Title
Negative feedback confers mutational robustness in yeast transcription factor regulation

Author
Denby, Charles

Publication Date
2012

Peer reviewed|Thesis/dissertation

Negative feedback confers mutational robustness in yeast transcription factor regulation

by

Charles Morrison Denby

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Rachel B. Brem, Chair

Professor Michael Eisen

Professor Douglas Koshland

Professor Adam Arkin

Spring 2012

Negative feedback confers mutational robustness in yeast transcription factor regulation

by Charles Morrison Denby

Abstract

Negative Feedback Confers Mutational Robustness in Yeast Transcription Factor Regulation

Charles Morrison Denby

Doctor of Philosophy in Molecular and Cell Biology University of California, Berkeley

Designated Emphasis in Computational and Genomic Biology

Professor Rachel Brem, Chair

Organismal fitness depends on the ability of gene networks to function robustly in the face of environmental and genetic perturbations. Understanding the mechanisms of this stability is one of the key aims of modern systems biology. Dissecting the basis of robustness to mutation has proven a particular challenge, with most experimental models relying on artificial DNA sequence variants engineered in the laboratory. In this work, we hypothesized that negative regulatory feedback could stabilize gene expression against the disruptions that arise from natural genetic variation. We screened yeast transcription factors for feedback, and used the results to establish the hypoxia regulator Rox1 as a model system for the study of feedback in circuit behaviors and its impact across genetically heterogeneous populations. Mutagenesis experiments revealed the mechanism of Rox1 as a direct transcriptional repressor at its own gene, enabling a regulatory program of rapid induction during environmental change that reached a plateau of moderate steady--state expression. Additionally, in a given environmental condition, Rox1 levels varied widely across genetically distinct strains; the ROX1 feedback loop regulated this variation, as the range of expression levels across genetic backgrounds showed greater spread in ROX1 feedback mutants than among strains with the ROX1 feedback loop intact. Our findings indicate that the ROX1 feedback circuit is tuned to respond to perturbations arising from natural genetic variation, in addition to its role in induction behavior. We suggest that regulatory feedback may be an important element of the network architectures that confer mutational robustness across biology.

Table of Contents

Acknowledgments

I would like to thank Dr. Rachel Brem for her mentorship, guidance and support over the last six years. The sustained effort you put into my education was remarkable. I would also like to thank my thesis committee members, Dr. Adam Arkin, Dr. Michael Eisen, Dr. Douglas Koshland, and Dr. Jeremey Thorner, who challenged me and helped me to contextualize my work. In addition, Dr. Adam Arkin's lab meetings, seminars, and classes have provided a constant intellectual inspiration throughout graduate school. More thanks go to the folks at the Molecular Sciences Institute for technical support and providing me with the tools I needed to work through my project.

I would like to thank Dr. Jon Cooper and Dr. Shiv Grewal for taking me into their labs as an intern. Both these experiences inspired me to go to graduate school, and each experience was instrumental to my success as a graduate student. Also, thanks to Dr. Meghan Maurer and Dr. Kenichi Noma for excellent mentorship during these internships.

I would like to thank my MCB classmates and the other members of the Brem lab. I had a blast spending time with you all, and it has been a pleasure to get to know you. From outdoor adventures to afternoon coffee, it has been fantastic. Special thanks go to Joo Hyun Im for her hard work helping me with my project.

I would like to thank all my friends here in the Bay Area. You have made my time in Berkeley a much richer experience, and the support you have provided is immense. If it were not for you all, I may not have survived. Lakshmi, AJ and Bekah, Anand, Joel, Lev, Nat, An-Chi, Corey, Greg, Adam Williamson, Dan Richter, Elisabeth and Nick, Christine and Steve Carr, East-bay Goaltimate friends, to name a few.

Finally, I want to thank my family. My mother has had a huge influence on me, encouraging me to pursue my intellectual curiosity at every step along the way. My father, having paved the way for a family of PhDs, has offered me so many supportive words of encouragement. Mom and Dad, I cannot thank you enough. My sister and Anush have been so loving—having them down in California for the last three years of my studies has been spectacular. Nick and Camille, you are amazing people—your love and care make our family a wonderful thing to be a part of. The stability you offer has been invaluable to me, and without it, graduate school would have been far more difficult.

Chapter 1

An introduction to biological robustness

Living systems show a remarkable ability to maintain stable phenotypes despite diverse perturbations. This principle, broadly referred to as robustness, can be defined as the stability of a system's characteristic phenotype in the face of environmental and genetic variation[1]. Biological robustness was first theoretically formalized by Waddington[2]. He argued that phenotypic variation in conditions similar to those of wild populations tends to be small compared to the variation observed after major environmental insults. To illustrate this point, he showed that developing *Drosophila* embryos subjected to heat shock showed more morphological variation in *Drosophila* wing-vein pattern compared to embryos grown at milder temperatures[3]. In this way, *Drosophila* development at temperatures characteristic of wild habitats is more robust than at higher temperatures.

Robustness is a complicated concept. To characterize a system as robust, one has to define the phenotype of interest and the underlying source of variation. In the above example, the phenotype is the presence or absence of a cross-vein formed between two specific wing veins of adult *Drosophila* subjected to heat shock during development. The variation in this experiment derives from two sources: Firstly the strain used for these experiments was a wild isolate. Due to the heterozygosity of wild strains, recombinant progeny contain genetic variation. Secondly the stochastic nature of biochemical events and low copy number of biochemical species results in stochastic variation[4]. It is important to note that the variation between developmental temperatures in this experiment is part of the phenotype under consideration, and is not the source of variation. By increasing the temperature of development, the system is sensitized to the effect of stochastic and genetic variation on wing morphology, and demonstrates that wing morphology variation is buffered during development at lower temperature. That is, *Drosophila* wing cross-vein development at a temperature characteristic of a wild habitat is more robust than at higher temperatures.

Waddington observed that conditional robustness could have profound evolutionary consequences[2]. On the one hand, robustness ensures a consistent developmental program that is well suited for conditions characteristic of wild habitats despite the accumulation of genetic variation in the population. On the other hand, if the stability breaks down, that genetic variation manifests as diverse phenotypes on which natural selection can act. In other words, conditional robustness gives rise to a general mechanism where genetic variation can be tolerated in some conditions and serve as the substrate of evolutionary adaptation in others, potentially giving rise to greater evolvability.

Has conditional robustness been selected to increase evolvability? Some have argued that a single mechanism confers robustness to both environmental and genetic variation. This leads to the possibility that robustness to genetic variation is simply a side effect of selection for robustness to environmental variation[5]. It is unknown whether all mechanisms of genetic robustness also confer environmental robustness. One of the major motivations for the continued study of robustness is the expectation that ultimately, these hypotheses can be tested. First, however, we need a better understanding of what

mechanisms underlie either class of variation, and where these mechanisms are used in biology.

Robustness in networks

Since Waddington's pioneering work, the principle of biological robustness has been widely appreciated. However, the underlying molecular mechanisms that give rise to robust behaviors have been emerging slowly ever since. The first and now canonical example of a molecular determinant of robustness is HSP90, a chaperone protein involved in maturation and maintenance of signal transducers[6]. In wild-type flies grown under non-stressful conditions, recombinant progeny rarely display morphological variation. However, strains that are mutant for HSP90 give rise to recombinant progeny with diverse and pervasive morphological variation. The morphological variation was proposed to be due to preexisting cryptic genetic variants, although recent work suggests that it could have been due to transposon-based lesions[7]. Regardless of the source of genetic variation, the original study proposed a molecular mechanism by which a single locus could buffer phenotypic variation in some conditions, while in other conditions could allow genetic variation to manifest as phenotypic variation on which natural selection could act.

Another instructive, but more opaque example of robustness lies in the yeast genetic network. For 80% of the ~6000 yeast genes, deletion mutant strains grow indistinguishably from wild-type strains on rich media[8]. This suggested that the large majority of yeast genes either serve no function in growth on rich media, or the network is robust to individual deletions. Subsequent studies of synthetic lethality—where two non-essential genes are shown to be lethal in combination—distinguished between these possibilities: nearly all non-essential genes have many synthetic lethal interactions[9], suggesting that each of these genes has a function. Growth on rich media is therefore robust to genetic variation in the form of single gene deletions[9,10].

The major question that arises from these studies is what mechanisms encode robustness in this genetic network? A great deal of theoretical work has sought an answer. One simple strategy is redundancy—where multiple components or pathways encode a single network function. There are two types of redundancy mechanisms. The first type is genetic redundancy, where multiple genes encode the same activity. The propensity for gene duplications in living systems provides the potential for genetic redundancy. Yitzhak Pilpel's group showed that for a class of yeast paralogs, deletion of one paralog leads to upregulation of the corresponding paralog, supporting genetic redundancy as a mechanism for robustness[11]. However, genetic redundancy cannot account for a significant proportion of observed robustness. First, only 13% of yeast genes have paralogs, and second, only 30% of paralogs have the capacity to back each other up[12]. Therefore, given the observed robustness in this model system, there must be other mechanisms of robustness that cannot be accounted for by genetic redundancy.

The second type of redundancy mechanism that encodes robustness is pathway redundancy, where an essential metabolite is synthesized through an alternate series of

chemical reactions. There are a number of known systems that illustrate how pathway redundancy confers robustness. For example, in wild-type E.coli strains, pyruvate kinase connects glycolysis with the Krebs cycle by converting phosphoenolpyruvate to pyruvate. When pyruvate kinase is deleted, this step is bypassed through the metabolic intermediates oxaloacetate and malate[13]. However, pathway redundancy also cannot account for the bulk of observed robustness, as *in silico* analysis of the yeast metabolic network suggests that only a small fraction of gene deletions are likely to be compensated by metabolic flux redistribution[14].

Another line of reasoning argues that robustness is an inherent property of complex regulatory networks subject to stabilizing selection. Bergman and Siegal performed simulations of complex regulatory networks illustrating this point[15]. Starting with an *in silico* gene network composed of 10 factors with random connections of varying sign and strength, they simulated mutation, quantified steady state expression levels, and selected new genotypes most closely matching a given expression state phenotype. Over the course of simulated evolution, the network connections diverged between individuals, while the phenotype remained stable. At the end of simulations, all strains closely resembled the optimal phenotype, while single gene deletions showed considerable variation in their distance from the optimal phenotype. In sum, in genetic networks subject to stabilizing selection, genetic variation accumulated within a population without a loss of fitness. But when a single gene was deleted from the network of each individual, the accumulated genetic variation manifested in large phenotypic variation between individuals.

In support of the Bergman-Siegal model, a subsequent study showed that deleting at least 5% of yeast genes results in greater morphological variation across different environments compared with a wild-type strain[16,17]. In contrast with the Bergman-Siegal model, the source of variation in these experiments is environmental and not genetic. However, the authors argue that because mechanisms conferring robustness to genetic variation or environmental variation are often congruent, these results reflect how such genes are capable of buffering genetic variation.

Case studies of regulatory networks illustrating robust behaviors

In addition to systematic gene deletion studies, case studies have proven useful for determining mechanisms of robustness. Barkai and Leibler provided an example where mathematical modeling predicted how a complex regulatory network confers robust adaptation in bacterial chemotaxis[18]. Here, "adaptation" refers to how the chemotaxis network adapts to changes in concentration of chemostimulants instead of absolute chemostimulant levels. In this system, the authors found that some properties of chemotaxis are sensitive to variation of component protein concentrations, while others maintain stability across a wide range of variation.

Bacterial chemotaxis has been the subject of extensive genetic and biochemical study for many decades, and thus provided an excellent case for systems analysis. Bacteria can sense chemical gradients by continuously monitoring chemical

concentrations as they swim. If they sense increasing concentrations of attractant over time (or decreasing repellent), they continue swimming. If they do not sense a gradient, they randomly reorient their direction of swimming—a behavior called tumbling. Chemical sensing in *E. coli* is mediated by receptor proteins that bind attractant or repellent molecules. These receptors relay information about the environment to flagellar motor proteins through a complex biochemical network, the output of which controls the frequency of tumbling and swimming. When an organism is placed in an environment with fixed concentration, it exhibits constant rates of switching between swimming and tumbling. When the organism is placed in an attractant concentration gradient and begins swimming up the gradient, the network senses the increase and responds by reducing the transition rate between swimming and tumbling, thereby increasing the chance that it reaches the attractant. Shortly after the organism reaches the higher concentration, it adapts to the new concentration and resumes the original rates of swimming and tumbling. An amazing feature of *E. Coli* chemotaxis is that it functions over a broad range of chemical concentrations. This type of system—where the output responds to changes in concentration rather than absolute concentration—is said to exhibit perfect adaptation.

Initial models describing the output of the genetic network with changing attractant concentrations explained the observed perfect adaptation behavior[19]. However, whereas E. coli adaptation is robust to various modifications to network components, the models failed to exhibit robustness when a corresponding modification was added to the model. A key insight came when Barkai and Leibler revisited the models and showed that a minor modification—restricting a particular chemical modification to a specific receptor conformation—gave rise to a system that is robust to parameter modifications[18]. Their new model was then validated experimentally: when various component concentrations were modified, the model mimicked the observed effects on behavior where previous models did not[20]. Interestingly, while behavior of perfect adaptation was robust to major network perturbations, the steady-state behavior and timescale of adaptation were not. One possibility is that the network architecture has been selected to ensure stability of adaptation while affording plasticity in component expression levels and adaptation time scale.

Another example of a network architecture that confers robustness lies in biosynthetic pathway metabolism. In a foundational paper on metabolic control analysis, Kacser and Burns put forth the flux summation theorem to show that, for a given enzymatic pathway, the amount of control that each enzyme imparts on total pathway flux sums to one[21]. Here, control is defined as the ratio between the fractional change of flux and the fractional change of enzyme concentration for a given enzyme in a pathway: If a 1% increase in an enzyme's concentration brings about a 1% increase on pathway flux, that enzyme is said to have perfect control (i.e. control = 1). The flux summation theorem implies that flux control can be either *i)* evenly distributed throughout the pathway, or *ii)* restricted to a small number of control points. In either situation, the result is that, for at least some of the enzymes in the pathway, small changes in enzyme concentration will yield little or no effect on pathway flux—in other words, the pathway is robust to concentration changes for those proteins. To demonstrate that the theory

applies to biological systems, the dose of three enzymes in the terminal steps of the arginine biosynthesis pathway of Neurospora were altered. As predicted, reducing the genetic dose of each enzyme to as little as ¼ of wild-type gave nearly equivalent pathway flux[22].

Another interesting conclusion from Kacser and Burns' work is that end-product feedback regulation—where a pathway product represses production of an upstream enzyme—can result in near complete control of flux by a single enzymatic step[21]. That is, the levels of all other enzymes in the pathway are robust to modification since changes in their activity can be compensated by the feedback mechanism. This architecture is extremely common for metabolite homeostasis in biological systems[23].

Negative feedback on gene expression as a mechanism for robustness

In addition to feedback acting on metabolic outputs, a number of studies have shown that direct transcriptional feedback on gene expression is a recurrent "motif" of biological networks[24,25]. This suggests that negative feedback may encode robustness for expression of particular genes. Many regulators are known to have dose-dependent signaling properties, such that proper cellular concentration may be of critical importance for fitness; one hypothesis is that appropriate levels are maintained via negative feedback. However, it remains unclear whether the widespread incidence of feedback is the product of selection or not. Supposing that feedback has been broadly selected to confer gene expression robustness, it is unknown what type of variation—whether environmental/stochastic or genetic—it would have been selected to buffer. In addition, regardless of the selective pressures underlying the widespread incidence of feedback, if feedback buffers genetic variation, this may have evolutionary importance: If feedback buffers variants that would be deleterious in some set of conditions, feedback may allow for such variants to persist in the population until a distinct set of conditions arise in which the variants are favorable. In this sense, feedback may contribute to evolvability[26].

Negative feedback as a robustness mechanism for stochastic variation

Despite considerable effort, whether feedback functions to reduce stochastic variation of gene expression in natural circuits is unclear. Theoretical work predicted that negative feedback would stabilize expression in response to stochastic variation[27,28], and a prominent study used a synthetic circuit to show that negative feedback can in fact decrease stochastic variation in gene expression[29]. However, subsequent work suggested that the expression variation observed in the latter experiment was likely due to variation in plasmid copy number[30], and therefore the result did not necessarily bear on feedback loops encoded in the genome. Follow up studies using the same system showed that stochastic models could only explain the observed gene expression behavior when the effect of plasmid variation was included[31]. Additional studies using various computational modeling techniques suggest that negative feedback is not predicted to reduce expression variation under a physiologically relevant parameter regime[32,33]. An additional study argues that using signaling-based mechanisms to cope with molecular fluctuation associated with low molecule number is a

metabolically expensive solution for reducing variation, and that cells generally reduce molecular variation associated with small molecule number by transcribing regulatory genes tens of thousands of times per cell cycle[34]. Furthermore, studies of circuits with naturally occurring negative feedback loops have yielded mixed reports of whether feedback reduces stochastic variation of gene expression[35,36].

In contrast, a recent synthetic study reported that feedback at a chromosomally integrated transcription factor reduces stochastic variation of downstream gene expression at a chromosomally integrated downstream reporter[37]. Interestingly, the most pronounced noise reduction occurred at intermediate expression levels, consistent with earlier studies showing that noise strength from a reporter expressed from GAL1 promoter is maximal at intermediate gene expression levels[38]. This phenomenon may be limited to the synthetic system: the GAL1 promoter is regulated by the SAGA complex, the targets of which show particularly high expression noise[27,39]. It seems unlikely that a transcription factor gene selected for stable expression would fall under the regulation of the most unstable known regulatory mechanism observed in yeast.

In sum, it remains to be seen whether the high incidence of endogenous feedback circuits do in fact reduce stochastic variation in naturally occurring biological networks, and whether feedback was widely selected to reduce stochastic variation.

Negative feedback as a robustness mechanism for genetic variation

Despite the impressive effort that has gone towards studying whether gene expression feedback buffers stochastic variation, what effect gene expression feedback has on genetic variation has received considerably less attention. If stability of certain transcription factors is critical for organismal function over the course of evolution, genetic variation in regulatory factors that affect that transcription factor's expression may not be tolerated. However, feedback on transcription factors could buffer variation affecting their expression, thereby allowing for variation in their regulators. Much like the case where *E. coli* exhibit robustness for chemotactic adaptation while allowing for plasticity of other properties, feedback may enable robust transcription factor expression while allowing for plasticity of their regulatory genes' activity. Therefore, feedback may facilitate accumulation of genetic variants upon which selection can act, and could have a crucial role in evolution of adaptive genotypes.

To begin to address which functions feedback may have in the yeast regulatory network, in Chapter 2, we identify cases of feedback in yeast transcription factors. In Chapter 3, we show that feedback can in fact confer gene expression robustness in the face of naturally arising genetic variation. We also provide evidence that feedback may have evolved to prevent transcription factor misexpression.

Negative feedback as a robustness mechanism for environmental variation

The relationship between feedback-mediated robustness and environmental variation is in some ways similar to that of stochastic variation. Just as individual cells

are subject to internal environmental changes due to stochastic processes, they are also subject to external environmental changes in nutrients and stresses. Feedback may confer robustness to variation in the external environment much the same as it may confer robustness to stochastic variation.

On the other hand, in some cases a transcription factor has evolved to vary with specific environmental changes. There are two potential functions for feedback in this case. First, when a cell transitions between environments the time it takes for a transcription factor to reach its new steady state concentration can be accelerated by feedback[27,40]. This principle has been shown to be true in both synthetic[36,40] and biological systems[36,41], and, in chapter 3, we show that this is true of the yeast transcription factor ROX1.

The other function for feedback in responding to environmental change lies in linearizing dose response. The relationship between input (i.e. promoter activation) and output (i.e. gene expression level), referred to as "dose-response," can be made more linear by feedback, and may prove selectively advantageous. Dose response linearization was first conceptualized in engineering systems[27,41] and was later extended to biological systems in theory[27,37,42], in synthetic circuits[37,42-45], and in natural regulatory networks[43-45]. Importantly, dose-response linearization can also improve the fidelity of information transmission. For example, if a transcription factor gene is controlled by an activity that reports an environmental signal, a more linear dose-response can encode more information about the signal to downstream components.

Is dose-response linearization likely to provide a selective advantage in natural systems? The authors of Yu *et al.* make a compelling argument for how linearization may be important for fitness: Differences in pheromone concentration are important for haploid yeast cells' mating partner choice—they show a remarkable ability to preferentially mate with partners producing the most pheromone[46]. Surprisingly, when Yu *et al.* disrupt the putative feedback mechanism, although the dose response curve shifts significantly, it does not change its linearity. This finding illustrates how it can be difficult to predict the function of a negative feedback loop embedded inside a large regulatory network with many unaccounted interactions.

As we note in Chapter 2, many of the transcription factors found in negative feedback loops are themselves embedded in larger metabolic homeostasis feedback loops. In Chapter 4 we show preliminary evidence that feedback linearizes the ROX1 transcription factor's dose-response, and discuss the potential importance of linearization in the context of metabolite homeostasis.

Chapter 2

A method for measuring feedback for yeast transcription factors

Introduction

Cells rely on regulatory networks to modify the expression of genes in response to information about their environment. This allows cells to make the proteins they need at the correct times and amounts. Much is known about how regulatory networks are connected, however, it is not clear whether there are principles that explain why natural networks have evolved particular topologies. One possibility is that naturally selected topologies offer evolutionary advantages over alternate topologies.

In recent years, new technologies have enabled regulatory interactions to be catalogued on a genomic scale in a variety of model systems. One regulatory structure common to all model systems examined to date is that of direct transcriptional feedback, where a gene product regulates its own expression[1,2]. Theoretical work has proposed that feedback may perform information-processing functions, and a number of case studies have illustrated its potential functional relevance[3]. For example, negative feedback on gene expression has been shown to confer gene expression homeostasis in the face of various perturbations[4-6], and can improve information transmission by linearizing a transcription factor's dose response[7,8]. On the other hand, positive feedback on gene expression has been shown to drive phenotypic diversity[9,10]. While such case studies have firmly established that feedback circuits have the capacity to confer the proposed functions, it is difficult to judge whether the proposed functions have been selected in biological systems. A first step towards addressing the general relevance of feedback in naturally evolved systems would be to test how often negative and positive feedback occurs in regulatory networks. A second step would be to ask what cellular processes negative and positive feedback are associated with. If negative or positive feedback is associated with a common cellular process, this could drive hypotheses for what information-processing functions feedback may have been selected to confer.

