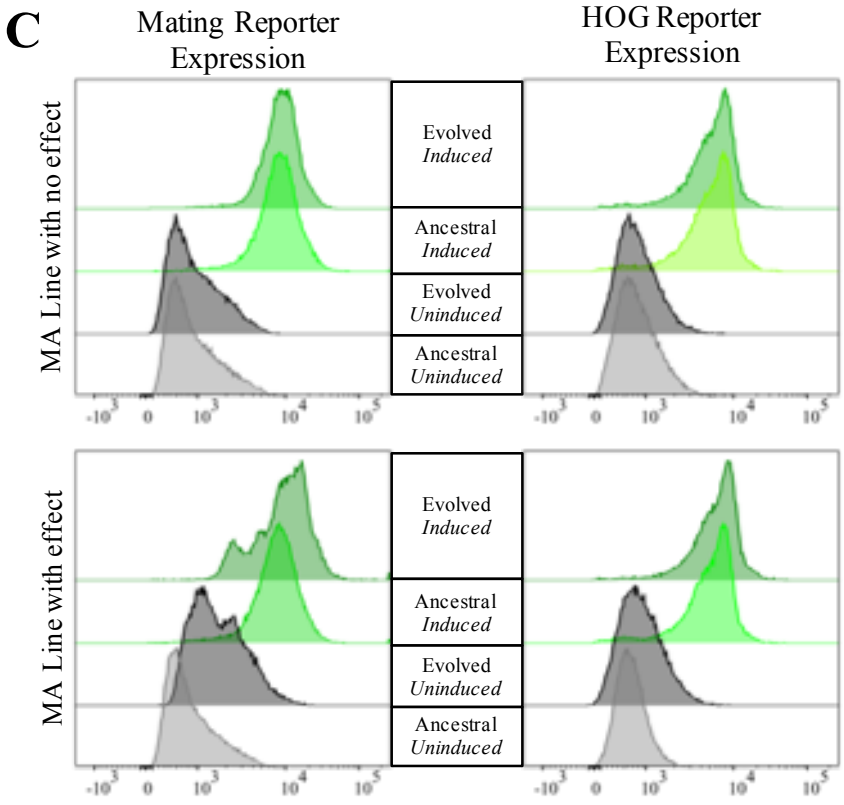

C
Mating Reporter Expression
HOG Reporter Expression
MA Line with no effect
MA Line with effect
Evolved Induced
Ancestral Induced
Evolved Uninduced
Ancestral Uninduced
-10^3
0
10^3
10^4
10^5

Figure 3-1: Schematic of yeast MAPK pathways and experimental setup. A) The yeast mating pathway (left) responds to pheromone and turns on mating related genes and the HOG pathway (right) responds to salt and turns on osmoregulation related genes. The HOG pathway consists of two, partially redundant, osmosensing branches. Both pathways consist of the three-tiered MAPK cascade - Ste11-Ste7-Fus3 in mating and Ste11/Ssk2/22-Pbs2-Hog1 in HOG pathway. The pathways consist of various feedback mechanisms (negative interactions are indicated with red lines) and shared components (orange lined circles). B) The two strains used in the mutation accumulation study contain GFP expressed from a mating responsive promoter, pFUS1, (left) or a HOG responsive promoter, pSTL1, (right) reporting mating and HOG pathway responses, respectively. C) Distributions of mating and HOG reporter expression after 240 minutes of 1uM mating pheromone induction and 30 minutes of 0.4M KCl induction. Evolved and ancestral lines for two mutation accumulation lines are shown, one with no effect on reporter distribution (top) and one where reporter distributions change in shape (bottom).

Similar to other MAPK and signaling pathways (Seaton and Krishnan 2011), mating and HOG pathways use some of the same components to regulate different processes (Figure 3-1A, orange circles). Despite these shared components, the HOG and mating pathways maintain signaling fidelity as cells respond only to the appropriate stimulus (Patterson *et al.,* 2010). Mechanisms such as scaffold proteins (Albert and Oltvai 2007), docking interactions (Saito 2010), negative-feedback loops (Hao *et al.,* 2008; Nagiec and Dohlman 2012; O'Rourke and Herskowitz 1998; Yamamoto *et al.,* 2010) maintain the signaling specificities of the two pathways. At physiological levels of stimulus, the two pathways are insulated in their response (McClean *et al.,* 2007; Patterson *et al.,* 2010). Mutations that disrupt this insulation may be identified by measuring the pathway responses to the inappropriate stimulus.

To control for mutational target size, we introduce mutations into a shared component of mating and HOG pathways, Ste50, an adapter protein responsible for recruiting the MAPKKK Ste11 to the membrane for activation by Ste20. Ste50 interacts with Ste11 via its Sterile Alpha Motif (SAM) domain and with membrane proteins Opy2, Sho1 and Cdc42 via its Ras-associating (RA) domain (Hao *et al.,* 2008; Yamamoto *et al.,* 2010). There is also a cluster of MAPK consensus phosphorylation sites in the intrinsically disordered region (IDR) that are known to modulate MAPK signaling (Zarin *et al.,* 2017). Mutations in the SAM domain of Ste50 decrease osmoresistance and increase crosstalk with the mating and filamentation pathways under hyperosmotic conditions (Jansen *et al.,* 2001). While Ste50 is required for both mating pathway activation and activation of the Sho1 branch of the HOG pathway, there are some differences in the role of Ste50 in the two pathways. For example, in addition to Fus3 and Hog1, another protein kinase, a casein kinase I, phosphorylates Ste50 at a specific threonine residue in the SAM domain (Chen and Thorner 2007). This phosphorylation is required for the mating response but not HOG response indicating the importance of Ste50 interactions in ensuring Ste11 activates the appropriate MAPK. To ensure the effects of mutations in Ste50 are not masked by redundancy of the HOG pathway, we inactivated the Sln1 branch by deleting the MAPKKKs SSK2 and SSK22. While the partially redundant branch contributes the HOG pathway robustness, the tight regulation required for the transient activation (adaptation) of the HOG pathway MAPK even in the absence of the redundant branch, suggests that HOG pathway may be more robust to mutations in Ste50 than the mating pathway. To monitor the effects of the same mutations on both pathways, we engineered a strain with two fluorescent reporters, each under the control of a

pathway-specific promoter. This strain also allows us to simultaneously study the effects of mutations on pathway insulation.

Point mutations are an important source of genetic variation and many systems-level studies on robustness assess the effects of deleting genes or altering expression levels (Bauer *et al.,* 2015). Here, in addition to studying the effects of whole genome spontaneous mutations by experimental evolution, we explore the effects of random point mutations in a shared component on the robustness of different signaling pathway responses and measure the responses by flow cytometry. The two approaches complement one another allowing both a higher-level understanding of the effects of mutations and a targeted, more comparable and mappable study of the effects of mutations. With both approaches, we investigate how random mutations affect the system at the level of the pathway as our previous work (Kompella *et al.,* 2017) and the work of others (Furukawa and Hohmann 2013; Levchenko *et al.,* 2000; Suderman and Deeds 2013; Thomson *et al.,* 2011) have shown that mutations in signaling pathways must be interpreted in the context of the whole pathway, rather than the effects on the protein. For example, in previous work, we showed that stop codons in the yeast mating MAPK can lead to wild-type signaling in the context of an over-expression mutant.

Our results show that both spontaneous whole genome mutations and random mutations in the shared adapter protein, Ste50, have a stronger effect on the mating pathway than on the HOG pathway signaling. We show that spontaneous mutations decrease the information capacity of the mating pathway without affecting the HOG response. We also find that these mutations increase the variance in mating signaling. After controlling for mutational target size and redundancy, we find that the HOG pathway shows decreased, but higher than mating signaling, robustness. Finally, we show that mutations in Ste50 can have opposite effects on the amount of information in mating and HOG signaling crosstalk.

3.3 Results

3.3.1 Whole-genome spontaneous mutations affect mating reporter activity more than HOG reporter activity

To test the phenotypic effects of spontaneous whole genome mutations on MAPK pathways in S. cerevisiae, we performed a MA experiment whereby we propagated 15 replicate lines of wild-

type strains with a pathway reporter either for mating or HOG signaling response, pFUS1-GFP and pSTL1-GFP respectively (Figure 3-1B). Since we are limiting the influence of selection, we expect deleterious mutations to accumulate and result in an eventual loss of signaling response. We measure the effects of MA on signaling responses for each pathway by measuring pathway reporter activities of the replicate lines at the end of the evolution experiment to the isogenic, ancestral lines from the beginning, before and after induction with mating pheromone or salt (Figure 3-1C).

To analyze the effects of spontaneous mutations on the two pathway activities, we first looked at the median fluorescent reporter (GFP) activities, as is common for flow cytometry data. Across all of the lines, we find that mutations have a stronger effect on the variance in the mating reporter expression than that of the HOG reporter (Figure 3-2A, Student's F-test p-values < 0.05 comparing all time points of ancestral and evolved mating reporter lines with and without induction; Student's F-test p-values < 0.05 for only the first four time points with induction for HOG reporter lines, rest are not significant differences). For the evolved lines with large effects, some of the distributions of reporter expression are no longer unimodal and therefore, the median does not represent the distribution well. To better compare the differences in the distributions without using population metrics such as median, we turned to information theory.

One way to directly compare the entire distributions is by calculating Kullback-Leibler (KL) Divergence (Perez-Cruz 2008). The larger the difference between the two distributions, the larger the KL Divergence value. We quantified the KL divergence value of the mating and HOG responses between the evolved and ancestral lines. We find that the basal mating signaling and the mating response to pheromone are more affected by whole genome mutations than the basal HOG signaling or the HOG response to salt (data not shown). While this value is better at summarizing the effects of the mutations on the distributions, i.e., we can see the effects on basal signaling response, it is still not a comprehensive way of describing the signaling output as it requires separate quantification of the basal reporter expression levels and the induced reporter expression levels. We therefore sought a measure of signaling that considers the changes to the basal level as well as to the induced level.

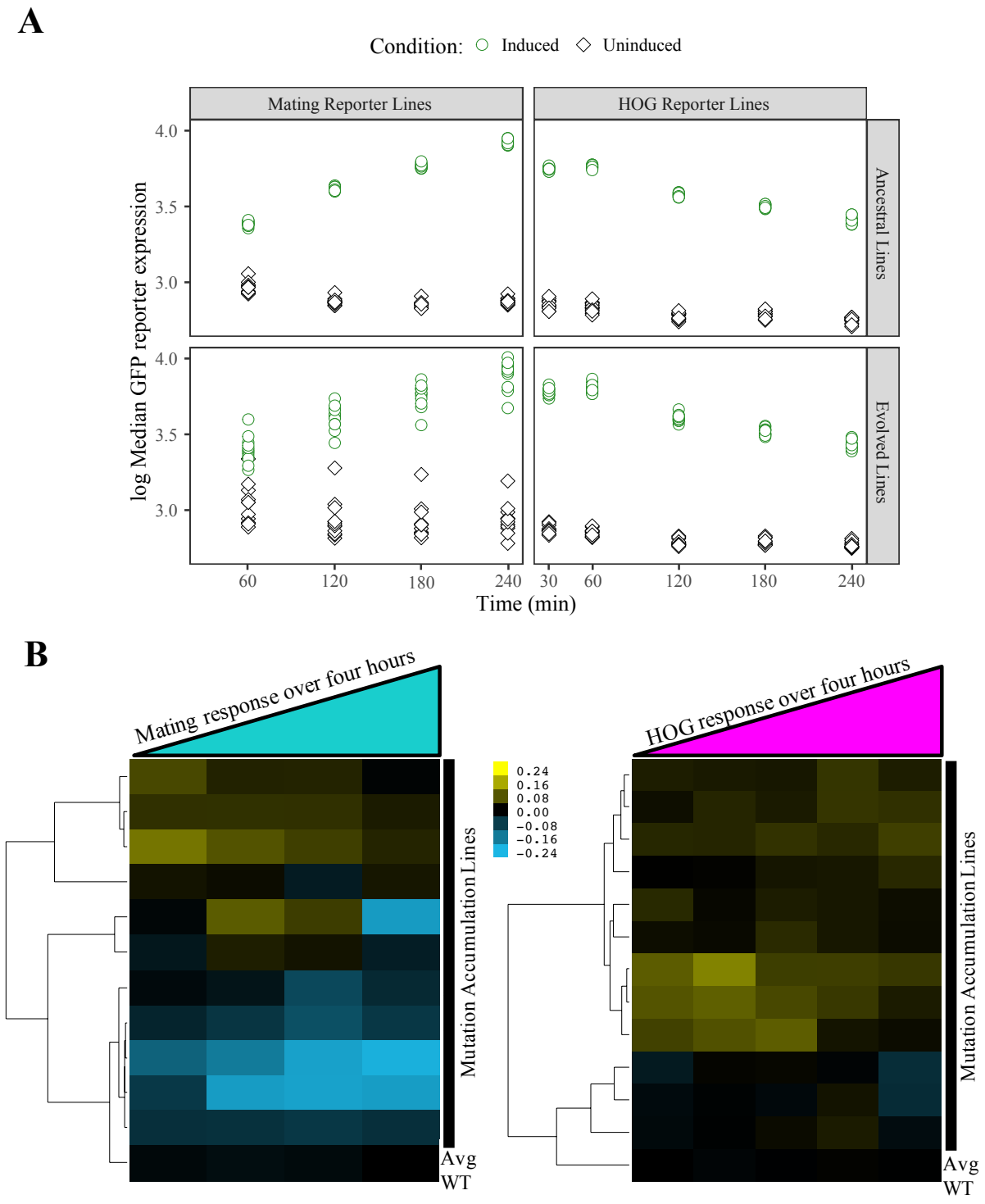


Figure 3-2: Effects of whole genome mutations on reporter expression and a clustered heat map of change in mutual information due to spontaneous mutations. A) Each point

represents the log median GFP expression value at the indicated time of signaling pathway response for one line of the mating (left) and HOG (right) reporter strains with induction (green circle) and without induction (black diamond). Mating pathway was induced with 1uM mating pheromone and HOG pathway was induced with 0.4M KCl. Top panels represent the isogenic, ancestral lines and bottom panels are the evolved lines. B) This heat map represents the effect of mutations on mutual information as represented by the change in mutual information (evolved-ancestral). Each row shows the change in mutual information for one line of the mating (left) or HOG (right) reporter lines. The last row is the average change in mutual information in the ancestral lines ("Avg WT") The columns represent the change in mutual information over four hours of signaling response. Positive values are yellow and negative values are blue. The intensity of the color corresponds to the magnitude of the change in mutual information, as shown in the color bar.

Measuring signaling fidelity by mutual information (MI) is a more holistic way of comparing the responses of signaling pathways (Filippi *et al.,* 2016; Mc Mahon *et al.,* 2015). To calculate an overall measure of information transmission of a communication channel, we can ask how much knowing the input determines the output (or vice versa) and MI is a measure of this reduction in uncertainty. For cell signaling responses, if there is a large overlap in the distributions of responses with and without a stimulus (or multiple doses of a stimulus or multiple stimuli), then knowing the response distribution does not aid in discerning the state of the stimulus (Granados *et al.,* 2018). Specifically, cells with responses in the overlap fail in determining the state of the stimulus causing the loss of information. Quantifying the signaling responses in terms of information transmission not only allows us to observe all kinds of effects of mutations on the distributions but also takes into account the correlation of the signal and output including the intrinsic differences in the noise in the two pathway reporter expressions such as the higher levels of noise in the mating pathway reporter expression than the HOG pathway reporter expression (Waltermann and Klipp 2011). To measure the efficiency of a signaling pathway in transducing the signal to a given pathway, we need to consider the biological noise in the response to the signal (as measured by the distribution of responses of isogenic cells to the same stimulus). MI captures the relationship of both the mean and shape of the distribution of the output with the input, is increasingly being adopted as a measure of signaling phenotypes (Granados *et al.,* 2018).

Here, we wish to quantify the effect of mutations on signaling phenotypes. We therefore introduce the change in MI, (i.e., the amount of transmitted information gained/lost due to mutations,) as the measure of the effects of the mutations on the signaling fidelity of the mating and HOG pathways.

To calculate the change in MI due to mutations, we subtracted the amount of MI (see Methods for details), in the ancestral lines from evolved lines. If mutations do not have an effect, the change in MI would be zero. If mutations decreased the information capacity of the pathway response, the change in MI would be less than zero. Across 12 MA lines for each pathway reporter, we found that mutations knockout mating signaling more than HOG signaling (5/12 mating reporter lines showed a decrease in MI at all time points tested, compared to 0/12 HOG reporter lines, Fisher's Exact test p-value = 0.03727, Figure 3-2B). We also found that mutations decrease the amount of information transmitted through the mating pathway more than that of

the HOG pathway (10/12 mating reporter lines showed a decrease in MI at one or more of the time points tested compared to only 3/12 HOG reporter lines, Fisher's Exact test p-value = 0.01228).

We quantified the variance in MI across the 12 MA lines for each strain and compared the variance in the ancestral lines and evolved lines. We found a significant increase in the variance of the mating signaling at 3 of the 4 time points (60 min, 180 min, and 240 min) and a significant decrease in the variance of the HOG response at one of the 5 time points (180 min) (F-test p-value < 0.05, Figure 3-3). This leads us to conclude that mutations had a stronger effect on the mating signaling variance than on that of the HOG signaling. In general, variance in MA lines increases with time in mutation accumulation studies, likely due to the accumulation of deleterious mutations (Eyre-Walker and Keightley 2007). To our knowledge, our study is the first to suggest that signaling phenotypes follow the same general pattern as other phenotypes.

Since the HOG pathway has a second, partially redundant branch, it is possible the effects of mutations are being masked due to the redundancy. The HOG pathway is also known to show adaptation (Saito and Posas 2012), and it is possible that the pathway is simply adapting to aberrant signaling levels caused by mutations in each MA line. In either case the differences between the pathways would be due to the differences in pathway structure and dynamics. On the other hand, it is possible that the differences we observed are simply due to a difference in mutational target size: protein components of the mating pathway may be more sensitive to mutations and there may simply be more proteins, such that mutations are more likely to fall within the mating pathway.

