Analysis of Robustness and Stochasticity in Biochemical Networks

By

Mei-Lyn Ong

B.Eng. Mechanical Engineering
Nanyang Technological University, Singapore, 2005

SUBMITTED TO THE PROGRAM IN COMPUTATIONAL AND SYSTEMS BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN COMPUTATIONAL AND SYSTEMS BIOLOGY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

Signature of author _____
Program in Computational and Systems Biology
October 19, 2011

Certified by _____
Alexander van Oudenaarden, Ph.D.
Professor of Physics and Biology
Thesis Supervisor

Accepted by _____
Christopher B. Burge, Ph.D.
Professor of Biology and Biological Engineering
Director, Computational and Systems Biology Graduate Program

Analysis of Robustness and Stochasticity in Biochemical Networks

By

Mei-Lyn Ong

Submitted to the Computational and Systems Biology Program
on December 1st, 2011 in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Systems Biology

Abstract | Cells are constantly faced with the challenge of functioning reliably while being subject to unpredictable changes from within and outside. Here, I present two studies in which I analyze how biochemical circuits that regulate signaling and gene expression can generate robustness or phenotypic variability between otherwise identical yeast cells.

Using the osmosensing signaling pathway which consists of a phosphorelay connected to a MAPK cascade, we predict signaling robustness to changes in kinetic rate constants by employing a computational sensitivity analysis. Consistent with the model predictions, we find that the input-output relation of signaling activation is severely impacted by protein coding sequence changes in the MAPK cascade genes, but not the phosphorelay genes. By decoupling the network into two separate modules, we show that an input-output analysis of each of the modules can generate the observed disparity in their tolerance to kinetic parameter variations. Our analysis suggests that the input-output relation of catalytic signaling pathways i.e. MAPK cascade are intrinsically sensitive to kinetic rate perturbations. By contrast, signaling governed by stoichiometric biochemical reactions i.e. phosphorelay exhibit robust input-output functions. We further find that cells challenged to alter their input-output function mostly recovered by gaining mutations in the MAPK cascade genes, which further supports our model.

We next explore how *HAC1* RNA splicing contributes to heterogeneity in the unfolded protein response (UPR). We adapt the single molecule FISH (sm-FISH) method to count endogenous spliced and unspliced *HAC1* transcripts in single cells. We use a stochastic bursting-transcription-and-splicing model to determine the kinetic rates from the single cell measurements. We find that the cell-to-cell variability in the degree of splicing is tightly regulated in the presence of a UPR-inducing chemical agent, but is compromised under heat stress. By considering models including extrinsic noise at the splicing or transcriptional level, we show that the increased variability in the degree of splicing under heat stress can be generated by increased fluctuations in the splicing rate.

Lastly, we present an approach using sm-FISH and protein synthesis inhibitors to measure translation and we show preliminary results suggesting its feasibility.

Thesis Supervisor: Alexander van Oudenaarden, Ph.D.
Title: Professor of Physics and Biology, M.I.T.

ACKNOWLEDGEMENTS

I am deeply indebted to many colleagues, friends and family for their tireless support over the years, without which, this thesis work would not have been possible. First and foremost, I have to thank my advisor Alexander van Oudenaarden for creating such a diverse, exciting and stimulating environment in which I was given the freedom to pursue my ideas and was offered generous support whenever I needed it. His foresight, modesty and optimism are exemplary qualities I admire and hope to emulate. I can never thank him enough for making my experience as a PhD student challenging and fun, and for the personal growth I have achieved through this experience.

I thank my thesis committee members, Angelika Amon and Douglas Lauffenburger for their insightful feedback and advice on my work, and for generously sharing with me their expertise, all of which I greatly appreciate. I am grateful to Eduardo Sontag for the spontaneous and helpful discussions on control theory and modeling.

The work in this thesis could not have been done single-handedly. I have to thank these talented people in the van Oudenaarden lab whom I have had the privilege of working closely with – Shankar Mukherji and Qiong Yang on the osmosensing signaling project, and Stefan Semrau on the single cell translational profiling project. I am grateful to everyone in the lab for making it such a wonderful and supportive place to do research in. I must especially thank Bernardo Pando, for serving as my 2nd mentor, Dale Muzzey for his support and advice, and Jay Mettetal for being inspiring.

I could not have done this without the great friendships and unwavering support from my 2nd family at MIT - Leah Octavio, Jaimie Lee and Dmitry Kashlev.

Finally, I must thank my family - I would not be the person I am without them.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2:

Chapter 3:

LIST OF FIGURES

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 1

INTRODUCTION

Unifying principles in biological and engineered systems

Cellular biochemical networks are tasked with the complex challenge to function reliably in noisy external and internal environments given imperfect components. Errors in fundamental processes such as signal transduction or gene expression can lead to serious outcomes, affecting growth and development, and consequently causing diseases. While robustness is a long recognized concept in biology, it is also infamous for being difficult to evaluate (its causes, extent and functional impact). The fact that biological networks need to be robust to such perturbations imposes constraints on their design. Exactly how the cell achieves this feat has been elusive until recently.

One of the first insights came from a theoretical study of the well characterized bacterial chemotaxis signaling network (Barkai and Leibler, 1997), which suggested that adaptation is robust to changes in biochemical parameters of the system. Adaptation is a well-established feature of chemotaxis i.e. the signal output returns to the prestimulus level and accordingly, the steady-state tumbling frequency is insensitive to ligand concentration (Berg, 1975). Experimental analyses later confirmed this result (Alon et al., 1999). Using techniques from control theory, Yi *et al.* further showed that an integral feedback control generates the robustness of perfect adaptation to changes in intracellular parameters or enzyme concentrations. These studies provide an example of an elegant design of bacterial signaling circuits that allows cells to consistently achieve their precise functions despite uncertainties in

the environment (varying stimulus levels) and system components (varying protein concentrations).

Subsequently, design principles including feedback control, redundancy and modularity, were found to give rise to robustness of biological networks in diverse organisms from bacteria to humans (Stelling et al., 2004; Kitano, 2004 and 2007), suggesting that they form the basic building blocks of complex networks. Recently, Shinar and Feinberg formulated a theorem for the structural requirements of a class of biochemical networks which exhibit "absolute concentration robustness" (ACR) (concentration of a species remains exactly the same in any positive steady states of the system). The theorem connects the structure of networks with their capacities for ACR and holds for all reaction networks in the cell, which provides a general physical framework in which one can study biological robustness.

In Chapter 2, I explore the robustness of the osmosensing signaling pathway in *S. cerevisiae* to kinetic rate constant changes of its pathway proteins. I will show that signaling sensitivity is highly varied, with signaling being most impacted by MAPK cascade variations and is robust to changes in the phosphorelay genes, consistent with a computational sensitivity analysis of the pathway. By decoupling the network into two separate modules, I will show using theoretical analysis that the input-output relation of the catalytic MAPK cascade is intrinsically sensitive to kinetic rate perturbations. By contrast, signaling governed by the stoichiometric mechanism of the phosphorelay exhibits robust input-output functions.

The stochastic nature of gene expression

The inherent stochasticity of all biochemical events in the cell involving small numbers of molecules necessarily results in fluctuations in their levels. These random fluctuations can be detrimental to the precision of cellular functions. Due to limited experimental capabilities to formally characterize these fluctuations, biological noise has been largely ignored previously. The first experiments to investigate the sources of gene expression noise came from the ground-breaking studies by Elowitz et al. and Ozbudak et al. Elowitz et al. provided a conceptual framework for analyzing gene expression variability in terms of extrinsic and intrinsic components (derived mathematically by Swain et al., 2002). In their experiments in *E. coli*, they introduced a dual reporter assay (having two copies of the same promoter each driving the expression of either a yellow fluorescent protein or cyan fluorescent protein in the same cell) capable of quantifying and distinguishing between the different sources of expression variability. Under this scheme, intrinsic fluctuations arise from the stochastic events during gene expression, and cause uncorrelated variations in the levels of YFP and CFP. Extrinsic fluctuations result from fluctuations in the levels of upstream regulatory molecules, and cause correlated fluctuations.

Ozbudak et al. showed in *B. subtilis* that gene expression variability depended on the underlying transcriptional and translational rates, and the experiments confirmed an earlier theoretical analysis (Thattai and van Oudenaarden, 2001) that predicted that noise as expressed by the coefficient of variation (standard deviation divided by mean) scales inversely with transcription rate, but is independent of translation (due to translational bursts, since multiple proteins are produced from a

single transcript), thus introducing the concept of bursts in gene expression. Subsequent studies using single-molecule techniques demonstrated the bursty dynamics of protein (Cai et al., 2006) and mRNA production (Golding et al., 2005). These studies dismiss the assumption that the expected size of fluctuations of a species (standard deviation) is equal to the square root of its copy number (Schroedinger, 1944), and show that the fluctuations can be significantly larger. Subsequently, bursting has been observed in a variety of systems including bacteria, yeast, mammalian cells and *Drosophila*.

Global studies of protein expression noise in *S. cerevisiae* showed that the degree of noise correlated with protein function i.e. proteasomal genes showed low variation while stress-response genes were highly variable (Bar-Even et al., 2006; Newman et al., 2006). One of the first pieces of evidence for the functional consequences of noise came from a study which demonstrated that the expression noise of essential genes is minimized (due to their deleterious effects on cellular functions), suggesting that variation is subject to selection (Fraser et al., 2004). While it is generally expected that noise is a barrier that cells have to overcome to achieve robust functions, there are contrasting examples where noise is advantageous. In unicellular organisms, noise can create a diversity of phenotypes in genetically identical populations. Noise in a stochastic state-switching system can be used to implement a bet-hedging strategy under fluctuating environments (Nachman et al., 2007; Acar et al., 2008), or be used to trigger stochastic cell-fate specification decisions (Eldar and Elowitz, 2010). Even without any feedback, the generation of

increased phenotypic heterogeneity among otherwise genetically identical cells can be a survival strategy under stress.

In Chapter 3, I explore how *HAC1* RNA splicing contributes to heterogeneity in the unfolded protein response (UPR) in *S. cerevisiae*. I will show that the cell-to-cell variability in the degree of splicing is tightly regulated in the presence of a UPR-inducing chemical agent, but is compromised under heat stress. By considering models including extrinsic noise at the splicing or transcriptional level, I will show that the increased variability in the degree of splicing under heat stress can be generated by increased fluctuations in the splicing rate. And I will argue that splicing mis-regulation *in trans* can generate substantial variability in splicing outcomes. In Chapter 4, I will present an approach using single-molecule fluorescence *in situ* hybridization and protein synthesis inhibitors to profile translation in single cells. I will describe preliminary results where we applied our method to monitor translation in exponentially growing yeast, and to explore the changes in translational regulation upon switching the cells from nutrient rich to starvation conditions.

REFERENCES

Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. Nat Genet *40*, 471-475.

Alon, U., Surette, M. G., Barkai, N. and Leibler, S. (1999) Robustness in bacterial chemotaxis. Nature *397*, 168-171.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., Barkai, N. (2006) Noise in protein expression scales with natural protein abundance. Nat Genet *38*, 636-643.

Barkai, N. and Leibler, S. (1997) Robustness in simple biochemical networks. Nature *387*, 913-917.

Berg, H. C. (1975) Chemotaxis in bacteria. Annu. Rev. Biophys. Bioeng. *4*, 119-136.

Cai, L., Friedman, N., Xie, X. S. (2006) Stochastic protein expression in single cells at the single molecule level. Nature *440*, 358.

Eldar, A. and Elowitz, M. B. (2010) Functional roles for noise in genetic circuits. Nature *467*, 167-173.

Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002) Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J., Eisen, M. B. (2004) Noise minimization in eukaryotic gene expression. PLoS Biol. *2*, e137.

Golding, I., Paulsson, J., Zawilski, S. M. and Cox, E. C. (2005) Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025-1036.

Kitano, H. (2004) Biological robustness. Nat Rev Genet *5*, 826-837.

Kitano, H. (2007) Towards a theory of biological robustness. Mol. Syst. Biol. *3*, 137.