The organism *Saccharomyces cerevisiae* has proven an excellent model system for studying eukaryotic gene expression. Given the small gene size and compactness of the genome, yeast has been an ideal system for applying genomic technologies. Microarray and RNA-seq studies have been used to characterize the relationship between genetic perturbations and gene expression by modulating expression of a regulatory protein (i.e. deleting or overexpressing), and measuring the effect on gene expression[11,12]. This approach has proven successful in revealing relationships between regulatory proteins and other genes; however, it does not reveal cases where regulatory proteins affect their own expression. Two strategies have emerged that circumvent this shortcoming. First, genome-wide ChIP-on-chip studies have determined which transcription factors bind to their own encoding genes. This strategy has its own major shortcomings, the most notable of which is that it does not distinguish between positive and negative feedback. An alternative strategy is to use a transcriptional reporter to test for feedback. The general

approach measures expression from a reporter, while varying copy number of the gene of interest. If a promoter drives higher expression of a β–Galactosidase gene in a deletion strain than in a strain bearing a wild-type gene copy, this is taken as evidence of a negative feedback loop. Such experiments, however, have only been performed on a limited scale or on a gene-by-gene basis[13-19].

Two recent studies surveyed the yeast genome for feedback on non-regulatory genes, and reported conflicting results. In a first study, gene expression was measured in strains with different gene copy number. The authors reasoned that if genes were dosage compensated (i.e. subject to negative feedback), then each strain would show similar total expression levels. They observed dosage compensation for 13 of 16 genes, and concluded that the majority of yeast genes are subject to feedback[20]. In a second study, gene expression from a single GFP-tagged locus was compared between diploid strains carrying either a deletion or an untagged gene at the second locus. If these genes were dosage compensated, then expression from the tagged locus should increase by two-fold when the copy number is reduced by half. In contrast with the first study, less than 5% of 730 genes examined showed dosage compensation[21].

Neither of these studies tested transcription factors for feedback, despite the fact that feedback in transcription factors could have particular importance for ensuring homeostatic expression. Given that small differences in transcription factor concentrations can have a major impact on expression of a multitude of genes, small deviations from their optimal levels can have strong deleterious effects on cellular physiology. Furthermore, there are many known cases of transcription factor feedback from previous studies, and a transcription factor's ability to directly regulate transcription provides a simple mechanism by which feedback can act.

In this study, we systematically tested transcription factors for feedback. Previous studies have avoided testing transcription factors because many are below the threshold of detection using standard high-throughput methods[21]. Here, we developed a highly sensitive fluorescence microscopy method that enabled us to measure expression of many transcription factors that are below the detection limit of other methods. We used this method to test for feedback by measuring expression from a GFP-tagged locus in strains with varying untagged gene copy number (Figure 2.1). This strategy improved on previous ChIP-on-chip studies because it identified cases of *functional* feedback, and informs the sign of feedback. We found evidence that many transcription factors are subject to negative feedback, consistent with our expectation that feedback functions widely in expression homeostasis. We also found evidence that a number of transcription factors were subject to positive feedback. Each of these transcription factors is involved in yeast stress response, consistent with our expectation that positive feedback may promote phenotypic variation when cells are faced with stressful conditions.

Results

Establishing a reporter system for measuring feedback

Transcription factors are expressed at low levels from their endogenous loci[1,22]—the majority of the transcription factors whose expression we set out to measure were not detectable using a flow cytometer[23], which is the industry standard for sensitive high-throughput fluorescence quantification. Thus, our first step towards assaying for feedback was to develop a method for quantifying transcription factor expression. As a first strategy, we attempted to avoid the difficulty of measuring low endogenous expression by using a collection of multi-copy plasmids, each bearing a promoter of interest fused with GFP[24]. Each plasmid also contains a URA3 gene, which compliments the uracil auxotrophy of the parent strain. In a pilot experiment, we measured cellular fluorescence of a strain bearing the *prYHP1*-GFP plasmid by flow cytometry. The strain was grown in synthetic media lacking uracil in order to ensure the presence of the plasmid within all cells of the culture. We found that the distribution of cellular fluorescence intensities within the cultures varied between parent strains of different ploidy and different culture densities. We also found that population fluorescence measurements conformed to a bimodal distribution, with the higher expressing mode spread over many orders of magnitude (Figure 2.2a-d). We corroborated this distribution pattern using fluorescence microscopy (Figure 2.2e), observing that some cells exhibit low GFP expression and cells with higher expression vary widely in expression magnitude. The ratio of total number of cells conforming to each mode varied greatly between haploid and diploid strains (Figure 2.2a and b) and high and low density cultures (Figure 2.2b and d). The mean and median value of the plasmid-bearing mode differed substantially (~1.5 fold) between different culture densities (Figure 2b and d), suggesting that gating out non-expressing cells would not be sufficient to normalize expression.

We hypothesized that the low-expressing cells did not contain plasmid, and that the expressing cells contained the plasmid, but with highly variable copy number. We confirmed that some fraction of the culture had lost plasmid by sorting the cells based on fluorescence and measuring the viability of resulting cells grown on a media lacking uracil. All GFP positive cells grew on plates lacking uracil, but the large majority of GFP negative cells did not, indicating that they had lost the plasmid. We concluded that a reporter system with variable copy number was not ideal for making sensitive and reproducible measurements of gene expression.

Having ruled out the strategy of increasing copy number to improve signal, we resorted to a second strategy of developing a more sensitive method for quantifying low fluorescence signal. In reviewing data from a fluorescence microscopy based genome-wide protein localization screen[25], where each gene was tagged with GFP at their endogenous locus, we noticed that the large majority of genes we intended to measure had been reported as "nuclear localized." Inspection

of the image files showed that nuclear fluorescence signal for many of these factors was clear but weak. We reasoned that quantifying the ratio between nuclear fluorescence and background autofluorescence in microscopy images may enable us to accurately quantify fluorescence for factors that were previously undetectable by flow cytometry[23]. Thus, the anticipated advantage of fluorescence microscopy was not in sensitivity per se, but in the ability to measure fluorescence in subcellular structures to enhance signal-to-noise ratio. Previously reported quantification measurements using these expression constructs proved extremely reproducible[23], presumably because the reporter was stably integrated at a single locus. We next made strains with which we could assay feedback for 105 regulatory genes, most of which were transcription factors, corresponding to a collection of over-expression plasmids that we received from Bhupinder Bhullar (Figure 2.1). We successfully generated 95 sets of strains (Table 2.1).

Quantification Method

As a pilot experiment we imaged OPI1-GFP tagged strains using fluorescence microscopy, and analyzing images with Cell-ID software, developed at the Molecular Sciences Institute. The software takes as input a brightfield and fluorescence image, and uses the brightfield image to determine cell boundaries. We used the brightest fluorescence-channel pixel inside each cell's boundary as a proxy for the center of the nucleus, and calculated nuclear intensity by subtracting average background cellular fluorescence from average nuclear fluorescence (Figure 2.3a, see figure legend for details). As expected, a small increase in total cellular fluorescence relative to background (~15%; Figure 2.3b) corresponded to a much greater nuclear signal increase relative to background (~130%; Figure 2.3c). Furthermore, we found that expression differences between strains that were undetectable when comparing total cellular fluorescence measurements became substantial when comparing nuclear signals (compare Figure 2.3b and 2.3c).

We used fluorescence microscopy in conjunction with our image processing protocol to quantify fluorescence for all generated strains. Each strain was grown in rich media to mid log phase, cells were deposited on glass-bottom plates in rich media, and then rich media was replaced with a minimal media. Images were acquired within ~30 minutes of replacing the media. For imaging we chose to use a minimal media lacking sugar and amino acids because of its low autofluorescence. However, this had the additional effect of inducing the stress response, as illustrated by expression dynamics of the stress-induced transcription factor CIN5 (Figure 2.5c). 62 of the 95 factors showed mean nuclear fluorescence (averaged across varying doses) $2^{0.5}$-fold greater than background and were considered for further study (Figure 2.4; table 2.2). We found that our measured expression levels were significantly correlated with expression levels reported in Newman *et al.*[23] (Pearson's product-moment correlation; p-value=0.005; r=0.65). From this analysis, we confirm that our quantification method is both sensitive and accurate.

Assaying transcription factors for feedback

For each of the 62 sets of strains, we calculated the ratio of expression levels between hemizygote and wild-type and between wild-type and over-expression. The genes for which both ratios were above $2^{0.2}$ were considered as subject to negative feedback, while the one gene for which both ratios were below $2^{-0.2}$ was considered as subject to positive feedback (Figure 2.5a and Table2.3). We initially collected data for only a single biological replicate. To assess reproducibility of our measurements, we next performed the same experiments in biological triplicate for 17 of the 62 factors. Comparing replicate values revealed a modest amount of variation between biological replicates (Figure 2.5b). We used these data to generate an error distribution from which we estimated the false discovery rate of feedback (see Methods). We identified feedback for 10 of 62 factors at a cutoff where only one was expected by chance.

The effects indicating feedback by single replicate measurements were generally consistent with biological triplicate experiments (Figure 2.5c-e). The overlap between factors identified as subject to feedback in single replicate experiments and triplicate experiments is shown in Figure 2.7a. The three genes where single replicate measurements report feedback but triplicate measurements do not agree have all been inferred to have feedback in other studies (Table 2.5), suggesting that our triplicate measurements resulted in a high rate of false negatives. A high rate of false negatives is not ideal for a comprehensive screening method, so we next sought to modify our growth and imaging protocol to improve reproducibility.

In order to achieve higher measurement reproducibility, we imaged each strain in rich media instead of minimal media. Because rich media is more autofluorescent, and because some genes are not expressed as highly in rich media, we could only detect a subset of the factors that were detected in minimal media (Figure 2.6a). On the other hand, comparing expression values between replicates confirmed that imaging in rich media improved measurement reproducibility (compare r^2=0.82 with r^2=0.96 between Figure 2.5b and Figure 2.6c, see figure legends for details). In rich media, we identified feedback for 7 of 36 factors, with a FDR estimation of 0 false positives (Table 2.3), and genes identified as subject to feedback by single replicate measurements were generally consistent with biological triplicate experiments (Figure 2.6d-e and Figure 2.7b). Of the 36 factors we detected in rich media, only 13 were previously detected by flow cytometry in a similar synthetic complete media[23], suggesting that our method offers a substantial improvement in sensitivity for nuclear localized factors.

Discussion

We present a simple, accurate, and highly sensitive method for detecting functional feedback on yeast transcription factors. By using fluorescence microscopy with image processing we are able to interrogate functional feedback for a class of genes that could not be detected using whole-cell fluorescence quantification methods[21,23]. The results generated using our method, combined with results from previous whole-genome ChIP studies[2,26], provide a substantially richer view of feedback. There are a number of sources of uncertainty in ChIP studies that can be informed with functional feedback assays. First, functional feedback assays can determine if a gene that binds its promoter actually affects gene expression at that locus. Second, functional feedback assays can detect cases of indirect feedback. Third, when a gene is known to be auto-regulated, auto-activation can be distinguished from auto-repression.

By combining ChIP data with our functional feedback results, we are able to infer a mechanism of feedback for a number of factors. Of all the cases of feedback we identified in this study, we describe either a direct (i.e. a transcription factor binds its own gene) or semi-direct (i.e. a transcription factor forms a complex with other proteins, which in turn bind the transcription factor's gene) mechanism by which feedback is imparted for ~½ of the factors (Table 2.5). Furthermore, our results indicate that negative feedback is far more prevalent than positive feedback. This is consistent with previous meta-analysis results from *E. Coli* transcription factors[1] and also with previous reports focused on non-transcription factor genes[20,21].

Many of the transcription factors that we find to be subject to negative feedback are regulators of metabolism. Specifically, ROX1, MOT3 and SUT1 regulate heme and ergosterol biosynthesis genes; OPI1 and INO2 regulate inositol biosynthesis genes; NRG1 regulates glucose-repressed genes (such as the GAL genes involved in galactose metabolism); and ARG80 regulates arginine biosynthesis genes. A number of these factors transmit information from a metabolite sensor (or are themselves metabolite sensors) to the regulon encoding the control point for that metabolic pathway (Figure 2.8). In this way, they are involved in larger feedback loops that may serve to maintain homeostasis of a metabolic end product. Given their position within the control circuit, a compelling hypothesis is that the direct feedback loops we identified have evolved to increase information transmission between metabolite levels and metabolite production, thereby improving the precision of metabolite homeostasis.

Interestingly, the factors that display evidence for positive feedback—CIN5, YAP3, YAP5, and HOG1 (see Table 2.3 and Figure 2.5e)—are all related to stress response. There are two major functions for positive feedback that may explain this observation. First, positive feedback can drive expression diversity[27]. Expression diversity may in turn lead to phenotypic diversity[28], which can confer an advantage

in surviving stresses that cannot be precisely sensed[29]. It seems possible that positive feedback in yeast transcription factors has evolved to ensure diverse responses to various stresses. A second known function of positive feedback is in enhancement of cellular memory[30]. For a number of stresses, a population of yeast cells become more resistant to a particular type of stress if they have been preconditioned with that stress[31]. It seems possible that when a stress-responsive transcription factor is induced by exposure to a stress, positive feedback encodes memory of that environment by prolonging expression of that factor in some fraction of the cells, thereby conferring greater resistance to future instances of such a stress.

In sum, our results hint at general functions that positive and negative feedback may confer. It would be interesting if these trends could be statistically substantiated with a larger data set, though the low quality of functional annotations is currently a major limitation. For instance, the YAP protein family contains 8 stress response transcriptional activators that are known to mediate stress response, however there is no single Gene Ontology term that unites them. Finally, future studies will be needed to determine what functions each of these feedback loops confer. In the next two chapters, we investigate potential functions of negative feedback for the **R**epressor of hyp**OX**ia transcription factor ROX1.

An additional advantage of our using fluorescence microscopy as a method for quantifying gene expression is that we obtained expression levels for individual cells of an isogenic population. Because negative feedback has been proposed to reduce expression variation between isogenic cells, we had initially hypothesized that feedback would be correlated with lower variation between cells grown in log-phase cultures. This turned out not to be true, and there are a number of potential explanations for this. First, it is possible that feedback does not reduce expression variation at all—although a number of studies have predicted that negative feedback reduces expression variation[27,32], other studies have suggested that such a prediction is ill founded[33,34]. A second possibility is that negative feedback only reduces expression variation in certain conditions. In some cases, when a gene is activated at intermediate induction levels, variation is at its maximum[6,35]. This is due to the inherent switch-like activation behavior of some gene's expression; Intermediate induction results in some cells with active expression, and others without. The feedback effects we measured here may in fact reduce variation, but only in conditions where genes are intermediately activated. The conditions we used probably resulted in low or high induction of gene expression for most genes tested. A third possibility is that negative feedback does reduce expression variation in the conditions we used and our experimental design fails to measure it. This could be due either to measurement error, or to heterogeneity of cellular state in the given growth condition. For a number of yeast genes, the majority of cell-to-cell variation is due to differences in cellular state rather than stochasticity of gene expression[36,37]. For example, a number of studies have shown that differences in cell-cycle position result in considerably increased variation[36,38]. Even if gene expression were perfectly coordinated in every cell, because many genes are

expressed throughout the cell cycle, we would expect to see some level of variation between cells due to differences in cell-cycle positions. Therefore, it remains a formal possibility that we did not detect differences in expression variation for genes with feedback because we had not controlled for differences in cellular states.

Materials and methods

Yeast strains, and genetic manipulations. BY4741 was transformed with the URA3-marked plasmid containing YHP1 promoter fused with GFP[24] ("haploid" strain in Figure 2.2), and mated with BY4742 ("diploid" strain in Figure 2.2).

For each of 95 yeast transcription factors (Table 2.1), a strain bearing one copy of a GFP-fused gene[25] (Invitrogen) was transformed with either an overexpression plasmid[39] bearing URA3 and the corresponding open reading frame under the GAL1,10 promoter, or the backbone of the overexpression plasmid. Strains bearing the backbone plasmid were mated to either the corresponding gene deletion strain[40] (Invitrogen) to form "hemizygote" strains, or BY4742 to form "wild-type" strains. Strains bearing the overexpression plasmids were mated to BY4742 to form "overexpression" strains. All plasmid transformations and yeast cell matings were performed by standard techniques[41].

Feedback assays. For feedback assays, a culture of each yeast strain was inoculated into CSM –uracil media (MP Biomedicals) with 2% raffinose and grown in 2ml round bottom 96-well plates with a glass bead at 30°C with shaking to saturation. These plates were then refrigerated and served as stocks from which we could inoculate cultures for experiments. 10µl of re-suspended fridge-stock solutions were inoculated into 1ml fresh media and grown overnight to saturation. Strains were then back-diluted into CSM –uracil media with 2% galactose to an optical density of 0.1. Strains were imaged after ~6 hours of growth, having reached optical density 0.5-1.0. Upon preparation of 96-well plates (see below), 100 µl of cell suspension was added to each well and allowed to settle and bind for 30 minutes before imaging as described below. For replicate experiments three independent cultures of each strain were grown and prepared for imaging as above.

Fluorescence microscopy. 96-well glass bottom plates (Falcon) were coated with 100 µl of a 100 µg/ml solution of concanavalin A type V (Sigma-Aldrich), incubated 1 hour at room temperature and then washed three times with water. Wells were then washed and resuspended with media immediately before imaging. For imaging, we used a 60X PlanApo objective (N.A.=1.4) under oil immersion in a Nikon TE2000 inverted microscope equipped with a mercury lamp, a motorized stage and 512BFT MicroMax cooled CCD camera (Photometrix, Tucson, AZ). We imaged 3-6 observation fields per strain, with each field containing ~100 cells. For each field we acquired a 0.5-second GFP exposure and a defocused bright field image for cell identification and boundary determination, using Metamorph 7.0 software (Universal Imaging Corporation, Downingtown, PA). We used Cell-ID 1.0[42] to process images. We imposed the following gates to remove spuriously identified cells: Cells were removed if 1) the average pixel intensity of the edge of the cell was not darker than the inside of the cell in the brightfield image, 2) the fft statistic—an indicator of circularity—was greater than 0.5, or 3) cells were less than 200 pixels in

area, or 4) cells were greater than 600 pixels in area. We estimated nuclear fluorescence by calculating average pixel intensity in the area of 3-pixel radius around the brightest point. We estimated background fluorescence by calculating the average intensity of pixels immediately inside the cell's edge. Normalized nuclear fluorescence was calculated by subtracting the estimated background cellular fluorescence from the estimated nuclear fluorescence. Total cellular fluorescence was calculated for each cell by subtracting the average fluorescence of pixels surrounding the cell's boundary from the average fluorescence of pixels inside the cell's boundary.

Flow cytometry. An Epics XL-MCL Flow Cytometer Analyzer (Beckman Coulter, Brea, CA) was used to quantify total cellular fluorescence as in Figure 2.2a-d. Strains were grown overnight to saturation in complete synthetic media (CSM) lacking uracil, and were then back-diluted to OD 0.05 ("low-density" in Figure 2.2) or 0.2 ("high-density" in Figure 2.2) and grown for 8 hours before performing flow cytometry.

For cell sorting, a Influx Cell Sorter (BD Bioscience, San Jose, CA) was used to sort highly fluorescent cells and non-fluorescent cells. Resulting populations were then plated on Yeast Peptone Dextrose (YPD) media at a density ~ 100 cells/plate and were allowed to grow 2 days. Colonies were then replica plated onto CSM lacking uracil and the number of colonies that grew on this media was scored by hand.

Evaluating significance of feedback effects. For each of the 62 factors we tested for feedback in minimal media, we calculated the expression ratios between different untagged gene doses, and called the factors that showed consistent effects exceeding a given cutoff value as subject to feedback. Then we used the expression values from triplicate experiments to generate an error distribution—for each strain that we measured three replicate expression values for, we calculated the ratio between the measured expression values of a given replicate and the mean expression value of the three replicates. To estimate the number of genes that would be called as subject to feedback due to measurement error alone, we used the error distribution to generate mock expression values for 62 sets of strains, each set containing 3 different doses. The number of feedback effects observed over 100,000 permutations was averaged to generate a false discovery rate. The same procedure was used for experiments performed in rich media.

Figures

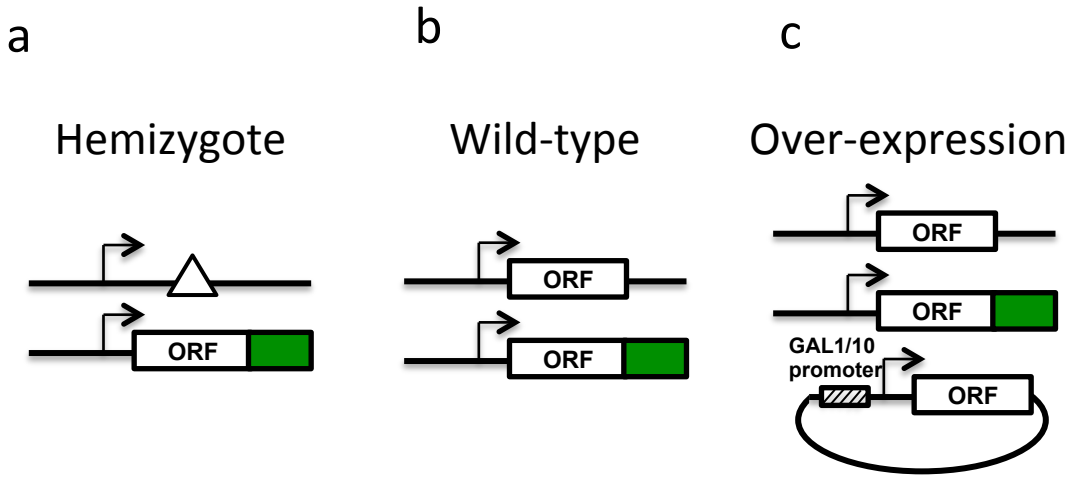

Figure 2.1. Design of assay for regulatory feedback. Each panel represents one yeast strain. Green rectangles indicate the coding sequence of Green Fluorescent Protein (GFP), and the oval in (c) represents a plasmid carrying an overexpression construct controlled by the GAL1-10 promoter. Δ represents a whole-gene deletion.