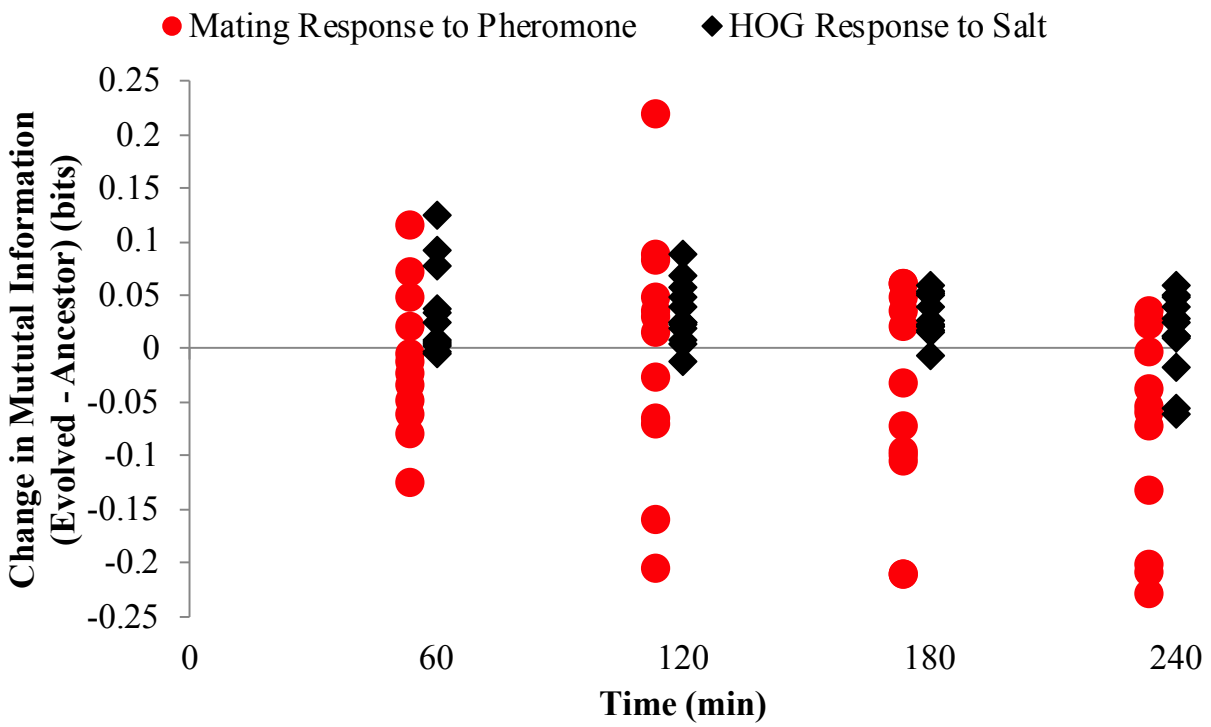
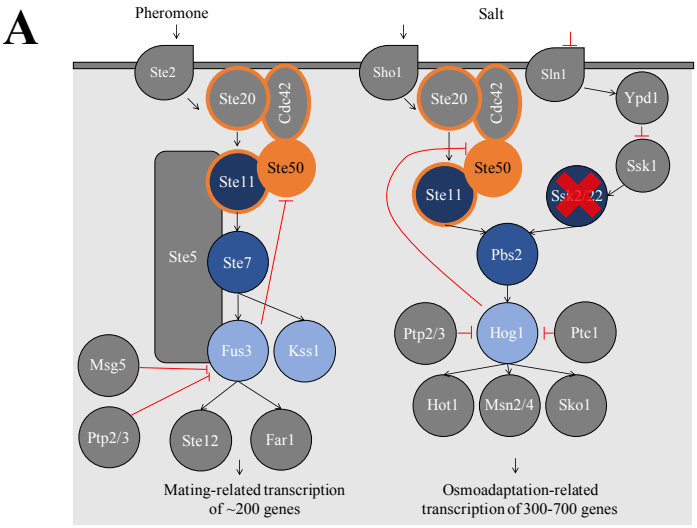


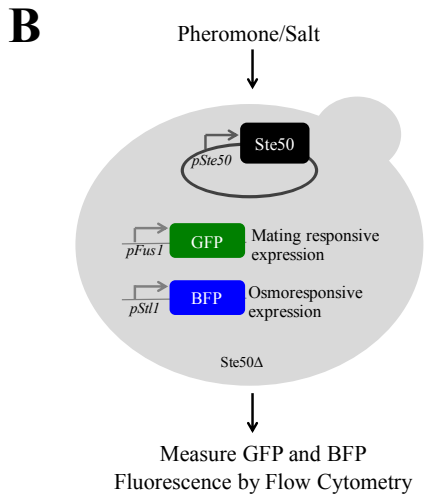
Figure 3-3: Effects of spontaneous mutations on amount of information transmitted to the HOG and mating pathways. Each symbol represents the change in mutual information (see Methods for details) for one mutation accumulation line at the indicated time after signaling pathway induction. HOG pathway reporter lines (black diamonds) were induced with 0.4M KCl and mating reporter lines (red circles) were induced with 1uM mating pheromone. Data shown are for twelve lines for each strain.

3.3.2 Mutational target size and redundancy explain most but not all of HOG signaling mutational robustness

To control for mutational target size, we next sought to quantify the robustness of mating and HOG pathways to the same mutations. Here, we took advantage of the promiscuity of the yeast MAPK pathways and introduced random mutations by error-prone PCR into Ste50, a shared component that is involved in both the mating and osmostress responses in a strain where the second branch of the HOG pathway is knocked out (Figure 3-4A). Ste50 recruits the mating and HOG MAPK kinase kinase Ste11 to the plasma membrane for activation by the p21-activated protein kinase Ste20 and is required for the proper function of both of HOG and mating MAPK pathways (Saito and Posas 2012; Truckses *et al.,* 2006). The MAPKs from both pathways phosphorylate Ste50 to downregulate the responses (Yamamoto *et al.,* 2010). Mutations in Ste50 have been known to change HOG and mating pathway signaling dynamics and basal signaling responses (Hao *et al.,* 2008; Yamamoto *et al.,* 2010)

To test the effects of same mutations in Ste50 on both mating and HOG pathway responses, we engineered a dual reporter *S. cerevisiae* strain with fluorescent mating and HOG pathway responsive reporters (Figure 3-4B). We expressed either wild-type (WT) Ste50 or a library of random mutants from pSte50 on a plasmid and measured the reporter expression of both pathways in parallel (Figure 3-4B). This strain design allows us to also measure both the signaling response ("signal") (using the reporter expression with the appropriate stimulus) and ("crosstalk") (using the reporter expression of the opposite stimulus) (Figure 3-4C).


A
Pheromone
Ste2
Ste20
Cdc42
Ste11
Ste50
Ste5
Ste7
Fus3
Kss1
Msg5
Ptp2/3
Ste12
Far1
Mating-related transcription
of ~200 genes
Salt
Sho1
Ste20
Cdc42
Ste50
Ste11
Sln1
Ypd1
Ssk1
Pbs2
Ptp2/3
Hog1
Ptc1
Hot1
Msn2/4
Sko1
Osmoadaptation-related
transcription of 300-700 genes


B
Pheromone/Salt
pSte50
Ste50
pFus1
GFP
Mating responsive
expression
pStl1
BFP
Osmoresponsive
expression
Ste50Δ
Measure GFP and BFP
Fluorescence by Flow Cytometry

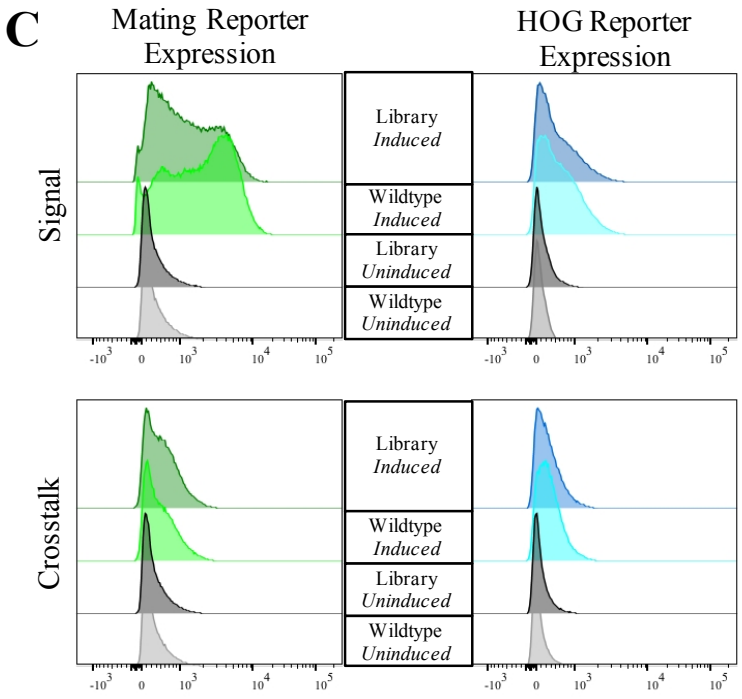

C
Mating Reporter
Expression
HOG Reporter
Expression
Signal
Crosstalk
Library
Induced
Wildtype
Induced
Library
Uninduced
Wildtype
Uninduced
-10^3
0
10^3
10^4
10^5

Figure 3-4: Experimental design to study effects of mutations in a shared component. A) Schematic diagram of the yeast mating and HOG response pathways highlighting the shared component Ste50 (orange circle) and the deletion of the Ssk2/Ssk22 (red cross) resulting in inactivation of the Sln1 branch of the HOG pathway. B) The yeast strain used in this experiment contains GFP expressed from a mating pathway responsive promoter, pFUS1, and BFP expressed from a HOG pathway responsive promoter, pSTL1. Wildtype Ste50 or a library of random variants generated by error-prone PCR (see Methods for details) is expressed on a plasmid from its endogenous promoter. C) Distributions of mating (left) and HOG (right) reporter expression with the appropriate stimulus ("signal", top panels) and the incorrect stimulus ("crosstalk", bottom panels) comparing library of Ste50 variants and wildtype Ste50. Top left panel shows distributions of mating reporter (GFP) expression after 360 minutes with and without 1uM mating pheromone. Top right panel shows distributions of HOG reporter (BFP) expression after 90 minutes with and without 0.4M KCl. Bottom left panel shows distributions of mating reporter (GFP) expression after 210 minutes with and without 0.4M KCl. Bottom right panel shows distributions of HOG reporter (BFP) expression after 270 minutes with and without 1uM pheromone.

First, to ensure that we are not introducing expression level differences when expressing Ste50 from a plasmid, we compared mating pathway response to pheromone of a strain carrying wild-type Ste50 and a Ste50 knock out strain (Ste50Δ) expressing Ste50 from its endogenous promoter (pSte50) on a plasmid and did not find a significant difference (Figure 3-5A). Protein stability and, thereby, function decreases exponentially with the number of substitutions (Bloom *et al.,* 2006). Therefore, we sequenced 49 random Ste50 variants from the random mutant library generated by error-prone and calculated the mutation rate to be 2.4 mutations per Ste50 (Figure 3-5B) (see Methods for details). We expect at least 26% of the mutations to have a wild-type like response (the fraction of the variants containing no amino acid changes) (Figure 3-5C).

To quantify the effect of these mutations on signaling response, we calculated the change in MI with the Ste50 library compared to wild-type Ste50. Unlike the effect of whole genome mutations, we find random mutations in Ste50 decrease information transmission in both HOG and mating signaling (38/44 time points taken over four replicate experiments show decreases in the HOG response to pheromone and 42/44 time points taken over the four replicate experiments show decreases in the mating response to pheromone) (Figure 3-6A). Since the two pathways are expected to have different durations in the response (HOG response adapts, and mating response shows sustained activity), we compared the maximum change in MI and found that the mating response to pheromone (at 360 minutes) is more severely affected by mutations in Ste50 than the HOG response to salt (at 120 minutes), (Student's t-test p-value < 0.05).

We quantified the effects of mutations on pathway insulation ("crosstalk") by calculating the changes in MI in the HOG response to pheromone and mating response to salt due to the library of Ste50 variants (Figure 3-6B). Mutations in Ste50 have opposite effects on the two crosstalk responses during the last three hours of response (Student's t-test p-values <0.01 for all time points from 180 minutes to 360 minutes). Mutations decrease HOG crosstalk more than mating crosstalk (35/44 time points tested over four replicate experiments showed decreases in HOG crosstalk compared to 10/44 for mating crosstalk). Mutations in Ste50 also have a small effect on the mating response to salt. Mutations increase mating crosstalk more than HOG crosstalk (34/44 time points tested over four replicate experiments showed increases in mating crosstalk compared to 9/44 for HOG crosstalk).

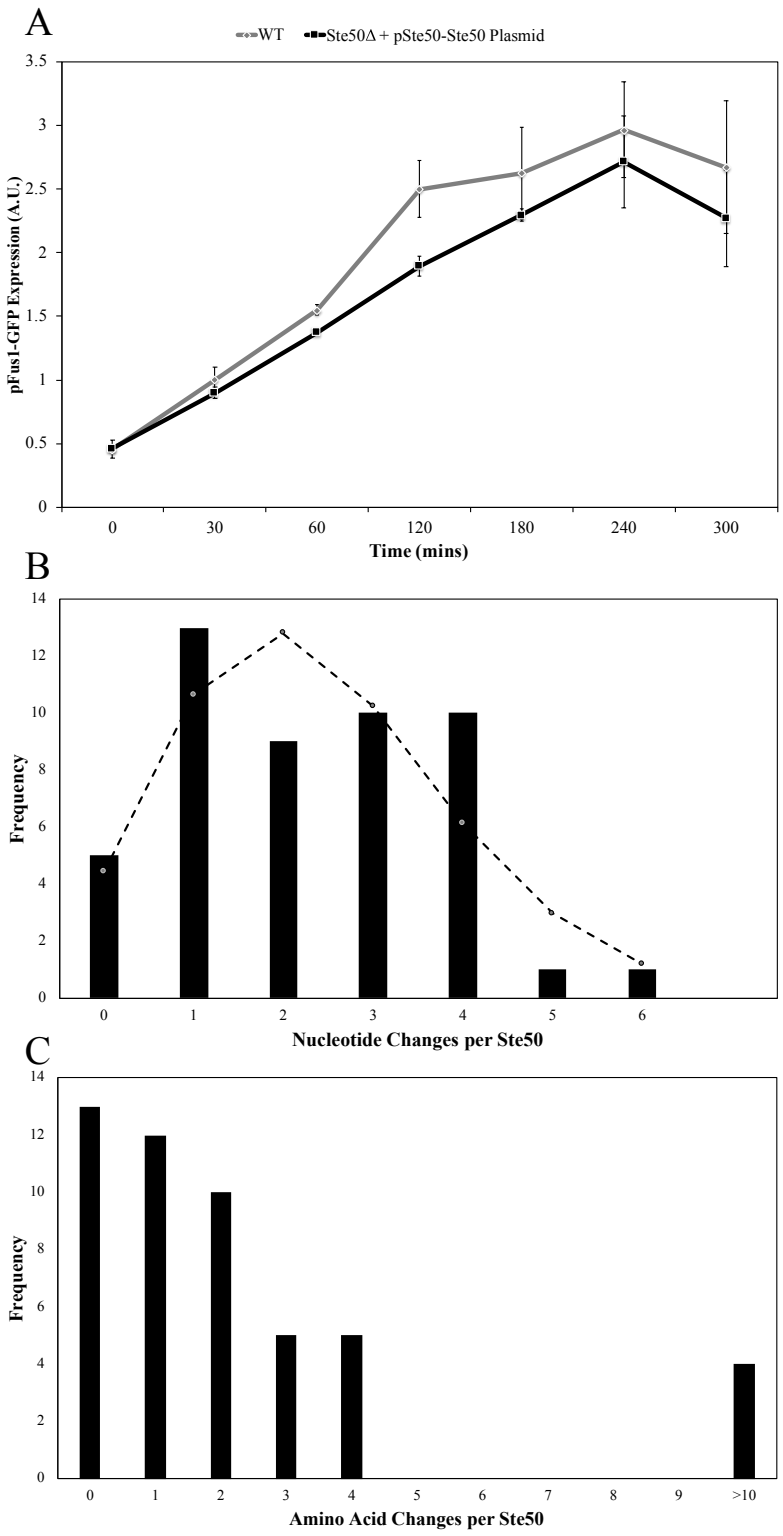


Figure 3-5: Construction of a Ste50 random mutant library. A) Average pFUS1-GFP levels obtained from flow cytometry after mating induction with 1uM mating pheromone for Ste50 expressed on a plasmid from its endogenous promoter in a Ste50 deletion strain (black trace) and

wildtype strain (grey trace). Error bars represent standard deviation with three replicates. B) Distributions of nucleotide changes in Ste50 random mutant library. Each black bar represents the frequency of Ste50 variants (y-axis) carrying the number of mutations listed on the x-axis. The dotted trace is the predicted percent variants under a Poisson fit, as determined by the method of least squares. C) Distributions of amino acid residue changes in Ste50 random mutant library, as in B).

3.3.3 Random variants carrying mutations in Ste50 SAM and RA domains decrease robustness of mating and HOG signaling

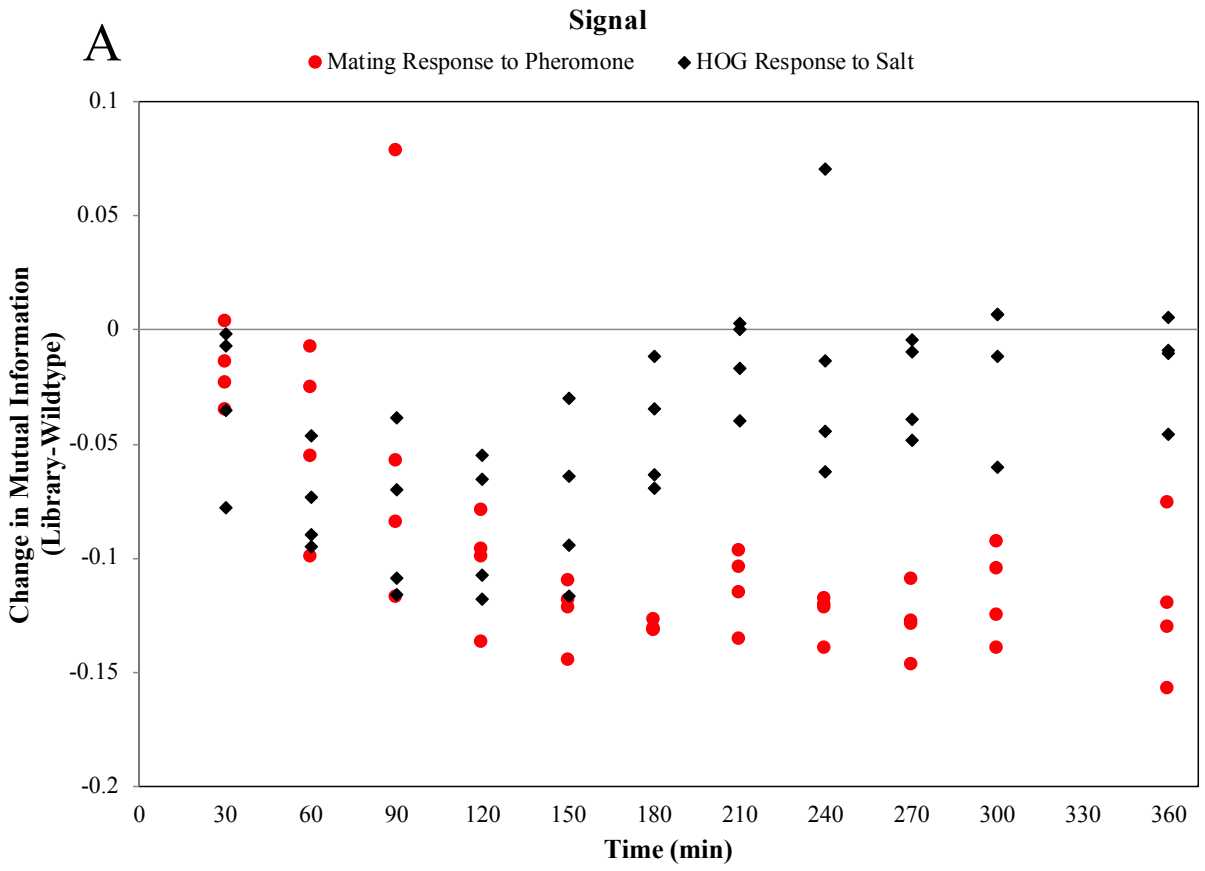
While we could measure the overall effect of mutations in Ste50 on mating and HOG signaling and crosstalk using the library of random variants, to more directly compare the Ste50 random variants to the lines obtained in the MA experiments, we isolated 49 individual Ste50 variants. We identified the mutations by sequencing (Table 3-1). We measured the pathway responses to pheromone and salt every hour for six hours and quantified the change in MI due to mutations (Figure 3-7). Five of the 49 variants had no changes to the nucleotide sequence ("WT") and we used these to understand the non-genetic sources of variance in MI (Figure 3-7, dashed lines). There were also 8 variants with only synonymous mutations ("WT-like"). We did not group these with the WT as synonymous mutations may have an effect on protein function (Plotkin and Kudla 2011).

We found that more variants affect HOG signaling (31/44) than mating signaling (16/44) at the earliest time point of 60 minutes (Fisher's Exact test p-value = 0.0026, Figure 3-7A). However, there are more variants that affect mating signaling than HOG signaling at the last three time points (240 minutes: 30/44 for mating compared to 16/44; 300 minutes: 31/44 for mating compared to 19/44 for HOG, and 360 minutes: 33/44 for mating compared to 23/44 for HOG; Fisher's Exact test p-value < 0.05).

We quantified the amount of variance in the effect of the mutations on MI. Similar to the MA lines, we found that the variance in the mating response increases more than the variance in the HOG response for four of the six time points (F test p-value < 0.05).

Mutations in Ste50 had opposite effects on crosstalk at all time points except the last one (at 60 minutes, 25/27 variants with effects decreased HOG crosstalk whereas only 5/24 variants with effects decreased mating crosstalk, at 120 minutes these numbers were 23/29 compared to 0/22, at 180 minutes these numbers were 14/19 compared to 4/21, at 240 minutes, these numbers were 10/21 compared to 1/22 and at 300 minutes, the number of variants with effects that decreased HOG crosstalk were 12/27 compared to 1/21 variants with similar effects on mating crosstalk, Fisher's Exact test p-value <0.05) (Figure 3-7B).

We mapped each mutation to one of the known functional domains in Ste50: the N-terminal Sterile Alpha Motif (SAM), intrinsically disordered region (IDR) and the C-terminal Ras-associating (RA) domain (Figure 3-8A, Table 3-1). To better visualize these effects for each variant, we clustered the change in MI (Figure 3-8B).