Nachman, I., Regev, A., Ramanathan, S. (2007) Dissecting timing variability in yeast meiosis. Cell *131*, 544-556.

Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., Weissman, J. S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840-846.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., van Oudenaarden, A. (2002) Regulation of noise in the expression of a single gene. Nat Genet *31*, 69-73.

Schrodinger, E. (1944) What is Life? Cambridge, UK: Cambridge University Press.

Shinar, G. and Feinberg, M. (2010) Structural sources of robustness in biochemical reaction networks. Science *327,* 1389-1391.

Stelling, J., Sauer, U, Szallasi, Z, Doyle F. J. 3rd and Doyle, J. (2004) Robustness of cellular functions. Cell *118*, 675-685.

Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. *98*, 8614-8619.

Yi, T. M., Huang, Y., Simon, M. I. and Doyle, J. (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc. Natl Acad. Sci. *97*, 4649-4653.

Chapter 2

CONTROL OF ROBUSTNESS AND TUNABILITY IN THE YEAST OSMOSENSING SIGNALING PATHWAY

ABSTRACT | Robustness is a widely observed property of biological systems. While the robustness of cellular processes to variations in gene expression has been extensively explored, robustness to perturbations in biochemical activities of proteins is poorly understood. Using the osmosensing signaling pathway in the budding yeast *Saccharomyces cerevisiae*, we measured the distribution of signaling sensitivity to genetic perturbations through systematic orthologous pathway gene substitutions and experimental evolution. We find that signaling sensitivity is highly varied across the network component genes, with signaling being most impacted by MAPK cascade variations and more robust to changes in the phosphorelay genes, consistent with a computational robustness analysis of the pathway. Results from our theoretical analysis show that the differential robustness pattern emerges from the distinct signaling mechanisms of the two-part pathway architecture, and identifies the stoichiometric phosphoryl-transfer mechanism as a means for buffering genetic variation.

INTRODUCTION

Robustness, the invariance of phenotypes to perturbations, has been observed across diverse levels of biological organizations including gene expression at the molecular level, physiological homeostasis at the cellular and organismal level, and development (Kitano, 2004; Stelling et al., 2004; Wagner, 2005; Barkai and Shilo, 2007). Investigations on the robustness of cellular functions have mainly focused on perturbations at the transcriptional level to understand how cellular functions remain remarkably robust despite the intrinsic stochastic fluctuations in gene expression (Barkai and Leibler, 1997; Alon et al., 1999; Little et al., 1999; Batchelor and Goulian, 2003; Kollmann et al., 2005; Moriya et al., 2006; Shinar et al., 2007; Krantz et al., 2009; Shinar and Feinberg, 2010; Lestas et al., 2010). Robustness to perturbations in biochemical activities of proteins, however, is poorly understood. Genetic perturbations can result in changes in protein-coding sequences that can affect the biochemical activities of proteins. These changes can be positive, for example by enhancing the systems-level fidelity or efficiency of processes mediated by protein-protein interactions in the cell, but typically, these will be deleterious to biochemical function. This thus raises the question of whether cells could have evolved the means for buffering genetic variation at the systems level.

To explore this, we used the high osmolarity glycerol (HOG) pathway in the budding yeast *Saccharomyces cerevisiae*, which forms the signaling module of the hyperosmotic shock response (Hohmann, 2009). The HOG pathway is especially well suited for robustness analysis because its molecular components and interactions have been well characterized (Brewster et al., 1993; Maeda et al., 1994; Posas et al., 1996;

Krantz et al., 2009). Moreover, its network input (extracellular osmolyte concentration) and output (Hog1 activity) can be quantitatively measured and manipulated. We focused on the Sln1 branch of the HOG pathway by deleting the other primary osmosensor Sho1, leaving Sln1 as the main activator of Hog1. Importantly, inactivating the Sho1 branch obviates crosstalk with other MAPK cascades (McClean et al., 2007; Schwartz and Madhani, 2004). The Sln1 branch of the HOG pathway consists of a phosphorelay chain of proteins (Sln1, Ypd1 and Ssk1) that acts on a downstream MAP kinase cascade (Ssk2/Ssk22, Pbs2 and Hog1) to ultimately modulate Hog1 activity (Figure 1a). To further insulate the pathway, we deleted the functionally redundant MAPKKK Ssk22. Upon salt stress, Hog1 translocates into the nucleus (Ferrigno et al., 1998) to initiate transcriptional changes in response to the osmotic shock (O'Rourke and Herskowitz, 2004).

Here, we combine experimental and computational approaches to investigate the robustness of the osmosensing signaling pathway in the budding yeast *Saccharomyces cerevisiae* to perturbations in the biochemical activities of its pathway proteins. By performing systematic orthologous pathway gene substitutions, we found that signaling was significantly altered by sequence variations in the downstream MAPK cascade genes, but remained relatively robust to changes in the upstream phosphorelay components. This agrees well with a computational robustness analysis which predicts that signaling is most sensitive to kinetic parameter changes involving the MAPK cascade proteins. We then showed that yeast cells challenged with hyperactive HOG signaling restored wild-type fitness and signaling mainly via point mutations in the MAPK cascade genes. Furthermore, we found that the growth defect of cells with

compromised HOG signaling under salt stress was rescued by sequence changes in the MAPK genes, but not the phosphorelay component genes. From a theoretical analysis, we showed that the skewed distribution of osmosensing signaling sensitivity can be achieved through the cascading of two biochemically different signaling mechanisms in the pathway. Our findings suggest that stoichiometric biochemical cascades, such as those found in metabolic pathways and in phosphorelay signaling, are fundamentally more robust to genetic changes than processive cascades.

RESULTS

Computational robustness analysis predicts that HOG signaling is most sensitive to MAPK cascade component parameter variations, and is least affected by changes in the phosphorelay genes

To computationally investigate the effects of kinetic rate constant changes in the HOG pathway genes on signaling dynamics, we performed sensitivity analyses on key dynamical properties of the signaling module, i.e. the peak Hog1 phosphorylation level M_{Hog1} and the initial Hog1 phosphorylation rate r_{Hog1}, using a simplified biochemical network model (Supplementary Data, Table S1). We observed a strikingly flat surface for Ypd1-associated parameter changes, indicating that r_{Hog1} remains almost unchanged over a wide range of parameter space (Fig. 1c). In contrast, Pbs2-associated parameter changes significantly altered the r_{Hog1} landscape (Fig. 1b). To systematically compare the effects of parameter variations across individual pathway proteins on signaling, we computed the local logarithmic gradient of the landscape evaluated at wild-type levels and we defined

this metric as our sensitivity measure (Supplemental Data). Figs. 1d and 1e summarize the sensitivities of r_{Hog1} and M_{Hog1} respectively for all pathway genes upstream of *HOG1*. Additionally, we formulated an alternative sensitivity metric that utilizes the full distribution of the model output, instead of only the region around the wild-type level, and measures the relative spread of this distribution for each parameter (Supplementary Data). Both sensitivity analyses predicted that r_{Hog1} and M_{Hog1} are most sensitive to kinetic rate constant changes involving the MAPK cascade genes, and are least affected by variations in the phosphorelay components (Fig. S1).

HOG signaling displays varied sensitivity to ortholog and allele substitutions

To measure the effects of sequence variation in the genes of the HOG pathway on signaling, we utilized the natural variation in the HOG pathway genes across different yeast species and systematically generated mutant strains in which each pathway gene except *HOG1* was replaced one at a time with its ortholog from two evolutionarily diverged yeast species i.e. *Candida glabrata* and *Candida albicans*. Then, we quantified their abilities to recapitulate wild-type signal propagation under a hyperosmotic shock. By using presumably functional orthologs rather than randomly mutated sequences, we more efficiently searched the space of sequences that had a reasonable chance of complementing wild-type behavior. Compared with *S. cerevisiae*, all *C. glabrata* pathway proteins had ClustalW (Thompson et al., 1994) sequence similarity scores between 50 and 60, except Ssk1 which scored 37. *C. albicans*, being evolutionarily more distant from *S. cerevisiae* than *C. glabrata*, displayed lower sequence conservation for all the pathway proteins, ranging from 22 to 46. To estimate the degree of protein functional

changes manifested by the sequence divergence of the orthologs, we computed a potential functional score for each ortholog. The functional score is defined as the percentage of amino acids changes at highly conserved residues identified from comparative genomic analyses of the HOG pathway proteins across various fungi species (Supplemental Data) (Krantz et al., 2006).

We measured the signaling activity by simultaneously monitoring the sub-cellular localization of the nucleus marked by the nuclear factor Nrd1 fused to a red fluorescent protein (Nrd1-RFP) and of Hog1 fused to a yellow fluorescent protein (Hog1-YFP). Upon a hyperosmotic shock, the Hog1 nuclear enrichment dynamics of the Sln1- and Ypd1-ortholog hybrid pathways from both yeast species were indistinguishable from that of the wild-type response despite their low functional scores (Fig. 2a). By contrast, the majority of Ssk2- and Pbs2-ortholog hybrid pathways displayed grossly defective signaling (Fig. 2a, bottom). Importantly, the ability of the hybrid pathways to approach wild-type signaling did not correlate in any simple way to sequence conservation. For example, *C. albicans* Ssk2 and Sln1 have similar functional scores indicating that each protein has a similar fraction of highly conserved amino acid residues changed, but clearly *C. albicans* Sln1 can complement its *S. cerevisiae* counterpart, while Ssk2 cannot.

To further substantiate this finding, we focused on the two architecturally distinct proteins Ypd1 and Pbs2, which belong to the phosphorelay and MAPK modules respectively. We generated strains with *PBS2* and *YPD1* orthologous substitutions from yeast species with increasing evolutionary distance from *S. cerevisiae*, including *Ashbya gossypii*, *Kluyveromyces lactis*, *Neurospora crassa* and *Debaryomyces hansenii*. Despite higher sequence divergence and lower functional scores of these Ypd1 proteins, all of

them still fully mimicked wild-type Hog1 signaling (Fig. 2b). In contrast, signaling performance decreased with increasing sequence divergence and decreasing functional scores of the Pbs2 protein. We also used well characterized *YPD1* alleles (i.e. K67A, R90A and Q86A) with varying degrees of changes in either the phosphotransfer rate $k_{Sln1P\text{-}Ypd1}$ or binding constant $Kd_{Sln1P\text{-}Ypd1}$ between phosphorylated Sln1 and Ypd1 (Janiak-Spens et al., 2005) (Table S2), and we measured the signaling abilities of these *YPD1* alleles under the same hyperosmotic shock. None of the mutants displayed significant changes in Hog1 signaling dynamics compared to wild-type, even in the case where $k_{Sln1P\text{-}Ypd1}$ was reduced by 17-fold (Figs. 2c and S2). Together, these experimental results support the computational predictions that HOG signaling is likely to be more effectively tuned by variations in parameters affecting the MAPK cascade than the phosphotransfer relay.

Rapid adaptive evolution of yeast cells underexpressing *YPD1*

We harnessed naturally occurring genetic variation by evolving yeast cells with hyperactive HOG signaling. We expect that only mutations that can significantly downregulate signaling would be able to rescue the growth defect of these cells. Thus, identifying the pathway genes in which the adaptive mutations lie would allow us to reconstruct the distribution of signaling sensitivity to parameter variations in the network component genes. Since deletion of the *YPD1* gene in the HOG pathway leads to hyperactive signaling and subsequent cell lethality (Posas et al., 1996), we underexpress Ypd1 using a TetO7-Ypd1 strain (Supplementary Data, Figs. S3-S5). From growth rate measurements at different doxycycline concentrations, we confirmed that the cells

suffered a severe growth defect at low doxcycline concentrations, where *YPD1* expression was repressed (Fig. S3). We observed that Hog1 was predominantly localized in the nucleus, therefore confirming that the pathway was indeed hyperactivated under *YPD1* underexpression (Fig. S5). In contrast, Hog1 was uniformly distributed throughout the cytoplasm in cells with high *YPD1* expression and in wild-type cells. Because Hog1 activation induces the expression of *GPD1* and *GPP2*, which encode proteins responsible for glycerol synthesis (Albertyn et al., 1994), we assessed the transcriptional readout of the signaling activity by measuring intracellular glycerol. We found that cells underexpressing *YPD1* had at least two-fold higher intracellular glycerol concentration than cells with high *YPD1* expression (Fig. S4), which was consistent with our observation that the pathway was hyperactivated under *YPD1* underexpression. Finally, by measuring Hog1 nuclear enrichment at different doxycycline levels, we established that growth rate was inversely correlated with Hog1 nuclear accumulation.