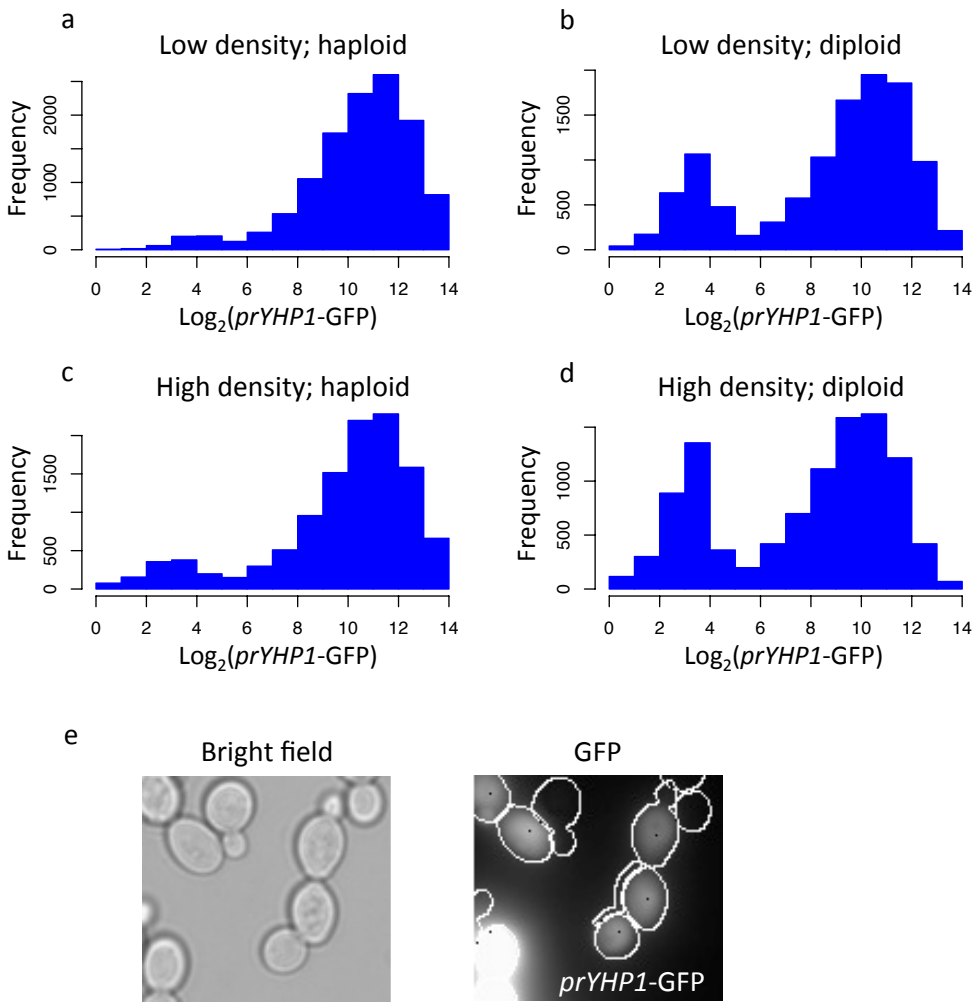

Figure 2.2 Promoter-GFP construct expression. (a-d) Each histogram displays a distribution of fluorescence intensities. Strains and growth conditions are as indicated. (e) Brightfield and fluorescence channels captured by fluorescence microscopy for the diploid *prYHP1*-GFP strain.

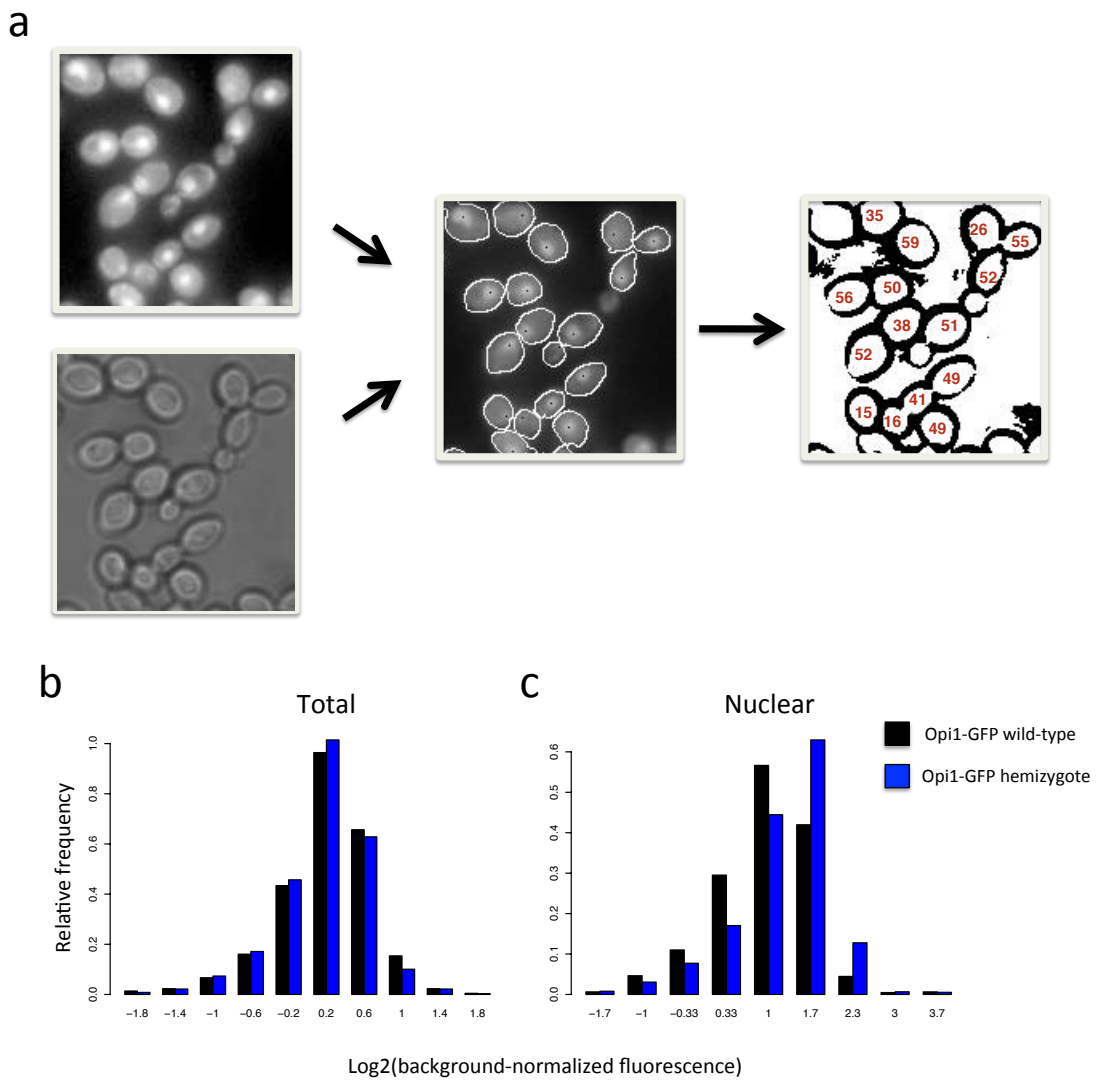

Figure 2.3 Quantifying Opi-GFP fluorescence. (a) Schematic illustrating Cell-ID process. A brightfield image is used to determine cell boundaries, which are transposed onto the fluorescence image (white ovals). Cell-ID generates nuclear expression values by subtracting the estimated background fluorescence intensity from the estimated nuclear intensity (see Methods for details). Quantities measured for 15 cells are in red text. (b) Histograms displaying distributions of total cellular fluorescence values for OPI1-GFP wild-type and hemizygote strains (see Methods for details). Expression levels are normalized by dividing by the median total cellular fluorescence of a strain without GFP. (c) As in b, except with nuclear fluorescence values.

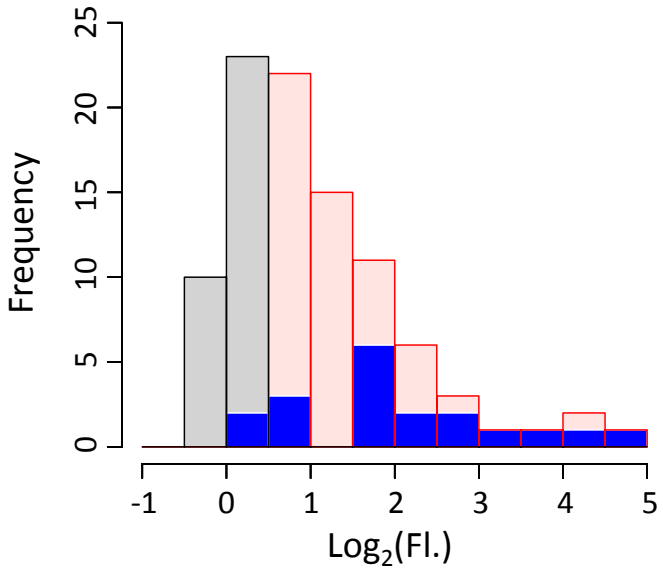

Figure 2.4 Distribution of fluorescence levels for all factors assayed in YNB. For each strain, the median expression value of ~500 cells was used to calculate the average expression across three strains of different untagged gene dose. This value was divided by the background fluorescence, and plotted on a $\log_2$-scale. The 62 strains exhibiting expression at least $2^{0.5}$-fold higher than background are represented in red, and the 33 strains exhibiting expression less than $2^{0.5}$-fold higher than background are represented in grey. The blue boxes indicate the distribution of the flow cytometry-detectable factors from Newman et al.[23], as quantified by our microscopy method.

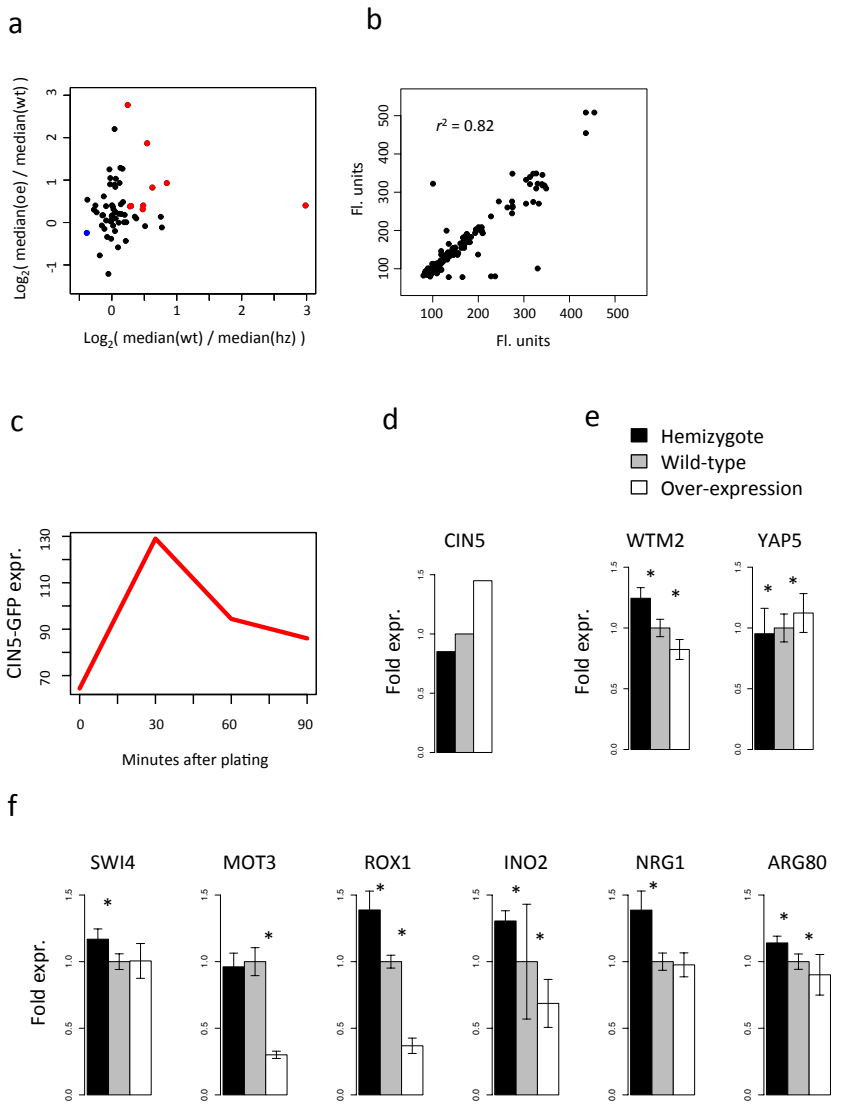

Figure 2.5 Assaying feedback in YNB. (a) The measured expression ratios between wild-type and hemizygote and overexpression and wild-type are plotted on the x and y axes, respectively. Each dot represents one of 62 strains. Red dots indicate negative feedback and the blue dot indicates positive feedback. (b) Correlation between replicate measurements. Expression values of each pair-wise combination of 3 replicates were plotted. r^2 value was calculated using the cor function in R. Axes are in arbitrary fluorescence units. (c) Cin5-GFP expression in cells transferred from rich media to minimal media. (d) CIN5-GFP strains illustrating reproducibility of CIN5 positive feedback effects. (e & f) Selected results from biological triplicate measurements. Expression effects for WTM2 and YAP5 strains exhibit negative and positive feedback respectively, though neither of these effects was detected in single replicate experiments. Genes in f exhibited feedback in single replicate experiments. Single replicate experiments were generally

recapitulated in triplicate experiments. Asterisks indicate effects that are significant at Wilcoxon $p < 0.1$.

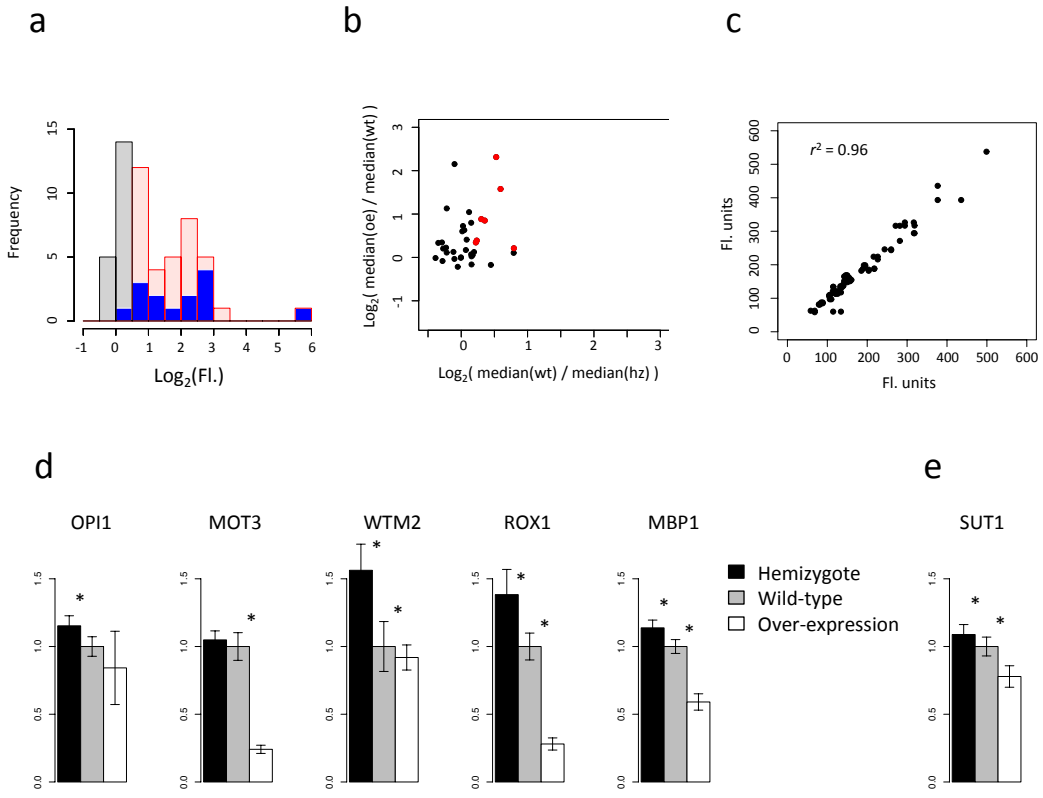

Figure 2.6 Assaying feedback in rich media. (a) Distribution of fluorescence levels for all factors tested in rich media. Color labeling as in 2.4. Here, there are 36 strains exhibiting expression at least $2^{0.5}$-fold higher than background, and 19 strains exhibiting expression less than $2^{0.5}$-fold higher than background. (b) Expression ratios between strains of differing gene dose as in 2.5a. (c) Plot illustrating replicate measurement correlations as in 2.5b. (d) Triplicate measurements of genes that showed feedback in single replicate measurements. (e) SUT1 exhibited negative feedback in biological triplicate experiments, though this effect was not present in single replicate experiments.

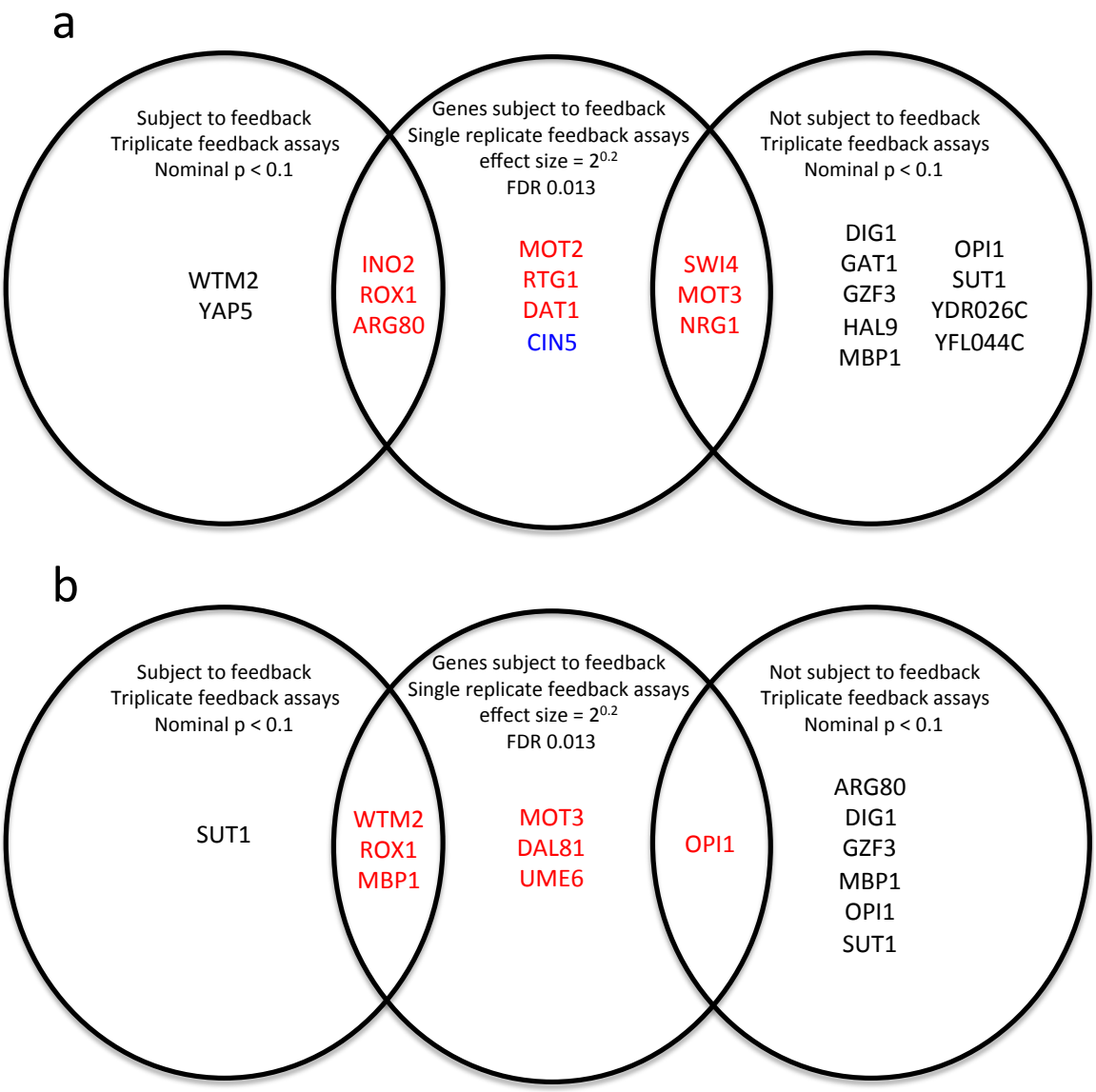

Figure 2.7 Overlap between single replicate and triplicate feedback assays. Genes in the center circles were determined to be subject to feedback by single biological replicate measurements performed in minimal media (a) and rich media (b). Genes in the left and right circles were classified as either exhibiting or not exhibiting feedback, based on biological triplicate experiments performed in the respective media.

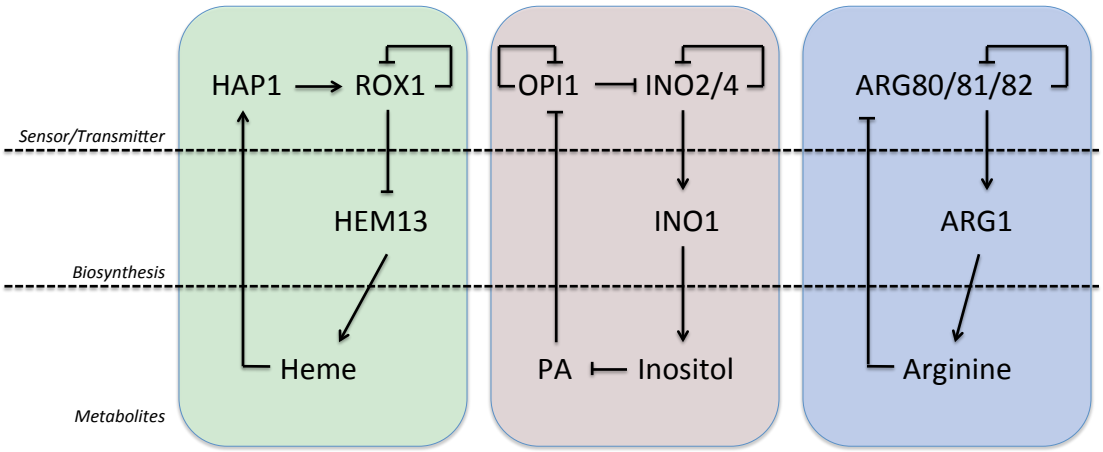

Figure 2.8 Feedback on factors transmitting information from metabolite sensors to biosynthesis pathways. Left: HAP1 encodes a transcription factor that directly senses/binds heme[43] and activates ROX1 in the presence of heme[44]. ROX1 represses HEM13[45], the gene encoding a rate-limiting step in heme biosynthesis[46]. Center: OPI1 encodes a transcriptional regulator that senses the signaling molecule phosphatidic acid. Phosphatidic acid on the endoplasmic reticulum binds the transcriptional repressor Opi1, thereby excluding it from the nucleus. When in the nucleus, Opi1 represses the INO2/4 complex. Ino2/4 is an activator of INO1, which codes for the rate-limiting enzyme in inositol synthesis[47]. Right: The arginine-repressed ArgR-Mcm1 complex, composed of Arg80, Arg81, Arg82, and Mcm1, activates[48] transcription of ARG1[49] and ARG3[50], which may be the rate-limiting steps in arginine biosynthesis.