A
Signal
Mating Response to Pheromone
HOG Response to Salt
Change in Mutual Information (Library-Wildtype)
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
0
30
60
90
120
150
180
210
240
270
300
330
360
Time (min)

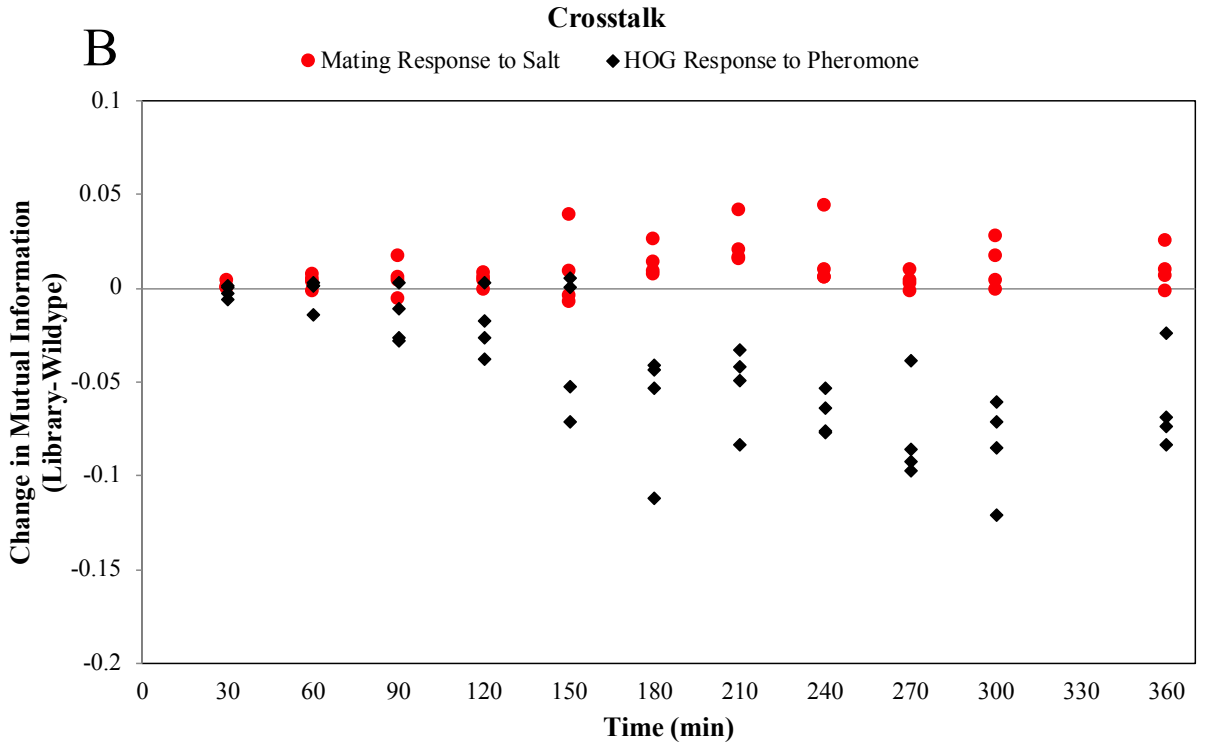

B
Crosstalk
Mating Response to Salt
HOG Response to Pheromone
Change in Mutual Information (Library-Wildtype)
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
0
30
60
90
120
150
180
210
240
270
300
330
360
Time (min)

Figure 3-6:. Effects of mutations in shared component on amount of information transmitted to the HOG and mating pathways. A. Each symbol represents the change in mutual information (see Methods for details) in the correct signaling response ("signal") with the library of Ste50 random variants to wildtype Ste50 at the indicated time. HOG pathway response (black diamonds) to 0.4M KCl and mating pathway response (red circles) to 1uM mating pheromone. B. Each symbol represents the change in mutual information in the incorrect signaling response ("crosstalk") with the library of Ste50 random variants to wildtype Ste50 at the indicated time. HOG pathway response (black diamonds) to 1uM mating pheromone and mating pathway response (red circles) to 0.4M KCl. Data shown are for four replicate experiments.

Table 3-1: Sequence data for 49 individual Ste50 random variants.

Variant ID	AA Change in SAM (30-104)	AA Change in IDR (151-251)	AA Change in RA (246-326)	Nonsense Mutation	1 or more Provean predicted Deleterious	Number of Synonymous Changes	Amino Acid Changes
7	No	No	Yes	No	Yes	0	I306F
8	No	Yes	Yes	No	Yes	0	R200S, S215P, L218S, D279Y
26[a]	No	No	No	No	No	1	WT
46	No	Yes	No	No	Yes	0	D20V, D38V, S196T, H214R
12	No	Yes	Yes	No	Yes	0	N121D, H192R, A258T
16[a]	No	No	No	No	No	2	WT
10[b]	No	No	No	No	No	0	none
11	No	Yes	No	No	No	2	S12P, N243Y
15[c]	No	No	No	No	No	2	N339Y
44[a]	No	No	No	No	No	1	WT
49[b]	No	No	No	No	No	0	none
14	Yes	Yes	No	No	No	1	L19Q, R59T, P156Q,
41	No	Yes	No	No	No	1	D221H
28[a]	No	No	No	No	No	2	WT
4	Yes	Yes	No	No	No	2	E49V, R199G
31[b]	No	No	No	No	No	0	none
19[b]	No	No	No	No	No	0	none
24[b]	No	Yes	No	No	No	0	H214N
22	No	No	Yes	No	Yes	0	I306N

29[c]	No	No	No	No	No	2	G23D
17	Yes	Yes	Yes	No	Yes	0	V50A, P249H, Q294H, E299V
1	No	No	Yes	No	Yes	2	E302G
23	No	Yes	Yes	No	Yes	0	Q195R, A258V, R324G
21	Yes	Yes	No	No	Yes	1	D77G, T125A, N243S
13	Yes	No	No	Yes	Yes	1	F92I, R100*
18	No	No	Yes	No	Yes	2	P304H
3	Yes	No	Yes	No	Yes	0	L69P, M321T
6	No	No	Yes	No	Yes	2	L268S
25	Yes	No	Yes	No	Yes	0	D54E, G315V
27	No	No	Yes	No	Yes	0	R296G, V305E
43	No	No	No	Yes	Yes	0	M1*
45	No	No	Yes	Yes	Yes	0	L316*
33	No	No	No	No	Yes	0	...V21*
40	No	Yes	Yes	Yes	Yes	1	Q228-A252, L253*
50	Yes	No	No	No	No	1	N13D, V50M, G84S
30	Yes	No	No	No	No	0	L95P
36	No	Yes	Yes	Yes	No	3	Q233L, K269E
51	No	Yes	Yes	No	Yes	0	R141G, G179S, L218M, R283G
20[b]	No	No	No	No	No	0	none
2[a]	No	No	No	No	No	1	WT
9	No	No	No	Yes	Yes	0	G336R, S337K, D338*
34	No	Yes	No	No	No	1	S217F, S222R
32	No	Yes	No	No	No	0	V191I, V206A
37[a]	No	No	No	No	No	1	WT
38[a]	No	No	No	No	No	2	WT
39	Yes	No	Yes	No	Yes	0	E51D, I320F
48[a]	No	No	No	No	No	1	WT
42	No	Yes	Yes	No	Yes	0	T165A, A278V, N310I, S337G
47	No	Yes	No	No	No	0	Y187C

[a] No mutations lead to non-synonymous amino acid substitutions

[b] No mutations found

[c] No Mutations occurred in the domains listed

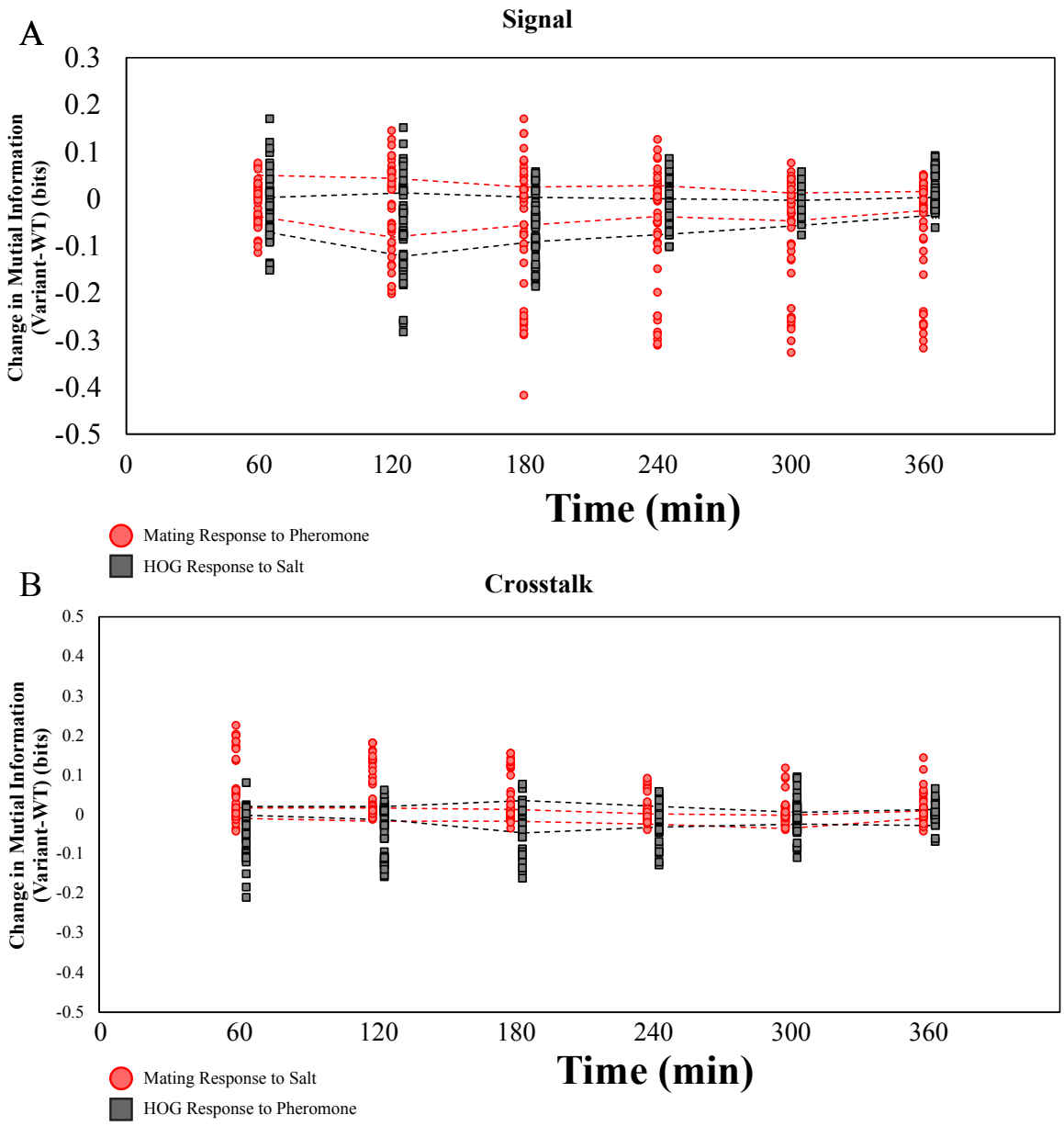
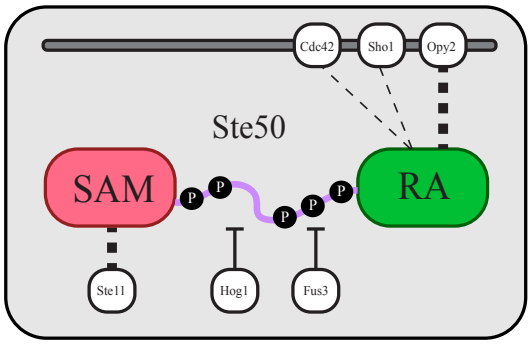


Figure 3-7: Direct comparison of the effects of mutations in shared component on amount of information transmitted to both the HOG and mating pathways. Each symbol represents the change in mutual information with a random Ste50 variant to wildtype Ste50 in the pathway responses after induction with 1uM mating pheromone or 0.4M KCl for the given times. A. The change in mutual information in the correct signaling responses ("signal", HOG pathway response to salt, black squares, and mating pathway response to pheromone, red circles). B. The change in mutual information in the incorrect signaling responses ("crosstalk", HOG pathway response to pheromone, black squares, and mating pathway response to salt red circles).

A


Cdc42
Sho1
Opy2
Ste50
SAM
RA
P
Ste11
Hog1
Fus3

B

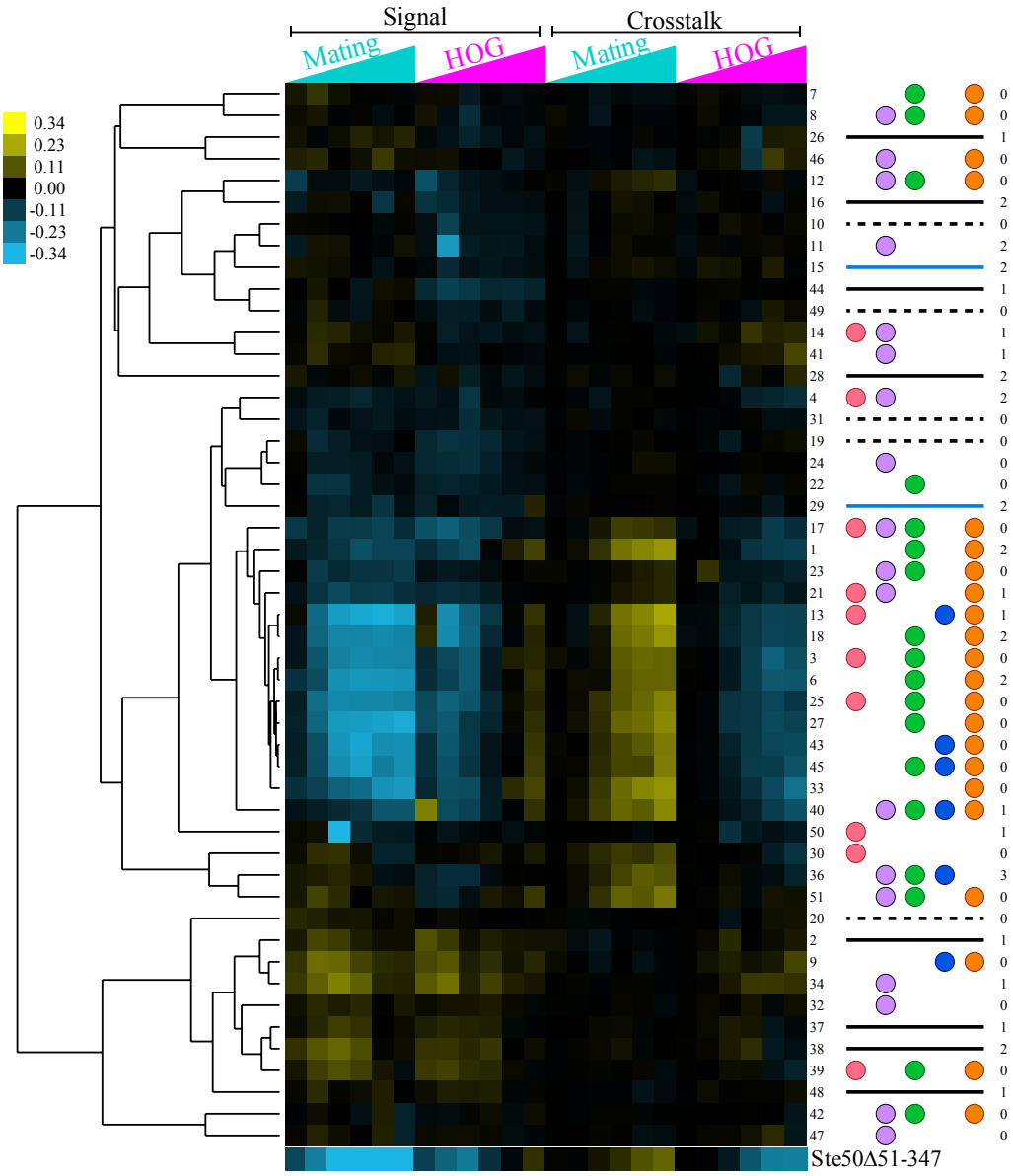

Signal
Crosstalk
Mating
HOG
Mating
HOG
0.34
0.23
0.11
0.00
-0.11
-0.23
-0.34
Ste50Δ51-347

Figure 3-8: Schematic of Ste50 domains and interactions and clustered heat map of change in mutual information due to mutations in Ste50. A. Ste50 recruits Ste11, MAPKKK, to the cell membrane (dark grey bar) via its N-terminal Sterile Alpha Motif (SAM) domain (light red rectangle). Ste50 interacts with membrane proteins, Opy2, Sho1 and Cdc42, via the C-terminal Ras-associating (RA) domain (green rectangle). The intrinsically disordered region (IDR, purple line) between the SAM and RA domains contains MAPK consensus phosphorylation sites (filled black circles). The mating and HOG pathway MAPKs, Fus3 and Hog1, phosphorylate Ste50 at these sites resulting in down regulation of the pathways. B. This heat map represents the effect of mutations on mutual information as represented by the change in mutual information (Ste50 variant - wildtype). Each row represents one Ste50 variant. The columns represent the change in mutual information over four hours of mating (teal triangle) and HOG (fuchsia triangle) signaling induction with the appropriate stimulus ("Signal", left) and the opposite stimulus ("Crosstalk", right). Positive values are yellow and negative values are blue. The intensity of the color corresponds to the magnitude of the change in mutual information, as shown in the color bar. The numbers to the right of the heatmap represent the variant identification number. The filled circles show the location and the type of the amino acid change in Ste50 protein sequence. Mutations in the SAM domain are indicated by light red circles, in the IDR by purple circles, and in the RA domain by green circles. Blue circles represent nonsense mutations. Orange circles represent mutations that PROVEAN, software used to predict the impact of amino acid substitutions, insertions and deletions on protein function, predicts to be deleterious for Ste50 function. The numbers to the right represent the number of synonymous substitutions in the Ste50 variant. Solid black lines indicate the variants with only synonymous substitutions ("WT-like") and dashed black lines indicate variants with no substitutions ("WT"). Blue lines indicate variants carrying missense mutations outside of the three domains.

We found a cluster of variants (ID#s: 17,1,23,21,13,18,6,25,27,43,45,33,40) that decreased mating and HOG signal, increased mating crosstalk and decreased HOG crosstalk (13/14 variants in this cluster displayed this phenotype across at least four out of the six time points tested compared to 3/35 outside this cluster, Fisher's Exact test p-value= 2.76×10^{-8}; 290/336 time points in the cluster, and 402/840 outside the cluster show this phenotype).

We find that this cluster is deprived of WT/WT-like variants (0/14 variants in this cluster are WT/WT-like compared to 13/22 outside the cluster, Fisher's Exact test p-value = 0.009759). The cluster is not enriched mutations in the SAM (5/14 variants in the cluster have mutations in the SAM domain compared to 5/35 outside the cluster. Fisher's Exact test p-value = 0.1235). The cluster is not enriched for mutations in the IDR (4/41 variants in the cluster have mutations in the IDR compared to 14/35 outside the cluster, Fisher's Exact test p-value = 0.5273). We do not expect mutations in IDR that do not change the charge of the phosphosites to have a strong effect on HOG or mating signaling (Zarin *et al.,* 2017).

The cluster is enriched for mutations in the RA domain (10/14 of these variants in this cluster had mutations in the RA domain compared to 7/35 outside the cluster, Fisher's Exact test p-value = 0.0001829). We used PROVEAN, a tool that predicts the effects of amino acid substitutions, insertions and deletions based on sequence and evolutionary conservation to predict the effects of the mutations in Ste50 on protein function (Choi *et al.,* 2012). We found that the cluster is enriched for mutations that PROVEAN predicts to be deleterious (14/14 of these variants in this cluster are predicted to be compared to 8/35 outside the cluster, Fisher's Exact test p-value = 4.736×10^{-7}). Variants in this cluster have a similar pattern of effects in MI in mating and HOG signaling as a truncated Ste50 variant (Ste50Δ51-347) suggesting that mutations in these variants are expected to have large impacts on Ste50 function and that the differences we are observing between the change in information between the mating and HOG signaling pathways are largely due to their different overall dependence on Ste50 function.

Although 4 out of 14 of these variants in this cluster carried non-sense mutations, the cluster was not enriched for them, as they were two additional variants carrying nonsense mutations outside of this cluster (Fisher's Exact test p-value = 0.1639). These additional two variants had similar effects on mating crosstalk as those in the cluster.