We then evolved nine independent lines of the yeast strain with reduced Ypd1 expression each with a population size on the order of 10^7 cells, and monitored their mean population growth rates using turbidostats (Acar et al., 2008). A turbidostat is a continuous culture device which maintains the culture at a constant optical density achieved via a feedback system between the turbidity of the culture and dilution rate. Rapid adaptation occurred shortly after five days, and qualitatively similar adaptation dynamics were observed in the nine experiments (Fig. 3a). At the end of the evolution experiments, five randomly selected single colonies were isolated from each of the nine adapted populations for further analyses. To determine if the hyperactivation of the HOG pathway had been resolved, we measured the evolved strains' basal Hog1 nuclear

enrichment and intracellular glycerol in two randomly selected colonies out of five from each of the nine adapted populations. In 17 out of 18 evolved strains, both basal Hog1 nuclear enrichment and intracellular glycerol content had restored to levels comparable with the ancestor in the unstressed condition (Figs. 3b and 3c), and most of the evolved strains were still capable of partially inducing HOG signaling upon a hyperosmotic shock (Fig. 3c). Thus, we established that the hyperactivation of the HOG pathway had been alleviated in almost all evolved strains.

PBS2 and _SSK2_ are preferentially mutated in independent evolution experiments and their changes are mainly responsible for the improved fitness

To identify the candidate molecular changes that led to the adaptation, we sequenced all six genes in the pathway including their promoter regions for the 45 isolated evolved strains. 40 out of a total of 45 evolved strains contained a single point mutation in one of the genes in the pathway (Figs. 4a and 4b). Strikingly, all 40 evolved strains, except 3 with mutations in only one of the phosphorelay module genes *SSK1*, had mutations in the MAPK cascade genes. We identified a total of 25 unique mutations and all except one were non-synonymous mutations, and more than half of them were in the protein kinase domains, which are highly conserved (Table S3). Almost all the mutations were predicted by the SIFT software (Ng and Henikoff, 2001) to affect protein function (Table S5).

After normalizing the number of unique mutations observed by gene length and the functional impact of each residue for individual genes, we consistently found that, among all the pathway genes, mutations were overrepresented in the MAPK cascade

genes *PBS2* and *SSK2* (Figs. 4c and S6). To test whether *PBS2* and *SSK2* mutations account for the adaptive phenotype, we replaced the endogenous *PBS2* or *SSK2* gene in the ancestral strain with 13 of the unique mutant alleles ("transformed strains"), and compared their growth dynamics to those of the ancestor. These 13 mutant alleles were selected to broadly represent mutations across various protein domains. Unlike the ancestral allele, almost all the mutations conferred a significant growth advantage when the cells were subjected to the original imposed selection (Fig. 4d). The growth increase conferred by the single mutations matched the fitness advantage of most of the evolved strains, confirming that *PBS2* and *SSK2* mutations were primarily responsible for the improved fitness.

For a majority of the transformed strains, the growth rates were similar to that of their respective gene deletion strain i.e. *pbs2*Δ or *ssk2*Δ under no doxycycline conditions i.e. (0.38 ± 0.07) hr^{-1} and (0.42 ± 0.05) hr^{-1}. Since HOG signaling was not completely abolished in the evolved strains, these data further supported that the mutations cause a partial loss-of-function of *PBS2* and *SSK2*, thereby mitigating signaling hyperactivation.

MAPK cascade alleles rescued the growth defect of *PBS2* underexpressing cells under salt stress, but not the phosphorelay *YDP1* alleles

One hypothesis as to why *PBS2* and *SSK2* mutations dominated the space of observed mutations is that in order to alleviate pathway hyperactivity in low salt conditions, selection needs to pick out either loss-of-function mutations in the MAPK cascade genes or gain-of-function mutations in the upstream phosphorelay genes, and that presumably these loss-of-function mutations are far more prevalent. To eliminate the

possibility that *PBS2* and *SSK2* mutations dominated the spectrum of mutations because of this gain- versus loss-of-function disparity, we performed experiments with a strain exhibiting pathway hypoactivity in high salt conditions, thus reversing the previously imposed selection pressure. We imposed pathway hypoactivity by placing the *PBS2* gene under the TetO7 promoter and growing the cells in the absence of doxycycline. In order to directly test whether loss-of-function alleles in the upstream phosphorelay genes could rescue the cells from pathway hypoactivity in high salt conditions, we utilized the *YPD1* alleles with defined reductions in phosphotransfer rate constants used in Fig. 2c. To test the ability of gain-of-function alleles in the downstream MAPK genes, we used constitutively active *PBS2* and *SSK2* alleles known to cause pathway hyperactivity in unstressed conditions (Maeda et al., 1995; Wurgler-Murphy et al., 1997). As shown in Fig. 4e, while the loss-of-function phosphorelay alleles are unable to rescue the growth defect seen in cells containing the hypoactive pathway, the gain-of-function alleles are readily able to repair the growth rate defect. Taken together, the experimental evolution outcomes and the results from the forward genetics and complementation experiments showed that the phosphotransfer relay module confers genetic robustness to osmosensing signaling activity.

Theoretical analysis identifies stoichiometric phosphoryl-transfer mechanism as a means for buffering genetic variation

To investigate the mechanism responsible for buffering genetic perturbations, we used our model and solved analytically the dependence of Hog1 signaling on the biochemical parameters associated with individual network component proteins. To this

end, we made two simplifications to the model underlying the simulations. First, we assumed that signal propagation is fast compared to changes in the osmotic pressure variable that drives pathway activity. Second, because in vitro studies have shown that the phosphotransfer reactions favor rapid product formation and thus limit the pool of phosphorylated Sln1 (Janiak-Spens et al., 2005), we assumed that the concentration of unphosphorylated Sln1 can be approximated by the total Sln1 concentration. Surprisingly, we found that the level of phosphorylated Hog1 depends only on the rate in which phosphoryl groups enter and exit the phosphorelay chain, which is governed by stoichiometric signaling, and is thus determined only by Sln1 and Ssk1 parameters. The rate constants and concentration of Ypd1, as long as they are in the parameter regime consistent with the assumption of rapid phosphoryl flow from Sln1 to Ssk1, do not play a role in establishing the quasi-steady state level of phosphorylated Ssk1 and consequently the level of phosphorylated Hog1 (Supplementary Data). A similar mechanism has been used to describe the robustness of the two-component osmosensing signaling system in *Escherichia coli* to variations in the concentrations of its components (Shinar et al., 2007; Shinar and Feinberg, 2010). By contrast, the MAPK cascade uses a catalytic signaling mechanism where each MAPKKK phosphorylates multiple MAPKKs, and as a result, the level of phosphorylated Hog1 is highly dependent on Pbs2-associated parameters. Furthermore, the quantitative dependence of the level of phosphorylated Hog1 is significantly lower for Sln1- and Ssk1-parameters compared to the MAPK cascade associated parameters. Our model thus suggests that the observed robustness pattern is achieved through the cascading of two biochemically distinct signaling mechanisms.

DISCUSSION

Since protein-protein interactions are crucial for many biological processes, could cells have evolved mechanisms to be robust against changes in protein-coding sequences in addition to gene expression? Our work provides a first foray into this question, and we expect that our work relating the biochemistry of signaling to genetic robustness can generalize to signaling pathways more broadly. For example, the pathway architecture we studied involving a stoichiometric phosphotransfer relay connected to a catalytic cascade is seen in systems as wide-ranging as the *Dictyostelium* sporulation pathway to ethylene signaling in *Arabidopsis thaliana* (Brown and Firtel, 1998; Thomason and Kay, 2000). More generally, we expect that our analytical result predicting robustness to changes in rate constants will apply to any stoichiometric system satisfying our assumptions. Metabolic pathways, many of which feature stoichiometric flows of material, in the form of metabolites rather than phosphoryl groups as was the case in our study, in principle can also display the patterns of robustness to genetic variation we uncovered in our study.

Furthermore, our results can be of practical importance in the design of anti-fungal agents. The osmoresponse pathway plays a critical role in regulating the virulence of fungal pathogens such as *Cryptococcus neoformans* (Bahn et al., 2007), but the yeast MAPK protein itself is not an optimal drug target due to its strong homology to the human p38 MAPK. As humans have no histidine phosphorelay, it is thought that the phosphotransfer proteins could potentially be valuable drug targets (Stephenson and Hoch, 2002; Stephenson and Hoch, 2004). Our findings that osmosensing signaling is robust to changes in the phosphorelay proteins suggest, however, that agents such as

competitive inhibitors of the phosphorelay components would have to entirely abolish interactions between the phosphorelay proteins to have any significant effect on downstream signaling and accordingly pathogenicity. Given these, our work underscores the importance of studying the organizational principles of biochemical circuits of protein networks in understanding cellular behaviors and their emergent properties.

MATERIALS AND METHODS

Strain background and construction

Our haploid ancestor strain used in the laboratory evolution experiment (DMY028) was derived from the DMY017 strain (Muzzey et al., 2009), the only difference being that it contained a plasmid bearing two TetO7 promoters, one of which drives the expression of CFP, while the other controls *YPD1* expression. In this strain, the *SHO1 ORF* was excised via standard PCR-based methods. The mutant strains referred to in this study were similarly derived from the DMY017 strain, except that the endogenous genes in the Sln1 branch of the HOG pathway were singly knocked out and replaced with its corresponding orthologs from various yeast species. Firstly, the endogenous genes were singly knocked out and replaced with the *Candida albicans URA3* gene using the pAG60 plasmid (Euroscarf). *SLN1* and *YPD1* gene deletions are lethal due to the hyperactivation of the pathway. To circumvent this, we knocked out these genes using a cassette containing both the *C. albicans URA3* gene and the Hog1 phosphatase *PTP2* placed under the control of the *ADH1* promoter. The orthologous genes from various yeast species were stitched to the 500-bp *S. cerevisiae* upstream and downstream gene flanking

sequences using overlap extension PCR. These final constructs were then transformed into the endogenous gene knockout strains described earlier, and single colonies were selected for the absence of *URA3* expression on 5-FOA plates. All integrations were subsequently confirmed by sequencing. The *PBS2* underexpressing strain was created by integrating the TetO7 promoter upstream of the endogenous *PBS2* gene. The two plasmids containing P_{GAL1}-*PBS2DD* and P_{GAL1}-*SSK2ΔN* were transformed separately into the P_{TetO7}-*PBS2* strain. The *YPD1* alleles were transformed into the P_{TetO7}-*PBS2* strain with the endogenous *YPD1* gene knocked out. A list of our yeast strains is provided in Table S4.

Growth and media conditions

Unless otherwise stated, all experiments were performed on exponentially growing cell cultures in synthetic dropout media with the appropriate amino acid supplements at 30 °C. The ancestral and evolved strains were grown consistently in 0.4 M NaCl for all experiments, except when their signaling abilities were analyzed upon a hyperosmotic shock of 1 M NaCl. In addition, all experiments involving the evolved strains were performed in the absence of doxycycline. Prior to the evolution experiment, the ancestral strain was grown overnight with doxycycline and the culture media was replaced with media without doxycycline before propagating them in the turbidostat (Acar et al., 2008). In experiments where cells were treated with doxycycline, a 5 μg/ml concentration was used.

Glycerol assays

Intracellular glycerol levels were measured using the Free Glycerol Reagent Kit (Sigma) as described (Muzzey et al., 2009). For details regarding the method and cell preparations, see the Supplemental Data.