Array 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	YNL139C	YER111C	YJL056C	YBL008W	YER040W	YDR049W	YHR006W	YDL048C	YPL202C	YML027W	YHL009C	YOR344C
B	YGL131C	YDR310C	YOR113W	YER169W	YGL071W	YHL027W	YJL110C	YBR083W	YHL020C	YGL209W	YFR034C	YEL009C
C	YML081W	YOL089C	YML099C	YKR099W	YPL089C	YKL185W	YBL054W	YMR070W	YGL181W	YKL043W	YFL044C	YGL237C
D	YGR097W	YKL015W	YDR213W	YDR520C	YHR084W	YKL032C	YDL170W	YOR229W	YDR423C	YPR065W	YDR123C	YIR018W
E	YJR127C	YBL005W	YDR207C	YMR164C	YMR075W	YER068W	YDR009W	YOR230W	YMR021C	YGR040W	YGL162W	YML113W
F	YDR216W	YIR023W	YLR228C	YKL222C	YIL101C	YDR463W	YGL035C	YPL049C	YER028C	YJR147W	YPL177C	YOL028C
G	YNL257C	YNL068C	YDL056W	YDR146C	YNL167C	YDR026C	YFL021W	YLR113W	YOL116W	YJR060W	YGL254W	YOR358W
H	YBR150C	YLR451W	YOR162C	YPR008W	YHR206W	YDL020C	YIL131C	YML051W	YBR267W	YDR451C	YOR028C	

Array 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	YMR182	YOR298C-A										
B	YDR253C	YKL020C										
C	YPL038W	YDR043C										
D	YIR017C											
E	YMR042W											
F	YOL067C											
G	YOL108C											
H	YBL021C											

Table 2.1 Array coordinates of genes tested for feedback. Black typeface indicates genes for which strains were successfully generated.

Detected in Newman et al.	Detected YNB	Detected Gal
YDR049W		
YPR008W		
YHL009C	YHL009C	
YHL020C	YHL020C	YHL020C
YMR070W	YMR070W	YMR070W
YKL032C	YKL032C	YKL032C
YOR229W	YOR229W	YOR229W
YPR065W	YPR065W	YPR065W
YER068W	YER068W	(not tested)
YOR230W	YOR230W	YOR230W
YML113W	YML113W	YML113W
YGL035C	YGL035C	YGL035C
YPL049C	YPL049C	YPL049C
YLR113W	YLR113W	YLR113W
YJR060W	YJR060W	YJR060W
YBR267W	YBR267W	(not tested)
YOR298C-A	YOR298C-A	(not tested)
YOL067C	YOL067C	YOL067C
YBL021C	YBL021C	YBL021C
	YIR018W	
	YDR207C	YDR207C
	YMR075W	YMR075W
	YML081W	
	YDR009W	(not tested)
	YOL089C	
	YGR040W	YGR040W
	YGL162W	YGL162W
	YKR099W	
	YDR216W	(not tested)
	YIR023W	YIR023W
	YLR228C	YLR228C
	YIL101C	YIL101C
	YDR463W	(not tested)
	YER169W	YER169W
	YFL044C	
	YJR147W	YJR147W
	YPL177C	YPL177C
	YOL028C	YOL028C
	YDL056W	YDL056W
	YNL167C	YNL167C

Table 2.2 continues on following page

YDR026C	
YFL021W	
YKL015W	
YOL116W	(not tested)
YDR213W	YDR213W
YGL254W	
YOR358W	YOR358W
YBR150C	
YOR162C	
YHR206W	
YML051W	YML051W
YJL110C	YJL110C
YDR451C	YDR451C
YOR028C	YOR028C
YMR182C	
YBR083W	
YDR043C	
YIR017C	
YMR042W	YMR042W
YOR113W	YOR113W
YOL108C	YOL108C
YDR123C	
YER111C	
YHR006W	YHR006W

Table 2.2 Genes detected by microscopy compared with previous study.

YNB (62 genes)				
Ratio cutoff (X in legend)	0.25	0.2	0.15	0.1
Genes above cutoff	INO2	INO2	INO2	INO2
	NRG1	NRG1	NRG1	NRG1
	ROX1	ROX1	ROX1	ROX1
	MOT2	MOT2	MOT2	MOT2
	RTG1	RTG1	RTG1	RTG1
	ARG80	ARG80	ARG80	ARG80
	DAT1	DAT1	DAT1	DAT1
	SWI4	SWI4	SWI4	SWI4
		CIN5	CIN5	CIN5
		MOT3	MOT3	MOT3
			HOG1	HOG1
			DIG1	DIG1
			OPI1	OPI1
			HAP5	HAP5
				OTU1
				GZF3
				YDR026C
				UME6
				YAP5
				MBP1
				YAP3
				RGM1
Number of genes called	8	10	14	22
Expected by chance	0	1	2	6
FDR	0.005	0.013	0.031	0.09

Table 2.3 Feedback results for single replicate measurements in YNB media. All genes for which expression ratios between both hemizygote and wild-type and wild-type and over-expression exceeded 2^X show evidence for negative feedback and are in red. Genes for which expression ratios are both less than 2^{-X} show evidence for positive feedback and are in blue.

Galactose (36 genes)				
Ratio cutoff	0.25	0.2	0.15	0.1
Genes above cutoff	MOT3	MOT3	MOT3	MOT3
	ROX1	ROX1	ROX1	ROX1
	MBP1	MBP1	MBP1	MBP1
	OPI1	OPI1	OPI1	OPI1
		WTM2	WTM2	WTM2
		DAL81	DAL81	DAL81
		UME6	UME6	UME6
			KSS1	KSS1
				AZF1
				HAP5
				IXR1
Number of genes called	4	7	8	11
Expected by chance	0	0	1	3
FDR	0.0003	0.012	0.032	0.074

Table 2.4 Feedback results in rich media.

Gene (YNB)	Previously reported	Mechanism	Means of inference
INO2	Y	Direct	ChIP-chip[26]
NRG1	Y	Direct	ChIP-chip[26]
ROX1	Y	Direct	Feedback measured directly[14] and by ChIP-chip[26]
MOT2	N		
RTG1	N		
ARG80	N		
DAT1	N		
SWI4	Y	Direct	Feedback measured directly[15] and by ChIP-chip[26]
CIN5	Y	Direct	ChIP-chip[26]
MOT3	N	Direct	Data not shown
WTM2	N		
YAP5	N		
(Galactose)			
MOT3	Y	Direct	Data not shown
ROX1	Y	Direct	Feedback measured directly[14] and by ChIP-chip[26]
MBP1	Y	Indirect	Inferred from ChIP-chip[26]: MBF complex (Mbp1/Swi6)[51] member Swi6 binds downstream intergenic region
OPI1	Y	Indirect	Feedback measured directly[17], mechanism inferred from ChIP-chip[26]: OPI1/INO2/INO4 complex member INO2 binds upstream intergenic region
WTM2	N		
DAL81	N		
UME6	N		
SUT1	Y	Direct	Chip-chip[26]

Table 2.5 Mechanism of feedback for genes called for feedback. Genes in red/blue showed evidence for negative/positive feedback at cutoff of $2^{0.2}/2^{-0.2}$ or were identified as subject to feedback in biological triplicate experiments. Feedback was inferred from ChIP, literature, or a combination of the two.

Plates	Names	Description	Source
YCDP13		BY4741 + ORF-GFP::HIS3 (Array 1)	Invitrogen
YCDP14		BY4741 + ORF-GFP::HIS3 (Array 2)	Invitrogen
YCDP15		YCDP13 transformed with pBY011 (Array 1)	This study
YCDP16		YCDP13 transformed with overexpression constructs (Array 1)	This study
YCDP17		YCDP14 transformed with pBY011 or overexpression constructs (Array 2)	This study
YCDP21		BY4742 ORF::Kan (Array 1)	Invitrogen
YCDP22		BY4742 ORF::Kan (Array 2)	Invitrogen
YCDP23	"Wild-type" strains	YCDP15 x BY4742 (Array 1)	This study
YCDP24	"Overexpression" strains	YCDP16 x BY4742 (Array 1)	This study
YCDP25	"Wild-type" and "Overexpression" strains	YCDP17 x BY4742 (Array 2)	This study
YCDP26	"Hemizygote" strains	YCDP15 x YCDP21 (Array 1)	This study
YCDP26	"Hemizygote" strains	YCDP17 x YCDP22 (Array 2)	This study

Table 2.6 Yeast strain plates generated in this chapter. All plates are ordered by the coordinates listed in Table 2.1. Note: pBY011 is the parental plasmid for the overexpression constructs. It contains the same prototrophy marker, but no over-expression gene.

Chapter 3

Negative feedback confers mutational robustness in yeast transcription factor regulation

Introduction

Robustness of organismal function in the face of perturbations is critical for fitness. Since the seminal work of Waddington[1], biologists have remarked on the stability of phenotypes against environmental and genetic variation, and understanding how organisms achieve robustness remains one of the major challenges in systems biology[2-4]. Much of the search for molecular mechanisms of robustness has focused on gene regulation. Characteristics of regulatory networks that confer robustness include pathway redundancy and master regulatory organization[5], phenotypic capacitors[6-8], paired activating and inhibiting input, and cooperative and feed-forward regulation[9]. Additionally, negative regulatory feedback, in which a biomolecule represses its own abundance, can buffer variation in gene expression[10,11], and negative feedback loops have been shown to underlie robustness to variable environmental conditions and to stochastic intracellular change[12-14]. Negative feedback may also confer network stability against the effects of mutations[3,15], but evidence for negative feedback as a driver of mutational robustness in vivo has been at a premium[16]; the relevance of this principle to natural genetic variation remains largely unknown.

In this work, we focused on negative feedback in yeast hypoxia regulation, motivated by the extensive evidence for feedback in oxygen response pathways across biology[17]. We characterized the feedback loop at the yeast hypoxia regulator ROX1 in molecular detail, and we harnessed this system as a test bed to study how feedback confers stability against naturally occurring mutations. Given the precedent for negative feedback as a determinant of quantitative behaviors of inducible circuits[18-22], we also investigated the role of Rox1 feedback in expression regulation during oxygen response.

Results

We set out to establish a tractable model system for the study of feedback and robustness using yeast transcription factors. For this purpose, we first screened transcription factor genes for feedback on protein abundance. We used fluorescence microscopy[23] to measure protein expression from a single genomic copy of each factor tagged with GFP in a diploid strain[24], while varying levels of an untagged copy of the factor (Figure 3.1a). To maximize signal to noise, we analyzed the 23 most highly abundant factors in rich medium (Table 3.1), and found evidence for negative feedback in four cases (Figure 3.1b). As expected[24,25], these screen hits included Rox1 and Swi4; we also found as yet uncharacterized feedback loops by Mbp1 and Mot3. Independent replicate experiments confirmed the ability of each factor to repress its own abundance (Figure 3.1c).

To investigate the role of feedback in mutational robustness and systems-level network behaviors, we focused on the transcription factor Rox1. This master regulator is repressed in hypoxic conditions and induced under normoxia to regulate biosynthetic pathways that use molecular oxygen as a substrate[26]. Anticipating that Rox1 would act directly at the ROX1 promoter[24], we identified four candidate Rox1 binding sites in the 500 bp upstream of the ROX1 coding start site (Figure 3.2). Mutagenesis confirmed the role of these sites in ROX1 feedback, with each site contributing incrementally to the strength of auto-regulation in a *ROX1* transcriptional reporter (Figure 3.3). Promoter response to changes in dose of ROX1 was markedly reduced when all four sites were mutated in combination, indicating a near-complete abrogation of feedback (Figure 3.3). We used these mutations to engineer a version of *ROX1* in which the feedback-mutant promoter drove expression of Rox1 fused to GFP; to avoid potentially confounding effects from elevated Rox1-GFP steady-state levels in the presence of the feedback mutations, we manipulated the usage of optimal codons[27] in the ROX1-GFP coding region. The suboptimized sequence, in conjunction with mutated Rox1 binding sites in the ROX1 promoter, gave rise to steady-state expression levels comparable to those of the wild-type (Figure 3.4). In what follows, we refer to this version of ROX1 as the suboptimized feedback mutant; a strain harboring this gene grew indistinguishably from the wild-type across environmental conditions (Figure 3.5).

To analyze the role of feedback during *ROX1* induction, we grew wild-type and feedback-mutant strains under hypoxic conditions and measured Rox1-GFP expression upon oxygenation. For the feedback-mutant promoter driving expression of a wild-type reporter, levels during induction far overshot the steady state of the wild-type (Figure 3.6). Such elevated expression was toxic to cells, as this strain displayed growth defects under a variety of environmental conditions (Figure 3.5). As expected, in the suboptimized feedback mutant, protein levels reached those of the wild-type at steady state but with slower kinetics, owing to the reduced translation efficiency of the reporter construct (Figure 3.6). We conclude that the ROX1 locus harbors regulatory information encoding strong activation

upon oxygenation. In wild-type cells, repression by Rox1 serves as a brake on this induction signal, avoiding the deleterious effects of elevated expression. Thus, the feedback loop tunes the kinetics of ROX1 induction, enabling a rapid approach to a moderate level of steady-state expression during the transition from hypoxia to normoxia.

We next sought to evaluate Rox1 feedback as a mechanism for robustness of gene expression to naturally occurring genetic variation. For this purpose, we developed an assay to interrogate the effects on *ROX1* expression of the spectrum of variants present in a set of divergent yeast strains of environmental and laboratory origin. For each such tester strain, we crossed it to a laboratory strain bearing the wild-type ROX1-GFP reporter, and we performed an analogous cross using the suboptimized feedback mutant. Haploid recombinant progeny, each a mosaic of inheritance from the tester parent and the laboratory parent, served as a panel of genetically distinct strains among which Rox1 expression could vary. For a given cross, we measured the median Rox1-GFP levels in a culture of each progeny strain. The results, shown in Figure 3.7, revealed deviation in Rox1-GFP abundance of up to eight-fold across strain cultures, reflecting the impact of naturally occurring genetic variation on Rox1 expression. Eliminating feedback compromised robustness to these variants, with a wide spread of median Rox1-GFP levels across genetic backgrounds in feedback-mutant strains relative to wild-type; the coefficient of variation across strains was 2 to 5-fold higher in the presence of the feedback mutation (Figure 3.7). Control experiments ruled out codon suboptimization as a predominant cause of this effect (Figure 3.8). Interestingly, the extent of expression variation across recombinant progeny was a function of the tester strain parent, indicating that some testers harbored alleles with more dramatic consequences for Rox1-GFP expression than others (Figure 3.7). We conclude that Rox1 auto-repression buffers the effects of natural genetic variation on ROX1 expression, establishing regulatory feedback as a determinant of mutational robustness in this system.

We hypothesized that the ability to buffer gene expression could be a driver of the prevalence of feedback circuits across yeast transcription factors. In particular, we reasoned that evolutionary pressures for buffering expression levels, and thus for feedback, would be strongest among factors for which deviations from homeostatic expression levels gives rise to fitness defects. In the case of Rox1, both increases and decreases in dosage were toxic to yeast cells (Figure 3.5). To test the relationship between expression homeostasis and feedback more generally, we used growth rates of yeast strains manipulated to overexpress each transcription factor in turn[28]. We first integrated these data with the results of our reporter-based screen for feedback among transcription factors (Figure 3.1). Conforming to our model, upon overexpression, factors subject to feedback conferred a fitness defect ~70% more severe than factors with no evidence for feedback (Wilcoxon $p = 0.02$; Figure 3.9a). We next sought to expand our analysis beyond the transcription factors detectable in our experimental screen. For this purpose, we used a bioinformatic approach, described in Materials and Methods, to predict instances of

direct transcriptional feedback by transcription factors based on the presence of their binding sites in promoters of their own encoding genes. Even using this unvalidated set of feedback inferences, evidence for auto-regulation was again associated with overexpression toxicity, albeit to a less dramatic extent: toxicity effects were 13% more severe for factors with inferred feedback relative to the remainder of the set (Wilcoxon $p = 0.04$; Figure 3.9b). To address the possibility that computationally predicted binding sites were better specified for factors with stronger overexpression toxicity, we tested the relationship between overexpression growth rate for a factor and the number of its predicted targets genome-wide and found no effect (regression $p = 0.2$). Taken together, our results highlight feedback by transcription factors as a correlate of the toxic effects of overexpression, lending credence to the notion that many such feedback loops function in vivo as a control against misregulation.

Discussion

Landmark studies have identified negative feedback loops that tune the quantitative properties of gene circuits[18-22,29]. Negative feedback has also been implicated in the robustness of gene networks to environmental and stochastic change[12-14], but the role of native feedback circuits as buffers against natural genetic variation has remained unknown. Addressing the question requires detailed molecular analysis of feedback regulation and its impact across genetically heterogeneous populations, for which we have established an experimental paradigm using yeast Rox1 as a model system.

Rox1 regulates the expression of genes involved in oxygen-dependent sterol-biogenesis and respiratory pathways. We have shown that cell growth is remarkably sensitive to changes in this activity even in normoxic conditions, with overexpression and deletion of Rox1 each conferring distinct defects. Consistent with a requirement for tight regulatory control of Rox1, our analysis has revealed two related ways in which Rox1 feedback limits deviations from steady-state expression optima. In a constant genetic background, negative feedback enables rapid ROX1 activation during oxygen exposure, minimizing the time spent in intermediate expression states and avoiding toxic effects of overexpression in a manner that dovetails with similar roles for auto-repression in other networks[10,30]. Additionally, across genetic backgrounds, Rox1 feedback serves as a buffer against the perturbations arising from naturally occurring sequence changes. A primary implication of our findings is thus that the Rox1 negative feedback circuit is tuned to respond quantitatively to subtle up- and down-regulating effects of natural genetic variation as well as to dramatic shifts in regulatory input when conditions change. In each case, *trans*-acting input which upregulates *ROX1* would be counterbalanced by increased Rox1 occupancy and repression at its own encoding locus, and input which represses *ROX1* would be counterbalanced by a reduction in Rox1 occupancy at its own gene.

Will regulatory feedback prove to underlie mutational robustness as a general mechanism across biology? Pathway-level feedback is common in yeast: in many cases, the network can detect the perturbation of an artificially introduced genetic lesion and upregulate functionally related genes to compensate[16]. Whether gene-level feedback will be of similar importance for robustness on a genomic scale depends in part on the prevalence of auto-regulation in gene networks. Most highly expressed genes in yeast are not subject to complete dosage compensation when mutated[31], but feedback at the gene level may be particularly common among transcription factors[32,33]. In light of our evidence that feedback may serve to constrain the deleterious effects of misregulation among transcription factors, a model invoking especially strong pressures for such feedback control would be consistent with the extreme overexpression toxicity observed across transcription factors in yeast[28]. The emerging picture is one in which both gene- and pathway-level feedback, at transcription factor genes and elsewhere in the yeast network,

may be key elements of the architecture that mediates buffering of genetic change. Understanding which gene circuits are buffered, and how, will be of critical interest in the effort to interpret the extent and phenotypic penetrance of regulatory variation[34-37], and the evolution of robustness mechanisms[15,38]. Addressing these questions, and application of the emergent principles to bioengineering[39] and human disease treatment[40-42], will serve as continued motivation for the study of feedback and robustness in regulatory circuits.

Has feedback at ROX1 likely been selected for robustness to natural genetic variation or by another selective pressure? Previous work suggests that the circumstances in which mutational robustness is a direct target of selection are limited[43]. In particular, evolution of mutational robustness is constrained to organisms with high mutation rates and large population sizes. Recent work suggests that outcrossing in *S. cerevisiae* occurs only once every 50,000 generations[44], and estimates of heterozygosity in wild populations have confirmed that outcrossing is rare compared to inbreeding and asexual reproduction[45]. Both these observations argue that *S. cerevisiae* has a small effective population size, and it is therefore unlikely that mutational robustness is under direct selection. However, these observations do not preclude the possibility that feedback plays an important role in shaping genome evolution.

Materials and Methods

Yeast strain construction. For each of 67 yeast transcription factors (Table 3.1), strains bearing one copy of a GFP protein fusion of the respective factor gene[46] (Invitrogen) were mated to strains deleted for the respective gene [47](Invitrogen) and imaged by microscopy as described below. For the 23 factors of highest abundance (nuclear fluorescence >30000 in arbitrary units; Table 3.1), GFP-tagged haploid strains were then mated to strains bearing URA3-marked plasmids encoding each open reading frame under the GAL1,10 promoter[48] for the complete screen as shown in Figure 3.1. All plasmid transformations and yeast cell matings were performed by standard techniques[49].

ROX1 knockout strains were generated by replacing the endogenous *ROX1* gene and the upstream 467 bp with a *URA3* cassette[50] or kanamycin resisance cassette[51]. To generate constructs for transcriptional reporters, a portion of the *ROX1* promoter was amplified from yeast genomic DNA and fused to *CFP* (Addgene) flanked by a kanamycin resistance cassette[51] *via* ligation-independent cloning (LIC)[52]. To generate constructs for protein-fusion reporters with wild-type coding sequences, the *ROX1* promoter and coding sequence were amplified from yeast genomic DNA and fused to *GFP* (Addgene) flanked by a kanamycin resistance cassette[51] *via* LIC. To generate constructs for protein-fusion reporters with suboptimized codons, we first generated each coding sequence separately; sequences are provided in Figure 3.10. A suboptimized *ROX1* sequence was synthesized and incorporated into a cloning vector (GeneOracle). We also generated by PCR a fusion of commercial oligonucleotides (IDT) corresponding to suboptimized *GFP* regions with overlapping ends and ends containing LIC sites; the fused product was introduced into a LIC vector. For reporter fusions bearing wild-type *ROX1* and suboptimized *GFP*, we then used primers containing LIC sites to amplify two regions with overlapping ends, one containing the wild-type *ROX1* promoter and *ROX1* coding sequence, and the other containing suboptimized *GFP*; we generated a fusion of these regions by PCR and introduced the resulting construct into a LIC vector. For reporter fusions bearing suboptimized *ROX1* and suboptimized *GFP*, we used primers containing LIC sites to amplify three regions with overlapping ends: one containing the wild-type *ROX1* promoter, a second containing the *ROX1* coding sequence, and a third containing suboptimized *GFP*. We fused these regions into one combined fragment and cloned as above. Yeast strains bearing each reporter were ultimately generated by transforming ROX1::URA3 strains with PCR-amplified construct products. Integrants were selected by growing on YPD media supplemented with 300μg/ml G418 (Cellgro).

To generate feedback mutant strains, Rox1 binding sites in the *ROX1* promoter were identified as perfect matches to the Rox1 consensus sequence[53]. Site-directed mutagenesis of these sites on cloned promoter constructs was performed with QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). Sanger sequencing

confirmed DNA sequences of the resulting constructs and ROX1::URA3 strains were transformed with PCR products and selected for integration as above.