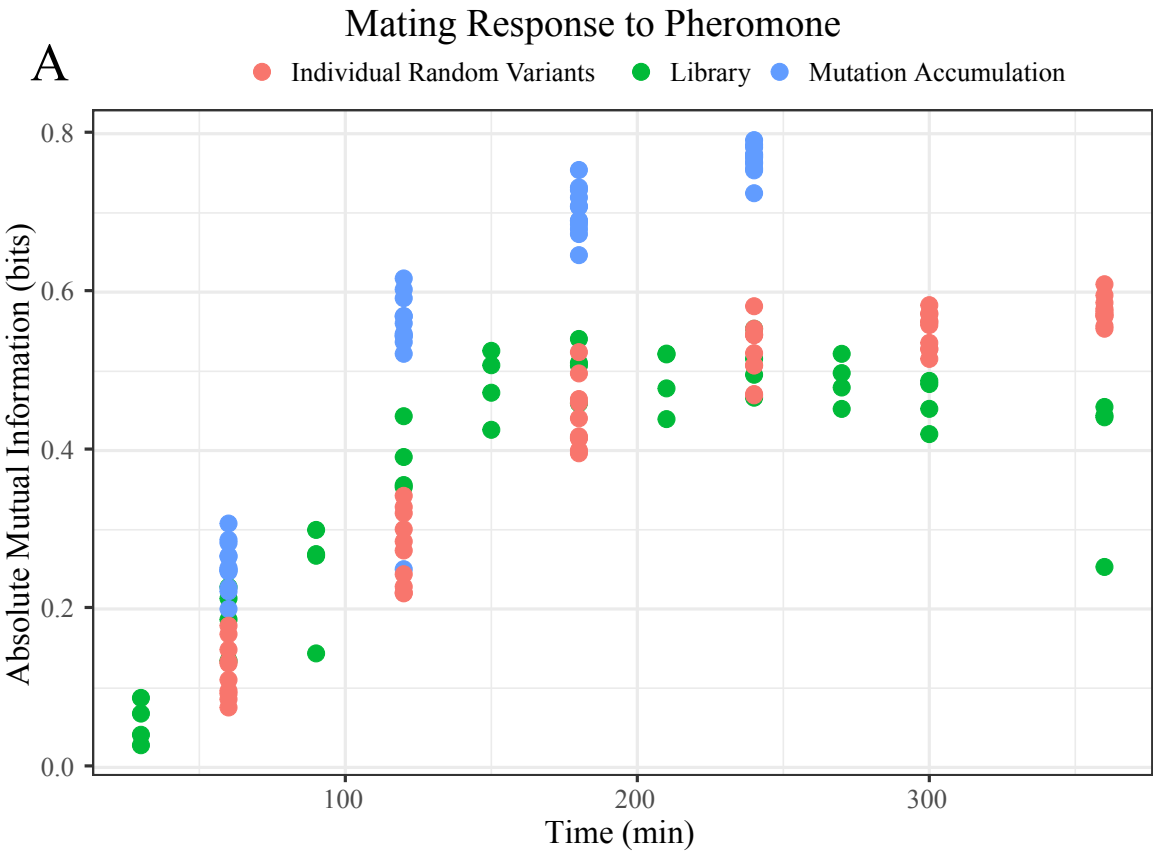
If we consider the effects on mating crosstalk alone, we find that in addition to the 14 variants in this cluster, there are four additional variants (ID#s: 12, 30, 36,51) that increase mating crosstalk. None of these 18 variants are WT or WT-like compared to 13 of the 31 other variants (Fisher's exact test p-value = 0.001593). These 18 variants are enriched for nonsense mutations (5/18 of these variants carry nonsense mutations compared to 1/35 outside of these variants, Fisher's Exact test p-value = 0.02032). They are also carrying mutations predicted to be deleterious by PROVEAN (16/18 of these variants have mutations predicted to be deleterious compared to 7/31 other variants, Fisher's Exact test p-value = 1.348×10^{-5}).

There are many ways in which the cluster of 14 or the 18 variants may be affecting signaling. Since they are enriched for mutations predicted to be deleterious, they are most likely reducing Ste50 function. Ste50 SAM domain mutations are known to affect mating and HOG signaling in opposite ways (Jansen *et al.,* 2001). Specifically, mutations in this domain can increase mating response to pheromone and salt and decrease the HOG response to salt or affect mating but not HOG signaling. Ste50 interacts with the membrane proteins Sho1 and Opy2 to direct the salt signal to the HOG pathway. The interaction between the Ste50 RA domain and the membrane protein, Opy2, is vital for osmoresistance (Yamamoto *et al.,* 2010). Mutations in the RA domain that affect this interaction can be rescued by tethering Ste50 to the membrane. Improper activation of mating signaling can occur if the normal binding partners of Sho1 are mutated such that the salt stress is incorrectly routed to the mating pathway (Marles *et al.,* 2004).

3.3.4 Strain background contributes more to differences in the absolute MI in the two pathways than growth conditions

In addition to differences in the strain background between the Ste50 library or individual random variant and MA experiments, there were some differences in the growth conditions between the library and individual random variant or MA experiments. Briefly, the library experiments were large-scale, requiring the pooling of hundreds of thousands of transformants compared to the other two experiments where a single colony was used to inoculate cultures (see Methods for details.)

To test if the different growth conditions affected reporter expression, we compared the amount of MI in the mating response to pheromone ("Mating Signal") and the HOG response to salt ("HOG signal"), absolute MI, in the three experiments (Figure 3-9). For the mating responses, where we measured this signal with the same fluorescent protein (GFP) in all experiments, we found one time point (240 min, around maximum mating response), where the differences in absolute MI between the library and individual random variant experiments were not significant (Student's t test p-value = 0.44, Figure 3-9A). In contrast, measurements at all four time points (60, 120, 180, 240 minutes) were significantly different between the MA and individual random variant experiments (where the growth conditions were the same) (Student's t test p-value < 1.3×10^{-7}). This suggests that growth conditions did not contribute to the differences observed in the absolute mating MI as much as strain background. For the HOG responses, where we measured the signal with BFP in the library and random variants experiment and GFP in the MA experiments, there were many time points where the growth conditions did not affect signal (Student's t test p-value = not significant for 120, 180, 240, 300 and 360 minutes comparing library and individual random variant experiments; only time point where the difference was significant is 60 minutes, Student's t test p-value = 0.01, Figure 3-9B). In comparison, the differences were significant for all the time points compared between the MA and individual random variants experiments where the growth conditions were the same (Student's t test p-value < 3.78×10^{-17}). This suggests that strain background contributed more to the differences in the absolute HOG MI than growth conditions.


A
Mating Response to Pheromone
Individual Random Variants
Library
Mutation Accumulation
Absolute Mutual Information (bits)
0.0
0.2
0.4
0.6
0.8
100
200
300
Time (min)

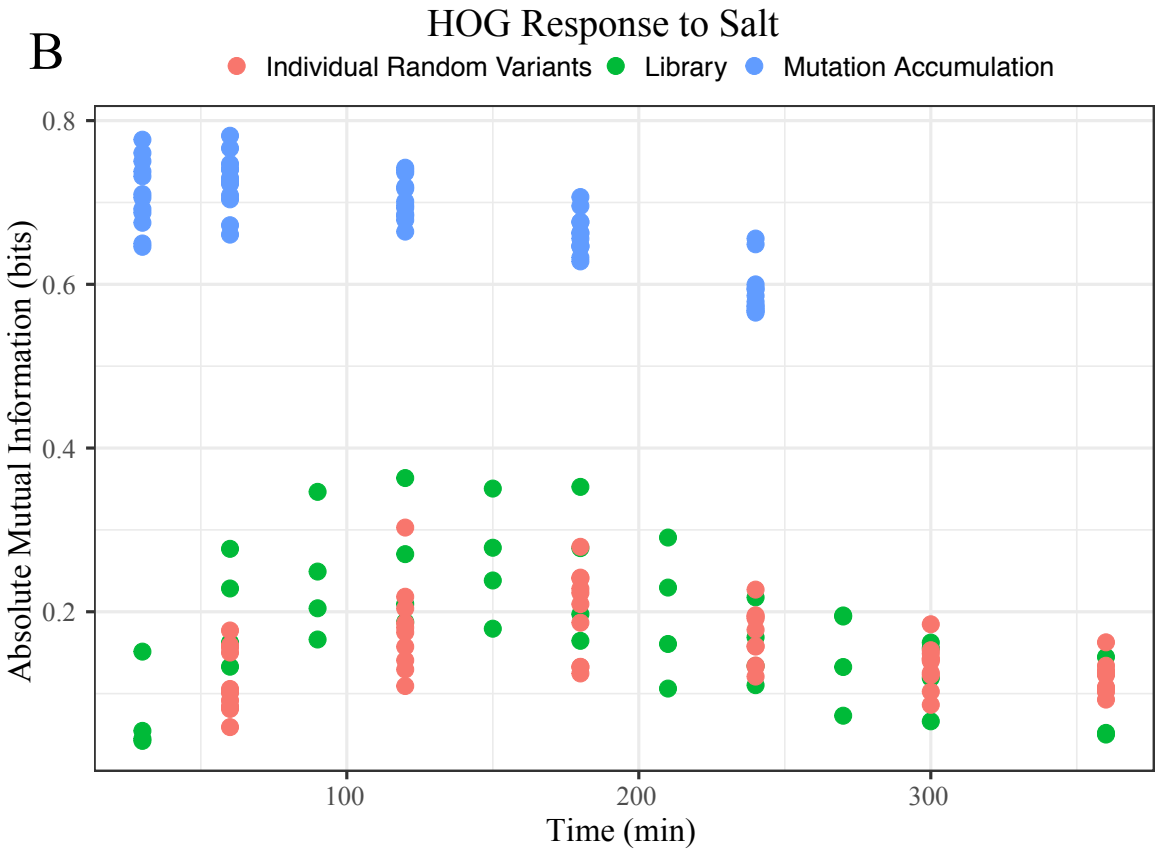

B
HOG Response to Salt
Individual Random Variants
Library
Mutation Accumulation
Absolute Mutual Information (bits)
0.2
0.4
0.6
0.8
100
200
300
Time (min)

Figure 3-9: Comparison of the information transmission capacity of the mating and HOG pathways with the three experimental methods in this study. A. The amount of information transmitted (absolute mutual information) in the mating response to pheromone with the three experimental methods employed in this study. Large-scale experiment ("Library") where many colonies of transformants expressing wildtype Ste50 were pooled are indicated with green circles. Ste50 random variant experiment ("Individual Random Variants") where a single colony of transformants expressing wildtype Ste50 were tested are indicated with light red circles. Single colonies of the ancestral lines in the mutation accumulation ("Mutation Accumulation") experiment are indicated with blue circles. Mating response to pheromone is reported by pFUS1-GFP in all experiments. In the library and individual random variant experiments, wildtype Ste50 is expressed from a centromeric plasmid in a Ste50Δ strain where the second HOG pathway branch has been inactivated by Ssk2/22 deletions. The mutation accumulation lines do not have deletions in either Ste50 or the second branch of the HOG pathway. Each point represents the absolute mutation for one sample after induction with 1uM mating pheromone for the given time. Data shown are the four replicates from the "library" experiments, the five wildtype controls plus the five Ste50 variants without any nucleotide changes from the "Individual Random Variants" experiments and the twelve ancestral lines from the "mutation accumulation" experiments. B. The amount of information in the HOG response to salt with the three experiments listed in A). The pSTL1-fused HOG reporter in the library and individual random variants strain is BFP and in the mutation accumulation lines is GFP. Each point represents the absolute mutation for one sample after induction with 0.4M KCl for the given time.

There are two confounding differences between the strains used in the MA experiments and the library/individual variant experiments – the fluorescent reporters and HOG pathway redundancy. To test how differences in the fluorescent reporter affected the measurements, we compared the maximum absolute MI for the mating and HOG signals in the MA experiment (where GFP reported both pathway activities) and the individual random variants experiment (where the growth conditions were the same as the MA and two different fluorescent proteins reported the two pathway activities). The absolute MI in HOG and mating signals were more comparable in the MA experiments than in the individual variant experiments (maximum mating absolute MI at 180-240 minutes compared to maximum HOG absolute MI at 30-60 minutes in the MA experiment were both around 0.8 bits, Student's t test p-value = 0.15 (not significant.); maximum mating absolute MI at 300-360 minutes, ~ 0.6 bits, compared to maximum HOG absolute MI at 120-180 minutes, ~0.2 bits, in the individual variant experiments, Student's t test p-value = 2.74×10^{-27}). Although this difference is likely due to the strain background (the strain used in the MA experiment has the second, partially redundant branch intact unlike the strain used in the individual random variants experiment). When both pathways are near wild-type, like in the MA experiments, the maximum amount of MI is comparable, around 0.8 bits, therefore, the change in mutual information due to the mutations represents similar effects. For example, with similar wildtype MI of 0.8 bits for both pathways, a decrease in mutual information of 0.2 bits represents a 25% decrease in MI for both pathways. However, if the wildtype MI for one of the pathways was 0.2 bits and the other 0.8 bits, a similar decrease in MI of 0.2 bits would represent 100% decrease for the former and 25% for the latter.

To test the contributions of the second branch to absolute MI of the HOG signal, we compared maximum absolute MI of the HOG signal in the strain with the branch to the strain without the second branch. We calculated a 3.78x fold reduction in absolute MI due to the missing branch (average maximum absolute MI in the MA experiment between, between 30-60 minutes, divided by the average maximum absolute MI in the individual random variant experiment between, between 120-180 minutes). This is comparable to the previously reporter estimate of 3.79x fold reduction (0.51 bits with two branches/0.14 bits with one branch) (Waltermann and Klipp 2011). Based on this, we believe that the strain background accounts for more of the differences in the MA and individual variants experiments, rather than the differences in the fluorescent reporter.

3.4 Discussion

We measured the effects of mutations on the amount of mutual information in mating and HOG signaling using two approaches, mutation accumulation to study effects of spontaneous mutations and random mutagenesis of a shared component to limit mutational target size and study effects of the same mutations on the two pathways, and with both methods we found that mutations had a stronger effect on the amount of mutual information in mating signaling than on HOG signaling.

The STL1 and FUS1 reporters we used in this study have previously been used for detection of the HOG and mating pathway activities, respectively (Furukawa and Hohmann 2013; Kompella *et al.,* 2017; Metzger *et al.,* 2017; Pelet *et al.,* 2011; Sato *et al.,* 2014; Zarin *et al.,* 2017). Upon activation, Hog1 induces the transcription of many stress-response genes including the strong expression of STL1, which is widely studied as model for hyperosmotic stress activated gene expression (Aymoz *et al.,* 2016; Bai *et al.,* 2015; Zhao *et al.,* 1994). Induction with mating pheromone changes the expression profiles of many genes, including a strong increase in FUS1 expression (McCaffrey *et al.,* 1987). Although these promoters are routinely used for detecting these pathway responses, it is likely that some of the differences we observe could be specific to these promoters. For a more detailed understanding of these complex cellular responses, other pathway-specific promoters should also be studied.

We observed higher variance in mating reporter activities than in HOG reporter activities. Previous work shows that in general, the longer the MA lines are propagated, the greater the increase in variance – likely due to the accumulation of deleterious mutations (Eyre-Walker and Keightley 2007). Our observed increase in variance could be due to the mating pathway components accumulating more deleterious mutations or fewer compensatory mutations than the HOG pathway, or not being able to buffer the effects of these mutations as well as the HOG pathway. In any case, our results suggest that it is easier for mutations to change mating signaling than HOG signaling.

It was not obvious that the HOG pathway would be affected by fewer mutations as Landry *et al.*, show in a MA study that genes with TATA boxes are more sensitive to genetic perturbation and

are more likely to accumulate mutations that affect their expression (Landry *et al.,* 2007). While one fifth of yeast genes contain a TATA box, stress-response genes are particularly enriched for these cis-regulatory elements. Even if the stress response genes in our MA lines had accumulated more mutations, our results suggest that the HOG pathway is more robust to the effects of these mutations more than the mating pathway. One reason for this is the redundancy in the pathway. We controlled for mutational target size and redundancy and found that mutations are able to affect HOG signaling (albeit still not as much as mating signaling). We suspect that this loss of information is due to the missing branch. In support of our hypothesis, previous work has shown that the amount of MI transmitted via a pathway that can detect an input using two channels is more than that of the pathway where one of these channels is not used (Waltermann and Klipp 2011).

We measure fluorescent reporter expression by flow cytometry as it is a high-throughput method that captures single cell data generating a distribution of fluorescent reporter expression levels representing the variability in gene expression that even isogenic populations display. Typically, the median reporter fluorescence intensity is used as a metric of measurement. While this metric is useful in summarizing the population behavior, reducing fluorescence distributions to population level values results in loss of information about the complex cellular response distribution that comparing the distributions directly can avoid (Handly *et al.,* 2016). One method for comparing distributions that makes intuitive sense for signaling response is information theory, typically used in communications but also to quantify cell signaling. Previous work has shown that presence of feedback loops and other mechanisms that reduce signal interference and direct the signal to the appropriate branch, increase the information transfer through that branch (Mehta *et al.,* 2009). Specifically, for the mating pathway the presence of a negative feedback loop has been shown (albeit not by information theory analysis) to increase the information transfer in the pathway (Yu *et al.,* 2008). Furthermore, using MI to quantify cellular responses allows us to understand the biological relevance of physiological changes such as an increase in basal activity of a pathway (Voliotis *et al.,* 2014).

Since MI is a correlation metric, if mutations affect the absolute signaling levels but not the relative signaling with and without induction, change in MI will not capture those affects. Furthermore, as change in MI summarizes the effects of mutations on the signaling response with and without induction, to distinguish between effects on basal activation and signaling

response after induction, we did a thorough revision of reporter distributions. In the mutation accumulation experiment, the decrease in the amount of information due to mutations is largely caused by increases in basal signaling. Although mutations in Ste50 are known to affect basal signaling of both mating and HOG pathways (Hao *et al.,* 2008; Yamamoto *et al.,* 2010), we did not observe these effects in our random mutagenesis experiment. Furthermore, we did not observe any increases in basal HOG signaling in any of the experiments. Although basal responses of both pathways are tightly regulated as both the mating and HOG response pathways trigger cell cycle arrest upon stimulation, albeit transiently in the case of the HOG pathway (Chen and Thorner 2007), our observation suggests that basal HOG signaling is more tightly regulated. Previous work shows that point mutations and expression level changes of mating pathway components can increase basal mating signaling output through the breakdown of feedback loops that tightly regulate these pathways (Hao *et al.,* 2008; Thomson *et al.,* 2011).

Different traits can display a spectrum of responses to mutations - from robustness to fragility - therefore the robustness of a trait is more appropriate as a relative measurement (Felix and Barkoulas 2015). Quantifying the effects of mutations introduced by mutation accumulation which introduce all types of mutations across the whole genome, and the more precisely controlled mutation of a shared component of HOG and mating pathway by error-prone PCR, we find that information transmission in HOG signaling is more robust to mutations than that in mating signaling. My work in studying the effects of mutations on one component of a signaling network may be relevant to understanding the mutational robustness of the network as previous work has shown that the dynamics of mutations in multiple nodes (such as those that might occur in complex human diseases) of a signaling network can be understood by studying mutations in a single node (Kwon *et al.,* 2016).

With this work, we hope to add to our understanding of the differences in the mutational robustness of different signaling pathways and the phenotypic significance of genetic variation across the whole genome and in promiscuous components of signaling pathways.

3.5 Future Work

To understand the extent to which the Sln1 branch of the HOG pathway contributes to the mutational robustness of the HOG response, we proposed to study the effects of mutations in Ste50 in a strain where this branch is intact.

To understand how intrinsic differences in our fluorescent reporters such as maturation rate and stability contribute the differences observed in the two pathways, we propose to repeat the measurements of the effects of random mutations in Ste50 on mating and HOG responses, with a strain with the pathway responsive fluorescent reporters expressed from the reciprocal promoters (pFUS1-BFP and pSTL1-GFP). In lieu of this, we propose to express GFP or BFP from the mating reporter (pFUS1) or osmolarity reporter (pSTL1) on a plasmid in a wild-type strain and quantify any differences in fluorescence levels when expressed from the same promoter. We will also use the correlation of the different fluorescent proteins expressed from the same promoter to measure the contributions of intracellular (intrinsic) and cell-to-cell (extrinsic) variability to the overall variation (Elowitz *et al.,* 2002).