Fluorescence microscopy and image analysis

Cell preparation and immobilization, and image acquisition and segmentation were performed as described (Mettetal et al., 2008). For our signaling experiments involving mutant strains with the orthologous pathway proteins, we corrected for any possible effects from outside the HOG pathway by measuring signaling in the respective pathway gene knockout strains in response to the same hyperosmotic shock ("basal signal"), and we subtracted this basal signal from that of the mutant strain's mean Hog1 trace. In addition, the reported Hog1 nuclear enrichment here represents the measured signal subtracted by the nuclear enrichment level prior to hyperosmotic shock.

SUPPLEMENTARY INFORMATION

Glycerol Assays

DMY028 and DMY028-derived evolved strains were grown overnight in 10 ml selective minimal media. Log-phase cells were spun at 2000 rpm for 2 minutes, and washed with 1 ml fresh media at the same [NaCl] as the original media in order to prevent internal glycerol leakage during washing. After another flash spin, cells were resuspended in 1 ml fresh media, and a 200 µl sample was incubated at 95 °C for 10 minutes and then spun at

13,000 rpm for three minutes to pellet cell debris. A sample of the supernatant was mixed with the Free Glycerol Reagent Kit (Sigma) as directed and then the OD_{540} was measured. The reported per optical density intracellular glycerol measurements are in units of OD_{540}/OD_{600}.

Sanger sequencing of genomic DNA sequences

50 µl PCR reactions targeting 1 kb upstream of the coding sequence of each of the HOG pathway genes and the synthetic constructs i.e. P_{MYO2}-*rtTA* and P_{TetO7}-*YPD1* were performed using Platinum TAQ DNA Polymerase High Fidelity (Invitrogen). A standard PCR protocol was used for all regions as recommended (Invitrogen). 1 µl of the eluted PCR product was added to 1 µl of a 5 µM forward sequencing primer and diluted with 10 µl of pure ddH_20. The forward primers were designed to tile the entire sample DNA sequence. Sequencing was performed using the Big Dye Terminator Cycle Sequencing kit (Applied Biosystems). Subsequently, DNA Baser (HeracleSoftware) was used to assemble and view the data, and to detect SNPs.

Mutation reconstruction in the ancestral strain

To reconstruct the confirmed mutations in the ancestral strain, the mutant alleles and flanking sequences were amplified by PCR and transformed into the ancestral strain with the endogenous gene replaced with *URA3*. The transformants were then selected on 5-FOA plates. Sequencing was performed to check that the mutant alleles were properly integrated, and were of the correct sequence.

Functional score metric calculations

We calculated functional score as the percentage of amino acid changes in the orthologous sequence at conserved residues identified through multiple sequence alignment of orthologous genes from twenty fungal species (Krantz et al., 2006). Here, we consider a residue as being conserved if either all the residues at that position are identical across all sequences in the alignment, or if conserved or semi-conserved substitutions are observed. To analyze the number of mutations found in our laboratory evolution experiment, we accounted for both gene length and the functional impact of each residue by computing the effective number of functionally important residues for each gene using a weighted method. We assume a 0.9 probability that changes in a conserved residue would impact function, and a 0.1 probability that a less conserved residue would change function. We then normalized the number of unique mutations found in the evolution experiment against this effective number of functionally important residues for all the HOG pathway genes (Fig. S6).

Relating mutational robustness to local biochemistry via throughput analysis

One possible mechanism that could explain the pattern of mutational robustness we observe experimentally is that the biochemical circuitry of the phosphorelay network renders the terminal phosphorelay protein insensitive to changes in kinetic parameters of its upstream pathway components. To mathematically determine the contribution of this effect, consider a chain of signaling proteins where the steady state phosphorylation level of any cascade protein consists of a basal phosphorylation level independent of pathway

activity, and an additional component that is inducible by the steady state phosphorylation level of its immediate upstream activator:

$$x_2' = x_2 + \frac{\partial x_2}{\partial x_1}(x_1' - x_1) \qquad [3]$$

Here, x_1 and x_2 are the basal phosphorylation levels of the 1st and 2nd proteins in the cascade, and primed symbols represent the total protein phosphorylation levels, while the partial derivative denotes the amount of phosphorylated 2nd proteins derived from every phosphorylated 1st protein. Extending these equations for the 3rd protein in the cascade yields:

$$x_3' = x_3 + \frac{\partial x_3}{\partial x_2}(x_2' - x_2) \qquad [4]$$

Substituting [3] into [4] we obtain:

$$x_3' = x_3 + \frac{\partial x_3}{\partial x_2}\frac{\partial x_2}{\partial x_1}(x_1' - x_1) \qquad [5]$$

Extending the analysis for the j^{th} protein in the cascade, we obtain:

$$x_j' = x_j + \prod_{i=2}^{j}\frac{\partial x_i}{\partial x_{i-1}}(x_1' - x_1) \qquad [6]$$

From [6], it is clear that the biochemical details of signal transmission are buried mathematically in the chain of derivatives i.e. they represent how the activity of the cascade protein furthest upstream is transduced into changing the activity of the j^{th} cascade protein. For example, the contribution of the k^{th} protein to the chain of derivatives arises from two factors i.e. the effect of the $(k-1)^{th}$ protein on the activation of the k^{th} protein and the effect of the k^{th} protein on the activation of the $(k+1)^{th}$ protein:

$$\frac{\partial x_{k+1}}{\partial x_k}\frac{\partial x_k}{\partial x_{k-1}} = \frac{\partial x_{k+1}}{\partial x_{k-1}} \equiv \xi_k \qquad [7]$$

where we term ξ_k the steady state throughput of the kth protein.

The central claim of the throughput analysis is that the sensitivity of ξ_k to changes in parameters describing the kth protein can predict to what extent sequence changes in the kth protein will be tolerated by the system. An important corollary to this claim is that if ξ_k is invariant under parameter variations, then sequence changes in the kth protein will not affect signaling unless the sequence changes completely inactivate the protein altogether. To put this analysis into effect, we used a simplified model of the HOG pathway (Klipp et al., 2005):

$$\begin{aligned}
\frac{d[Sln1P]}{dt} &= k_1\left(\frac{\Pi(t)}{\Pi_0(t)}\right)^2[Sln1] + k_{-2}[Ypd1P][Sln1] - k_2[Ypd1][Sln1P] \\
\frac{d[Ypd1P]}{dt} &= k_2[Ypd1][Sln1P] - k_{-2}[Ypd1P][Sln1] - k_3[Ypd1P][Ssk1] \\
\frac{d[Ssk1P]}{dt} &= k_3[Ypd1P][Ssk1] - k_{-3}[Ssk1P]
\end{aligned} \qquad [8]$$

$$\begin{aligned}
\frac{d[Ssk2P]}{dt} &= k_4[Ssk2][Ssk1] - k_{-4}[Ssk2P] \\
\frac{d[Pbs2P]}{dt} &= k_5[Pbs2][Ssk2P] - k_{-5}[Pbs2P] \\
\frac{d[Hog1P]}{dt} &= k_6[Hog1][Pbs2P] - k_{-6}[Hog1P]
\end{aligned}$$

Since the signaling dynamics are fast relative to the osmotic pressure variable, separation of timescales allows one to treat the signaling system as if it were in steady state at every moment in time (the signaling pathway adiabatically follows the osmotic pressure dynamics, readjusting itself to the osmotic pressure variable at every point in time). To

determine the effect of local biochemistry on ξ_k, we examined the two most architecturally distinct proteins i.e. Pbs2 and Ypd1 whereby Pbs2 is a kinase sandwiched between similar kinase proteins, while Ypd1 is a phosphotransfer protein sandwiched between similar phosphotransfer proteins.

$$\xi_{Pbs2} \equiv \frac{\partial[Hog1P]}{\partial[Ssk2P]} = \frac{k_5 k_{-5} k_6 k_{-6}\, Pbs2_T\, Hog1_T}{[k_5 k_6 [Ssk2P]\, Pbs2_T + k_{-6}(k_5[Ssk2P] + k_{-5})]^2} \qquad [9]$$

From this expression, we observe that ξ_{Pbs2} depends on Pbs2 interaction parameters i.e. phosphorylation rate of Pbs2 and Hog1 etc. Changes in Pbs2 sequence can alter these rates and affect the steady state throughput, and can impact Hog1 phosphorylation levels. On the other hand, the throughput of Ypd1 is:

$$\xi_{Ypd1} \equiv \frac{\partial[Ssk1]}{\partial[Sln1]} = -\frac{k_1 (\frac{\Pi(t)}{\Pi_0(t)})^2}{k_{-3}} \qquad [10]$$

Remarkably, ξ_{Ypd1} is independent of Ypd1 parameters. This implies that, as a direct consequence of the local architecture of the network of biochemical reactions, Hog1 phosphorylation is shielded from potential changes in Ypd1 rate constants.

Sensitivity analysis of a model of the HOG pathway

We implemented the following steps: i) model changes in sequence as changes in kinetic rate constants, ii) define a sensitivity metric that captures how HOG signaling changes as kinetic rate constants are varied using the model presented in [8]. We examined several methods to execute step (ii). The first analysis involved computing the magnitude of the local logarithmic gradient about the wild-type parameter set from the model outputs namely initial Hog1 phosphorylation rate and the steady state Hog1 phosphorylation

level. To directly compare the different model outputs, we utilized logarithmic gradient calculations to render our analysis dimensionless:

$$\left\|\tilde{\nabla}\varphi\right\| = \sqrt{\sum_{i=1}^{3}\left(\frac{\partial \ln \varphi}{\partial \ln k_i}\bigg|_{wt}\right)^2} \qquad [11]$$

where ϕ is the model output whose sensitivity we are computing, and the k's represent the rate constants that are being varied. The wild-type parameters are obtained from Klipp *et al.* although similar results are obtained in a model with wild-type rate constants set equal to one another. The results of this analysis for ϕ are summarized in Figs. 1d and 1e.

To overcome the uncertainty in the wild-type parameters used in [11], we formulated a 2nd metric that is less dependent on the choice of the particular wild-type parameters. This method utilizes the full distribution of ϕ, instead of the only region around the wild-type level, and measures the relative spread of this distribution to determine the effects of variations in rate constants on ϕ. Using the same model outputs, we computed the following modified deviation metric:

$$\tilde{\sigma} = \sqrt{\frac{1}{V}\sum_{\vec{k}}\left(\varphi(\vec{k}) - \varphi_{wildtype}\right)^2} \qquad [12]$$