ROX1::CFP strains were generated in the S288c background. The background of *ROX1-GFP* strains was as follows. The W303 derivative JRY2334[54] was mated to a *HEM1D* DY150 derivative[55], kind gifts from J. Rine and J. Kaplan, respectively, and a haploid recombinant wild-type for *HEM1* was isolated from the progeny. This recombinant was then crossed to BY4742 (Open Biosystems), and a haploid recombinant bearing the W303 allele of *HAP1* rather than the S288C allele[56] was isolated.

Analysis of genetic variation. Crosses for analysis of genetic variation were generated by mating ROX1-GFP strains with BY4741 (Open Biosystems); SK1[57]; YPS606[57]; and 22:3:b[58]. In the case of YPS606 and SK1, HO was replaced with *URA3* by cloning and transformation as above. For each cross, hybrid diploids were sporulated on solid minimal sporulation medium[49], haploids were isolated by spore enrichment[59], and 60-90 GFP+ recombinants were selected by growth on YPD medium[49] supplemented with G418. To compare the variances of the feedback mutant and wild-type distributions for a given cross, we first calculated the *F* statistic for differential variance using the var.test function in R (www.r-project.org). We then permuted mutant and wild-type strain expression values 10,000 times, repeated the *F* calculation for each permuted data set, and evaluated the statistic from the real data against this null distribution to yield a one-sided empirical *p*-value.

Feedback assays. For the feedback screen, one culture of each yeast strain was inoculated into CSM –uracil media (MP Biomedicals) with 2% raffinose and grown in 2ml 96-well plates at 30°C with shaking to saturation. Strains were back-diluted into CSM –uracil media with 2% galactose to an optical density of 0.1. Strains were imaged after ~6 hours of growth, having reached optical density 0.5-1.0. Upon preparation of 96-well plates (see below), 100 µl of cell suspension was added to each well and allowed to settle and bind for 30 minutes before imaging as described below. In the screen, only transcription factors with putative feedback effects consistent between overexpression and deletion assays were considered for further study (Figure 3.11). For confirmation of screen hits in Figure 3.1c, three independent cultures of each strain were grown and prepared for imaging as above.

Hypoxia time course. For induction experiments, a culture of each yeast strain was grown in CSM with 2% galactose, back-diluted into deoxygenated media at OD 0.002, and grown for 40 hours at 30°C in a 16 x 125mm Hungate tube with septum stopper and screw cap (Bellco). From each culture, a 10-mL aliquot was spun down and cells were resuspended in 200 µL oxygenated media; 100µl was added to concanavalin-treated glass bottom plate wells and allowed to settle for 1-5 minutes, before imaging as described below, at timepoints as indicated in Figure 3.6. To deoxygenate media, 10mL was added to a stoppered tube and boiled for 5 minutes

with a syringe through the stopper. After boiling, the syringe was removed, and each tube of media was cooled to < 50°C and purged with N_2.

Fluorescence microscopy. 96-well glass bottom plates (Falcon) were coated with 100 µl of a 100 µg/ml solution of concanavalin A type V (Sigma-Aldrich), incubated 1 hour at room temperature and then washed three times with water. Wells were then washed and resuspended with media immediately before imaging. For imaging, we used a 60X PlanApo objective (N.A.=1.4) under oil immersion in a Nikon TE2000 inverted microscope equipped with a mercury lamp, a motorized stage and 512BFT MicroMax cooled CCD camera (Photometrix, Tucson, AZ). We imaged 3-6 observation fields per strain, with each field containing ~100 cells. For each field we acquired a 0.5-second GFP exposure and a defocused bright field image for cell identification and boundary determination, using Metamorph 7.0 software (Universal Imaging Corporation, Downingtown, PA). Using Cell-ID 1.0[23], for each cell, we estimated nuclear fluorescence by calculating average pixel intensity in the area of 3-pixel radius around the brightest point. We estimated cytoplasmic fluorescence by calculating the average intensity of pixels immediately inside the cell's edge. We then calculated normalized nuclear fluorescence by subtracting the estimated cytoplasmic fluorescence from the estimated nuclear fluorescence.

Flow cytometry. Three independent cultures of each strain were grown in CSM with 2% galactose media to saturation, and were back-diluted to OD 0.1 6-8 hours prior to measurement. Fluorescence measurements of ~10,000 cells/sample were acquired with an LSRFortesa (BD Biosciences) and analyzed in R with the flowCore package (www.bioconductor.org). A given strain was eliminated from analysis if the majority of side scatter values across cells in its culture sample did not fall within a range typical of a non-flocculent lab strain growing at log phase. The median background fluorescence from cells of a strain bearing no GFP gene was subtracted from the median across cells of each culture of interest.

Growth condition screen. Strains were grown overnight in CSM with 2% glucose at 30°C. 1:10 serial dilutions were performed in microtiter plates. Dilutions were transferred to plates using a multipronged inoculating device (frogger). Plate media were CSM with one of the following added: 2% glucose, 2% galactose, 3% glycerol, or 3.2% ethanol; 2% glucose with 20 µg/mL fluoconazole (Sigma-Aldrich); or YPD[49].

Bioinformatic predictions of feedback. Sequences corresponding to the 1000 base-pairs upstream of coding start for each of 5922 yeast genes were downloaded from[60]. Position weight matrices to score binding preferences for each yeast transcription factor were downloaded from http://fraenkel.mit.edu/Harbison/release_v24/final_set/Final_InTableS2_v24.motifs[61]. We used MCAST (http://meme.sdsc.edu/meme) to predict binding to each upstream sequence for each factor, and we retained for analysis all matrix matches with $-\log2(p/0.0005) > 3.45$, where the p-value estimates the significance of predicted binding, calculated with respect to a background model generated from the complete set of nucleotide frequencies in all yeast upstream sequences[62].

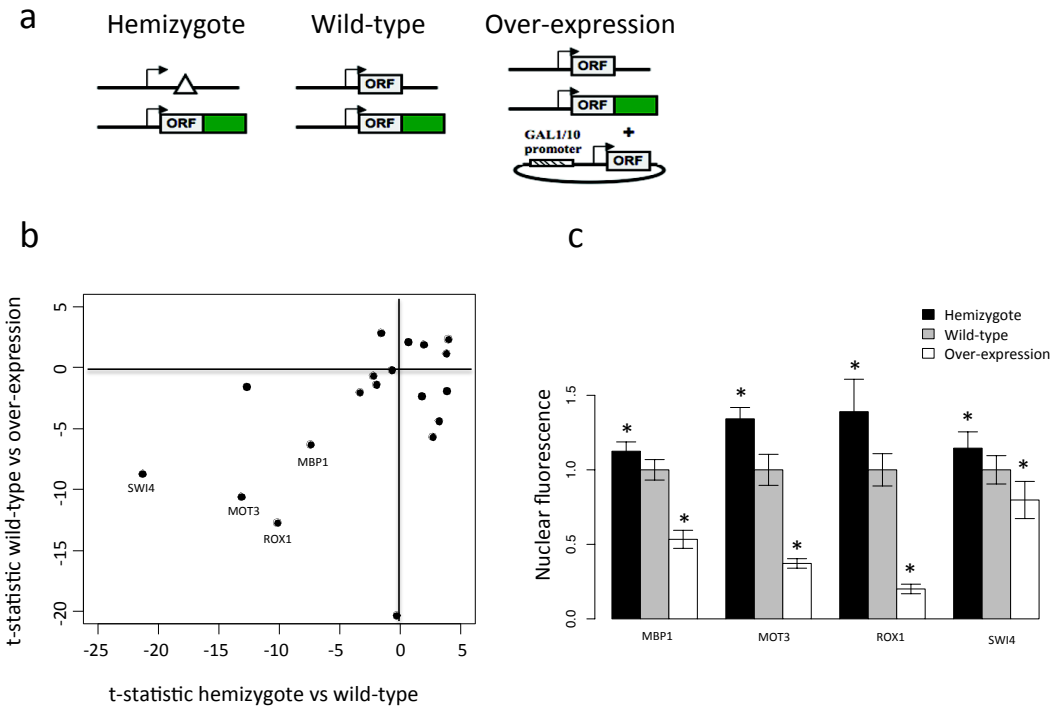

Figure 3.1 Screening yeast transcription factors for regulatory feedback. (a) Each panel represents one schematized yeast strain. Green rectangles indicate the coding sequence of Green Fluorescent Protein (GFP), and the oval at right represents a plasmid carrying an overexpression construct controlled by the GAL1-10 promoter[48]. Δ, whole-gene deletion. (b) Each data point represents the results of two analyses of nuclear abundance of a transcription factor fused to GFP[46], measured via quantitative microscopy in a diploid strain. Each analysis, as indicated on the x or y axis, compared fluorescence from a given tagged factor in two strains encoding different doses of the untagged version of the factor, with strains named according to the schematic in (a). (c) Each set of bars reports nuclear fluorescence measurements of the indicated factor as a fusion with GFP, measured via quantitative microscopy and normalized with respect to wild-type levels. Each bar reports measurements from one strain, with names as in (a); error bars represent the standard deviation over biological replicates and microscope fields. Asterisks represent comparisons relative to wild-type which are significant at Wilcoxon $p < 0.001$.

GGAGAGCTCT TTAATTAAGC GGCCGCCCTG CAGGACTCGA GTTCTAGAAA TAATTTTGTT

TAACTTTAAG AAGGAGATAT AGATCTTGGC CTGTCGAGGT ATCGGCCGCG TGGAACTACC

Site 1

GGGAATTACT ATGCAAAACA ATTGGAAATC TGGTAGGAAA ACCTTGTTCT AGAACTTGGC

Site 2

GATTGCTGAC AAAGAAGAAA AGGGCCTATT GTTGCTGCCT CTTTTGTTGT TCTTCCTCGT

Site 3

ATTGTCTTGC CGGTGTTCTT TGTGTCTTTT GTGTGTAGGT TCTTACTATT ATAGTGCTCT

TTGCTATTAT ATTTTCTTCG TTTTCACTTT GCGTAATGTA ACGGTCTTAA ACAAAGTTTT

Site 4

TTTTTTTTCG CTCTTGCATT TTCCTTTTCT GCTCTATCTT ATTTGCTAAT TGTAGTTTCA

GAAGTTTTAC TTAAATATAG CACTATTTTC CAGTTTTAAT GTTTCTTCTC ATTGCTTTCT

TTTATAATTT TCGCATATAA TTATACATTT ACGGTGTCTT AACTCTCCCT CTTCACCCCT

CATTATTCCA GAAAATACTA ATACTTCTTC ACACAAAAGA ACGCAGTTAG ACAATCAACA

ATG

Figure 3.2 Predicted binding sites for the Rox1 protein in the *ROX1* promoter. Shown is the yeast genome sequence between the *ROX1* start codon (bold) and 600 base pairs upstream; predicted Rox1 binding sites are in yellow.

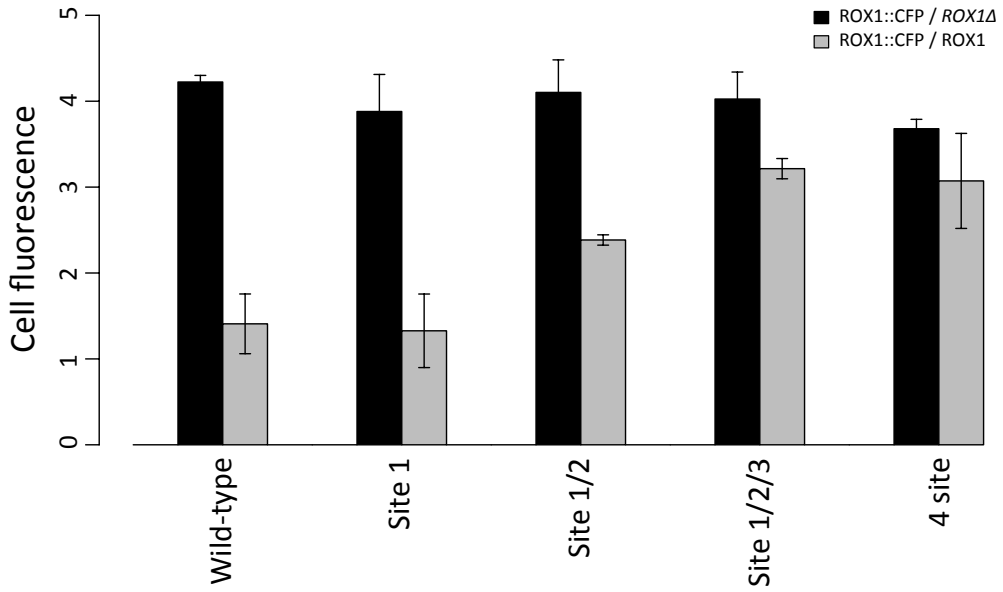

Figure 3.3. Transcription factor binding sites in the ROX1 promoter are required for transcriptional feedback. Each set of bars reports expression from one ROX1 transcriptional reporter in a diploid yeast strain, measured by flow cytometry. Shading represents the presence or absence of a single untagged copy of ROX1 in the strain background. Each set of bars labeled with "Site" corresponds to a reporter with the indicated Rox1 binding sites (Figure 3.2) mutagenized in the ROX1 promoter; "4site" indicates mutagenesis of all four sites. CFP, cyan fluorescent protein. Δ, whole-gene deletion of ROX1.

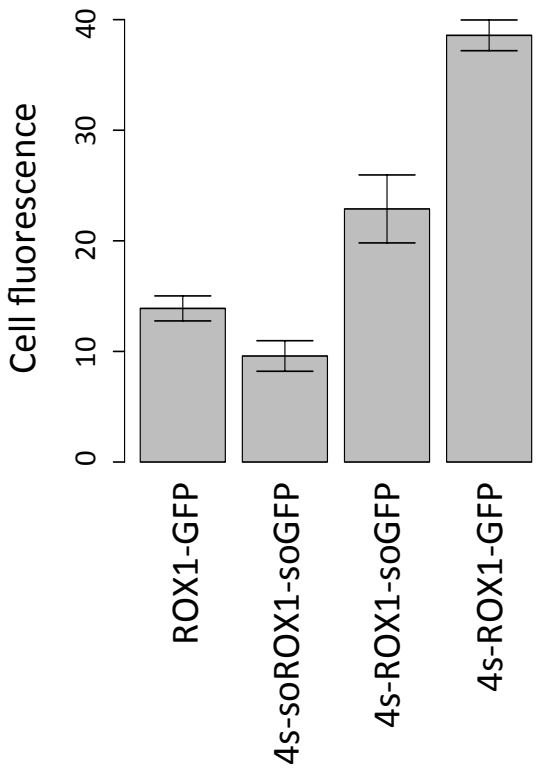

Figure 3.4 Suboptimized *ROX1* expression reporters. Each bar reports expression from a *ROX1-GFP* fusion reporter in a haploid yeast strain. Each bar labeled with "so" corresponds to a reporter construct with the indicated sequence component encoded with suboptimized codons; bars labeled with "4s" correspond to reporters where the four Rox1 binding sites in the *ROX1* promoter are mutagenized as in Figure 3.3. "4s-soROX1-soGFP" corresponds to the suboptimized feedback mutant in Figures 3.6 and 3.7, and "4s-ROX1-GFP" corresponds to the feedback mutant with wild-type *ROX1* and *GFP* sequences in Figure 3.6.

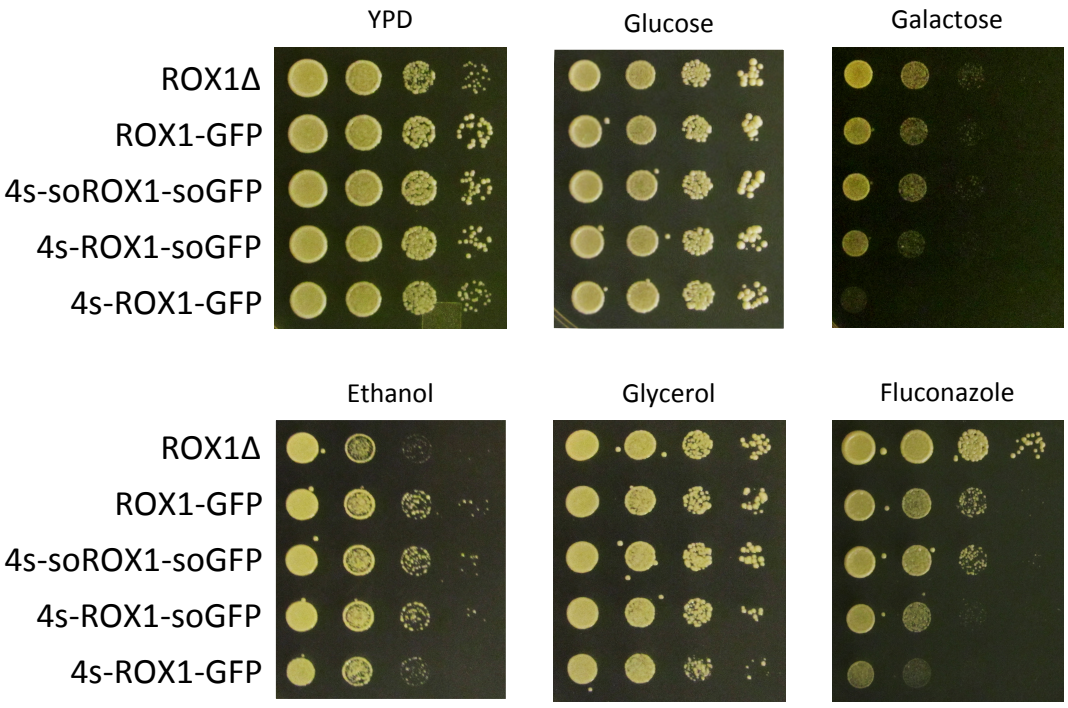

Figure 3.5 Perturbing *ROX1* expression affects yeast growth. Each panel reports a comparison of growth across yeast strains in one environmental medium. Each row represents a haploid strain with a *ROX1-GFP* fusion bearing the indicated modifications, grown from inocula of varying densities; naming is as in Figure 3.4. D, whole-gene deletion.

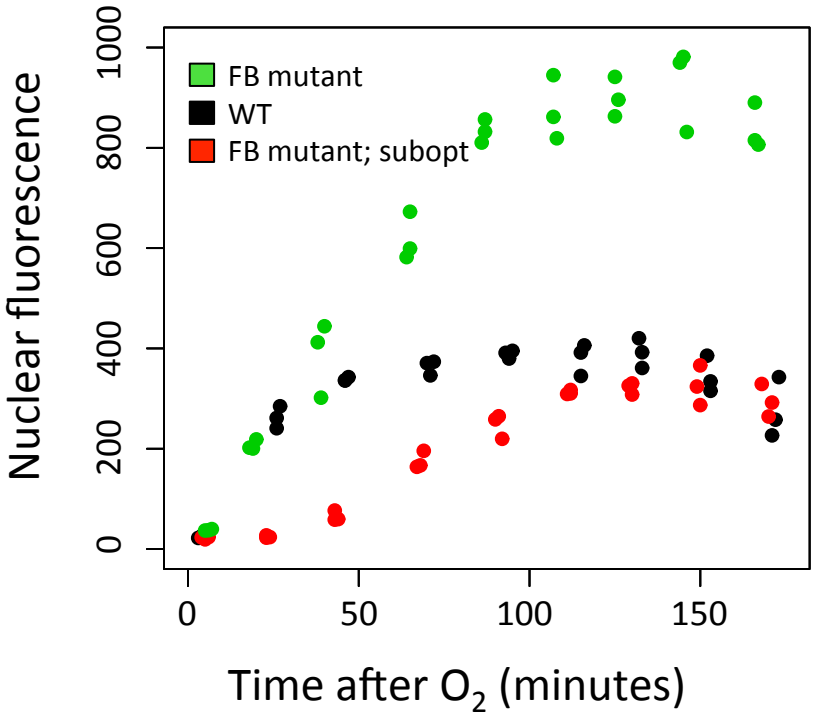

Figure 3.6 The ROX1 locus confers strong induction during oxygen exposure. Each set of points reports fluorescence of a haploid yeast strain bearing a ROX1-GFP fusion reporter gene, measured by quantitative microscopy after a transfer of a culture from hypoxia to normoxia. Each data point represents median nuclear fluorescence across cells of one culture at the indicated time after oxygenation. Each color represents one reporter: "FB mutant" indicates a feedback mutant ROX1 promoter with all four Rox1 binding sites mutagenized, and "subopt" indicates that the ROX1 and GFP sequences were encoded with suboptimized codons. WT, wild-type.

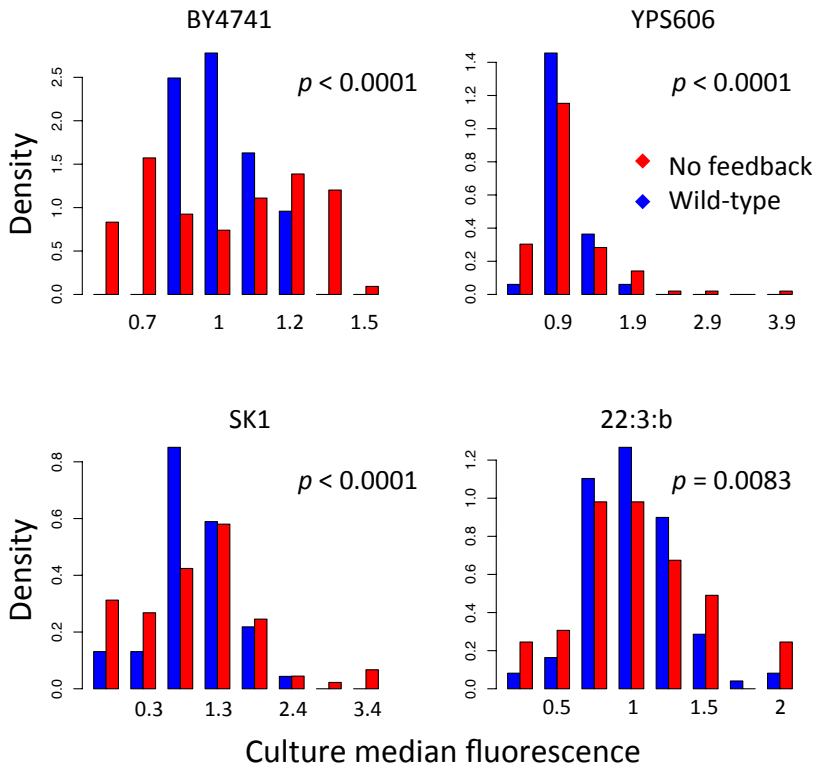

Figure 3.7. Rox1 feedback as a mechanism for mutational robustness. Each distribution in each panel shows ROX1 expression levels in haploid segregants from crosses between a Rox1-GFP strain and a tester strain: 60-90 recombinant strains from each cross were grown to log phase, flow cytometry was performed to determine expression levels for ~30,000 cells in each culture, and the median expression level of each culture was calculated. Each distribution is a probability density function of normalized median expression levels for 60-90 strains. We used a probability density function—where the y-axis is rescaled so the total area under the curve is equal to 1—instead of a frequency distribution for the purpose of comparing two distributions with different sample number. In other words, if the two distributions had the exact same number of strains (i.e. both represented median expression levels of 90 strains), a frequency histogram would look identical to the probability density functions, with the y-axis rescaled. Each color represents strains bearing a ROX1-GFP fusion with the indicated modification: "No feedback" indicates the suboptimized feedback mutant (Figures 3.3, 3.4, and 3.6). WT, wild-type. p, resampling p-value evaluating the F statistic for differential variance between the feedback mutant and wild-type distributions. Coefficients of variation of distributions of wild-type and feedback-mutant strains are 0.01 and 0.07, respectively, for the BY4741 cross; 0.07 and 0.31 for YPS606; 0.26 and 0.75 for SK1; and 0.10 and 0.18 for 22:3:b. BY4741, is a laboratory strain derived from a fig tree isolate; YPS606, an isolate from an oak tree[57]; SK1, a laboratory strain derived from a soil isolate[57]; and 22:3:b, a strain of hybrid laboratory and vineyard origin[58].