It would also be interesting to track the impact of the accumulation of mutations across evolutionary time for the two strains by analysis of the intermediate lineages. We would be able to identify if the observed robustness of the HOG pathway at generation 1520 is due to compensatory mutations or if it is able to buffer the effects of the mutations throughout the evolutionary time. Similarly, it would also be interesting to analyze the generation at which the mating response is affected and to observe if by the latest, 2000th, generation the mating pathway signaling response is rescued by compensatory mutations or buffering.

3.6 Material and Methods

3.6.1 Library Construction

Random mutagenesis was carried out error-prone PCR using Ste50 (SGD: S000000537) as a template and the Agilent GeneMorph II Random Mutagenesis Kit. The random mutants were constructed using AarI restriction sites and expressed from centromeric plasmids with His

selection, under the control of the endogenous Ste50 promoter (pSte50, 434bp upstream of STE50 start site, introduced using PspOMI and XhoI restriction sites), and an Adh1 transcription terminator. The mutation rate was confirmed by sequencing 51 independent variants from the unselected library at The Center for Applied Genomics (TCAG) DNA Sequencing Facility in Toronto, Canada. The variants were isolated by purifying plasmids from yeast by treating saturated yeast cultures with zymolase (Zymo Research) overnight at 4oC followed by standard plasmid miniprep protocol (Qiagen). This was followed by transforming into *E. coli* for amplification and purifying the plasmid by standard miniprep protocol (Qiagen).

3.6.2 Yeast Transformation

Yeast strain (W303 MATa, bar1::NatR, far1Δ, Ste50::Trp1, Ssk2::Leu2, Ssk22::Ura3, mfa2::pFUS1-GFP, ura3::pSTL1-BFP-KanMX, his3) was transformed by specific Lithium-Acetate based high efficiency, large scale transformation method described by Gietz *et al.*, with minor modifications (Gietz and Schiestl 2007). Ssk2 and Ssk22 were deleted (by homologous recombination using the markers listed above) to knock out the second, partially redundant, branch of the HOG pathway.

3.6.3 Mutation Accumulation

CB009 was generously provided by Dr. Wendell Lim. We made CW001 by introducing pSTL1-GFP-KanMX at the HO locus in BY4741 (Strome *et al.,* 2018) using the standard Lithium Acetate method.

While the mating and HOG reporter strains have some differences in their strain background, we expect the whole genome mutation rate to be the same. The mating reporter strain is W303 MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 mfa2::pFUS1-GFP*,* bar1::NatR, far1Δ and the HOG reporter strain is S288C MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 HO::pSTL1-GFP-KanMX. The two parental strains of CB009 and CW001, W303 and S288C, respectively, have about 9500 single nucleotide variations resulting in non-synonymous changes in ~700 genes (about 10% of the yeast genome) (Matheson *et al.,* 2017). The length and structure

of the 16 chromosomes in the two strains is the same. This is important as there is strong linear correlation between chromosome size and the number of single nucleotide mutations per chromosome (Zhu *et al.,* 2017). Therefore, although the strains used in this experiment have some differences in the strain backgrounds, since they were evolved under the same conditions (single cell bottleneck, every 2 days for 74 passages), we don't expect significant differences in mutation rates between the two strains.

Each strain was streaked out on rich media (YPD) to obtain single colonies. Fifteen colonies for each strain were stored as the ancestral lines and propagated by plating on YPD plates for 2 days at 30^{o}C or room temperature. After two days of growth, single colony bottlenecks were implemented by choosing a single colony (closest colony to a previously marked line, to reduce bias in colony size or morphology) and streaking it on YPD plates. This process was continued for 100 days for CB009 and 82 days for CW001. Intermediate lineages were stored by freezing every other passage (even number passages). The colony chosen to propagate was also inoculated in YPD liquid media overnight and 500uL of the overnight culture was mixed with 500uL 50% glycerol and stored at -80oC. One line for each strain dropped out before the end of the MA study and two lines for each strain did not grow upon streaking from stocks.

To calculate the number of generations in our MA study, we assumed the number of generations the cells were allowed to grow prior to imposing the single-cell bottleneck to be 20 generations (for ~48 hours) (Zhu *et al.,* 2014), leading to our estimate of 1480 generations for 74 passages.

There are various estimates of the haploid single-nucleotide mutation rate in yeast:

- 1.67×10^{-10} per base per generation (Zhu *et al.,* 2014) resulting in 0.002 mutations/generation (1.67×10^{-10} x number of bases in yeast nuclear genome, Saccharomyces Genome Database (SGD), 12,071,326)
- 4.04×10^{-10} per base per generation (Sharp *et al.,* 2018) resulting in 0.005 mutations/generation
- 0.0041 per generation (Lynch *et al.,* 2008).

We expect 2.96-7.4 base substitutions (1480 generations x 0.002 or 0.005, using the minimum and the maximum estimate from above) to accumulate in each line. In addition, we expect 458 homopolymer mutations (0.3094 per generation), 2.81 microsatellite mutations (0.0019 per

generation), 0.296 small indels (0.0002 per generation), and between 0.7104-1.628 deleterious mutations (0.00048-0.0011 per generation) (Lynch *et al.,* 2008).

3.6.4 Flow Cytometry

3.6.4.1 Ste50 random mutant library

To ensure the diversity of the library was represented adequately, at least 100,000 transformants were harvested from fresh (less than two weeks old) plates by scraping in synthetic dropout medium without histidine and mixed by inverting. Similar number of transformants expressing wildtype Ste50 were also harvested by the same method. An aliquot of the transformant mixture was used to inoculate the selective medium to achieve the desired cell density (OD_{600}=0.002). This culture was grown overnight at 30^{o}C.

3.6.4.2 Ste50 random variants

Single colonies representing transformants expressing either wildtype Ste50 or a random Ste50 variant were inoculated synthetic dropout medium without histidine and grown overnight at 30°C.

3.6.4.3 MA lines

CB009 (passage 0 and passage 74) and CW001 (passage 0 and passage 74) were streaked on YPD plates from frozen glycerol stocks and single colonies were grown in YPD liquid media overnight at 30^{o}C.

For all experiments, overnight cultures were diluted to OD_{600} of 0.1 and grown until OD_{600} of 0.2 (about 4 hours). An aliquot of cells was spun down (4000 RPM, 1 minute) and resuspended in equal amount of fixing buffer, 0.5% Paraformaledhyde (Sigma) in 1X PBS at pH 7.4. Samples were then incubated with 1μM α-factor (Zymo Research) or 0.4M Potassium Chloride (KCl, Sigma) for 250 minutes or 360 minutes. Samples were taken every 30 minutes to 1 hour and also fixed using the above stated method. Samples were stored at -4oC until they were processed by flow cytometry (maximum 24 hours later).

A BD BioSciences LSRII Fortessa flow cytometer (at the Medical Sciences Flow Cytometry Facility in the University of Toronto) was used (with the high throughput sampler attached) to count over 100,000 cells in the library experiments or 10,000-50,000 cells in the other experiments. During acquisition, we employed a gate based on forward and side scatter to count the desired number of cells in the gate of interest.

Experiments on the Ste50 library were repeated 4 times and replicates are shown. Experiments on the MA lines were repeated twice, and the data from one experiment is shown. Results obtained from the second experiment were similar. Experiments on the 51 Ste50 variants were performed only once because of the complexity and size of this experiment.

3.6.5 Data Analysis

3.6.5.1 Sequence Analysis

Ste50 sequence was downloaded from the SGD, https://www.yeastgenome.org/locus/S000000537). Random variant sequence data (obtained from the TCAG was analysed using CLC Sequence Viewer v7.5 (Mac OSx).

The mutation rate of the Ste50 random mutant library generated by error-prone PCR was calculated by fitting the observed frequency distribution to the closest Poisson Distribution using http://vassarstats.net. This website computes the closest Poisson distribution by the method of least squares. The expected frequency under the Poisson probability distribution is calculated multiplying the probability distribution by the number of sequences (49).

The SAM and RA domains were delimited using the coordinates from the Pfam database (Finn *et al.,* 2016) the IDR using previously published work (Zarin *et al.,* 2017).

The effects of mutations on protein function were predicted using PROVEAN PROTEIN web server (Choi *et al.,* 2012).

3.6.5.2 Flow Cytometry Data

Flow cytometry data was analyzed using the flowCore library, ggplot2 and colorspace R packages with in-house scripts to calculate the log median GFP and BFP values. We did not

employ post-acquisition gating as tests of post-acquisition gating in FlowJo (v10.4.2, MacOSX) showed no differences.

3.6.5.3 Information Theory

3.6.5.3.1 Entropy

Information theory provides ways to quantify the uncertainty of a random variable (Rhee et al., 2012). One popular metric is Shannon's entropy ("entropy"), which is calculated by multiplying the probability of the variable taking on a certain value by the log of that probability (Equation 1). When the base 2 log is used, the measurement is in bits.

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log_2\, p(x_i)$$

Equation 1

Entropy is always non-negative, and it is largest when the variable can take on all the values at equal probabilities (a wide, uniform distribution) and smallest when the variable can take on only one value (a narrow distribution). Therefore, maximum entropy is $\log_2$ of number of values/states that a variable can take on, and minimum entropy is zero.

3.6.5.3.2 Kullback-Leibler Divergence

One way to compare distributions is by measuring the Kullback-Leibler divergence (KL divergence), also called relative entropy. KL divergence is a measure of how much one probability distribution varies from another and specifically quantifies the amount of information lost when one distribution is used to approximate the other. If the distributions are identical, then no information is lost, and KL divergence is zero. If the distributions are not identical, then some information is lost, and KL divergence is greater than zero and the larger the value, the more the distributions diverge from one another.

KL divergence is defined above on discrete distributions, and our flow cytometry data is continuous. We therefore used the method of linear extension (Perez-Cruz 2008) to compute the KL divergence without binning. KL divergence was calculated by modifying a script written by Ian Hsu for MATLAB to run on Octave v4.0.3. The functions ecdf.m and nansum.m were

downloaded from https://searchcode.com/codesearch/raw/9533414/ and https://www.apt-browse.org/browse/debian/wheezy/main/all/octave-statistics/1.1.31/file/usr/share/octave/packages/statistics-1.1.3/nansum.m, respectively.

3.6.5.3.3 Mutual Information

Mutual Information (MI) is calculated as the entropy of the output plus the entropy of the input subtracted by the joint entropy of the output and the input (Equation 2).

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$= H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X)$$

$$= \sum_{x,y} p(x,y) \; log_2 \frac{p(x,y)}{p(x)p(y)}$$

Equation 2

We can define certain properties of MI with this equation. First, MI is non-negative. This is because the entropy of the output given a known input cannot be greater than the entropy of the output without knowing the input (i.e., knowing the input cannot result in greater entropy in the output) and vice versa making H(X) H(X|Y) or H(Y) H(Y|X) and I(X;Y)≥0.

Next, we can determine the limits of MI. If knowing the input completely determines the output, then MI is maximum (no uncertainty about x once y is known, so H(X|Y)=0 and I(X;Y)=H(X). If knowing the input partially determines the output, then MI is greater than zero and less than maximum (H(X|Y)>0 so some uncertainty remains). If the input and output are independent (knowing the input does not determine the output at all) then MI is zero (H(X) =H(X|Y), I(X;Y)=0 or the joint probability distribution of the output and the input is equal to the probability distribution of the output multiplied by the probability distribution of the input; $\log_2 1 = 0$)

MI can be described as the KL divergence between the joint probability distributions and the product of marginal probability distributions (Zhao *et al.,* 2016).

In our flow cytometry experiments, the signal, say, X, corresponds to the presence or absence of either mating factor or salt in the media. Since this is a discrete (yes or no) we represent this as a 0 or 1 binary variable. Since the response, Y, corresponds to the pathway reporter output as measured by flow cytometry, this is actually a continuous measurement and in order to compute the formula above we must discretize the data. Discretization of continuous data inevitably leads to a loss of information, but we tried to minimize the loss of information due to binning the data by binning the data into 1000 bins for the MA and Ste50 variant experiments where we counted at least 10,000 cells and 10,000 bins for the Ste50 library experiments where we counted at least 100,000 cells. To address the choice of the number of bins, we computed the MI as a function of the number of bins. We found that as the number of bins increased, MI saturated as long as the number of bins was above 100. MI was calculated using an in-house script written in Python (see Appendix 1). To ensure bins are not empty, the bin width was set to the minimum and maximum values for each experiment.

The change in MI due to mutations was calculated by subtracting the MI with wildtype (ancestral lines in the MA experiments or wildtype Ste50 in the Ste50 library and variants experiments) from the MI with yeast strains carrying mutations (evolved lines in MA experiments, Library of Ste50 random variants in the Ste50 library experiments and individual, random variants for the Ste50 variant experiments).

The change in MI was clustered using Cluster 3.0 (for Mac OSx) using the following settings: Hierarchical clustering of genes with similarity metric of uncentered correlation using clustering method of average linkage (de Hoon *et al.,* 2004). The resulting cluster (.cdt file) was visualized using Java Treeview v1.1.6r4 (for Mac OSx) (Saldanha 2004).

3.6.5.3.4 Data Availability

The flow cytometry data is on Flow Repository (Spidlen *et al.,* 2012), Repository ID: FR-FCM-ZYRZ.

The Mutual Information script is included as Appendix 1. This and other scripts used to analyze the data are included in this thesis are available on GitHub at https://github.com/purnimakompella/MutualInformation.

4 Chapter 4: Discussion and Future Directions

4.1 Summary

Signaling networks play a vital role in a cell's ability to sense and respond to its environment. Biological processes exhibit pervasive robustness in the face of many types of perturbations including genetic changes (Wagner 2005). Point mutations are the ultimate source of genetic variation and those affecting signaling reveal how cells process information and are implicated in many human diseases (Kitano 2004). The goal of my thesis work was to quantitatively assess the effects of mutations on signaling function.

First, I identified variants of the yeast mating MAPK that rescued an overexpression driven decrease in mating signaling. I hypothesized that I would find variants that would increase kinase activity to rescue signaling. In addition to these, I found that premature stop codons at permissive residues in a protein kinase can rescue an over-expression driven decrease in yeast mating signaling by reducing the levels of the protein through inefficient readthrough of the premature stop codon. This work highlighted the importance of measuring the effects of mutations on signaling at the level of pathway function and not on protein function. Motivated by this surprising rescue of pathway function, I sought to quantify the robustness of signaling pathways to mutations.

To this end, I quantitatively assessed the effects of mutations on HOG and mating signaling by mutation accumulation and random mutagenesis. I hypothesized that the HOG pathway would have higher mutational robustness than the mating pathway based on the dynamics of the HOG pathway response and the pathway architecture. I used the information theory measure of mutual information to more comprehensively describe the effects of mutations on signaling response considering changes to the reporter distributions and the signal-to-noise ratio of the pathway. I showed that the HOG pathway is more robust to mutations. This may be due to the redundancy in the pathway, other mechanisms of buffering the effects of variation or a lower mutational target size. I showed that removing the second branch of the pathway and controlling for mutational target size can explain most of HOG pathway robustness but not all, suggesting that HOG signaling is more robust to mutations than mating signaling. I compared HOG and mating robustness more directly by studying the effects of the same mutations on the two signaling

responses. I found that the same mutations can have opposite effects on mating and HOG signaling crosstalk. The mutations with large effects on Ste50 protein have large effects on signaling in both pathways. However, I was not able to explain the opposite effects on mating and HOG signaling crosstalk. Thus, even in these well-characterized pathways, the complexity of the signaling response makes mapping genotype to phenotype non-trivial.

Overall, in this thesis work, I used various mutagenesis methods to test the effects of mutations on signaling in living cells. In the first part, I used an engineered random mutant library of a protein kinase to assess the functional significance of mutations on a yeast MAPK pathway. In the second part, along with genetic engineering, I used experimental evolution to quantify the mutational robustness of two yeast MAPK pathways. We found both non-intuitive and intuitive results and ways to quantify both. Applying the tools of systems biology and genetic engineering is a comprehensive and complementary approach to answering evolutionary questions.

4.2 Discussion

I began this thesis work to address the lack of high-throughput functional assays to understand the functional significance of genetic variation.

Understanding how variations in protein sequence affect function is crucial to our understanding of many aspects of biology including pathology and evolution (Baker *et al.,* 2012). A common experimental approach to studying the protein sequence-structure relationship is to mutate the protein, perform a functional assay and sequence the protein to analyze the mutation. In the past, the effectiveness of this method was limited to analyzing only a few mutations at once, due to the laborious and expensive Sanger sequencing process. With the advent of next generation sequencing (NGS) technologies, sequencing is no longer a bottleneck, thus in principle, one could evaluate thousands of mutations simultaneously. In fact, current sequencing studies are cataloguing genetic variation at a rate far higher than we are able to assign function to the uncovered genetic variation. While enumerating genetic variation is valuable, the real utility comes in understanding the functional significance of these mutations – if and how they contribute to physiological differences between individuals or the onset and progression of

disease (Greenman *et al.,* 2007). Hence, there is a need for functional assays that can match the throughput of NGS by interrogating the function of millions of mutations at once.

High-throughput research efforts aimed at mapping protein sequence-function relationships, also known as deep mutational scanning, had also relied on bioinformatics, specifically using either pattern recognition or 3D structure prediction algorithms (Torkamani and Schork 2008). The most popular programs for predicting the sequence-function relationship were Sorting Intolerant From Tolerant (SIFT) (Kumar *et al.,* 2009) and Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei *et al.,* 2010). Both algorithms predict the impact of protein sequence variants (Baker *et al.,* 2012). Polyphen-2 is a Bayesian classifier that uses sequence and structure-based features to predict whether a mutation will be neutral or deleterious. SIFT is a sequence homology-based tool that uses evolutionary conservation of amino acids to predict whether a mutation will be tolerated. This approach is based on the idea that substitutions at conserved residues are expected to impact the function of a protein more than those at non-conserved residues. There are several other such tools, including Protein Variation Effect Analyzer (PROVEAN), a sequence alignment-based approach that can also predict the effects of multiple amino acid substitutions and indels (Choi and Chan 2015). Although using these tools in combination can improve prediction sensitivity and specificity, there are still drawbacks to these bioinformatics analyses. For example, due to the predictive nature of these algorithms, complementary functional analyses are required to confirm the effects of the mutations (Torkamani and Schork 2008). In addition, to use the structure-based features in PolyPhen-2 requires a solved protein structure, and SIFT needs orthologous proteins. Furthermore, these methods do not predict the effect of synonymous substitutions, which are known to affect gene expression levels (Zhou *et al.,* 2016). Finally, since these predictions are based on our understanding of the relationship between genotype and phenotype, novel relationships such as those that can be found through functional analyses will be missed.

High-throughput experimental methods for deep mutational scanning were limited to assessing protein function by measuring binding affinity or survival in the presence of a challenge. Fowler *et al*. developed a large-scale functional assay that uses phage display to investigate the effects of genetic variation on protein function (Fowler *et al.,* 2010). A protein of interest is mutated, and variants are then displayed on the phage surface and presented to a ligand. Function of the variants is assessed based on binding affinity. After repeated washes to exclude low affinity

variants, the bound fraction is recovered and deep-sequenced. This method, while restricted by the intrinsic limitations of phage display, which include proteins incompatible with surface display and susceptibility to proteolysis during phage propagation, is promising in terms of throughput. However, because it is an affinity-based assay, it is only useful to characterize binding functions, rather than catalysis, and even then, it is of limited use with proteins with weak binding affinities or mutations that cause functional changes without altering binding affinity. Finally, the effect of a mutation on protein function while interesting, only tells part of the story. It is also important to understand how these changes affect the function of the network of interacting proteins.

Another high-throughput experimental model uses an evolutionary approach to assessing protein function. Protein variants are screened in the presence of an environmental (such as temperature) or chemical (such as an antibiotic) challenge with the expectation that mutations will have a beneficial, neutral or detrimental effect on protein function and thereby, survival. For example, Hietpas *et al.*, combined saturation mutagenesis with growth competitions to assess the effect of point mutations on survival (Hietpas *et al.,* 2011). They engineered a yeast strain with a temperature sensitive heat shock protein 90 (Hsp90), an essential chaperone protein in eukaryotes, to analyze all possible substitutions in a nine-amino acid region of Hsp90. Survival was measured, as a function of growth, at a non-permissive temperature for both the mutants and wild-type. With this model, they were able to functionally assess over 500 codon variants. Due to their interest in generating as close to all amino acid substitutions as possible, they chose saturation mutagenesis over random mutagenesis, which required them to restrict their study to only a nine-amino acid region.