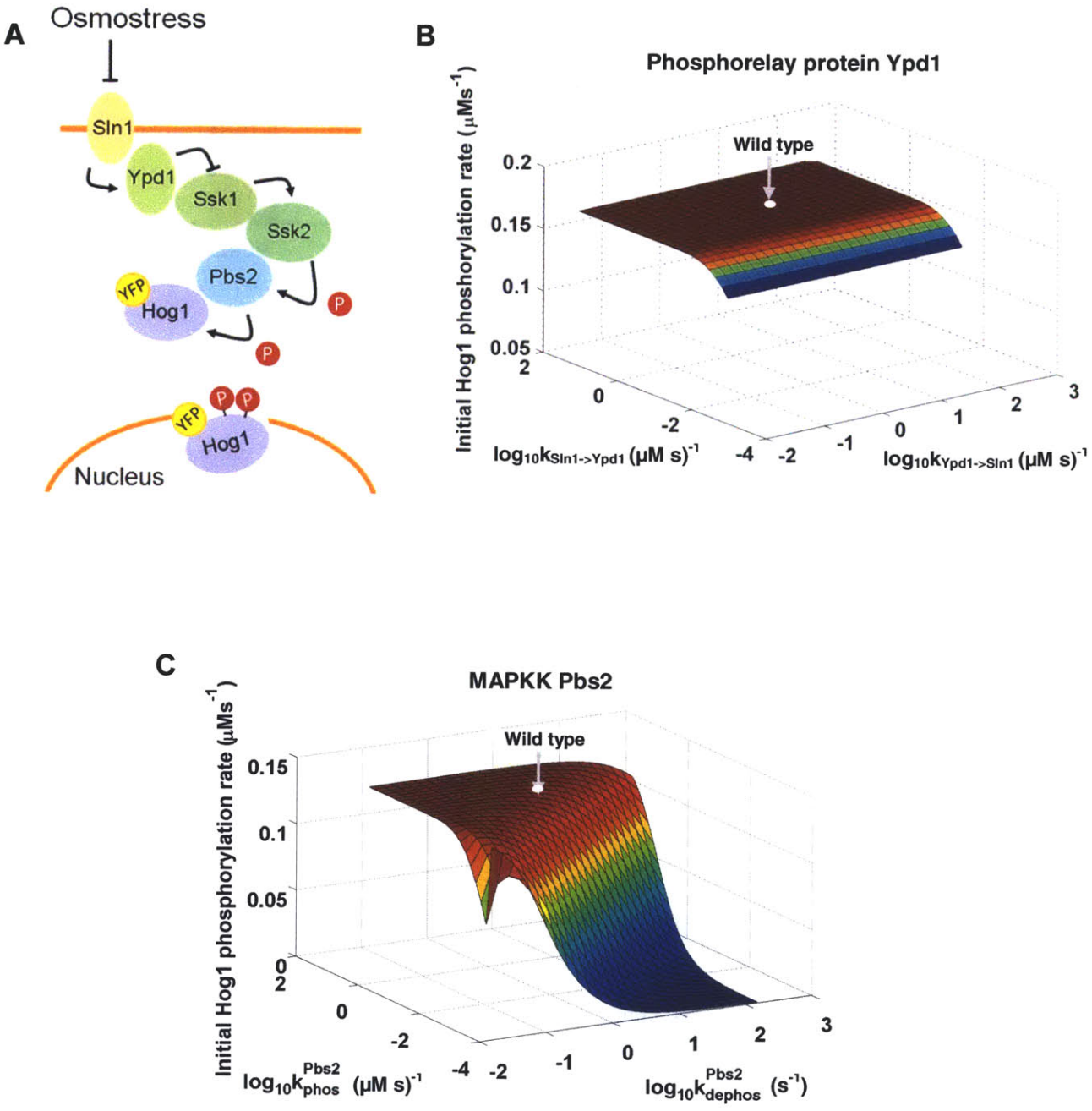
where V is the phase space volume over which the parameters are swept. Similar to the local logarithmic gradient, large values of the modified standard deviation indicate greater sensitivity to parameter variations, while smaller values indicate greater robustness to parameter variations. The results of this analysis are shown in Fig. S1. In summary, both analyses highlighted above yielded the same qualitative answer i.e. HOG

signaling is most affected by changes in the rate constants of the downstream MAPK proteins and least by the upstream phosphorelay proteins.

All the rate equations and parameters used are provided in Table S1.

FIGURES AND TABLES

Fig 1



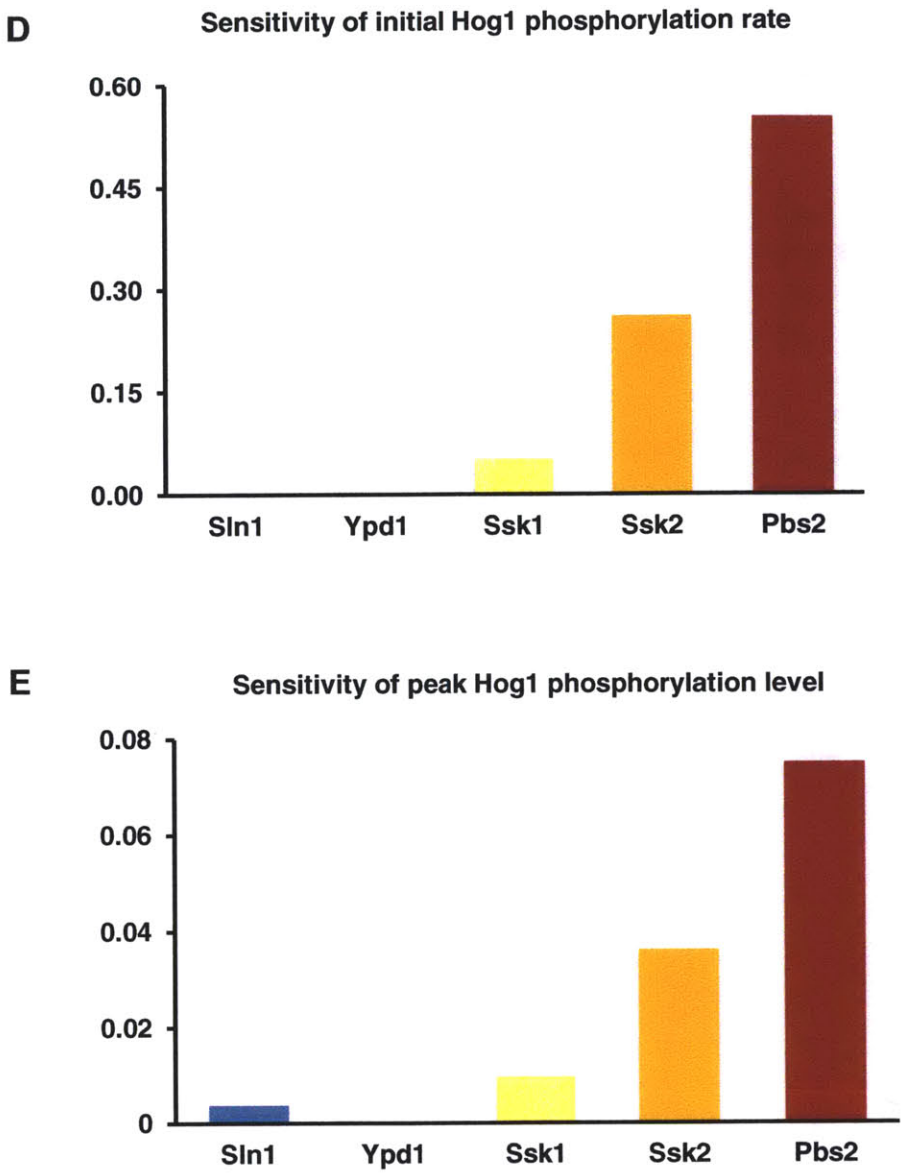
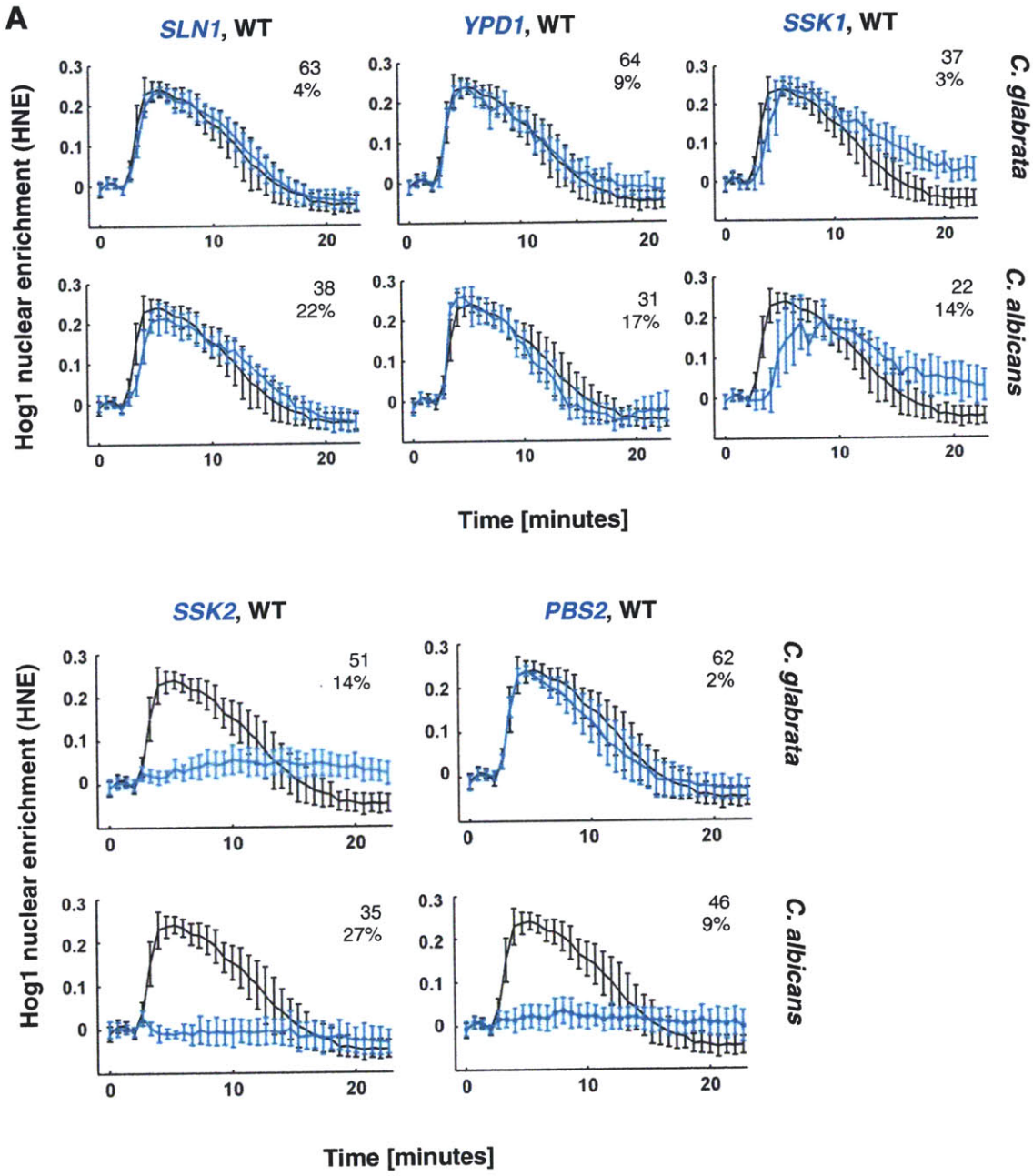
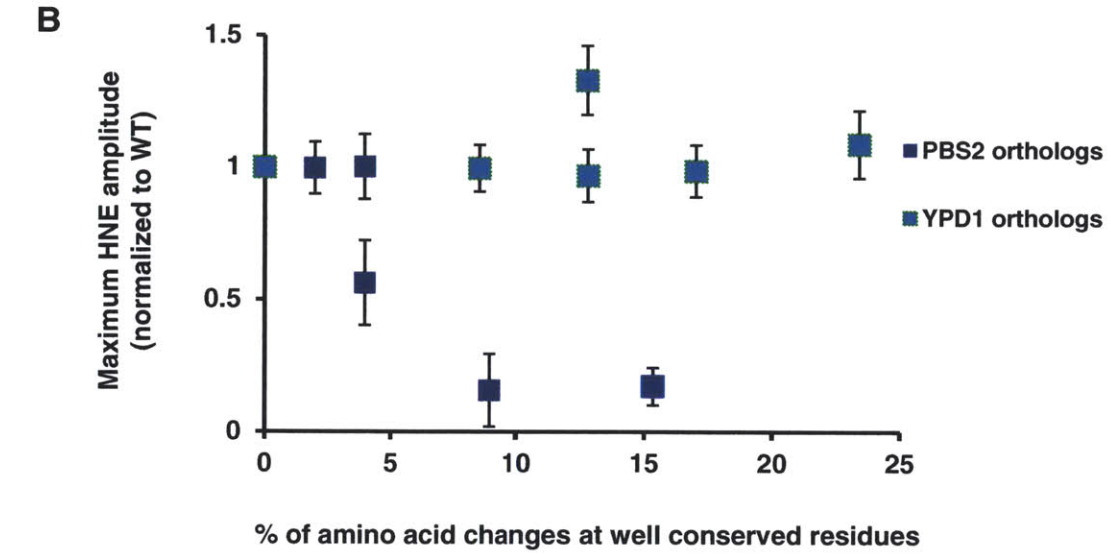


Figure 1 | Computational analysis of sensitivity of HOG pathway signaling to kinetic parameter changes. A, Schematic representation of the SLN1 branch of the HOG signaling pathway. Sln1 is the main osmosensor in our strain since *SHO1* is deleted. **B**, Changes in initial Hog1 phosphorylation rate upon varying two of the kinetic rate constants associated with Ypd1 (i.e. Sln1-to-Ypd1 and Ypd1-to-Sln1 phosphotransfer rates) over two orders of magnitude about wild-type levels. **C**, Changes in initial Hog1 phosphorylation rate upon varying two of the kinetic rate constants associated with Pbs2

(i.e. Pbs2 phosphorylation and dephosphorylation rate constants) over two orders of magnitude about wild-type levels. **D**, Distribution of sensitivity (i.e. magnitude of local logarithm gradient of the surface shown in A-B evaluated at wild-type levels; see supplemental data) of initial Hog1 phosphorylation rate across the HOG pathway genes upstream of *HOG1*. **E**, Distribution of sensitivity of peak Hog1 phosphorylation level across the HOG pathway genes upstream of *HOG1*.