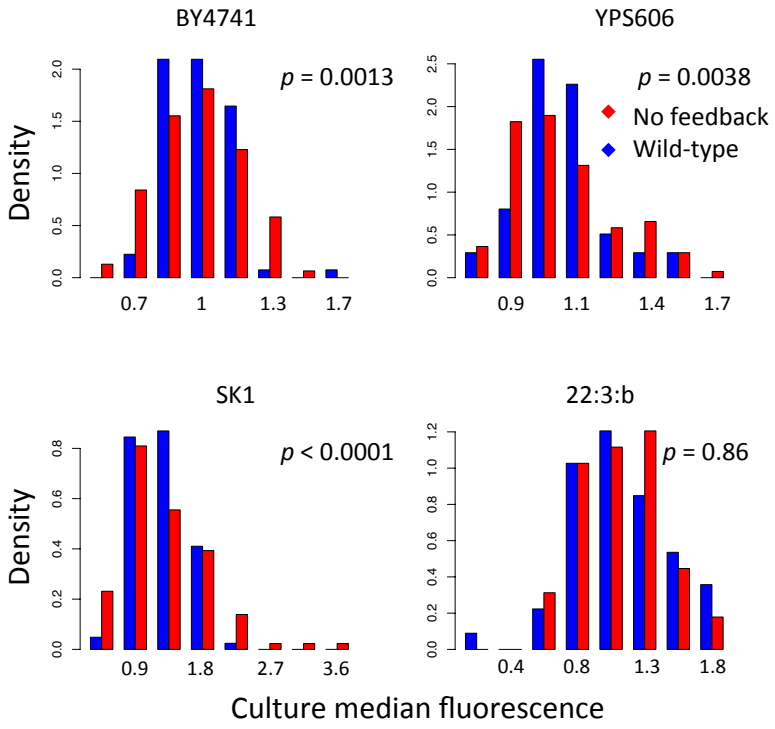

Figure 3.8 Eliminating feedback in an otherwise wild-type Rox1 reporter compromises mutational robustness. Data are as in Figure 3.7, except that "No feedback" indicates a *ROX1-GFP* reporter construct in whose promoter all four Rox1 binding sites are mutagenized, and whose coding sequence bears wild-type codons (Figures 3.3, 3.4 and 3.6).

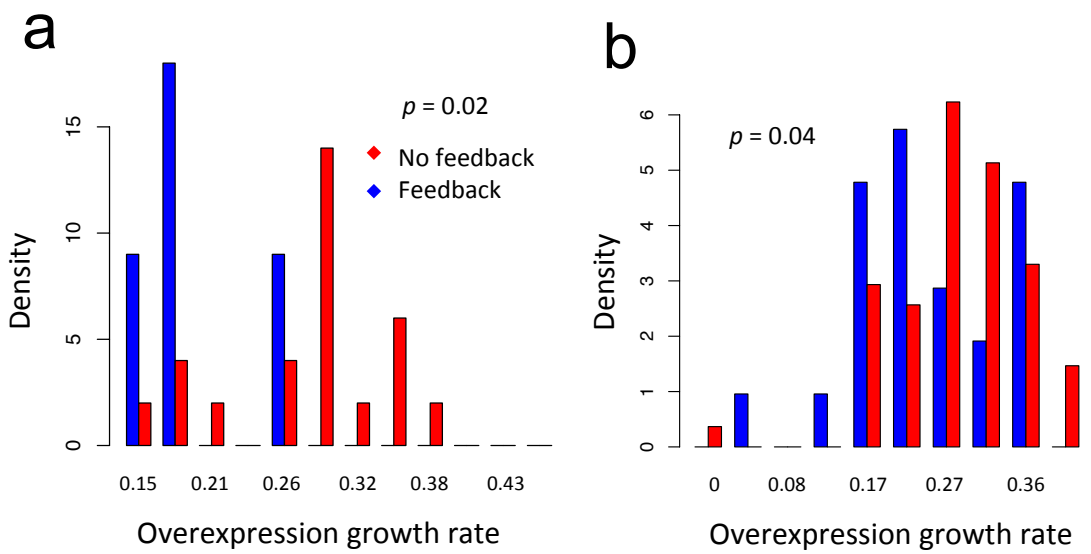

Figure 3.9 Transcription factors subject to feedback are toxic when overexpressed. Each panel represents analysis of growth rates across a panel of yeast strains, each overexpressing a single transcription factor[28]. Each distribution represents growth rates across the set of genes with or without evidence for feedback as indicated. The x-axis reports growth rate in doublings/hour. (A) Red represents factors emerging as screen hits in Figure 3.1; blue reports all other screened factors (n = 22). Distribution medians are 0.29 and 0.17, respectively. (B) Red represents factors with significant sequence matches to their own binding motif in their own promoters; blue reports all other tested factors (n = 83). Distribution medians are 0.27 and 0.24, respectively. p, Wilcoxon p-value evaluating the difference in growth rates between factors with and without evidence for feedback.

ATGaacccga agtcgtcgac gccgAAGatc ccgcggccgA AGAACgcgtt catcctcTTC
cggCAGCACT ACCACcggAT CctcatcGAC gagTGGacgg cacaggggGT Ggagatcccg
cacaactcgA ACatctcgaa gatcatcggg ACGAAGTGGA AGgggctcca gCCGgaggac
AAGgcgCACT GGgagaacct cGCGGAGAAG GAGaagctcg agcacgagcg gAAGtacccg
gagTACaagT ACAAGCCGgt gcggAAGtcg AAGAAGAAGc agctcctcct cAAGgagATC
GAGCAACAGC AGCAGCAACA ACAGaaggag CAGCAGCAGC AGaagCAGtc gcagCCGcag
ctccagCAGc cgttcAACAA Caacatcgtg ctcATGaagc gggcgcactc gctctcgccg
tcgtcgTCGG TGtcgtcgTC GAACtcgtac CAGTTCcagc tcAACaacga cctcAAGcgg
ctcccgatcc cgtcggtgaa cacgtcgAAC tacATGgtgt cgcggtcgct ctcggggctc
ccgctcACGc acgacAAGAC GgcgcggGAC ctcccgCAGc tctcgtcgca gctcaactcg
atcccgtacT ACtcggcgcc gCACGACccg tcgACGcggc accacTACCT CAACgtggcg
caggcgcagc cgcgggcgAA CTCGacgccg cagctcccgt tcatctcgtc gatcATCAAC
AACtcgtcgc agacgCCGgt gacg*ACT*ACT accacgtcg*A CT*acgacgGC Gacgtcgtcg
ccgGGGaagT TCtcgtcgtc gCCGAACtcg tcggtgctcG AGAACAACcg gctcAACtcg
ATCAACaact cgaaccagta cctcccgccg ccgctcctcc cgtcgctcca ggacttccag
ctcgacCAGT ACCAGCAGct cAAGCAGATG gggccgacgt acatcgtgaa gccgctctcg
CACacgcggA ACaacctcct ctcgacgacg acgccgACGc accacCACat cccgcacatc
ccgAACcagA ACatcccgct ccaccagatc atcAACtcgt cgAACacgGA Ggtgacggcg
aagacgtcgc tcgtgtcgCC GaagGGTGAC GGTGCTGGTt taATTAACAT Gtcgaagggg
gaggagctcT TCacgggggt ggtgccgatc ctcgtggagc tcgacgggga cgtgaacggg
CACaagttct cggtgtcggg ggaggggggag ggggacgcga cgTACgggaa gctcacgctc
aagttcatct gcacgacggg gaagctcccg gtgccgTGGc cgacgctcgt gacgacgctc
acgtacgggg tgcagtgctt ctcgcggTAC ccggaccacA TGaagcagca cGACttcTTC
AAGtcggcgA TGccggaggg gtacgtgcag gagcggacga tcttcTTCaa ggacGACggg
AACTACAAGa cgcgggcgga ggtgAAGttc gaggggggaca cgctcgtgaa ccggATCgag
ctcaagggga tcgacttcaa ggaggacggg AACatcctcg ggCACaagct cgagTACAAC
tacAACtcgC ACaacgtgTA CATCATGgcg GACaagcagA AGaacgggAT CaaggtgAAC
TTCaagatcc ggCACAACat cgaggacggg tcggtgcagc tcgcgGACca ctaccagcag
aacacgccga tcggggacgg gccggtgctc ctcccgGACA ACcacTACct ctcgacgcag
tcggcgctct cgaaggaccc gAACgagAAG cggGACCACA TGgtgctcct cgagttcgtg
acggcggcgg ggatcacgca cgggATGgac gagctcTACa ag**TAA**

Figure 3.10 Suboptimized *ROX1-GFP* sequence. Lowercase letters represent codons that have been modified from the wild-type sequence to a codon corresponding to a lower codon index value[27]. Uppercase letters represent unchanged codons. Uppercase italicized letters represent optimal non-wild-type codons. Blue shaded codons were modified from sub-optimal codons to prevent synthesis errors due to repeat sequence. Yellow shaded sequence represents the linker between *ROX1* and *GFP* incorporated to improve expression level[51]. Green shaded sequence represents the *GFP* coding sequence. Bold text represents the stop codon.

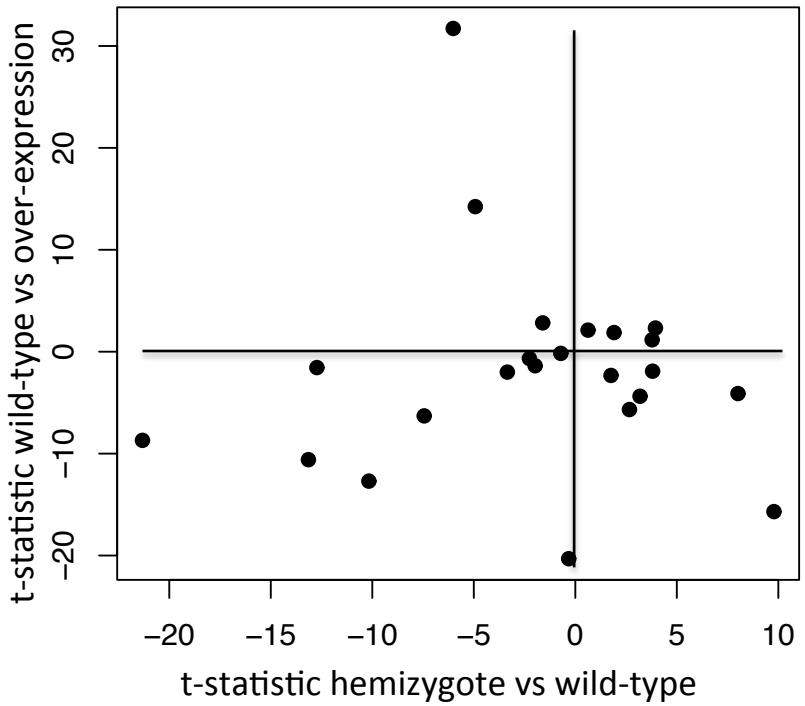

Figure 3.11 Complete data set from the screen of regulatory feedback across transcription factors. All symbols and values are as in Figure 3.1.

SWI4	35186.53
ZAP1	26578.11
HIR1	26386.28
GLN3	26353.92
STP2	27919.37
YAP3	26555.68
AZF1	27732.27
RPH1	26549.68
GZF3	26801.2
TEC1	26060.8
OPI1	28371.46
YML081W	26105.32
HAL9	27385.85
ARG81	26149.43
BAS1	26061.96
ASH1	26845.78
MOT3	31713.58
PHD1	25435.17
YFL044C	26348.15
PUT3	24942.47
UPC2	31527.14
YDR520C	24605.21
IXR1	41224.61
UGA3	24337.57
WTM2	27119.07
ROX1	34123.42
INO2	29926.03
YAP5	24166.04
ZMS1	24004.79
PDR3	23735.8
UME6	27712.91
RCO1	46841.26
WTM1	135981.4
MAC1	44675.11
KSS1	51426.07
SUT1	46944.43
DAT1	49860.42
DAL81	54671.8
ECM22	49117.57
XBP1	53446.87
MIG1	54576.54

Table 3.1 continues on following page

DIG1	52827.92
HMS2	48718.59
CUP9	48009.46
YAP7	45950.59
MBP1	55450.88
SKO1	48905.9
YDR026C	23436.41
GAT1	24449.59
HOG1	29015.13
CBF1	34552.59
FZF1	24368.94
HAP5	25887.41
TBS1	22157.88
LEU3	23521.49
YRR1	22629.15
SKN7	23984.04
GAL80	28860.63
YHP1	49703.81
CIN5	20915.41
RGM1	20419.31
NRG1	21071.44
MET28	20420.98
ARG80	20768.52
RTG1	26675.7
INO4	20302.5
HAP3	21632.88

Table 3.1 Abundances of yeast transcription factors. Each row reports the total cellular fluorescence in arbitrary units of the indicated transcription factor as a GFP protein fusion in a diploid hemizygote, measured *via* quantitative microscopy.

Chapter 4

The importance of feedback for dose response linearization

Introduction

Transcription factor negative feedback can confer a number of information-processing functions in regulatory networks: Previous work has demonstrated that negative feedback confers homeostasis of gene expression against stochastic noise[1-3]. In Chapter 3 we provide evidence that direct ROX1 negative feedback can speed response time and increase gene expression homeostasis in the face of natural genetic variation. In this chapter I will explore the possibility that negative feedback has been selected for a different function—to linearize transcription factor dose response and thereby improve the fidelity of information transmission. The effect of dose response linearization on information transmission was first conceptualized in engineering systems[4,5] and recent work has demonstrated its relevance in genetic networks[6].

The basic idea as applied to genetic networks is as follows: Many genes exhibit step-like responses—a gene that is controlled by an inducer is only expressed when inducer concentration exceeds a particular set point. The relationship between inducer signal input and expression output is commonly referred to as dose response. A step-like dose response distinguishes between a limited number of inducer levels; for a perfect step function, the gene responds to two distinct signal states (i.e. ranges of inducer levels): one state where inducer signal is below a particular set point and one state where inducer is above that set point. On the other hand, incorporating negative feedback into a simple genetic circuit can convert a step-like dose response to a more linearized dose response. Whereas a step-function distinguishes two states, a linear dose response can distinguish any number of states. In information theory, "information" is defined as the number of distinct states that can be transmitted between a source and a destination[7]. In this sense, a linearized dose response conveys much more information about the inducer level to the downstream gene than a step-like dose response.

There is, of course, one additional confounding factor that requires consideration: Biological systems exhibit variation of their molecular component parts due to the stochasticity of cellular biochemical events. As a result, there are limits to the amount of information that can be captured by and transmitted between them. Which is to say that a linear dose response in a biological system does not in fact transmit unlimited information. However, a linearized dose response is expected to greatly improve the fidelity of information transmitted between noisy components[8]. Consistently, a study by Nevozhay *et al.*[6] reported that a synthetic step-like dose response converted to a more linear dose response by incorporating direct transcriptional negative feedback. In addition, information transmission between inducer signal and expression level of a downstream gene is greatly improved. In fact, the feedback-free circuit is capable of distinguishing fewer than two inducer signal states, while the feedback-linearized circuit is capable of distinguishing a number of distinct signal states.

An additional key finding that emerged from this study is that the feedback-free system exhibited greater noise in downstream gene expression at intermediate inducer levels. This can be explained by comparing how noise is propagated through a step-like dose response versus a linear dose response. For a step-like response, as inducer level approaches intermediate inducer levels, even small variation in inducer level will cause some cells to respond strongly, while others will respond weakly, resulting in greatly amplified variation in gene expression. In contrast, for a linear dose response, small variation in signal intensity will result in proportionally small variation in gene expression (Figure 1, red lines).

In what situations is either a step-like dose response or a linearized dose response better suited for the cellular functions mediated by gene networks? A step-like dose response can be advantageous for genes involved in many processes, particularly in situations where an organism benefits when a cell commits to a particular state. Commitment to on-or-off gene expression states can be important in a number of biological contexts[9,10]. For example, multi-cellular organisms avoid developmental errors by enforcing commitment to a differentiated state with on-or-off gene expression[11]. On the other hand, there are situations in which a step-like response may not be the ideal strategy. For example, many inducible bacterial operons involved in metabolism exhibit expression levels that are proportional to inducer levels[12]. In this case, it may be advantageous for the genetic network to match expression of catabolic enzymes with the level of precursor present in the environment over a wide range of precursor levels. Interestingly, in most of these cases, the operon is subject to direct negative transcriptional feedback[12], consistent with the notion that negative feedback linearizes transcription factor activity.

In Chapter 2 we identified many cases of direct transcriptional negative feedback on yeast transcription factors, and noticed that these factors often regulate metabolic genes. As an added layer of complexity, the factors are often also involved in larger indirect metabolic homeostasis feedback loops. One possible explanation for this observation is that high fidelity of information transmission is generally important for metabolic homeostasis across different environments. In this chapter I will propose prospective experiments to test this hypothesis using the heme homeostasis feedback loop as a model. Heme is an important metabolite for yeast physiology: It is thought to serve as an indicator of cellular oxygen levels, and it functions as a prosthetic group for several enzymes, most of which are involved in the electron transport chain of oxidative phosphorylation. Therefore, heme homeostasis is likely of critical importance for cellular physiology. The heme activator protein HAP1 directly senses heme and activates both cellular respiration genes and the ROX1 transcription factor. ROX1 encodes a repressor of cellular respiration and metabolite biosynthesis genes, and has been shown to be subject to direct transcriptional negative feedback[13]. One of its targets is a terminal step and key control point in heme biosynthesis, the copropophyrinogen III oxidase encoded by HEM13[14]. In sum, ROX1 expression levels communicate activity of the heme-induced sensor protein to a control point for heme production, forming a heme homeostasis feedback loop (Figure 4.2). In this chapter I will present preliminary

data indicating that negative feedback on ROX1 is important for fidelity of information transmission within the heme homeostasis feedback loop, and propose further experiments to rigorously test this hypothesis.

Free heme levels not only indicate environmental oxygen concentration, but also reflect the balance between production and demand of heme. However, the importance of heme homeostasis for organismal fitness has not been studied. Heme is an essential prosthetic group of the cytochrome proteins that compose the electron transport chain. In aerobic conditions, the electron transport chain harnesses the energy from carbon metabolism to generate a proton gradient that in turn drives ATP synthase to produce ATP. Thousands of copies of each cytochrome complex are embedded in the folded inner membrane of each mitochondrion in order to meet each cell's energy demand. Heme demand changes drastically between fermentative (no respiration), non-fermentative (pure respiration), and respiro-fermentative metabolism. As cellular demand for heme changes, high or low free heme levels may be communicated by the heme homeostasis feedback loop so that production can be adjusted to match demand. I propose that high fidelity of information transmission associated with direct transcriptional feedback at ROX1 is important for precisely matching heme production with heme demand. In this chapter I will test whether feedback on ROX1 serves to optimize ROX1 expression for respirative growth over a range of heme pathway fluxes.

Results and proposed experiments

ROX1 linearization and information transmission

Previous studies have taken advantage of heme biosynthesis mutants to study the function of mitochondrial biogenesis[15]. I reasoned that I could use this approach to generate a ROX1-heme dose response curve: Deleting an early enzymatic step in the heme biosynthesis pathway and supplementing growth media with different concentrations of a complementary metabolic intermediate would provide indirect control of free heme levels, which would in turn lead to differential activation of ROX1 by HAP1. Deleting the gene encoding the first committed step in heme metabolism, HEM1, and supplementing media with different levels of the pathway intermediate 5-aminolevulinate, has been shown to result in different levels of mitochondrial function and biogenesis[1]. I generated an analogous strain where ROX1 was tagged with YFP (Figure 4.3a). By supplementing growth media with different amounts of 5-aminolevulinate, I observed a linear increase in ROX1 expression with increasing 5-aminolevulinate (Figure 4.3b).

In order to evaluate whether feedback affects the linearity of ROX1 dose response, I set out to generate a strain lacking feedback. A previous study obtained a structure for the ROX1 HMG domain by aligning its sequence with the human SRY protein, the structure of which has been determined[16]. Using this structure, they predicted that I18 directly interacts with the first two thymine residues of the ATTGT DNA consensus motif recognized by ROX1. They went on to show that Rox1 containing the I18T mutation lost the ability to auto-repress ROX1 gene expression *in vivo*, presumably because it cannot bind to the ATTGT consensus sequences in its promoter (see Chapter 3). I generated a strain containing an I18T mutation in a ROX1-YFP tagged strain in the *hem1* background and tested the effect of the mutation on dose response (Figure 2.3b). I observed a much more step-like dose response function of I18T-ROX1-YFP, suggesting that feedback does function to linearize dose response (Figure 2.3d). In addition, I observed a much higher expression level of I18T-ROX1-YFP than wild-type ROX1-YFP across a range of 5-aminolevulinate concentrations, as expected if the I18T mutation disrupts autoregulation. The major caveat of this experiment is that the I18T mutation probably eliminates all ROX1 activity, as a strain carrying this mutation shows similar phenotype to a ROX1 null strain in a number of conditions (data not shown). Although unlikely, it is possible that loss of linearization is not due to loss of direct feedback but rather some other function of ROX1. For example, it is possible that the indirect heme homeostasis feedback loop plays a role in linearizing the ROX1 dose response. In order to rule out this possibility, I will repeat these experiments with a set of strains that instead have mutations in the autoregulatory binding sites of the ROX1-GFP promoter and have a range of expression levels modulated by codon composition (see Chapter 3). These strains will be referred to as "direct feedback mutant" throughout this chapter. Methods for constructing and validating these strains are presented in Chapter 3. The important point is that the strains lack

direct feedback, and their steady-state expression levels span a range that includes wild-type expression.