Kato *et al.*, used site-directed mutagenesis to generate over 2000 single point mutation substitutions in the human tumor suppressor protein, p53 (Kato *et al.,* 2003). p53 is a transcriptional transactivator that binds to short genomic sequences to induce specific target genes. Kato and coworkers analyzed the function of mutants by measuring their ability to transactivate a fluorescent reporter in yeast. They found that most of the variants with mutations in the DNA binding domain were inactive. They also tested three novel mutations isolated from their yeast screening in human cells, confirming their findings. Their results further validate yeast as a suitable model of studying human genes. This assay, while high-throughput, is specific

to p53 function. Therefore, it cannot be used to detect mutations in other genes. Furthermore, their assay also measures only individual protein function and not overall pathway activity.

There were experimental methods to assess the impact of mutations on signaling pathway functions in human cells, however, they were not high-throughput. For example, Fröhling *et al*. used tissue culture-based methods to find individual driver mutations in receptor tyrosine kinase FLT3 (Frohling *et al.,* 2007). While this method is promising, due to the limitations of their cell-culture model system, in particular its low throughput, and the relative prevalence of passengers to drivers, they identified only nine alleles that affected function. Thus, while large-scale sequencing projects are cataloguing vast numbers of single nucleotide polymorphisms (SNPs), currently, there is no matching high-throughput functional assay to assign phenotypic significance to these SNPs.

My former PhD advisor developed a high-throughput method for studying the evolution of signaling networks, using the yeast mating pathway as a model system (Peisajovich *et al.,* 2010). Using a fluorescent reporter expressed under the control of a pathway-responsive promoter, they analyzed the role of protein domain recombination in the evolution of signaling pathways. They showed that some chimeric proteins that included fusions of domains from conserved MAPK proteins lead to constitutive pathway activation, as reflected in an increase in the basal levels of reporter expression level. Mutations that activate genes without extracellular stimulation are excellent candidates for cancer driver mutations. This system constitutes an excellent starting point for high-throughput protein sequence-function analysis.

I proposed an experimental method, based on the highly conserved, MAPK-mediated signaling pathway, to study the effect of mutations at the systems-level, for distinguishing between activating, neutral and inactivating mutations, to identify mutations that may promote aberrant pathway activation as well as for exploring the robustness of MAPK pathways by which resistance to therapy may be gained. Since the readout of the model system that I planned to use was activation of a pathway rather than of a single gene, it was potentially applicable to multiple yeast genes homologous to human genes. This is important, as there are disease-related proteins such as G protein-coupled receptors, G proteins, and small GTPases, among others, for which the functional analysis of mutations was limited and exploration through bioinformatics methods

was difficult. Importantly, this method would have the capability of interrogating millions of mutations at once, and thus it would match the throughput of NGS.

In the first part of my thesis work, I expected to find mutations that would rescue pathway activity by increasing kinase activity. While I did find some variants with these mutations, I also found mutations that rescued pathway function in non-intuitive ways. One of the limitations of the proposed system was the inability to distinguish between pathway overactivating mutants and those with wildtype levels of pathway activity when the mating MAPK was expressed from a weak promoter. This was not due to meeting the detection limit of the flow cytometer, concentrations of the pheromone (saturating concentrations were used), measuring the response too soon after induction (measurements were made at previously confirmed times of two or four hours after induction). Motivated by a desire to identify pathway overactivating mutations, I expressed the protein from a strong promoter which allowed me to distinguish variants that overactivated the pathway from those with lower levels of pathway activity. Overexpression of proteins and protein kinases is not an unusual phenotype for diseases (Cohen 2001). Therefore, I continued my efforts to identify ways in which the diminished pathway activity could be rescued by coding region mutations. By overexpressing the protein and limiting the mutational target size to the coding region, I found variants of the protein kinase such as those with premature stop codons, that could rescue pathway function. However, in the absence of overexpression, the premature stop codons would be expected to be deleterious and pathway function would be rescued by coding region mutations that reverted to the wildtype residue. Similarly, if the mutational target size were not limited, one trivial way to decrease or compensate for the overexpression would be through mutations in cis- or trans- regulatory regions of the protein. Therefore, our experimental setup forced us to find uncommon solutions to the problem of overexpression driven decrease in pathway activity.

In the next part of my thesis, my interest in studying the evolution of signaling pathways led me to design the experiment to allow understanding of the natural workings of the system. To quantify the extent to which MAPK pathways tolerate point mutations, I ensured that I did not introduce other types of variation, such as expression level changes, that severely affected pathway function. In my first attempt at introducing mutations into a shared component of the mating and HOG pathways, I turned to the MAPKKK Ste11. I tested endogenous promoters of different lengths, chose and optimized one by introducing a point mutation, and made random

mutant libraries with varying mutation rates. However, numerous strategies for cloning the random variants into the plasmid vector were unsuccessful and in the interest of time, I moved on to another shared component, Ste50. These trials illustrate that science often involves detours and requires creative and timely solutions to overcome experimental roadblocks.

With the individual engineered random variants of Ste50, I expected to find variants with small effects on signaling. However, similar to the first part of my thesis work, I found mostly variants that either had no effects or large effects. This suggests that mutations with small effects on signaling are not discernible in these laboratory experiments. In the second part of my thesis work, I used the change in mutual information as a test for the effects of mutations. Similar to my approach, previously the change in information content of individual genetic sequences due to mutations was used in bioinformatic analysis to quantify the effects of mutations (Zia and Moses 2011). Although information theory has been applied to biological processes for over three decades, using mutual information to understand and quantify cellular signaling is a current direction in signaling research (Bluthgen and Legewie 2013; Bowsher and Swain 2014; Granados *et al.,* 2018; Schneider *et al.,* 1986; Tabbaa and Jayaprakash 2014; Voliotis *et al.,* 2014; Waltermann and Klipp 2011). Mutual information captures the effects of biological variation in cellular responses and estimating the mutual information between time series and responses eliminates the need to make assumptions about which features of the responses - duration, magnitude, or some other feature - are biologically relevant (Granados *et al.,* 2018).

Some technical considerations include the importance of not limiting the number of successive genetic manipulations for yeast strains by leaving selective markers during gene deletions and appreciating the sources of variation that affect reproducibility of results. To quantify the effects of mutations on multiple signaling responses concurrently, I needed to use a strain with two different fluorescent proteins. To account for the intrinsic differences in fluorescent proteins, I sought to create a strain where the fluorescent proteins are expressed from the opposite promoters. I was not able to obtain the strain with the desired genotype despite numerous attempts. The difficulty in creating this strain highlights the importance of thoughtful strain design. For example, knocking out genes without leaving behind selection markers such as via Delitto Perfetto (Storici *et al.,* 2001) or newer, CRISPR-based methods (Generoso *et al.,* 2016) ensures future changes to the strain are not limited by exhausted selection markers. The strains

used the mutation accumulation experiment have slight differences that could have been avoided with some consideration to strain background.

In order to get reproducible results, there exist many sources of variation that need to be minimized. For example, growth conditions such as growing cells in 96-well deep well blocks compared to culture tubes or flasks, inoculating transformants directly into liquid culture prior to flow cytometry analysis compared to plating on media and scraping the cells, densities of cells used to inoculate overnight cultures and cultures prior to induction with various stimuli can each affect reproducibility. Especially when measuring the stress response, sources of stress such as temperature, need to be kept in mind. Differences in flow cytometers can also affect reproducibility and can be mitigated by including appropriate controls in the experiment.

4.3 Future Directions

4.3.1 Tracking the effects of mutations on robustness through evolutionary time

We can test the robustness of the mating and HOG responses with the mutation accumulation lines from various generations, thereby testing different mutation rates. We found that the HOG pathway response at generation 1480 is about the same as that of the beginning of the experiment. We can test the intermediate lineages to test if the mutational robustness of the pathway is constant or if it fluctuates. Similarly, we found that the variance in the mating pathway signaling increases due to mutations accumulated over 1480 generations. By quantifying the amount of variance in this trait over many generations, we could estimate the mutational variance, the amount of phenotypic variance added by mutations each generation (Duveau *et al.,* 2017; Landry *et al.,* 2007).

4.3.2 Quantification of the contributions of feedback loops to robustness

One of the advantages to working with a well-characterized system such as the yeast mating and HOG signaling network is the wealth of information available about the regulation of the signaling responses. For example, we know that single point mutations, such as Sho1 S166E and

Ste5 T287V, can disrupt feedback mechanisms responsible for the signaling fidelity of each pathway (Bhattacharyya *et al.,* 2006; Hao *et al.,* 2007). One exciting direction for the quantification of mutational robustness would be quantifying the contributions of various feedback mechanisms to signaling fidelity. Comparing the mutational robustness of a pathway using a strain with and another without a specific feedback loop will help allow the quantification of the contribution of that feedback loop. By comparing the effects of removing a feedback mechanism in one pathway on the mutational robustness of another pathway, we could also test if that mechanism plays a more general role in buffering the effects of mutations or if is pathway specific.

4.3.3 Effects of selection on mating and HOG signaling

Studies in *Escherichia coli* (Cooper and Lenski 2000) and yeast (Kvitek and Sherlock 2013) show that evolution favours gaining a fitness advantage by losing unused functions; for example, yeast cells lose the ability to sense and respond to extracellular cues through signaling pathways when under glucose-limitation in a constant environment. Therefore, if we want to study the effects of selection on signaling pathways, we should ensure that the signaling pathways are not lost, by either making the cells use them (by exposing them to the inducer) or selecting for the presence of pathway response. We can do the latter by FACS. For example, we can grow a wild-type strain with pathway responsive fluorescent reporter(s) and select, by gating populations of cells with specific fluorescence levels, cells that display the phenotypes of interest. We could impose various types of selection including stabilizing selection, directional selection or diversifying selection, by gating and sorting populations with a narrower reporter expression distribution, those that like at either tail of the distribution, or both tails of the distribution, respectively. This way, we can isolate various types of mutations such as deleterious (lower or slower than wild-type pathway response), neutral (wild-type response) and beneficial (higher or faster than wild-type pathway response).

4.3.4 Quantifying information capacity of signaling pathways

To more accurately quantify the information capacity of the mating and HOG pathways, it would be worthwhile to do a dose-response study assessing the output of the pathway at various levels of input.

Furthermore, previous work has shown that the amount of variance in a trait due to mutations can change depending on the environment (Duveau *et al.,* 2017). To tease out the effects of environmental on mutational robustness, the change in mutual information due to mutations could be quantified for the HOG response under various stresses such as different salts, other osmolytes, and temperature, or after exposure to mating pheromone. Complementary analysis of the effect of prior exposure to osmotic stress on the mutational robustness of the mating pathway would also be interesting. To understand how exposure to multiple signals affects the information capacity of the two pathways, cells could be exposed to multiple stimuli simultaneously (multiple osmolytes or an osmolyte and mating pheromone), as they are likely to encounter in natural environments, and the information capacity of the pathway(s) could be quantified.

While this thesis work has focused on the effects of point mutations on signaling pathways, another common type of change is expression level variation. We could study the effects of expressing various mating and HOG pathway components at different levels (with a library of promoter mutations or by expressing wildtype proteins from promoters of different strengths) on the information capacity of the two pathways.

References

Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S. and Sunyaev S.R. A method and server for predicting damaging missense mutations. *Nat Methods* **7,** 248-9 (2010).

Albert R. and Oltvai Z.N. Shaping specificity in signaling networks. *Nat Genet* **39,** 286-7 (2007).

Artieri C.G. and Fraser H.B. Evolution at two levels of gene expression in yeast. *Genome Res* **24,** 411-21 (2014).

Aymoz D., Sole C., Pierre J.J., Schmitt M., de Nadal E., Posas F. and Pelet S. Timing of gene expression in a cell-fate decision system. *Mol Syst Biol* **14,** e8024 (2018).

Aymoz D., Wosika V., Durandau E. and Pelet S. Real-time quantification of protein expression at the single-cell level via dynamic protein synthesis translocation reporters. *Nat Commun* **7,** 11304 (2016).

Azeloglu E.U. and Iyengar R. Signaling networks: information flow, computation, and decision making. *Cold Spring Harb Perspect Biol* **7,** a005934 (2015).

Bai C., Tesker M. and Engelberg D. The yeast Hot1 transcription factor is critical for activating a single target gene, STL1. *Mol Biol Cell* **26,** 2357-74 (2015).

Baker C.R., Booth L.N., Sorrells T.R. and Johnson A.D. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151,** 80-95 (2012).

Baltanas R., Bush A., Couto A., Durrieu L., Hohmann S. and Colman-Lerner A. Pheromone-induced morphogenesis improves osmoadaptation capacity by activating the HOG MAPK pathway. *Sci Signal* **6,** ra26 (2013).

Bardwell L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26,** 339-50 (2005).

Bardwell L. Mechanisms of MAPK signalling specificity. *Biochem Soc Trans* **34,** 837-41 (2006).

Bashor C.J., Helman N.C., Yan S. and Lim W.A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319,** 1539-43 (2008).

Basson M.A. Signaling in cell differentiation and morphogenesis. *Cold Spring Harb Perspect Biol* **4,** (2012).

Baudin-Baillieu A., Legendre R., Kuchly C., Hatin I., Demais S., Mestdagh C., Gautheret D. and Namy O. Genome-wide translational changes induced by the prion [PSI+]. *Cell Rep* **8,** 439-48 (2014).

Bauer C.R., Li S. and Siegal M.L. Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. *Mol Syst Biol* **11,** 773 (2015).

Bekaert M. and Conant G.C. Transcriptional robustness and protein interactions are associated in yeast. *BMC Syst Biol* **5,** 62 (2011).

Berggard T., Linse S. and James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7,** 2833-42 (2007).

Bergstrom A., Simpson J.T., Salinas F., Barre B., Parts L., Zia A., Nguyen Ba A.N., Moses A.M., Louis E.J., Mustonen V., Warringer J., Durbin R. and Liti G. A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* **31,** 872-88 (2014).

Bhattacharyya R.P., Remenyi A., Good M.C., Bashor C.J., Falick A.M. and Lim W.A. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* **311,** 822-6 (2006).

Blanchet S., Cornu D., Argentini M. and Namy O. New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae. *Nucleic Acids Res* **42,** 10061-72 (2014).

Bloom J.D., Labthavikul S.T., Otey C.R. and Arnold F.H. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103,** 5869-74 (2006).

Bluthgen N. and Legewie S. Robustness of signal transduction pathways. *Cell Mol Life Sci* **70,** 2259-69 (2013).

Bowsher C.G. and Swain P.S. Environmental sensing, information transfer, and cellular decision-making. *Curr Opin Biotechnol* **28,** 149-55 (2014).

Brill J.A., Elion E.A. and Fink G.R. A role for autophosphorylation revealed by activated alleles of FUS3, the yeast MAP kinase homolog. *Mol Biol Cell* **5,** 297-312 (1994).

Brown C.M., Dinesh-Kumar S.P. and Miller W.A. Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. *J Virol* **70,** 5884-92 (1996).

Cahill M.A., Ernst W.H., Janknecht R. and Nordheim A. Regulatory squelching. *FEBS Lett* **344,** 105-8 (1994).

Carlson J.M. and Doyle J. Complexity and robustness. *Proc Natl Acad Sci U S A* **99 Suppl 1,** 2538-45 (2002).

Chen B.S. and Wu C.C. On the calculation of signal transduction ability of signaling transduction pathways in intracellular communication: systematic approach. *Bioinformatics* **28,** 1604-11 (2012).

Chen R.E. and Thorner J. Function and regulation in MAPK signaling pathways: lessons learned from the yeast Saccharomyces cerevisiae. *Biochim Biophys Acta* **1773,** 1311-40 (2007).

Choi Y. and Chan A.P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31,** 2745-7 (2015).

Choi Y., Sims G.E., Murphy S., Miller J.R. and Chan A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7,** e46688 (2012).

Cimino P.A., Nicholson B.L., Wu B., Xu W. and White K.A. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. *PLoS Pathog* **7,** e1002423 (2011).

Cohen P. The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur J Biochem* **268,** 5001-10 (2001).

Cooper V.S. and Lenski R.E. The population genetics of ecological specialization in evolving Escherichia coli populations. *Nature* **407,** 736-9 (2000).

Costanzo M., Baryshnikova A., Bellay J., Kim Y., Spear E.D., Sevier C.S., Ding H., Koh J.L., Toufighi K., Mostafavi S., Prinz J., St Onge R.P., VanderSluis B., Makhnevych T., Vizeacoumar F.J., Alizadeh S., Bahr S., Brost R.L., Chen Y., Cokol M., Deshpande R., Li Z., Lin Z.Y., Liang W., Marback M., Paw J., San Luis B.J., Shuteriqi E., Tong A.H., van Dyk N., Wallace I.M., Whitney J.A., Weirauch M.T., Zhong G., Zhu H., Houry W.A., Brudno M., Ragibizadeh S., Papp B., Pal C., Roth F.P., Giaever G., Nislow C., Troyanskaya O.G., Bussey H., Bader G.D., Gingras A.C., Morris Q.D., Kim P.M., Kaiser C.A., Myers C.L., Andrews B.J. and Boone C. The genetic landscape of a cell. *Science* **327,** 425-31 (2010).

Creixell P., Schoof E.M., Simpson C.D., Longden J., Miller C.J., Lou H.J., Perryman L., Cox T.R., Zivanovic N., Palmeri A., Wesolowska-Andersen A., Helmer-Citterich M., Ferkinghoff-Borg J., Itamochi H., Bodenmiller B., Erler J.T., Turk B.E. and Linding R. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* **163,** 202-17 (2015).

de Hoon M.J., Imoto S., Nolan J. and Miyano S. Open source clustering software. *Bioinformatics* **20,** 1453-4 (2004).

Desai M.M. and Fisher D.S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176,** 1759-98 (2007).

Dixon S.J., Costanzo M., Baryshnikova A., Andrews B. and Boone C. Systematic mapping of genetic interaction networks. *Annu Rev Genet* **43,** 601-25 (2009).

Draghi J.A., Parsons T.L., Wagner G.P. and Plotkin J.B. Mutational robustness can facilitate adaptation. *Nature* **463,** 353-5 (2010).

Ducrest A.L., Amacker M., Lingner J. and Nabholz M. Detection of promoter activity by flow cytometric analysis of GFP reporter expression. *Nucleic Acids Res* **30,** e65 (2002).

Duina A.A., Miller M.E. and Keeney J.B. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. *Genetics* **197,** 33-48 (2014).

Dujon B. Yeast evolutionary genomics. *Nat Rev Genet* **11,** 512-24 (2010).

Dunn J.G., Foo C.K., Belletier N.G., Gavis E.R. and Weissman J.S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *Elife* **2,** e01179 (2013).

Duveau F., Yuan D.C., Metzger B.P.H., Hodgins-Davis A. and Wittkopp P.J. Effects of mutation and selection on plasticity of a promoter activity in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114,** E11218-E11227 (2017).

Elowitz M.B., Levine A.J., Siggia E.D. and Swain P.S. Stochastic gene expression in a single cell. *Science* **297,** 1183-6 (2002).

Endicott J.A., Noble M.E. and Johnson L.N. The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem* **81,** 587-613 (2012).

Errede B., Vered L., Ford E., Pena M.I. and Elston T.C. Pheromone-induced morphogenesis and gradient tracking are dependent on the MAPK Fus3 binding to Galpha. *Mol Biol Cell* **26,** 3343-58 (2015).

Eyre-Walker A. and Keightley P.D. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8,** 610-8 (2007).

Felix M.A. and Barkoulas M. Pervasive robustness in biological systems. *Nat Rev Genet* **16,** 483-96 (2015).

Filippi S., Barnes C.P., Kirk P.D., Kudo T., Kunida K., McMahon S.S., Tsuchiya T., Wada T., Kuroda S. and Stumpf M.P. Robustness of MEK-ERK Dynamics and Origins of Cell-to-Cell Variability in MAPK Signaling. *Cell Rep* **15,** 2524-35 (2016).

Finn R.D., Coggill P., Eberhardt R.Y., Eddy S.R., Mistry J., Mitchell A.L., Potter S.C., Punta M., Qureshi M., Sangrador-Vegas A., Salazar G.A., Tate J. and Bateman A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44,** D279-85 (2016).

Firth A.E., Wills N.M., Gesteland R.F. and Atkins J.F. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Res* **39,** 6679-91 (2011).

Fowler D.M., Araya C.L., Fleishman S.J., Kellogg E.H., Stephany J.J., Baker D. and Fields S. High-resolution mapping of protein sequence-function relationships. *Nat Methods* **7,** 741-6 (2010).