Fig 2


A
SLN1, WT
YPD1, WT
SSK1, WT
63
4%
64
9%
37
3%
38
22%
31
17%
22
14%
C. glabrata
C. albicans
Hog1 nuclear enrichment (HNE)
Time [minutes]
SSK2, WT
PBS2, WT
51
14%
62
2%
35
27%
46
9%
C. glabrata
C. albicans
Hog1 nuclear enrichment (HNE)
Time [minutes]



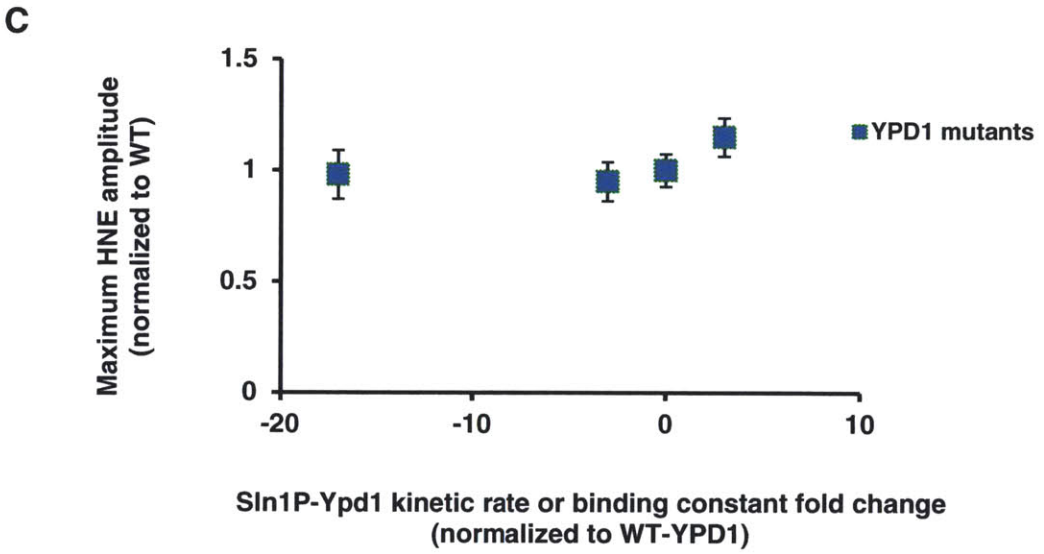
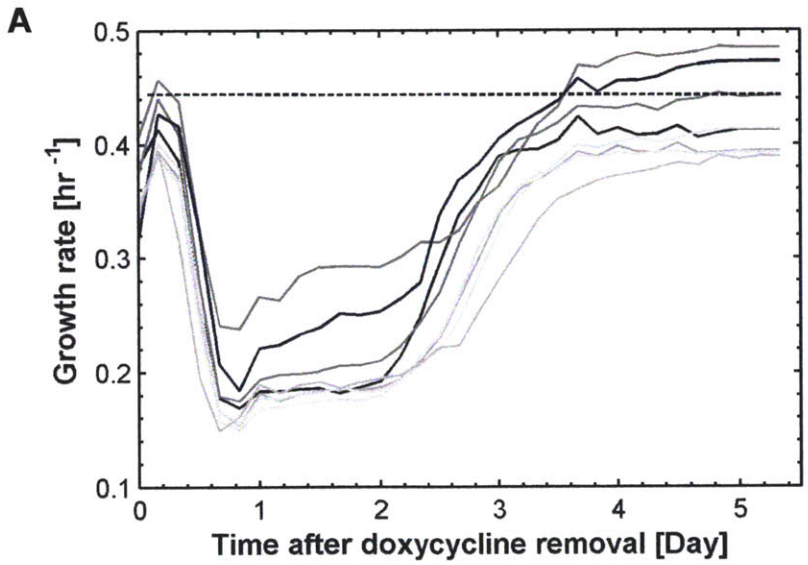


Figure 2 | Effects of ortholog substitutions of HOG pathway component genes on signal propagation. **A**, Hog1 nuclear enrichment dynamics in response to a 0.4 M NaCl hyperosmotic shock measured in the wild-type strain (in black) and mutant strains with the indicated pathway proteins (in cyan)replaced with its orthologs from *C. glabrata* and *C. albicans*. Shown in the upper right corner of each plot is the ClustalW score of the ortholog when aligned to the *S. cerevisiae* sequence. Right below the ClustalW score is the functional score for each ortholog, which represents the percentage of amino acid

changes at highly conserved residues identified via comparative genomics. The traces show the average response, obtained by taking the average of population averages from independent experiments (n = 3) ± SEM. **B**, Maximum Hog1 nuclear enrichment of mutant strains with orthologous *YPD1* and *PBS2* genes of varying degrees of functional scores under a 0.4 M NaCl hyperosmotic shock normalized against the wild-type response. Data point at 0 percentage change represents the wild-type response. Data depicts mean (n = 3) ± SEM. **C**, Maximum Hog1 nuclear enrichment of mutant strains with characterized *YPD1* alleles under a 0.4 M NaCl hyperosmotic shock normalized against the wild-type response. Two of the alleles exhibit a three- and seventeen-fold reduction in the Sln1-to-Ypd1 phosphotransfer rate $k_{Sln1P\text{-}Ypd1}$, while another has a three-fold increase in the binding constant $Kd_{Sln1P\text{-}Ypd1}$ compared to wild-type Ypd1 (Janiak-Spens et al., 2004) (Table S2). Data point at 1 fold change represents the wild-type response. Data depicts mean (n = 3) ± SEM.

Fig 3



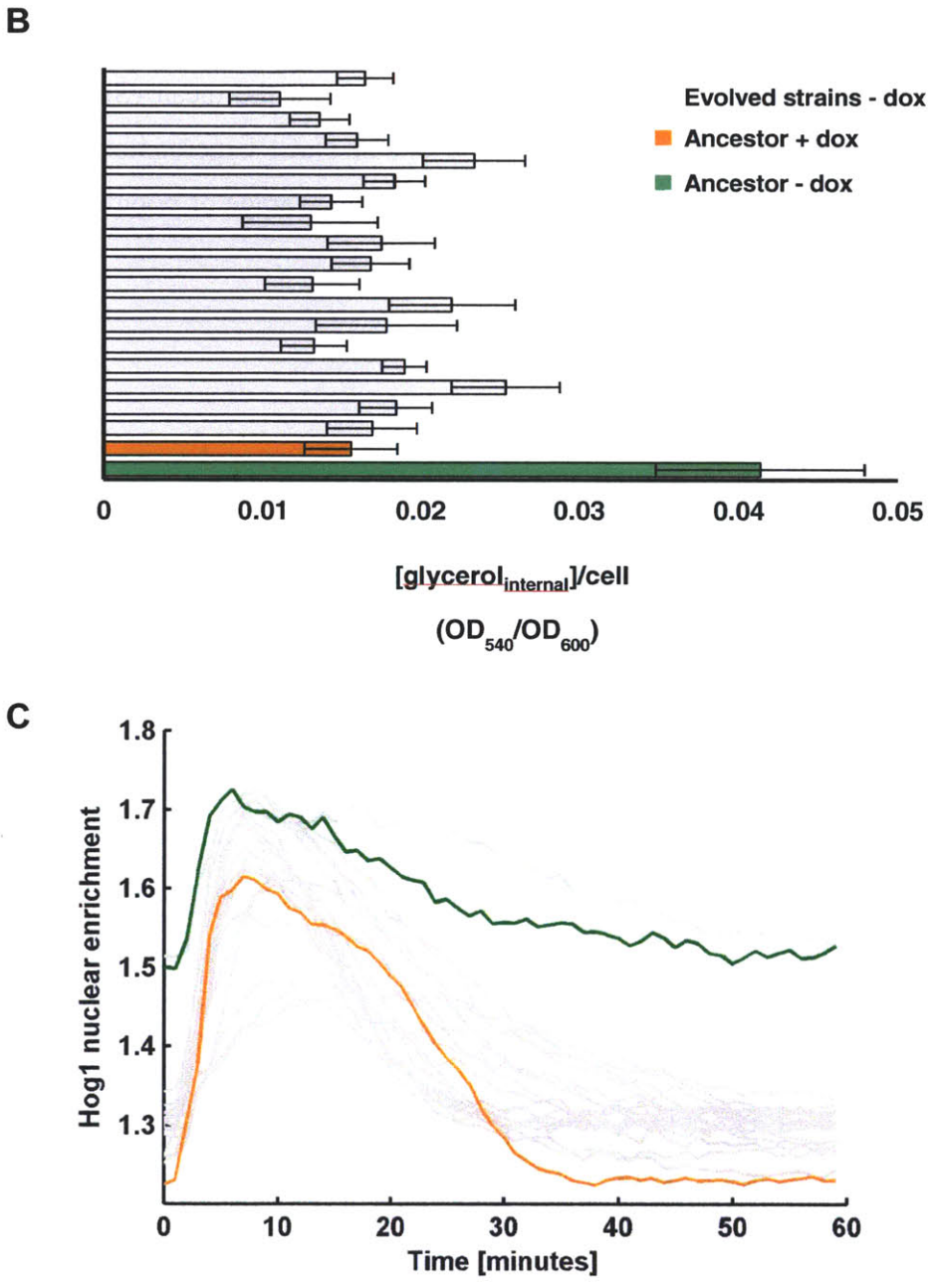
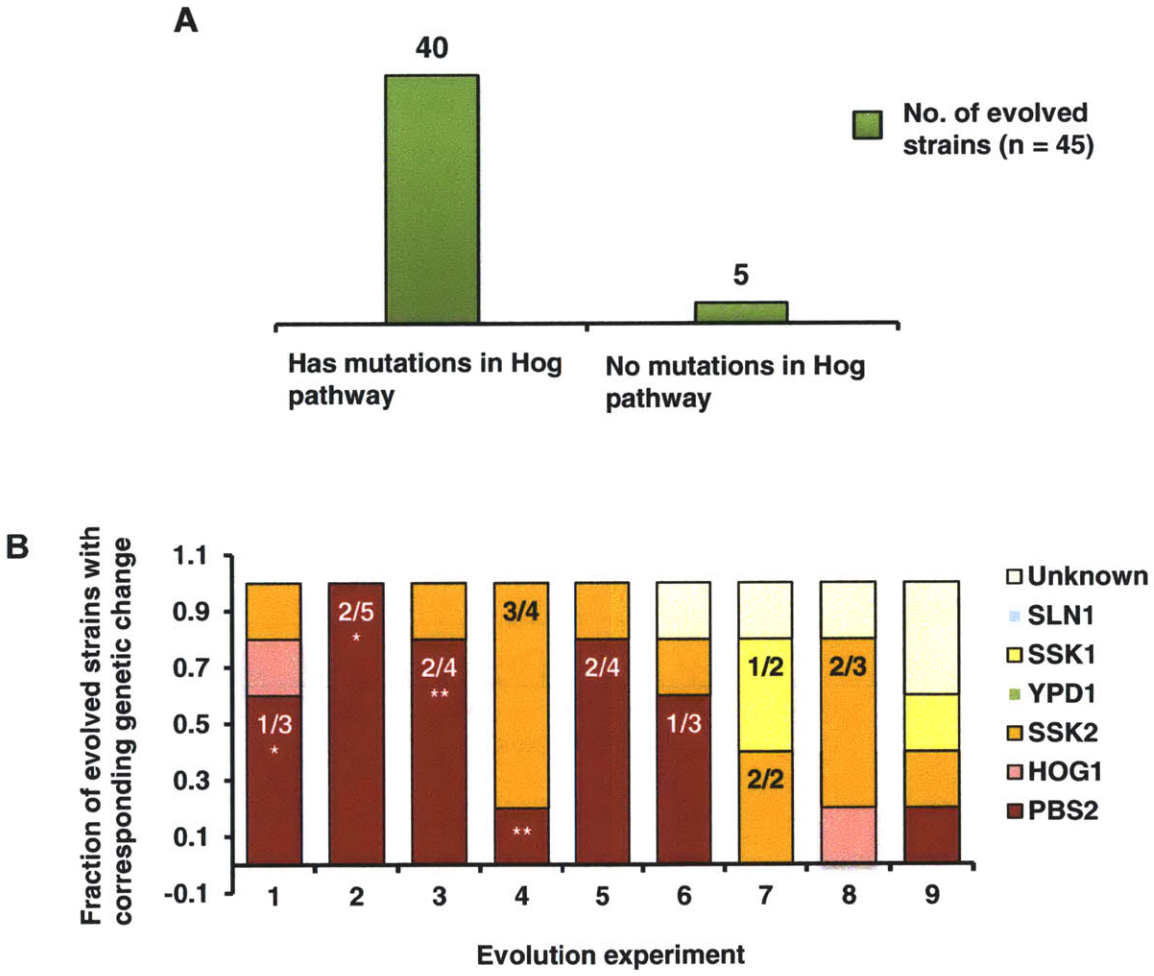