In order to further test the hypothesis that linearization of dose response at ROX1 is important for transmitting information to the metabolic control point HEM13, I will next incorporate a RFP tag at the HEM13 locus in the wild-type and direct feedback mutant ROX1-GFP strains. I will grow cells in synthetic media supplemented with ethanol and varying concentrations of 5-aminolevulinate, and measure single-cell expression of both ROX1-GFP and HEM13-RFP by flow cytometry. I expect that Hem13 dose response will be aligned with that of Rox1. This will manifest in an inverse proportional relationship between Rox1 and Hem13 dose response over a wide range of 5-aminolevulinate levels—where fractional changes in Rox1 level correspond to opposite fractional changes in Hem13 level—so as to maximize information transmission (Figure 4.4a). If this is the case, then a linearized ROX1-heme dose response should result in a linearized Hem13-heme dose response (black lines in Figure 4.4b and c). On the other hand, if Rox1 dose response is made more step-like, as is expected to be the case in feedback mutant strains, the Hem13 dose response will also become more step-like (colored lines in Figure 4.4b and c). A key prediction that follows is that loss of linearization at ROX1 will lead to increased Rox1 expression variation at intermediate 5-aminolevulinate concentrations, which will be propagated or even amplified when the Rox1 signal is passed on to Hem13[17]. I therefore expect to observe increased expression variation of Hem13 at intermediate levels of 5-aminolevulinate in direct feedback mutants compared with wild-type (Figure 4.4d-f).

With these data I will also quantify the capacity for each strain to transmit information between heme and HEM13 in terms of bits of information[18,19]. The schematic data in Figure 4.4d and e provide a qualitative illustration of how loss of feedback results in loss of fidelity of information transmission. When feedback is present, each cell's expression level at a given 5-aminolevulinate concentration falls within a range that is distinct from the range of each of the other 5-aminolevulinate concentrations. As a result, for each individual cell, the Hem13 expression level distinguishes which of the 5 concentrations the cells were grown in. On the other hand, when feedback is absent, Hem13 expression variation is increased at intermediate signal levels, and a cell's expression no longer distinguishes which of the five different concentrations the cells were grown in.

Information transmission and fitness

If ROX1 feedback improves fidelity of information transmission between the heme sensing protein HAP1 and the heme production control point, HEM13, and improved information transmission is important for precisely balancing heme demand with heme production, I would expect that loss of feedback would lead to impaired mitochondrial function. As a first step in testing this hypothesis, *hem1* yeast strains were grown on synthetic media supplemented with ethanol and varying concentrations of 5-aminolevulinate. Ethanol is a non-fermentable carbon source, so ATP production depends solely on mitochondrial respiration on this media. Each strain contained a copy of either I18T ROX1, wild-type ROX1, or direct feedback mutant ROX1 (effectively serving as an overexpression construct) at the ROX1 locus. The wild type strain grew at rates similar to HEM1 strains (data not shown) over a wide range of 5-aminolevulinate concentrations. However, either increasing or decreasing Rox1 activity relative to wild-type resulted in slower growth (Figure 4.5). These data are consistent with the hypothesis that wild-type ROX1 expression is closer to the optimal expression level for growth on ethanol over a range of 5-aminolevulinate concentrations. However, they do not demonstrate whether linearization or improved information transmission associated with feedback achieves closer-to-optimal expression levels than can be achieved without feedback.

Next I will test whether the wild-type ROX1 direct feedback circuit achieves a closer-to-optimal expression than is possible of ROX1 without direct feedback. In order to test this, I will first establish the relationship between ROX1 levels and non-fermentative growth rates over a range of free heme levels: I will generate strains where the endogenous ROX1 promoter is replaced with a battery of constitutive yeast promoters so that each strain will express ROX1 at different but fixed levels. These strains will be grown in synthetic media supplemented with ethanol and varying concentrations of 5-aminolevulinate. For each strain, growth rate and expression will be measured in log-phase cultures. These data will compose a fitness function, where growth is considered with respect to both 5-aminolevulinate concentration and Rox1 expression (Figure 4.5a). The strain that exhibits the fastest growth for each of the tested 5-aminolevulinate concentrations will be considered to express ROX1 at the optimal level for that condition. To assess whether the wild-type ROX1 direct feedback circuit achieves a closer-to-optimal expression level than the direct feedback mutant strains, I will use the measured fitness function to evaluate how closely the wild-type and direct feedback mutant Rox1 expression levels coincide with the optimal expression levels over varying 5-aminolevulinate concentrations (Figure 4.5a, black line versus colored lines). If the wild-type curve matches the measured optima more closely, I will conclude that the negative feedback loop facilitates closer-to-optimal expression over a range of free heme levels.

An additional question is whether gene expression noise associated with intermediate induction levels will give rise to deleterious phenotypic effects. Having

measured growth for strains with a wide variety of promoters at ROX1, I will likely observe a range in magnitudes of expression variation between strains. If expression variation does impact fitness, I expect that the growth rates of strains with similar mean expression levels will decrease with increasing expression variation. More specifically, I will test whether the expression variation reduction associated with direct transcriptional feedback at ROX1 impacts growth: At the 5-aminolevulinate concentration where wild-type and a given direct feedback mutant's ROX1 expression levels intersect, the two strains are expected to exhibit the same mean expression but with different expression variation (Figure 4.3b and f). If direct feedback does impact the capacity for information transmission, and such information is important for respirative growth, I expect that the lowered capacity for information transmission at the intersecting mean expression levels will result in a growth rate decrease in the direct feedback mutant strain (Figure 4.5b).

Discussion

It has long been appreciated that transcription factors and signaling pathways control the metabolic state of the cell. However, the reciprocal relationship, where the metabolic state feeds back on the transcription factors and signaling molecules to impose itself on the regulatory state has not received equal attention[20]. Here, I propose to study how information about the metabolic state of the cell is transmitted through the network to adjust further metabolite production. In particular, I set out to assess the importance of direct negative feedback for transmitting such information.

How realistic are the proposed results in light of findings from previous studies? The anticipated findings rely on a number of assumptions. First of all, it is possible that loss of the direct feedback loop will be fully compensated by the indirect heme homeostasis feedback loop. Preliminary evidence suggests that this is not the case. The direct feedback mutant strain expresses ROX1 about four-fold higher than wild-type in rich media. If the indirect loop fully compensated for the broken direct loop, we would expect no change in the expression. Interestingly, the expression from the direct feedback mutant is considerably lower than the I18T mutant where ROX1 is likely totally non-functional, suggesting that the indirect loop may partially compensate the broken direct loop.

Second, it is possible that the indirect heme homeostasis feedback loop is important for homeostasis in a limited range. HEM13 has been shown to serve as a limiting step at low O_2 levels[14]. Because HEM13 uses molecular oxygen as a substrate, it is possible that it only serves as a limiting step when oxygen is scarce. If this is the case, I would not expect to observe the linear relationship between Rox1 and Hem13 expression levels (Figure 4.3a). On the other hand, HEM13 is the only gene in the terminal steps of heme biosynthesis subject to substantial regulation in response to varying heme demand[21]. Given the preliminary evidence suggesting that there is an indirect ROX1 feedback loop in addition to the direct loop suggest that the HEM13 loop is operational over a broad range of oxygen concentrations.

Finally, it is possible that the expression of wild-type ROX1 does not match the optimum of the fitness function. A previous study that measured a fitness function for expression of the lac operon at different concentrations of lactose did not observe alignment between wild-type expression and optimal expression[22]. Rather, they found that if strains were grown in a given lactose concentration for many generations, they evolved to achieve a steady-state expression level matching the optimum. In light of these findings, it seems reasonable to expect that ROX1 expression will not match the optimum of the fitness function. On the other hand, it is also possible that the selective pressures on lac operon expression are different than those of heme homeostasis such that ROX1 expression is under a more constant selective pressure to align with the optimal fitness function. A key observation supporting this possibility is that growth yield homeostasis in respiring yeast is maintained across different growth conditions and genotypes[23]. It has been

shown that growth yield homeostasis depends on tight adjustment of the cellular contents of the respiratory chain compounds to the growth rate, and that a defect in this adjustment leads to energy yield decrease. The fact that this phenotype is held constant across environments and genotypes suggests that it may have been under constant selection. I have not directly measured whether the respirative growth defects resulting from misexpressed ROX1 are due to loss of growth yield homeostasis, but it seems likely that by losing heme homeostasis so that mitochondrial component production is not coordinated with mitochondrial demand could lead to a lack of coordination between mitochondria and growth rate. It therefore seems feasible that wild-type ROX1 expression has been under selective pressures resulting in alignment with the optimal fitness function.

What do we stand to gain from these experiments? If direct feedback at ROX1 does function to improve information transmission in the heme homeostasis feedback loop, this could serve as a paradigm for how feedback functions in other metabolite homeostasis loops. Additionally, and perhaps more importantly, studying the basic biology of model systems can result in a deeper understanding of how analogous systems function in human biology. A key motivation for studying heme metabolism is its central role in coordinating the mode of energy metabolism between anaerobic and aerobic growth. Upregulation of genes that are normally induced by hypoxia is coming to be recognized as a classic feature of cancer[24], and genetic mutations that directly affect activity of the transcription factor hypoxia-inducible factor-1 (HIF-1)[25] have been shown to predispose people to highly angiogenic tumors[26]. The experiments proposed here stand to inform the regulatory mechanisms controlling metabolism in yeast, and may serve as a paradigm for understanding similar processes in human biology.

Materials and methods

Plasmids used for ROX1-YFP strains. To generate constructs for protein-fusion reporters, the *ROX1* promoter and coding sequence were amplified from yeast genomic DNA and fused to Y*FP* (Addgene) flanked by a kanamycin resistance cassette[27] *via* ligation independent cloning[28]. This plasmid was used to generate an I18T mutant ROX1 construct or to generate a direct feedback mutant construct. For the I18T mutant, the second base of the 18th codon was changed from T to C (codon ATT -> ACT) by using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). For the direct feedback mutant construct, Rox1 binding sites in the *ROX1* promoter were identified as perfect matches to the Rox1 consensus sequence[29], and mutated by converting ATTGT consensus sequence to ATGGT.

Generating ROX1-YFP strains. ROX1-YFP strains were generated as follows: A HEM1::LEU2 strain in the DY150 background, obtained as a gift from Jerry Kaplan, was used to generate a ROX1::URA3 strain by replacing the endogenous *ROX1* gene and the upstream 467 bp with a *URA3* cassette[30]. The resulting strain was mated to a W303 strain, sporulated, and a recombinant was then mated with BY4742 and sporulated. Both the LEU2 and URA3 prototrophies were maintained through both crosses/sporulations and the HAP1 locus in the resulting strain was confirmed to have the non-mutant allele[31]. The resulting strain was transformed with PCR products amplified from ROX1-YFP constructs, and selected by growing on YPD media supplemented with 150mg/L 5-aminolevulinate (Merck) and 300mg/L G418 (Cellgro).

Measuring expression of ROX1-YFP strains across varying 5-aminolevulinate concentrations. Synthetic complete media supplemented with galactose and varying concentrations of 5-aminolevulinate was prepared using a 25mg/ml stock solution of 5-aminolevulinate (Merck) dissolved in water. Wild-type, and I18T-mutant ROX1-YFP strians were grown in triplicate in complete synthetic media with galactose and 200mg/L 5-aminolevulinate in flasks at 30°C with shaking overnight to saturation. 50µl of each culture was transferred to 1ml synthetic complete media with galactose and 10mg/L 5-aminolevulinate in 2ml 96-well plates, and incubated at 30°C with shaking for 10 hours. Then cells were diluted to optical density 0.15 and 100 µl were added to 1ml synthetic complete media with galactose and varying 5-aminolevulinate concentrations in 2ml 96-well plates, and incubated at 30°C with shaking for an additional 10 hours. The 5-aminolevulinate concentrations used were, in mg/L: 1600, 800, 400, 200, 100, 50, 25, 12, 6, 3, 1.5, 0. Then, fluorescence was measured for each culture using a LSRFortesa (BD Biosciences) and analyzed in R with the flowCore package (www.bioconductor.org). Mean expression for each culture was calculated from cells that passed forward/side scatter gating. Gating was based on the methods of Newman *et al.*[1], where the ~1% of cells with forward/side scatter closest to the mean forward/side scatter for the culture passed the gate.

Serial dilution assays. Strains were grown in flasks overnight in complete synthetic media supplemented with 150mg/L 5-aminolevulinate at 30°C with shaking. Each culture was diluted in fresh media at an optical density 0.2 and grown to optical density 1.0. 1:10 serial dilutions were performed in microtiter plates. Dilutions were transferred to plates using a multipronged inoculating device (frogger). Plate media were synthetic complete supplemented with 3.2% ethanol, 0.5% Tween 80, 20mg/L ergosterol, and either 500, 50, 10 or 2 mg/L 5-aminolevulinate. A Tween/ergosterol stock solution was prepared as in Crisp *et al*.[32].

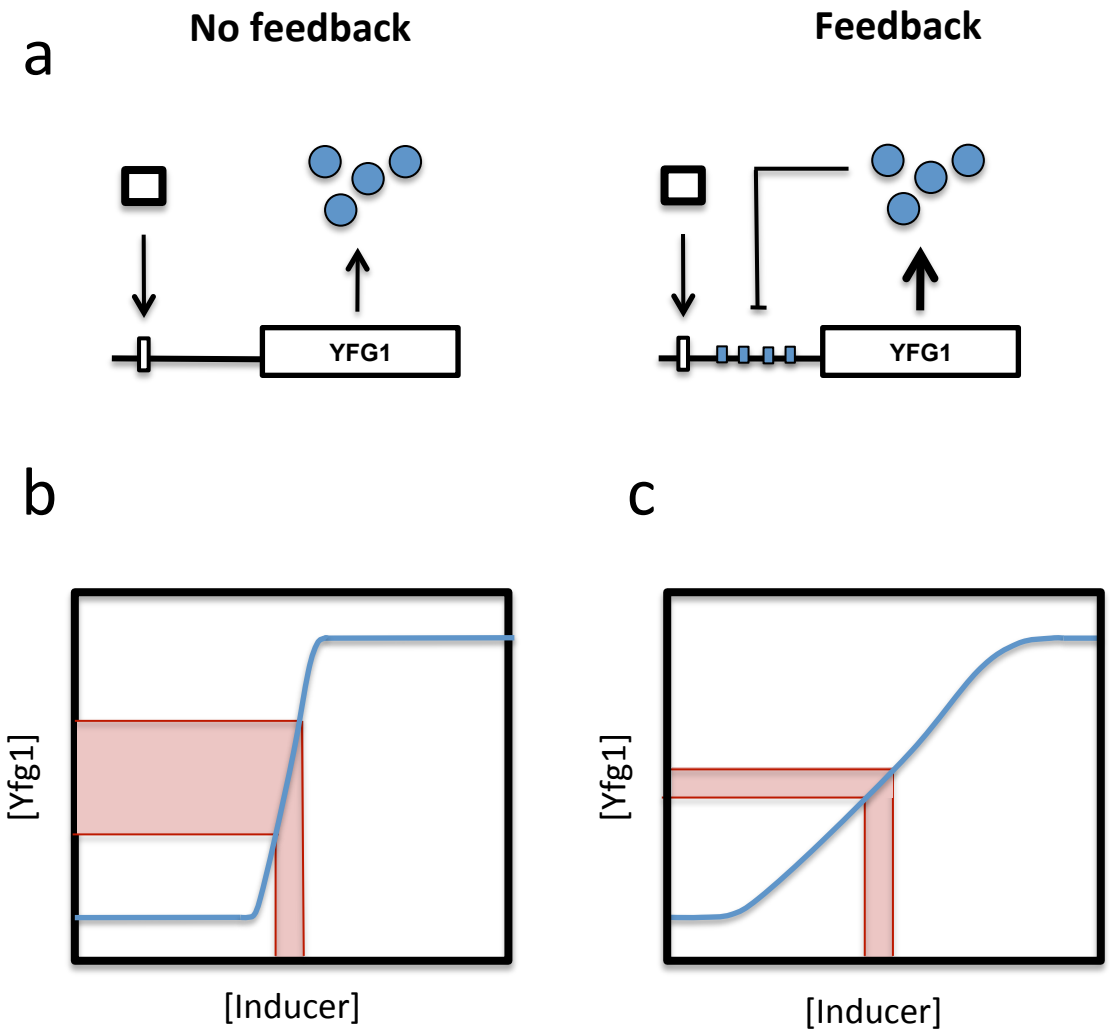

Figure 4.1 Feedback linearizes dose response and improves fidelity of information transmission. (a) Schematic illustrating alternate gene circuits. (b and c) Schematic illustrating anticipated effect of incorporating direct transcriptional feedback on gene expression dose response. Dose response (blue line) in (b) is much more step-like compared with the dose response in (c). Red shaded boxes indicate how small changes in inducer signal due to stochastic variation would result in a different magnitude of gene expression variation between step-like and linear dose responses at intermediate inducer concentrations.

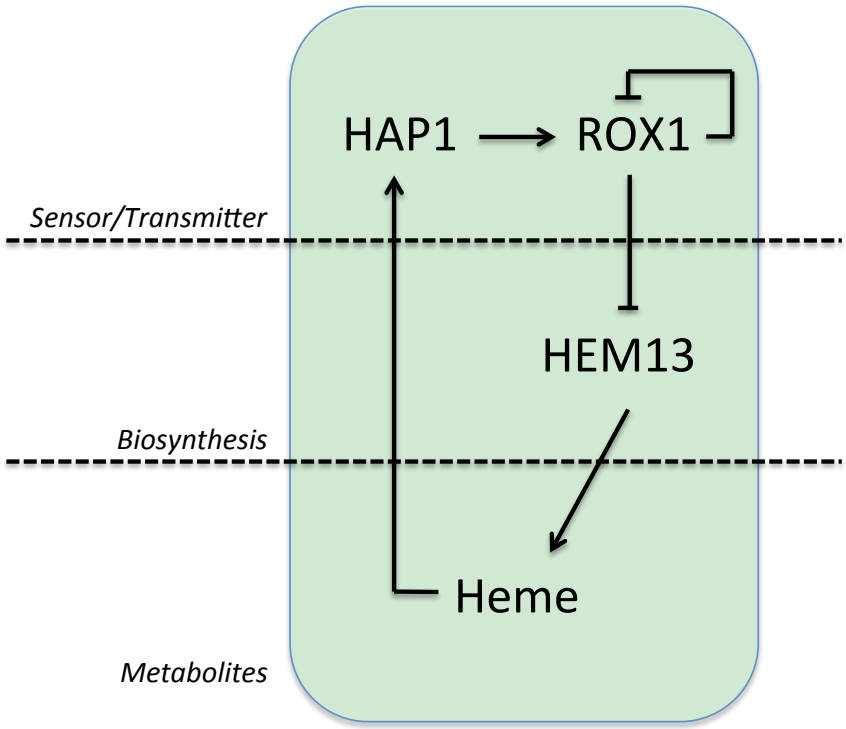

Figure 4.2 Direct ROX1 feedback loop is embedded in a heme homeostasis feedback loop. HAP1 encodes a transcription factor that directly senses/binds heme[27] and activates ROX1 in the presence of heme[28]. ROX1 represses HEM13[29], the gene encoding a rate-limiting step in heme biosynthesis[9].

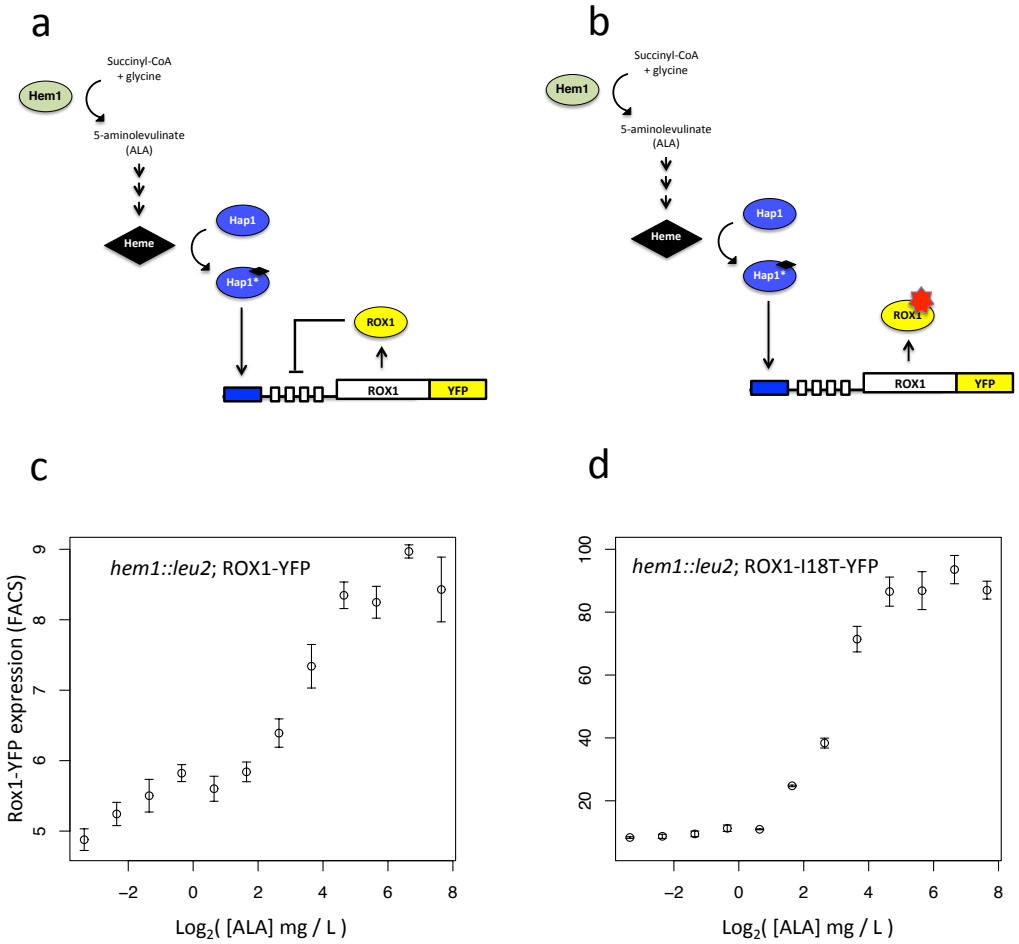

Figure 4.3 ROX1 dose response is linearized by feedback. (a and c) Hem1 catalyzes the first committed step in heme biosynthesis by converting succinyl CoA and glycine to 5-aminolevulinate. Heme levels were indirectly controlled by generating a HEM1 deletion strain and supplementing media with varying concentrations of 5-aminolevulinate. ROX1 was tagged with YFP and the effect of varying levels of 5-aminolevulinate was quantified by FACS. Error bars represent triplicate measurements of each population mean. (b and d) A mutation in I18T renders ROX1 non-functional, thereby eliminating the feedback loop. The ROX1 dose response curve is less linear in the feedback mutant compared with wild-type.