Frohling S., Scholl C., Levine R.L., Loriaux M., Boggon T.J., Bernard O.A., Berger R., Dohner H., Dohner K., Ebert B.L., Teckie S., Golub T.R., Jiang J., Schittenhelm M.M., Lee B.H., Griffin J.D., Stone R.M., Heinrich M.C., Deininger M.W., Druker B.J. and Gilliland D.G. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* **12,** 501-13 (2007).

Furukawa K. and Hohmann S. Synthetic biology: lessons from engineering yeast MAPK signalling pathways. *Mol Microbiol* **88,** 5-19 (2013).

Gasch A.P., Spellman P.T., Kao C.M., Carmel-Harel O., Eisen M.B., Storz G., Botstein D. and Brown P.O. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11,** 4241-57 (2000).

Generoso W.C., Gottardi M., Oreb M. and Boles E. Simplified CRISPR-Cas genome editing for Saccharomyces cerevisiae. *J Microbiol Methods* **127,** 203-205 (2016).

Gibson B., Geertman J.A., Hittinger C.T., Krogerus K., Libkind D., Louis E.J., Magalhaes F. and Sampaio J.P. New yeasts-new brews: modern approaches to brewing yeast design and development. *FEMS Yeast Res* **17,** (2017).

Gietz R.D. and Schiestl R.H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2,** 38-41 (2007).

Gill G. and Ptashne M. Negative effect of the transcriptional activator GAL4. *Nature* **334,** 721-4 (1988).

Givan A.L. Flow Cytometry: First Principles. *Wiley*. (1992).

Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., Louis E.J., Mewes H.W., Murakami Y., Philippsen P., Tettelin H. and Oliver S.G. Life with 6000 genes. *Science* **274,** 546, 563-7 (1996).

Good M., Tang G., Singleton J., Remenyi A. and Lim W.A. The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. *Cell* **136,** 1085-97 (2009).

Granados A.A., Crane M.M., Montano-Gutierrez L.F., Tanaka R.J., Voliotis M. and Swain P.S. Distributing tasks via multiple input pathways increases cellular survival in stress. *Elife* **6,** (2017).

Granados A.A., Pietsch J.M.J., Cepeda-Humerez S.A., Farquhar I.L., Tkacik G. and Swain P.S. Distributed and dynamic intracellular organization of extracellular information. *Proc Natl Acad Sci U S A* **115,** 6088-6093 (2018).

Greenman C., Stephens P., Smith R., Dalgliesh G.L., Hunter C., Bignell G., Davies H., Teague J., Butler A., Stevens C., Edkins S., O'Meara S., Vastrik I., Schmidt E.E., Avis T., Barthorpe S., Bhamra G., Buck G., Choudhury B., Clements J., Cole J., Dicks E., Forbes S., Gray K., Halliday K., Harrison R., Hills K., Hinton J., Jenkinson A., Jones D., Menzies A., Mironenko T., Perry J., Raine K., Richardson D., Shepherd R., Small A., Tofts C., Varian J., Webb T., West S., Widaa S., Yates A., Cahill D.P., Louis D.N., Goldstraw P., Nicholson A.G., Brasseur F., Looijenga L., Weber B.L., Chiew Y.E., DeFazio A., Greaves M.F., Green A.R., Campbell P., Birney E., Easton D.F., Chenevix-Trench G., Tan M.H., Khoo S.K., Teh B.T., Yuen S.T., Leung S.Y., Wooster R., Futreal P.A. and Stratton M.R. Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153-8 (2007).

Gresham D., Desai M.M., Tucker C.M., Jenq H.T., Pai D.A., Ward A., DeSevo C.G., Botstein D. and Dunham M.J. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* **4,** e1000303 (2008).

Handly L.N., Yao J. and Wollman R. Signal Transduction at the Single-Cell Level: Approaches to Study the Dynamic Nature of Signaling Networks. *J Mol Biol* **428,** 3669-82 (2016).

Hao N., Behar M., Parnell S.C., Torres M.P., Borchers C.H., Elston T.C. and Dohlman H.G. A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. *Curr Biol* **17,** 659-67 (2007).

Hao N., Zeng Y., Elston T.C. and Dohlman H.G. Control of MAPK specificity by feedback phosphorylation of shared adaptor protein Ste50. *J Biol Chem* **283,** 33798-802 (2008).

Hartwell L.H., Hopfield J.J., Leibler S. and Murray A.W. From molecular to modular cell biology. *Nature* **402,** C47-52 (1999).

Hietpas R.T., Jensen J.D. and Bolon D.N. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A* **108,** 7896-901 (2011).

Ho W.C. and Zhang J. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. *Mol Biol Evol* **31,** 1568-80 (2014).

Hoffman C.S., Wood V. and Fantes P.A. An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. *Genetics* **201,** 403-23 (2015).

Houle D. How should we explain variation in the genetic variance of traits? *Genetica* **102-103,** 241-53 (1998).

Ivanova N.N., Schwientek P., Tripp H.J., Rinke C., Pati A., Huntemann M., Visel A., Woyke T., Kyrpides N.C. and Rubin E.M. Stop codon reassignments in the wild. *Science* **344,** 909-13 (2014).

Jansen G., Buhring F., Hollenberg C.P. and Ramezani Rad M. Mutations in the SAM domain of STE50 differentially influence the MAPK-mediated pathways for mating, filamentous growth and osmotolerance in Saccharomyces cerevisiae. *Mol Genet Genomics* **265,** 102-17 (2001).

Jungreis I., Lin M.F., Spokony R., Chan C.S., Negre N., Victorsen A., White K.P. and Kellis M. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res* **21,** 2096-113 (2011).

Kato S., Han S.Y., Liu W., Otsuka K., Shibata H., Kanamaru R. and Ishioka C. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A* **100,** 8424-9 (2003).

Khurana V. and Lindquist S. Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nat Rev Neurosci* **11,** 436-49 (2010).

Kim E.K. and Choi E.J. Pathological roles of MAPK signaling pathways in human diseases. *Biochim Biophys Acta* **1802,** 396-405 (2010).

Kim J., Vandamme D., Kim J.R., Munoz A.G., Kolch W. and Cho K.H. Robustness and evolvability of the human signaling network. *PLoS Comput Biol* **10,** e1003763 (2014).

Kirschner M. and Gerhart J. Evolvability. *Proc Natl Acad Sci U S A* **95,** 8420-7 (1998).

Kitano H. Biological robustness. *Nat Rev Genet* **5,** 826-37 (2004).

Kompella P.S., Moses A.M. and Peisajovich S.G. Introduction of Premature Stop Codons as an Evolutionary Strategy To Rescue Signaling Network Function. *ACS Synth Biol* **6,** 446-454 (2017).

Kumar P., Henikoff S. and Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4,** 1073-81 (2009).

Kvitek D.J. and Sherlock G. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet* **9,** e1003972 (2013).

Kwon Y.K., Kim J. and Cho K.H. Dynamical Robustness against Multiple Mutations in Signaling Networks. *IEEE/ACM Trans Comput Biol Bioinform* **13,** 996-1002 (2016).

Lage K. Protein-protein interactions and genetic diseases: The interactome. *Biochim Biophys Acta* **1842,** 1971-1980 (2014).

Landry C.R., Lemos B., Rifkin S.A., Dickinson W.J. and Hartl D.L. Genetic properties influencing the evolvability of gene expression. *Science* **317,** 118-21 (2007).

Laub M.T. Keeping Signals Straight: How Cells Process Information and Make Decisions. *PLoS Biol* **14,** e1002519 (2016).

Leducq J.B., Henault M., Charron G., Nielly-Thibault L., Terrat Y., Fiumera H.L., Shapiro B.J. and Landry C.R. Mitochondrial Recombination and Introgression during Speciation by Hybridization. *Mol Biol Evol* **34,** 1947-1959 (2017).

Lenski R.E., Barrick J.E. and Ofria C. Balancing robustness and evolvability. *PLoS Biol* **4,** e428 (2006).

Levchenko A., Bruck J. and Sternberg P.W. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci U S A* **97,** 5818-23 (2000).

Levy S.F., Blundell J.R., Venkataram S., Petrov D.A., Fisher D.S. and Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519,** 181-6 (2015).

Li D. A special issue on cell signaling, disease, and stem cells. *Cell Res* **22,** 1-2 (2012).

Loughran G., Chou M.Y., Ivanov I.P., Jungreis I., Kellis M., Kiran A.M., Baranov P.V. and Atkins J.F. Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res* **42,** 8928-38 (2014).

Lynch M., Sung W., Morris K., Coffey N., Landry C.R., Dopman E.B., Dickinson W.J., Okamoto K., Kulkarni S., Hartl D.L. and Thomas W.K. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A* **105,** 9272-7 (2008).

MacArthur D.G., Balasubramanian S., Frankish A., Huang N., Morris J., Walter K., Jostins L., Habegger L., Pickrell J.K., Montgomery S.B., Albers C.A., Zhang Z.D., Conrad D.F., Lunter G., Zheng H., Ayub Q., DePristo M.A., Banks E., Hu M., Handsaker R.E., Rosenfeld J.A., Fromer M., Jin M., Mu X.J., Khurana E., Ye K., Kay M., Saunders G.I., Suner M.M., Hunt T., Barnes I.H., Amid C., Carvalho-Silva D.R., Bignell A.H., Snow C., Yngvadottir B., Bumpstead S., Cooper D.N., Xue Y., Romero I.G., Genomes Project C., Wang J., Li Y., Gibbs R.A., McCarroll S.A., Dermitzakis E.T., Pritchard J.K., Barrett J.C., Harrow J., Hurles M.E., Gerstein M.B. and Tyler-Smith C. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335,** 823-8 (2012).

Marles J.A., Dahesh S., Haynes J., Andrews B.J. and Davidson A.R. Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated signal transduction pathway in yeast. *Mol Cell* **14,** 813-23 (2004).

Martin A. and Courtier-Orgogozo V. Morphological Evolution Repeatedly Caused by Mutations in Signaling Ligand Genes:59-87. *Springer Singapore*, Singapore (2017).

Masel J. and Siegal M.L. Robustness: mechanisms and consequences. *Trends Genet* **25,** 395-403 (2009).

Masel J. and Trotter M.V. Robustness and evolvability. *Trends Genet* **26,** 406-14 (2010).

Matheson K., Parsons L. and Gammie A. Whole-Genome Sequence and Variant Analysis of W303, a Widely-Used Strain of Saccharomyces cerevisiae. *G3 (Bethesda)* **7,** 2219-2226 (2017).

Mc Mahon S.S., Lenive O., Filippi S. and Stumpf M.P. Information processing by simple molecular motifs and susceptibility to noise. *J R Soc Interface* **12,** 0597 (2015).

McCaffrey G., Clay F.J., Kelsay K. and Sprague G.F., Jr. Identification and regulation of a gene required for cell fusion during mating of the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **7,** 2680-90 (1987).

McClean M.N., Mody A., Broach J.R. and Ramanathan S. Cross-talk and decision making in MAP kinase pathways. *Nat Genet* **39,** 409-14 (2007).

McGary K.L., Park T.J., Woods J.O., Cha H.J., Wallingford J.B. and Marcotte E.M. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A* **107,** 6544-9 (2010).

Mehta P., Goyal S., Long T., Bassler B.L. and Wingreen N.S. Information processing and signal integration in bacterial quorum sensing. *Mol Syst Biol* **5,** 325 (2009).

Metzger B.P.H., Wittkopp P.J. and Coolon J.D. Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among Saccharomyces Species. *Genome Biol Evol* **9,** 843-854 (2017).

Mody A., Weiner J. and Ramanathan S. Modularity of MAP kinases allows deformation of their signalling pathways. *Nat Cell Biol* **11,** 484-91 (2009).

Moses A.M. and Landry C.R. Moving from transcriptional to phospho-evolution: generalizing regulatory evolution? *Trends Genet* **26,** 462-7 (2010).

Nagiec M.J. and Dohlman H.G. Checkpoints in a yeast differentiation pathway coordinate signaling during hyperosmotic stress. *PLoS Genet* **8,** e1002437 (2012).

Nagiec M.J., McCarter P.C., Kelley J.B., Dixit G., Elston T.C. and Dohlman H.G. Signal inhibition by a dynamically regulated pool of monophosphorylated MAPK. *Mol Biol Cell* **26,** 3359-71 (2015).

Namy O., Duchateau-Nguyen G., Hatin I., Hermann-Le Denmat S., Termier M. and Rousset J.P. Identification of stop codon readthrough genes in Saccharomyces cerevisiae. *Nucleic Acids Res* **31,** 2289-96 (2003).

Napthine S., Yek C., Powell M.L., Brown T.D. and Brierley I. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. *RNA* **18,** 241-52 (2012).

O'Rourke S.M. and Herskowitz I. The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in Saccharomyces cerevisiae. *Genes Dev* **12,** 2874-86 (1998).

O'Rourke S.M. and Herskowitz I. Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. *Mol Biol Cell* **15,** 532-42 (2004).

O'Connor J.-E. Flow Cytometry versus Fluorescence Microscopy:61-66. *Springer US*, Boston, MA (1996).

Patterson J.C., Klimenko E.S. and Thorner J. Single-cell analysis reveals that insulation maintains signaling specificity between two yeast MAPK pathways with common components. *Sci Signal* **3,** ra75 (2010).

Pawson T. Non-catalytic domains of cytoplasmic protein-tyrosine kinases: regulatory elements in signal transduction. *Oncogene* **3,** 491-5 (1988).

Peisajovich S.G., Garbarino J.E., Wei P. and Lim W.A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328,** 368-72 (2010).

Pelet S., Rudolf F., Nadal-Ribelles M., de Nadal E., Posas F. and Peter M. Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* **332,** 732-5 (2011).

Pelham H.R. Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature* **272,** 469-71 (1978).

Pereira-Leal J.B., Levy E.D. and Teichmann S.A. The origins and evolution of functional modules: lessons from protein complexes. *Philos Trans R Soc Lond B Biol Sci* **361,** 507-17 (2006).

Perez-Cruz F. Kullback-Leibler divergence estimation of continuous distributions. 2008 IEEE International Symposium on Information Theory, 1666-1670 (2008).

Plotkin J.B. and Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12,** 32-42 (2011).

Proulx S.R., Nuzhdin S. and Promislow D.E. Direct selection on genetic robustness revealed in the yeast transcriptome. *PLoS One* **2,** e911 (2007).

Ragan T., Kim K.H., Bahlmann K. and So P.T. Two-photon tissue cytometry. *Methods Cell Biol* **75,** 23-39 (2004).

Remenyi A., Good M.C., Bhattacharyya R.P. and Lim W.A. The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. *Mol Cell* **20,** 951-62 (2005).

Roberts C.J., Nelson B., Marton M.J., Stoughton R., Meyer M.R., Bennett H.A., He Y.D., Dai H., Walker W.L., Hughes T.R., Tyers M., Boone C. and Friend S.H. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287,** 873-80 (2000).

Robinson D.N. and Cooley L. Examination of the function of two kelch proteins generated by stop codon suppression. *Development* **124,** 1405-17 (1997).

Rockah-Shmuel L., Toth-Petroczy A., Sela A., Wurtzel O., Sorek R. and Tawfik D.S. Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. *PLoS Genet* **9,** e1003882 (2013).

Roy B., Leszyk J.D., Mangus D.A. and Jacobson A. Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. *Proc Natl Acad Sci U S A* **112,** 3038-43 (2015).

Sabbagh W., Jr., Flatauer L.J., Bardwell A.J. and Bardwell L. Specificity of MAP kinase signaling in yeast differentiation involves transient versus sustained MAPK activation. *Mol Cell* **8,** 683-91 (2001).

Saito H. Regulation of cross-talk in yeast MAPK signaling pathways. *Curr Opin Microbiol* **13,** 677-83 (2010).

Saito H. and Posas F. Response to hyperosmotic stress. *Genetics* **192,** 289-318 (2012).

Saldanha A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20,** 3246-8 (2004).

Sato P.M., Yoganathan K., Jung J.H. and Peisajovich S.G. The robustness of a signaling complex to domain rearrangements facilitates network evolution. *PLoS Biol* **12,** e1002012 (2014).

Schaber J., Baltanas R., Bush A., Klipp E. and Colman-Lerner A. Modelling reveals novel roles of two parallel signalling pathways and homeostatic feedbacks in yeast. *Mol Syst Biol* **8,** 622 (2012).

Schneider T.D., Stormo G.D., Gold L. and Ehrenfeucht A. Information content of binding sites on nucleotide sequences. *J Mol Biol* **188,** 415-31 (1986).

Schueren F., Lingner T., George R., Hofhuis J., Dickel C., Gartner J. and Thoms S. Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife* **3,** e03640 (2014).

Seaton D.D. and Krishnan J. The coupling of pathways and processes through shared components. *BMC Syst Biol* **5,** 103 (2011).

Shao D., Zheng W., Qiu W., Ouyang Q. and Tang C. Dynamic studies of scaffold-dependent mating pathway in yeast. *Biophys J* **91,** 3986-4001 (2006).

Sharp N.P., Sandell L., James C.G. and Otto S.P. The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc Natl Acad Sci U S A* **115,** E5046-E5055 (2018).

Solovieff N., Cotsapas C., Lee P.H., Purcell S.M. and Smoller J.W. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* **14,** 483-95 (2013).

Soyer O.S. and Bonhoeffer S. Evolution of complexity in signaling pathways. *Proc Natl Acad Sci U S A* **103,** 16337-42 (2006).

Spidlen J., Breuer K., Rosenberg C., Kotecha N. and Brinkman R.R. FlowRepository: a resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry A* **81,** 727-31 (2012).

Stiebler A.C., Freitag J., Schink K.O., Stehlik T., Tillmann B.A., Ast J. and Bolker M. Ribosomal readthrough at a short UGA stop codon context triggers dual localization of metabolic enzymes in Fungi and animals. *PLoS Genet* **10,** e1004685 (2014).

Storici F., Lewis L.K. and Resnick M.A. In vivo site-directed mutagenesis using oligonucleotides. *Nat Biotechnol* **19,** 773-6 (2001).

Strome B., Hsu I.S., Li Cheong Man M., Zarin T., Nguyen Ba A. and Moses A.M. Short linear motifs in intrinsically disordered regions modulate HOG signaling capacity. *BMC Syst Biol* **12,** 75 (2018).

Suderman R. and Deeds E.J. Machines vs. ensembles: effective MAPK signaling through heterogeneous sets of protein complexes. *PLoS Comput Biol* **9,** e1003278 (2013).

Tabbaa O.P. and Jayaprakash C. Mutual information and the fidelity of response of gene regulatory models. *Phys Biol* **11,** 046004 (2014).

Taher T.E.I. Monitoring Promoter Activity by Flow Cytometry. *Methods Mol Biol* **1651,** 65-73 (2017).

Thomson T.M., Benjamin K.R., Bush A., Love T., Pincus D., Resnekov O., Yu R.C., Gordon A., Colman-Lerner A., Endy D. and Brent R. Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range. *Proc Natl Acad Sci U S A* **108,** 20265-70 (2011).

Tian T. and Harding A. How MAP kinase modules function as robust, yet adaptable, circuits. *Cell Cycle* **13,** 2379-90 (2014).

Torkamani A. and Schork N.J. Predicting functional regulatory polymorphisms. *Bioinformatics* **24,** 1787-92 (2008).

Treusch S., Hamamichi S., Goodman J.L., Matlack K.E., Chung C.Y., Baru V., Shulman J.M., Parrado A., Bevis B.J., Valastyan J.S., Han H., Lindhagen-Persson M., Reiman E.M., Evans D.A., Bennett D.A., Olofsson A., DeJager P.L., Tanzi R.E., Caldwell K.A., Caldwell G.A. and Lindquist S. Functional links between Abeta toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334,** 1241-5 (2011).

Truckses D.M., Bloomekatz J.E. and Thorner J. The RA domain of Ste50 adaptor protein is required for delivery of Ste11 to the plasma membrane in the filamentous growth signaling pathway of the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **26,** 912-28 (2006).

True H.L. and Lindquist S.L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407,** 477-83 (2000).

van Leeuwen J., Pons C., Mellor J.C., Yamaguchi T.N., Friesen H., Koschwanez J., Usaj M.M., Pechlaner M., Takar M., Usaj M., VanderSluis B., Andrusiak K., Bansal P., Baryshnikova A., Boone C.E., Cao J., Cote A., Gebbia M., Horecka G., Horecka I., Kuzmin E., Legro N., Liang W., van Lieshout N., McNee M., San Luis B.J., Shaeri F., Shuteriqi E., Sun S., Yang L., Youn J.Y., Yuen M., Costanzo M., Gingras A.C., Aloy P., Oostenbrink C., Murray A., Graham T.R., Myers C.L., Andrews B.J., Roth F.P. and Boone C. Exploring genetic suppression interactions on a global scale. *Science* **354,** (2016).