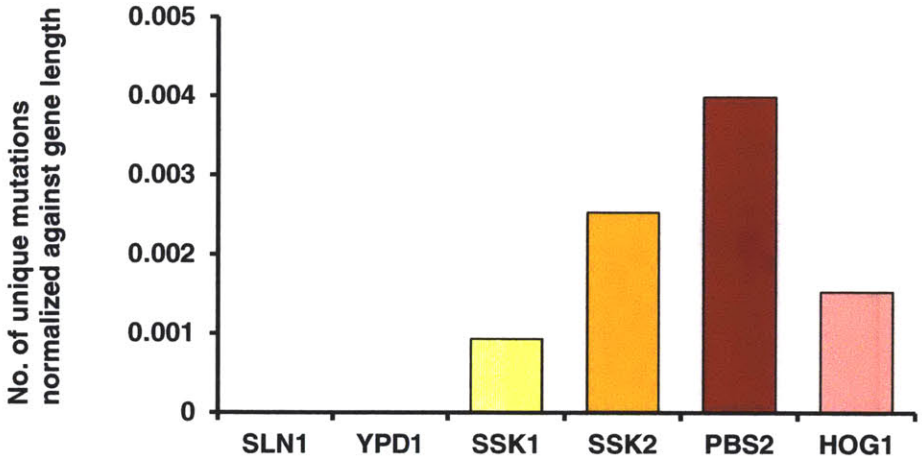
Figure 3 | Ypd1 underexpressing yeast cells with pathologically hyperactive HOG signaling rapidly evolve to restore wild-type growth and signaling. A, Adaptation dynamics across nine independent evolution experiments. The horizontal dotted line represents the growth rate of the ancestor cells induced with doxycycline. **B**, Intracellular glycerol concentrations (OD_{540} measurement; see Experimental Procedures) normalized

against cell growth (OD_{600} measurement) were measured in the ancestor cells in the presence and absence of doxycycline, and in the evolved strains. Data represents the mean of three independent experiments ± standard error of the mean (SEM). **C**, Hog1 nuclear enrichment dynamics in response to a 0.6 M NaCl hyperosmotic shock. The traces show the average response, obtained by taking the average of population averages from independent experiments (n = 3).

Fig 4

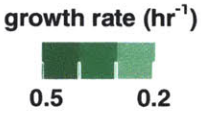


C



D

Gene/Mutation	γ [hr^{-1}] A	γ [hr^{-1}] B	Protein region
SSK2/V402A			Ssk1 binding domain, essential for Ssk2 activation
SSK2/W427C			Essential for Ssk2 activation
SSK2/C1172Y			Kinase domain
SSK2/P1393L			Kinase domain
SSK2/P1466L			Kinase domain
SSK2/G1471V			Kinase domain
SSK2/W1557C			Kinase domain
PBS2/Y43D			Docking site for Ssk2
PBS2/R61L			Docking site for Ssk2
PBS2/G423D			Kinase domain
PBS2/G509S			Kinase domain
PBS2/M526R			Kinase domain
PBS2/R640 to STOP			NLS?
SSK1/I504F			Close to response regulator receiver domain



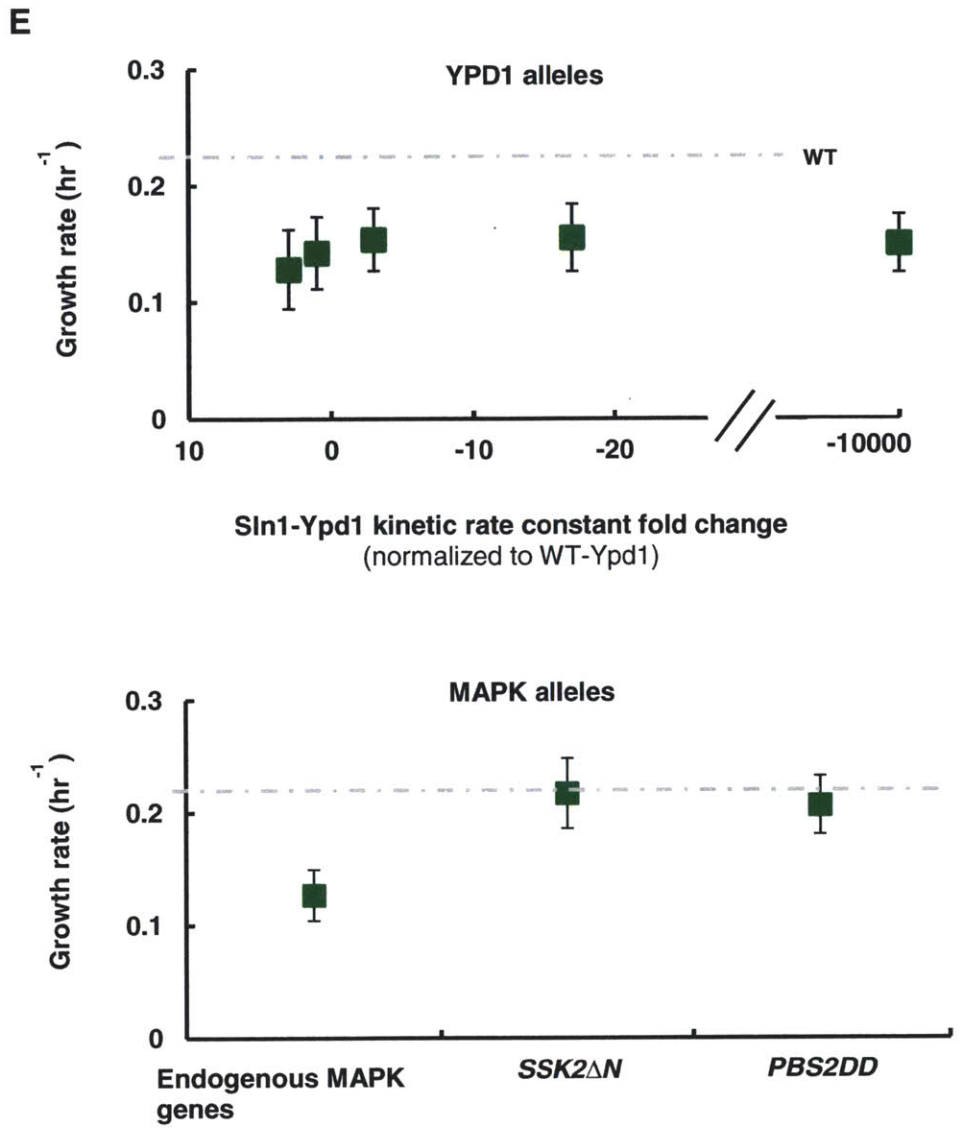
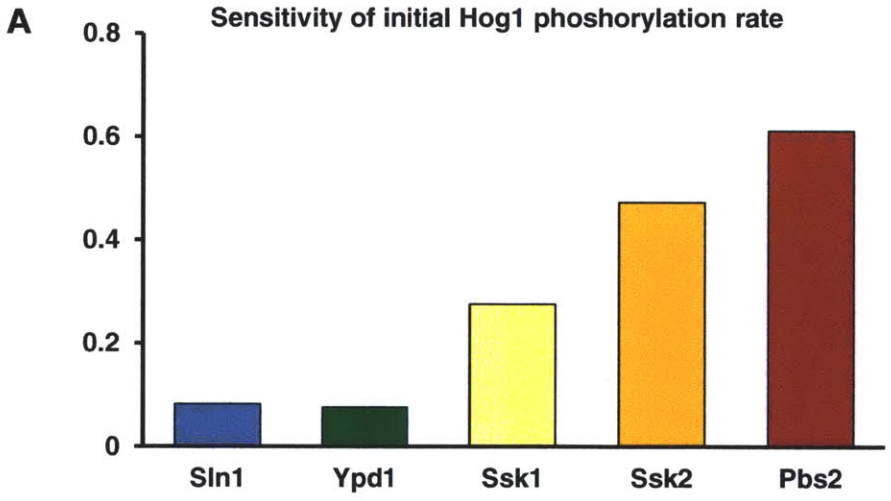


Figure 4 | Characterization of the molecular changes across independently evolved populations. A, Distribution of evolved strains with mutations in HOG pathway genes (n = 45). Data depicts five randomly selected evolved strains across nine independent experiments. **B**, Distribution of genetic changes in evolved strains corresponding to (a) across the nine experiments (n = 5). Indicated inside bars are the fractions of unique gene mutations observed in individual experiments. Notations * and ** represent two particular mutations which were found in independent experiments. **C**, Distribution of

unique HOG pathway gene mutations in evolved strains corresponding to panel **a** normalized against gene length. **D**, Growth rates γ of the "transformed strains" (i.e. with the endogenous ancestral gene replaced with fourteen of the randomly selected mutant alleles) (column B), and the corresponding evolved strains under no doxycycline conditions (column A). Data represents mean (n = 3) ± SEM. **E**, Comparison of growth rates of hypoactive pathway strains in 1.2 M NaCl containing loss-of-function *YPD1* alleles versus gain-of-function *PBS2* and *SSK2* alleles. Data represents mean (n = 3) ± SEM. Dotted line shows the growth rate of a wild-type strain in 1.2 M NaCl with *PBS2* fully expressed.

SUPPLEMENTARY FIGURES AND TABLES

Fig S1



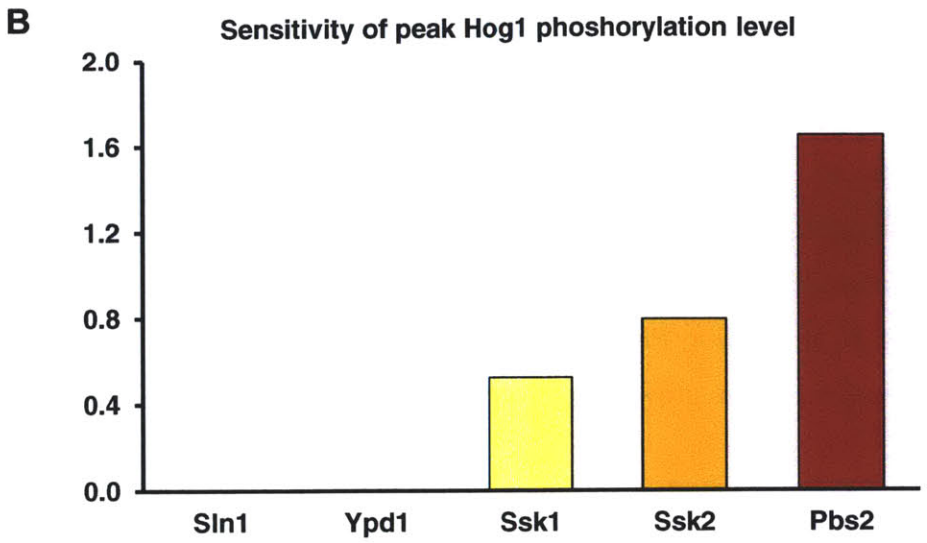


Figure S1 | Sensitivity analysis results obtained using modified standard deviation. Unlike the local logarithmic gradient measure, the modified standard deviation makes use of the full distribution of model outputs upon varying the rate constants. Sensitivity of the initial Hog1 phosphorylation rate (A) and sensitivity of peak Hog1 phosphorylation level (B) to rate constant changes in the HOG pathway proteins.