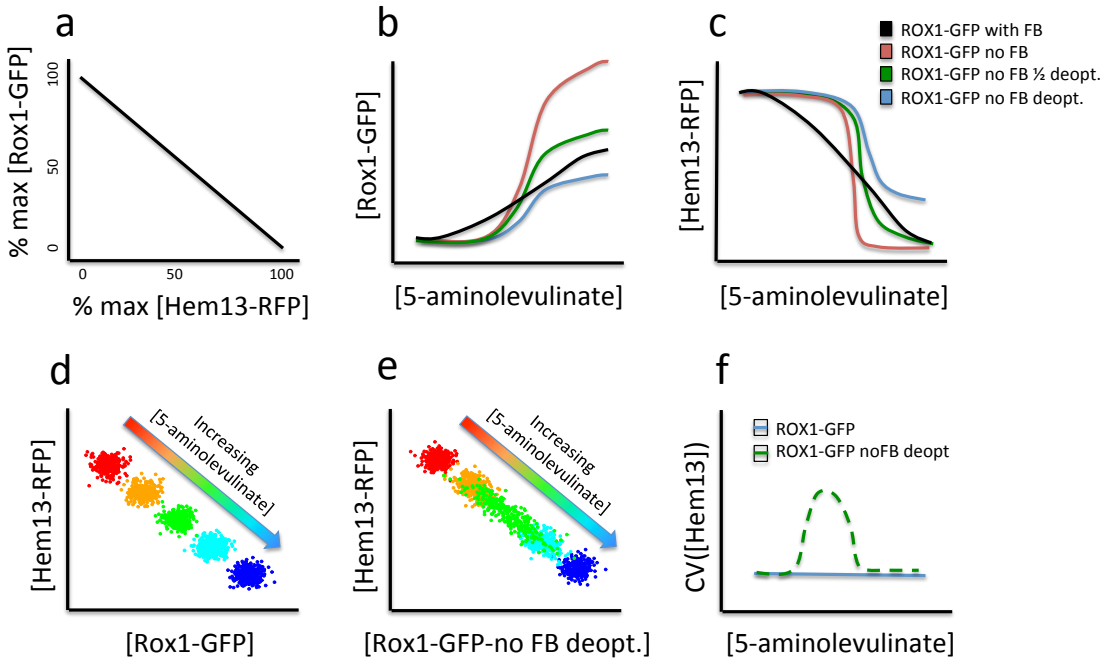

Figure 4.4 Anticipated results illustrating the loss of information transmission capacity from Rox1 to Hem13 in the absence of feedback. (a) The anticipated relationship between Rox1 and Hem13 expression if their dose response functions are aligned. (b and c) Anticipated results of deleting the 4 Rox1 binding sites in the ROX1 promoter. "no FB" (red curve) indicates the absence of Rox1 binding sites in the ROX1 promoter. "½ deopt." indicates that the codons in the GFP sequence are modified to reduce translational efficiency, and "deopt." indicates that the codons in the ROX1 and GFP sequences were modified to reduce translational efficiency. (d and e) Anticipated results showing the expression distributions of Rox1 and Hem13 at increasing concentrations of 5-aminolevulinate. (f) Coefficient of variation (variance/mean^2) is a mean-normalized measure of expression variation.

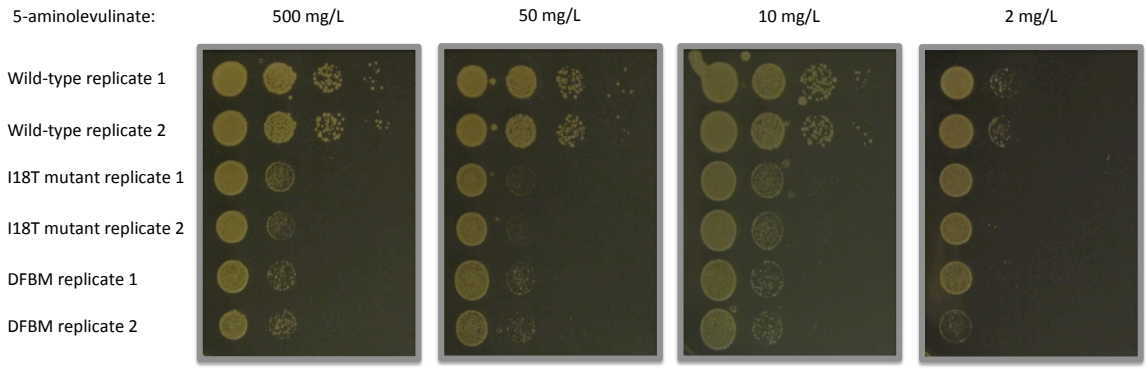

Figure 4.5 ROX1 mutants exhibit growth defect on ethanol over a range of 5-aminolevulinate levels. Wild-type strains contain ROX1 tagged with YFP; I18T strains contain ROX1 tagged with YFP and a mutation converting isoleucine to threonine at amino acid 18 in the ROX1 coding sequence; and direct feedback mutant (DFBM) strains contain ROX1 tagged with YFP and single base pair mutations in each of 4 auto-regulatory binding sites in the ROX1 promoter.

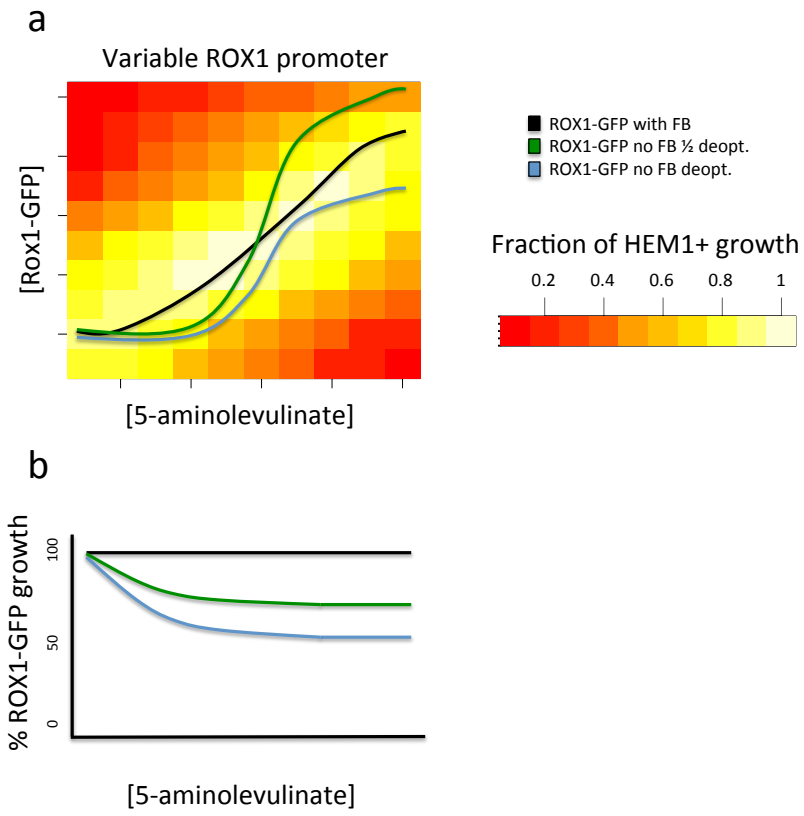

Figure 4.6 Anticipated growth rates over varying 5-aminolevulinate concentrations. (a) Anticipated growth rates of *hem1* strains with various yeast promoters driving expression of ROX1 plotted with respect to mean ROX1 expression and concentration of 5-aminolevulinate. Because each promoter will drive ROX1 expression with a different amount of variation, and strains with high variation are expected to grow collectively slower than strains with low variation, strains with low variation will be used to generate estimates of maximum growth. Maximum growth rates will be plotted as values relative to growth of a strain with an intact heme biosynthesis pathway (HEM1+). Lines represent expression levels for wild-type and direct feedback mutant strains. (b) Anticipated results illustrating growth of wild-type and direct feedback mutant strains at varying 5-aminolevulinate concentrations relative to wild-type ROX1-GFP; *hem1* strain.

References

1. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. & Doyle, J. Robustness of cellular functions. *Cell* **118**, 675–685 (2004).
2. Waddington, C. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 557–559 (1942).
3. Waddington, C. Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1953).
4. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814–819 (1997).
5. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol Evol* **17**, 468–473 (2002).
6. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
7. Specchia, V. *et al.* Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463**, 662–665 (2010).
8. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
9. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
10. Hartwell, L. GENETICS: Robust Interactions. *Science* **303**, 774–775 (2004).
11. Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* **37**, 295–299 (2005).
12. Li, J., Yuan, Z. & Zhang, Z. The cellular robustness by genetic redundancy in budding yeast. *PLoS Genet.* **6**, e1001187 (2010).
13. Sauer, U., Canonaco, F., Heri, S., Perrenoud, A. & Fischer, E. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli. *J. Biol. Chem.* **279**, 6613–6619 (2004).
14. Papp, B., Pál, C. & Hurst, L. D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
15. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**, 549–552 (2003).
16. Levy, S. F. & Siegal, M. L. Network hubs buffer environmental variation in Saccharomyces cerevisiae. *PLoS Biol.* **6**, e264 (2008).
17. Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 19015–19020 (2005).
18. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
19. Knox, B. E., Devreotes, P. N., Goldbeter, A. & Segel, L. A. A molecular mechanism for sensory adaptation based on ligand-induced receptor modification. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2345–2349 (1986).
20. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).

21. Kacser, H. & Burns, J. A. The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973).
22. Flint, H. J. *et al.* Control of the flux in the arginine pathway of Neurospora crassa. Modulations of enzyme activity and concentration. *Biochem. J.* **200**, 231–246 (1981).
23. Krishna, S., Semsey, S. & Sneppen, K. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20815–20819 (2007).
24. Lee, T. I. Transcriptional Regulatory Networks in Saccharomyces cerevisiae. *Science* **298**, 799–804 (2002).
25. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64–68 (2002).
26. Masel, J. & Siegal, M. L. Robustness: mechanisms and consequences. *Trends Genet.* **25**, 395–403 (2009).
27. Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546–549 (1974).
28. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614–8619 (2001).
29. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
30. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
31. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.* **2**, 41 (2006).
32. Marquez-Lago, T. T. & Stelling, J. Counter-intuitive stochastic behavior of simple gene circuits with negative feedback. *Biophys. J.* **98**, 1742–1750 (2010).
33. Stekel, D. J. & Jenkins, D. J. Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression. *BMC Syst Biol* **2**, 6 (2008).
34. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).
35. Tu, K. C., Long, T., Svenningsen, S. L., Wingreen, N. S. & Bassler, B. L. Negative feedback loops involving small regulatory RNAs precisely control the Vibrio harveyi quorum-sensing response. *Mol. Cell* **37**, 567–579 (2010).
36. Ramsey, S. A. *et al.* Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat. Genet.* **38**, 1082–1087 (2006).
37. Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K. & Balázsi, G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5123–5128 (2009).
38. Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
39. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals

the architecture of biological noise. *Nature* **441**, 840–846 (2006).
40. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
41. Black, H. Stabilized feed-back amplifiers. *… Institute of Electrical Engineers* (1934).
42. Madar, D., Dekel, E., Bren, A. & Alon, U. Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli. *BMC Syst Biol* **5**, 111 (2011).
43. Bhalla, U. S., Ram, P. T. & Iyengar, R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* **297**, 1018–1023 (2002).
44. Yu, R. C. *et al.* Negative feedback that improves information transmission in yeast signalling. *Nature* **456**, 755–761 (2008).
45. Paulsen, M., Legewie, S., Eils, R., Karaulanov, E. & Niehrs, C. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes development. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10202–10207 (2011).
46. Jackson, C. L. & Hartwell, L. H. Courtship in S. cerevisiae: both cell types choose mating partners by responding to the strongest pheromone signal. *Cell* **63**, 1039–1051 (1990).
47. Thieffry, D., Huerta, A. M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays* **20**, 433–440 (1998).
48. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
49. Alon, U. *An introduction to systems biology*. 301 (CRC Press: 2007).
50. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
51. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182 (2005).
52. To, T.-L. & Maheshri, N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327**, 1142–1145 (2010).
53. Hu, Z., Killion, P. J. & Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.* **39**, 683–687 (2007).
54. Mnaimneh, S. *et al.* Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**, 31–44 (2004).
55. Schüller, H. J., Schorr, R., Hoffmann, B. & Schweizer, E. Regulatory gene INO4 of yeast phospholipid biosynthesis is positively autoregulated and functions as a transactivator of fatty acid synthase genes FAS1 and FAS2 from Saccharomyces cerevisiae. *Nucleic Acids Res.* **20**, 5955–5961 (1992).
56. Deckert, J., Perini, R., Balasubramanian, B. & Zitomer, R. S. Multiple elements and auto-repression regulate Rox1, a repressor of hypoxic genes in Saccharomyces cerevisiae. *Genetics* **139**, 1149–1158 (1995).

57. Foster, R., Mikesell, G. E. & Breeden, L. Multiple SWI6-dependent cis-acting elements control SWI4 transcription through the cell cycle. *Mol. Cell. Biol.* **13**, 3792–3801 (1993).
58. Ashburner, B. P. & Lopes, J. M. Autoregulated expression of the yeast INO2 and INO4 helix-loop-helix activator genes effects cooperative regulation on their target genes. *Mol. Cell. Biol.* **15**, 1709–1715 (1995).
59. Nikawa, J.-I. & Kamiuto, J. The promoter of the yeast OPI1 regulatory gene. *J. Biosci. Bioeng.* **97**, 369–373 (2004).
60. Zhao, H. & Eide, D. J. Zap1p, a metalloregulatory protein involved in zinc-responsive transcriptional regulation in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **17**, 5044–5052 (1997).
61. Delahodde, A., Delaveau, T. & Jacq, C. Positive autoregulation of the yeast transcription factor Pdr3p, which is involved in control of drug resistance. *Mol. Cell. Biol.* **15**, 4043–4051 (1995).
62. Torres, E. M. *et al.* Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007).
63. Springer, M., Weissman, J. S. & Kirschner, M. W. A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* **6**, 368 (2010).
64. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
65. Dimster-Denk, D. *et al.* Comprehensive evaluation of isoprenoid biosynthesis regulation in Saccharomyces cerevisiae utilizing the Genome Reporter Matrix. *J. Lipid Res.* **40**, 850–860 (1999).
66. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
67. MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**, 113 (2006).
68. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
69. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).
70. Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *J. Theor. Biol.* **234**, 227–253 (2005).
71. Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228–232 (2005).
72. Berry, D. B. & Gasch, A. P. Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. *Mol. Biol. Cell* **19**, 4580–4587 (2008).
73. Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).
74. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
75. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864 (2006).
76. Butcher, R. A. *et al.* Microarray-based method for monitoring yeast

overexpression strains reveals small-molecule targets in TOR pathway. *Nat. Chem. Biol.* **2**, 103–109 (2006).
77. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
78. Burke, D. & Strathern, J. *Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual*. (2005).
79. Gordon, A. *et al.* Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods* **4**, 175–181 (2007).
80. Hickman, M. J. & Winston, F. Heme levels switch the function of Hap1 of Saccharomyces cerevisiae between transcriptional activator and transcriptional repressor. *Mol. Cell. Biol.* **27**, 7414–7424 (2007).
81. Keng, T. HAP1 and ROX1 form a regulatory pathway in the repression of HEM13 transcription in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **12**, 2616–2623 (1992).
82. Di Flumeri, C., Liston, P., Acheson, N. H. & Keng, T. The HMG domain of the ROX1 protein mediates repression of HEM13 through overlapping DNA binding and oligomerization functions. *Nucleic Acids Res.* **24**, 808–815 (1996).
83. Zagorec, M. *et al.* Isolation, sequence, and regulation by oxygen of the yeast HEM13 gene coding for coproporphyrinogen oxidase. *J. Biol. Chem.* **263**, 9718–9724 (1988).
84. Loewen, C. J. R. *et al.* Phospholipid metabolism regulated by a transcription factor sensing phosphatidic acid. *Science* **304**, 1644–1647 (2004).
85. Amar, N., Messenguy, F., Bakkoury, El, M. & Dubois, E. ArgRII, a component of the ArgR-Mcm1 complex involved in the control of arginine metabolism in Saccharomyces cerevisiae, is the sensor of arginine. *Mol. Cell. Biol.* **20**, 2087–2097 (2000).
86. Crabeel, M., Seneca, S., Devos, K. & Glansdorff, N. Arginine repression of the Saccharomyces cerevisiae ARG1 gene. Comparison of the ARG1 and ARG3 control regions. *Curr. Genet.* **13**, 113–124 (1988).
87. De Rijcke, M., Seneca, S., Punyammalee, B., Glansdorff, N. & Crabeel, M. Characterization of the DNA target site for the yeast ARGR regulatory complex, a sequence able to mediate repression or induction by arginine. *Mol. Cell. Biol.* **12**, 68–81 (1992).
88. Kirschner, M. & Gerhart, J. Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8420–8427 (1998).
89. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
90. Macneil, L. T. & Walhout, A. J. M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* **21**, 645–657 (2011).
91. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
92. Freeman, M. Feedback control of intercellular signalling in development. *Nature* **408**, 313–319 (2000).
93. Brandman, O. & Meyer, T. Feedback loops shape cellular signals in space and

time. *Science* **322**, 390–395 (2008).
94. Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S. & Carthew, R. W. A microRNA imparts robustness against environmental fluctuation during development. *Cell* **137**, 273–282 (2009).
95. Schwartsburd, P. M. Self-cytoprotection against stress: feedback regulation of heme-dependent metabolism. *Cell Stress Chaperones* **6**, 1–5 (2001).
96. Acar, M., Pando, B. F., Arnold, F. H., Elowitz, M. B. & van Oudenaarden, A. A general mechanism for network-dosage compensation in gene circuits. *Science* **329**, 1656–1660 (2010).
97. Kapuy, O., He, E., Uhlmann, F. & Novák, B. Mitotic exit in mammalian cells. *Mol. Syst. Biol.* **5**, 324 (2009).
98. Duong, H. A., Robles, M. S., Knutti, D. & Weitz, C. J. A molecular mechanism for circadian clock negative feedback. *Science* **332**, 1436–1439 (2011).
99. Justman, Q. A., Serber, Z., Ferrell, J. E., El-Samad, H. & Shokat, K. M. Tuning the activation threshold of a kinase network by nested feedback loops. *Science* **324**, 509–512 (2009).
100. Kwast, K. E. *et al.* Genomic analyses of anaerobically induced genes in Saccharomyces cerevisiae: functional roles of Rox1 and other factors in mediating the anoxic response. *J. Bacteriol.* **184**, 250–265 (2002).
101. Jansen, R., Bussemaker, H. J. & Gerstein, M. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res.* **31**, 2242–2251 (2003).
102. Yoshikawa, K. *et al.* Comprehensive phenotypic analysis of single-gene deletion and overexpression strains of Saccharomyces cerevisiae. *Yeast* **28**, 349–361 (2011).
103. Mettetal, J. T., Muzzey, D., Gómez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* **319**, 482–484 (2008).
104. Wall, M. & Hlavacek, W. Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* (2004).
105. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
106. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
107. Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250–253 (2011).
108. Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. *Science* **335**, 82–85 (2012).
109. Lynch, M. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* **8**, 803–813 (2007).
110. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 2006.0028 (2006).

111. Ideker, T. & Sharan, R. Protein networks in disease. *Genome Res.* **18**, 644–652 (2008).
112. Szalay, M. S., Kovács, I. A., Korcsmáros, T., Böde, C. & Csermely, P. Stress-induced rearrangements of cellular networks: Consequences for protection and drug design. *FEBS Lett.* **581**, 3675–3680 (2007).
113. Amit, I., Wides, R. & Yarden, Y. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Mol. Syst. Biol.* **3**, 151 (2007).
114. Wagner, A. Does evolutionary plasticity evolve? *Evolution* (1996).
115. Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* **38**, 1077–1081 (2006).
116. Magwene, P. M. *et al.* Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1987–1992 (2011).
117. Goldstein, A. L., Pan, X. & McCusker, J. H. Heterologous URA3MX cassettes for gene replacement in Saccharomyces cerevisiae. *Yeast* **15**, 507–511 (1999).
118. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast* **21**, 661–670 (2004).
119. Haun, R. S., Serventi, I. M. & Moss, J. Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *BioTechniques* **13**, 515–518 (1992).
120. Badis, G. *et al.* A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* **32**, 878–887 (2008).
121. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
122. Crisp, R. J. *et al.* Inhibition of heme biosynthesis prevents transcription of iron uptake genes in yeast. *J. Biol. Chem.* **278**, 45499–45506 (2003).
123. Gaisne, M., Bécam, A. M., Verdière, J. & Herbert, C. J. A "natural" mutation in Saccharomyces cerevisiae strains derived from S288c affects the complex regulatory gene HAP1 (CYP1). *Curr. Genet.* **36**, 195–200 (1999).
124. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
125. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1572–1577 (2005).
126. Rockmill, B., Lambie, E. J. & Roeder, G. S. Spore enrichment. *Meth. Enzymol.* **194**, 146–149 (1991).
127. Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda)* **1**, 11–25 (2011).
128. Bailey, T. L. & Noble, W. S. Searching for statistically significant regulatory modules. *Bioinformatics* **19 Suppl 2**, ii16–25 (2003).
129. Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).

130. Shannon, C. E., Weaver, W.Shannon *The Mathematical Theory of Communication*. 144 (University of Illinois Press: 1998).
131. Ptashne, M. & Gann, A. *A Genetic Switch, Third Edition: Phage Lambda Revisited*. 164 (Cold Spring Harbor Laboratory Press: 2004).
132. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics* **149**, 1633–1648 (1998).
133. Haslam, J. M. & Astin, A. M. The use of heme-deficient mutants to investigate mitochondrial function and biogenesis in yeast. *Meth. Enzymol.* **56**, 558–568 (1979).
134. Deckert, J., Khalaf, R. A., Hwang, S. M. & Zitomer, R. S. Characterization of the DNA binding and bending HMG domain of the yeast hypoxic repressor Rox1. *Nucleic Acids Res.* **27**, 3518–3526 (1999).
135. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
136. Cover, T. M. & Thomas, J. A. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. 776 (Wiley-Interscience: 2006).
137. McKnight, S. L. On getting there from here. *Science* **330**, 1338–1339 (2010).
138. *Biosynthesis of Heme and Chlorophylls*. 594 (Mcgraw-Hill (Tx): 1989).
139. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).
140. Devin, A. *et al.* Growth yield homeostasis in respiring yeast is due to a strict mitochondrial content adjustment. *J. Biol. Chem.* **281**, 26779–26784 (2006).
141. WARBURG, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
142. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
143. Kaelin, W. G. & Maher, E. R. The VHL tumour-suppressor gene paradigm. *Trends Genet.* **14**, 423–426 (1998).