Venkataram S., Dunn B., Li Y., Agarwala A., Chang J., Ebel E.R., Geiler-Samerotte K., Herissant L., Blundell J.R., Levy S.F., Fisher D.S., Sherlock G. and Petrov D.A. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell* **166,** 1585-1596 e22 (2016).

Voliotis M., Perrett R.M., McWilliams C., McArdle C.A. and Bowsher C.G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc Natl Acad Sci U S A* **111,** E326-33 (2014).

von der Haar T. and Tuite M.F. Regulated translational bypass of stop codons in yeast. *Trends Microbiol* **15,** 78-86 (2007).

Wagner A. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* **27,** 176-88 (2005).

Wagner A. Robustness, evolvability, and neutrality. *FEBS Lett* **579,** 1772-8 (2005).

Wagner A. Robustness and evolvability in living systems. *Princeton University Press. **Princeton, N.J. ; Woodstock*** (2007).

Wagner A. Robustness and evolvability: a paradox resolved. *Proc Biol Sci* **275,** 91-100 (2008).

Waltermann C. and Klipp E. Information theory based approaches to cellular signaling. *Biochim Biophys Acta* **1810,** 924-32 (2011).

Weng G., Bhalla U.S. and Iyengar R. Complexity in biological signaling systems. *Science* **284,** 92-6 (1999).

Winters M.J. and Pryciak P.M. Analysis of the thresholds for transcriptional activation by the yeast MAP kinases Fus3 and Kss1. *Mol Biol Cell* **29,** 669-682 (2018).

Wittkopp P.J., Haerum B.K. and Clark A.G. Evolutionary changes in cis and trans gene regulation. *Nature* **430,** 85-8 (2004).

Xu G., Jansen G., Thomas D.Y., Hollenberg C.P. and Ramezani Rad M. Ste50p sustains mating pheromone-induced signal transduction in the yeast Saccharomyces cerevisiae. *Mol Microbiol* **20,** 773-83 (1996).

Yamamoto K., Tatebayashi K., Tanaka K. and Saito H. Dynamic control of yeast MAP kinase network by induced association and dissociation between the Ste50 scaffold and the Opy2 membrane anchor. *Mol Cell* **40,** 87-98 (2010).

Yanagida H., Gispan A., Kadouri N., Rozen S., Sharon M., Barkai N. and Tawfik D.S. The Evolutionary Potential of Phenotypic Mutations. *PLoS Genet* **11,** e1005445 (2015).

Yu R.C., Pesce C.G., Colman-Lerner A., Lok L., Pincus D., Serra E., Holl M., Benjamin K., Gordon A. and Brent R. Negative feedback that improves information transmission in yeast signalling. *Nature* **456,** 755-61 (2008).

Zalatan J.G., Coyle S.M., Rajan S., Sidhu S.S. and Lim W.A. Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation. *Science* **337,** 1218-22 (2012).

Zarin T., Tsai C.N., Nguyen Ba A.N. and Moses A.M. Selection maintains signaling function of a highly diverged intrinsically disordered region. *Proc Natl Acad Sci U S A* **114,** E1450-E1459 (2017).

Zhao J., Zhou Y., Zhang X. and Chen L. Part mutual information for quantifying direct associations in networks. *Proc Natl Acad Sci U S A* **113,** 5130-5 (2016).

Zhao S., Douglas N.W., Heine M.J., Williams G.M., Winther-Larsen H.C. and Meaden P.G. The STL1 gene of Saccharomyces cerevisiae is predicted to encode a sugar transporter-like protein. *Gene* **146,** 215-9 (1994).

Zhou Z., Dang Y., Zhou M., Li L., Yu C.H., Fu J., Chen S. and Liu Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A* **113,** E6117-E6125 (2016).

Zhu Y.O., Sherlock G. and Petrov D.A. Extremely Rare Polymorphisms in Saccharomyces cerevisiae Allow Inference of the Mutational Spectrum. *PLoS Genet* **13,** e1006455 (2017).

Zhu Y.O., Siegal M.L., Hall D.W. and Petrov D.A. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A* **111,** E2310-8 (2014).

Zia A. and Moses A.M. Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. *BMC Bioinformatics* **12,** 299 (2011).

Appendix 1: Mutual Information script

```
#Title: MutualInformation.py
#Author: Purnima Kompella
#This code does:
    #1) discretizes continuous data by binning using linspace
    #2) adds a column of 0's and 1's to uninduced and induced, respectively;
        #this allows calculation of the entropy of the input, "input"
    #3) combines the data for uninduced, x, and induced, y, into "mergeddata"
    #4) calculates entropy using the formula entropyx = -sum(p(x)*log2 p(x))
    #5) calculates mutual information using the formula entropy mergeddata +
            #entropy input - [entropy (x)+entropy(y)]
         #6) This script requires files are named as such:
            #Date_StrainName_LineNumber_PassageNumber_Condition_x_x_x"


     import numpy as np
     import matplotlib
     #import matplotlib.pyplot as plt #can't import this when running on terminal (machine without display)
     import csv
     import glob
     import os
     import datetime #to save outfile with today's date

     matplotlib.use('Agg') #calling this to fix an error with opening on machines without display

     import matplotlib.pyplot as plt #calling after previous line to fix an error with opening on machines without display
     import matplotlib.mlab as mlab
     from matplotlib.patches import Rectangle

     #Use FindFewestRowsLowestandHighestGFP.R to find lower and upper limits for GFP values
     lowerlimit=0
     upperlimit=5.5
     #Number of bins should be at least ~0.5% of the number of data points. Example: at least 1000 bins for 20,000 cells
     numberofbins="1000"
```

```
34.     path = "" #insert path name here
35.
36.     now = datetime.datetime.now()
37.     todaysdate = now.strftime("%Y%m%d") #get's today's date in t
  he format yyyymmdd
38.
39.     outfilename = "MutualInformation_"+todaysdate+"_Range" + str
  (lowerlimit) + "-
  " + str(upperlimit) + "_" + str(numberofbins) + "bins.csv"
40.     outfile = open(outfilename, "w")  # opens and with "w", writ
  es, a new file in the same directory as path with the given name

41.     # write only takes one parameter so adding all the different
   column names together
42.     # (don't need to here but just doing it for consistency with
   variables later)
43.     # separating with a "," so the file can be csv
44.     # printing a new line at the end so the values can be propag
  ated in a new row
45.     outfile.write('Sample1,' + 'Sample2,' + 'MI_GFP,' + 'Date,'
  + 'Strain,' + 'Line,' + 'Passage,' + 'Time,\n') #add MI_BFP if ne
  cessary
46.
47.     #Calculates entropy of the merged data
48.     def calculate_entropy (data, cells, name, lowerlimit, upperl
  imit, numberofbins):
49.
50.         bins = np.linspace(lowerlimit, upperlimit, numberofbins)
   #bins values using the limits into the given number of bins
51.         digitized = np.digitize(data, bins) #Return the indices
  of the bins to which each value in input array belongs
52.         counts = np.bincount(digitized)[1:].astype(np.float32) #
  Count number of occurrences of each value in array of non-
  negative ints.
53.         normalized = counts / cells #calculates p(x)
54.
55.         sum = 0
56.         for i in normalized:
57.             if i != 0:
58.                 sum += i * np.log2(i)
59.             elif i == 0:
60.                 sum += i
61.
62.         if name != 'none':
63.             name_split = name.split("_")
64.             date=name_split[0]
65.             strain=name_split[1]
```

```
66.             line=name_split[2]
67.             passage=name_split[3]
68.             #condition=name_split[4]
69.             time=name_split[5]
70.             time=time+"min"
71.             reporter=name_split[12]
72.             figuretitle = date+" "+strain+" "+line+" "+passage+"
    "+reporter+" response \n+&- Induction after "+time+" ("+numberof
   bins+" bins)"
73.
74.             plt.hist(data, bins=bins, color='red', alpha=0.5, ed
   gecolor='red')
75.             plt.title(figuretitle)
76.             plt.ylabel("Count")
77.             plt.xlabel("Fluorescent Reporter Expression (A.U.)")

78.
79.         return -sum
80.
81.     #Calculates entropy of the input
82.     def calculate_entropy_input (data, cells):
83.         bins = np.linspace(0, 1, 2)
84.         digitized = np.digitize(data, bins)
85.         counts = np.bincount(digitized)[1:].astype(np.float32)
86.
87.         normalized = counts / cells
88.
89.         sum = 0
90.         for i in normalized:
91.             if i != 0:
92.                 sum += i * np.log2(i)
93.             elif i == 0:
94.                 sum += i
95.
96.         return -sum
97.
98.     #Calculates joint entropy
99.     def calculate_entropy_xy (data_x, data_y, cells, xname, ynam
   e, lowerlimit, upperlimit, numberofbins):
100.
101.        bins_x = np.linspace(lowerlimit, upperlimit, numberofbin
   s)
102.        digitized_x = np.digitize(data_x, bins_x)
103.        counts_x = np.bincount(digitized_x)[1:].astype(np.float3
   2)
104.
```

```
105.         bins_y = np.linspace(lowerlimit, upperlimit, numberofbin
   s)
106.         digitized_y = np.digitize(data_y, bins_y)
107.         counts_y = np.bincount(digitized_y)[1:].astype(np.float3
   2)
108.
109.         normalized_x = counts_x / cells
110.         normalized_y = counts_y / cells
111.
112.         sum = 0
113.         sum_x = 0
114.         for i in normalized_x:
115.             if i != 0:
116.                 sum_x += i * np.log2(i)
117.             elif i == 0:
118.                 sum_x += i
119.
120.         sum_y = 0
121.         for i in normalized_y:
122.             if i != 0:
123.                 sum_y += i * np.log2(i)
124.             elif i == 0:
125.                 sum_y += i
126.
127.         sum = sum_x+sum_y
128.
129.         name = yname
130.         name_split = name.split("_")
131.         date = name_split[0]
132.         strain = name_split[1]
133.         line = name_split[2]
134.         passage = name_split[3]
135.         #condition = name_split[4]
136.         time = name_split[5]
137.         time = time + "min"
138.         reporter = name_split[6]
139.
140.         #plt.figure()
141.         plt.hist(data_x, bins=bins_x, color='grey', alpha=0.8, e
   dgecolor='grey')
142.         #plt.title(figuretitlex)
143.         #plt.savefig(xname+".png")
144.         #plt.close()
145.
146.         #plt.figure()
147.         plt.hist(data_y, bins=bins_y, color='black', alpha=0.7)
```

```
148.         #plt.title(figuretitle)
149.         handles = [Rectangle((0, 0), 1, 1, color=c, ec="k") for
   c in ['grey', 'black', 'red']]
150.         labels = ["Uninduced", "Induced", "Combined"]
151.         plt.legend(handles, labels, loc=0)
152.         now = datetime.datetime.now()
153.         todaysdate = now.strftime("%Y%m%d")
154.         figurename = "Range"+str(lowerlimit)+"-
   "+str(upperlimit)+"_"+numberofbins + "bins_"+date + "_" + strain
   + "_" + line + "_"+ passage + "_" + reporter +"_"+ time+"_"+today
   sdate+".png"
155.         plt.savefig(figurename)
156.         plt.close()
157.
158.         return -sum
159.
160.     def open_file (path):
161.         file1 = open(path)
162.         list1 = csv.reader(file1, delimiter=',')
163.         matrix = np.array([row for row in list1])
164.         return matrix
165.
166.     #copied this function from: https://mail.python.org/pipermai
   l/tutor/2004-April/029019.html
167.     #Mac OS X inserts .DS_Store files and this function ignores
   any files that begin with a "."
168.     def mylistdir(directory):
169.         """A specialized version of os.listdir() that ignores fi
   les that
170.         start with a leading period."""
171.         filelist = os.listdir(directory)
172.         return [x for x in filelist
173.                 if not (x.startswith('.'))]
174.
175.     for filename in mylistdir(path):
176.         print("Filename: ", filename)
177.         strainName = None
178.         splitResult = None
179.         date = None
180.         line = None
181.         passage = None
182.         strain = None
183.         condition = None
184.         time = None
185.         rest = None
186.       path1name = None
187.         path1 = None
```

```
188.         path2name = None
189.         path2 = None
190.
191.         if filename.endswith(".csv"):
192.             strainName = filename
193.             splitResult = filename.split("_")  # separates the f
   ilename whereever there's an underscore
194.             date = splitResult[0]
195.             strain = splitResult[1]
196.             line = splitResult[2]
197.             passage = splitResult[3]
198.             condition = splitResult[4]
199.             time = splitResult[5]
200.             rest = ' '.join(["_",splitResult[6],"_",splitResult[
   7],"_",splitResult[8]])
201.
202.             if condition == 'Uninduced':
203.                 path1name = ' '.join([date, "_", strain, "_", li
   ne, "_",passage,"_Uninduced", "_", time, rest])
204.                 path1name = path1name.replace(" ", "")  # remove
    spaces
205.                 path1 = path1name
206.                 path2name = ' '.join([date, "_", strain, "_", li
   ne, "_",passage,"_Induced", "_", time, rest])  # add the string t
   ogether
207.                 path2name = path2name.replace(" ", "")  # remove
    spaces
208.                 path2 = path2name
209.
210.             else:
211.                 continue
212.
213.             print("Path1: ", path1)
214.             data1 = open_file(path1)
215.             print("Path2: ", path2)
216.             data2 = open_file(path2)
217.
218.             # Mutual Information (GFP)
219.         #Stores GFP values for uninduced cells from column 7 int
   o a new array and adds a column of 0s indicating induction status
   =uninduced
220.             data1_GFP = data1[1:, 6]  # gets GFP "B1-
   A" column, count starts at 0 so it's column 6
221.             data1_GFP_noblanks = list(filter(None, data1_GFP))
   # remove blanks
222.             data1_GFP_noblanks_array = np.asarray(data1_GFP_nobl
   anks)  # convert list to array
```

```
            data1_GFP_noblanks_array = data1_GFP_noblanks_array.astype(np.float32)
            zeros_GFP = np.zeros((len(data1_GFP_noblanks_array), 1))
            data1_GFP_noblanks_array_reshape = np.reshape(data1_GFP_noblanks_array, (len(data1_GFP_noblanks_array), 1))  # changes the shape of the array
            data1_GFP_noblanks_array_condition = np.append(data1_GFP_noblanks_array_reshape, zeros_GFP, axis=-1)

        #Stores GFP values for induced cells from column 7 into a new array and adds a column of 1s indicating induction status=induced
            data2_GFP = data2[1:, 6]
            data2_GFP_noblanks = list(filter(None, data2_GFP))
            data2_GFP_noblanks_array = np.asarray(data2_GFP_noblanks)
            data2_GFP_noblanks_array = data2_GFP_noblanks_array.astype(np.float32)
            ones_GFP = np.ones((len(data2_GFP_noblanks_array), 1))
            data2_GFP_noblanks_array_reshape = np.reshape(data2_GFP_noblanks_array, (len(data2_GFP_noblanks_array), 1))
            data2_GFP_noblanks_array_condition = np.append(data2_GFP_noblanks_array_reshape, ones_GFP, axis=-1)

        #Merging uninduced and induced
            mergeddata_GFP = np.concatenate((data1_GFP_noblanks_array_condition, data2_GFP_noblanks_array_condition))
            mergeddata_GFP = np.array(mergeddata_GFP)
            totalcells_GFP = len(mergeddata_GFP) #gets length of merged data array to find total number of cells

        #Gets file names without full path
            data1name = os.path.basename(path1)
            data1name = data1name.replace(".csv", "")
            data2name = os.path.basename(path2)
            data2name = data2name.replace(".csv", "")
            mergedname = data1name + "_" + data2name

        #Calculate entropy
            entropy1_GFP = calculate_entropy(mergeddata_GFP[:, 0], totalcells_GFP, (mergedname + "_GFP"), lowerlimit, upperlimit, numberofbins)
            entropy2_GFP = calculate_entropy_input(mergeddata_GFP[:, 1], totalcells_GFP)
```

```
252.             joint_entropy_GFP = calculate_entropy_xy(data1_GFP_noblanks_array, data2_GFP_noblanks_array, totalcells_GFP, (data1name + "_GFP"), (data2name + "_GFP"), lowerlimit, upperlimit, numberofbins)
253.
254.         #Calculate MI
255.             mutual_information_GFP = entropy1_GFP + entropy2_GFP - joint_entropy_GFP
256.             mutual_information_GFP = str(mutual_information_GFP)  # convert number to string so it can be stored (issue with storing mixed number and string array)
257.
258.     #       #Mutual Information (BFP)
259.         #Stores BFP values for uninduced cells from column 8 into a new array and adds a column of 0s indicating induction status=uninduced
260.     #       data1_BFP = data1[1:, 7]
261.     #       data1_BFP_noblanks = list(filter(None, data1_BFP))  # remove blanks
262.     #       data1_BFP_noblanks_array = np.asarray(data1_BFP_noblanks)  # convert list to array
263.     #       data1_BFP_noblanks_array = data1_BFP_noblanks_array.astype(np.float32)
264.     #       zeros_BFP = np.zeros((len(data1_BFP_noblanks_array), 1))
265.     #       data1_BFP_noblanks_array_reshape = np.reshape(data1_BFP_noblanks_array, (len(data1_BFP_noblanks_array), 1))  # changes the shape of the array
266.     #       data1_BFP_noblanks_array_condition = np.append(data1_BFP_noblanks_array_reshape, zeros_BFP, axis=-1)
267.     #
268.     #   #Stores BFP values for induced cells from column 8 into a new array and adds a column of 1s indicating induction status=induced
269.     #       data2_BFP = data2[1:, 7]
270.     #       data2_BFP_noblanks = list(filter(None, data2_BFP))
271.     #       data2_BFP_noblanks_array = np.asarray(data2_BFP_noblanks)
272.     #       data2_BFP_noblanks_array = data2_BFP_noblanks_array.astype(np.float32)
273.     #       ones_BFP = np.ones((len(data2_BFP_noblanks_array), 1))
274.     #       data2_BFP_noblanks_array_reshape = np.reshape(data2_BFP_noblanks_array, (len(data2_BFP_noblanks_array), 1))
275.     #       data2_BFP_noblanks_array_condition = np.append(data2_BFP_noblanks_array_reshape, ones_BFP, axis=-1)
276.     #
```

```
277.     #   #Merging uninduced and induced
278.     #       mergeddata_BFP = np.concatenate((data1_BFP_noblanks_
   array_condition, data2_BFP_noblanks_array_condition))
279.     #       mergeddata_BFP = np.array(mergeddata_BFP)
280.     #       totalcells_BFP = len(mergeddata_BFP)
281.     #
282.     #   #Calculate Entropy
283.     #       entropy1_BFP = calculate_entropy(mergeddata_BFP[:, 0
   ], totalcells_BFP, (mergedname + "_BFP"), lowerlimit, upperlimit,
    numberofbins)
284.     #       entropy2_BFP = calculate_entropy_input(mergeddata_BF
   P[:, 1], totalcells_BFP)
285.     #       joint_entropy_BFP = calculate_entropy_xy(data1_BFP_n
   oblanks_array, data2_BFP_noblanks_array, totalcells_BFP, (data1na
   me + "_BFP"), (data2name + "_BFP"), lowerlimit, upperlimit, numbe
   rofbins)
286.     #   #Calculate Mutual Information
287.     #       mutual_information_BFP = entropy1_BFP + entropy2_BFP
    - joint_entropy_BFP
288.     #       mutual_information_BFP = str(mutual_information_BFP)
     # convert number to string so it can be stored (issue with stor
   ing mixed number and string array)
289.
290.             # for testing:
291.             # print ('\tPath1:', os.path.basename(path1))
292.             # print ('\tPath2:', path2)
293.             # print ('\t\tmutual_information_GFP:', mutual_infor
   mation_GFP)
294.             # print ('\t\tmutual_information_BFP:', mutual_infor
   mation_BFP)
295.             # print ()
296.             #conditionname = condition.replace("No", "")
297.
298.         outfile.write(os.path.basename(path1) + ',' + path2 + ',
   ' + mutual_information_GFP +  ',' + date + ',' + strain + ',' + l
   ine + ',' + passage +',' + time + ',\n') #add mutual_information_
   BFP if necessary
299.             plt.close('all')
300.


301. outfile.close
```