Fig S2

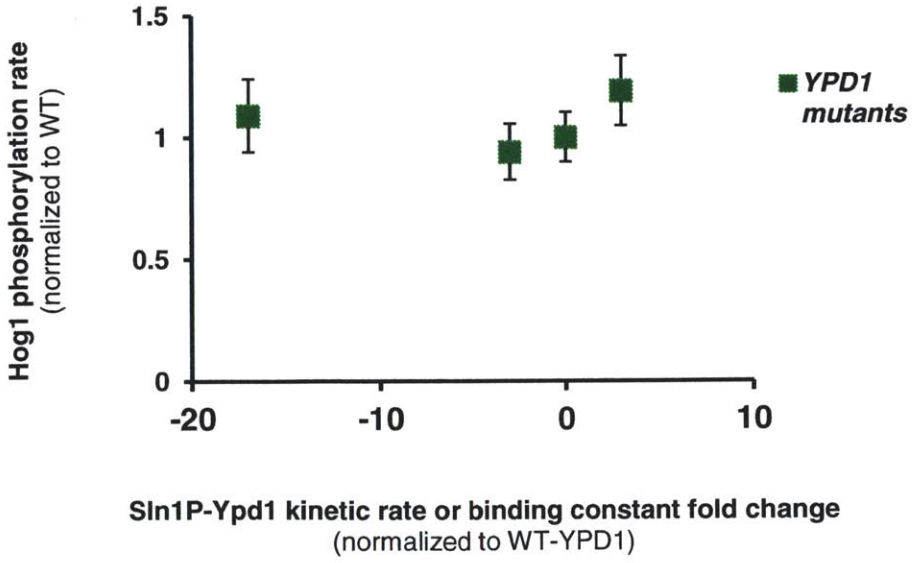


Figure S2 | Hog1 phosphorylation rate of mutant strains with characterized *YPD1* alleles under a 0.4 M NaCl hyperosmotic shock. Two of the alleles exhibit a three- and seventeen-fold reduction in the Sln1-to-Ypd1 phosphotransfer rate $k_{Sln1P\text{-}Ypd1}$, while another has a three-fold increase in the binding constant $Kd_{Sln1P\text{-}Ypd1}$ compared to wild-type Ypd1 (Janiak-Spens et al., 2005) (Table S2). Data point at 0 fold change represents the wild-type response. Data depicts mean (n = 3) ± SEM.

Fig S3

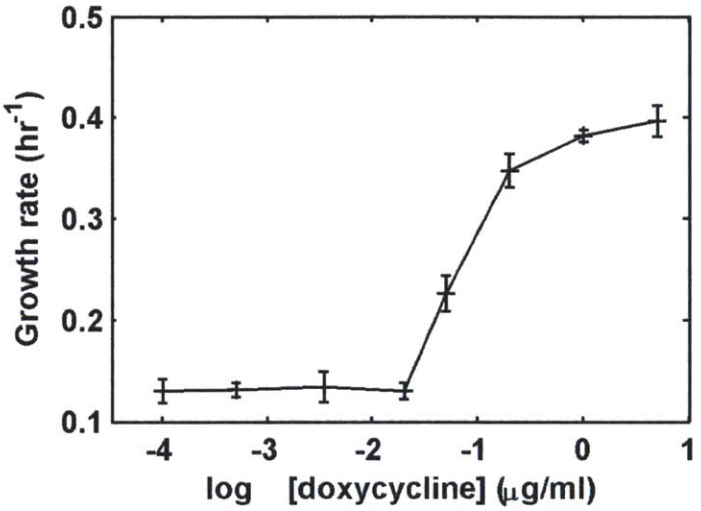


Figure S3 | Characterization of growth rates of ancestor yeast cells at different doxycycline concentrations. Cell densities at OD_{600} were measured. Data depicts mean (n = 3) ± SEM.

Fig S4

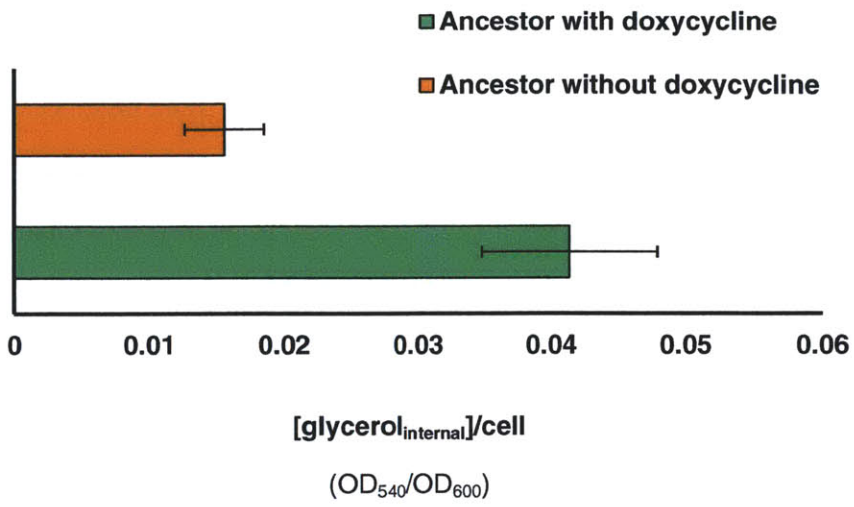


Figure S4 | Characterization of intracellular glycerol content of ancestor cells treated with and without doxycycline. Intracellular glycerol concentrations (OD_{540} measurement; see Experimental Procedures) normalized against cell growth (OD_{600} measurement) were measured in the ancestor cells in the presence and absence of doxycycline. Data depicts mean (n = 3) ± SEM.

Fig S5

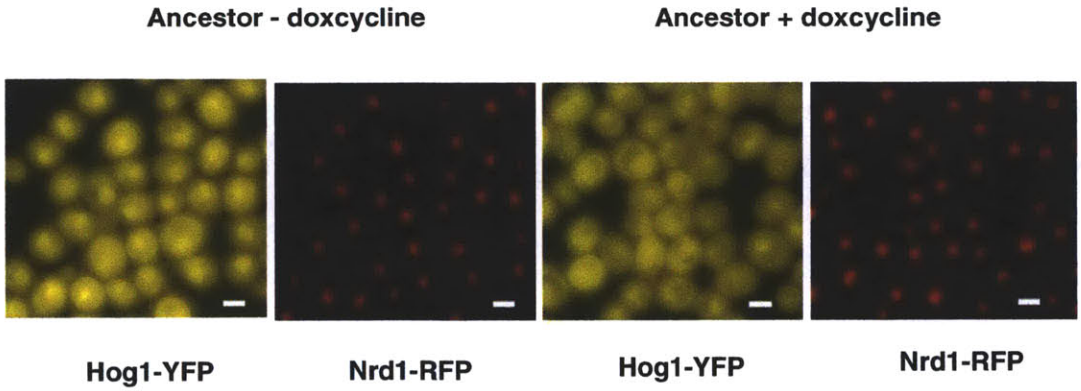


Figure S5 | Basal Hog1 nuclear enrichment at different doxycycline levels. Cells were grown in media at the respective doxycycline concentrations for approximately sixteen hours, after which they were imaged under the microscope. In cells treated with doxycycline (5 μg/ml), Hog1-YFP is distributed throughout the cytoplasm. By contrast, Hog1-YFP accumulates in the nuclei of cells grown in the absence of doxycycline. The Nrd1-RFP signal tracks the position of the nucleus. Scale bar = 2μm.

Fig S6

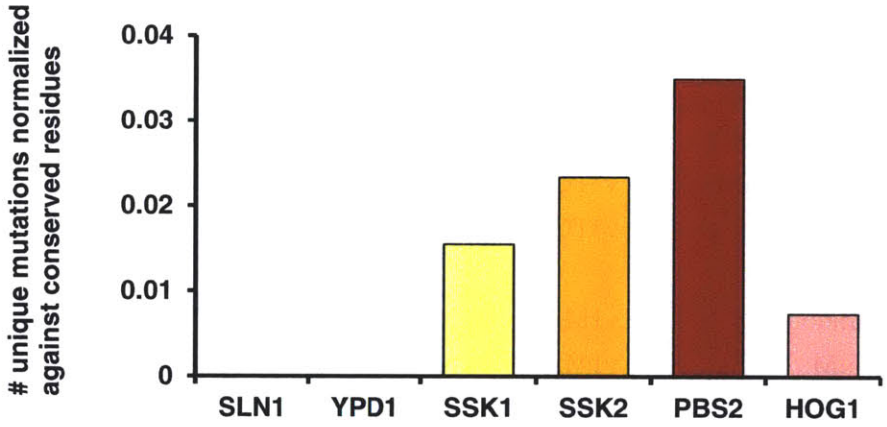


Figure S6 | Distribution of unique HOG pathway gene mutations in evolved strains normalized against the effective number of functionally important residues for each gene.

Table S1 | Rate equations and parameters

%parameters	%rate equations
k1TCS = 5 s^{-1};	Sln1P = y(1);
k2TCS_plus = 50 (μM s)$^{-1}$;	Ypd1P = y(2);
k2TCS_minus = 50 (μM s)$^{-1}$;	Ssk1P = y(3);
k3TCS_plus = 50 (μM s)$^{-1}$;	Ssk2P = y(4);
k4TCS_plus = 0.415 s^{-1};	Pbs2P = y(5);
	Hog1P = y(6);
k1MAP_plus = 1.538 (μM s)$^{-1}$;	
k1MAP_minus = 0.011 s^{-1};	v1TCS = k1TCS*(.5^2)*(Sln1T-Sln1P);
k2MAP_plus = 1.538 (μM s)$^{-1}$;	v2TCS = k2TCS_plus*Sln1P*(Ypd1T-Ypd1P) -
k2MAP_minus = 0.011 s^{-1};	k2TCS_minus*(Sln1T-Sln1P)*Ypd1P;
k3MAP_plus = 1.538 (μM s)$^{-1}$;	v3TCS = k3TCS_plus*(Ssk1T-Ssk1P)*Ypd1P;
k3MAP_minus = 0.011 s^{-1};	v4TCS = k4TCS_plus*(Ssk1P);
%total concentration	v1MAP_plus = k1MAP_plus*(Ssk2T-Ssk2P)*(Ssk1T-Ssk1P);
Sln1T = 0.016 μM;	v1MAP_minus = k1MAP_minus*Ssk2P;
Ypd1T = 0.156 μM;	v2MAP_plus = k2MAP_plus*(Pbs2T-Pbs2P)*Ssk2P;
Ssk1T = 0.029 μM;	v2MAP_minus = k2MAP_minus*Pbs2P;
Ssk2T = 0.0067 μM;	v3MAP_plus = k3MAP_plus*(Hog1T-Hog1P)*Pbs2P;
Pbs2T = 0.053 μM;	v3MAP_minus = k3MAP_minus*Hog1P;
Hog1T = 0.167 μM;	%equations
%initial concentration	dy(1) = v1TCS-v2TCS;
Sln1P^0 = 2.25 x 10^{-3} μM;	dy(2) = v2TCS-v3TCS;
Ypd1P^0 = 36 x 10^{-3} μM;	dy(3) = v3TCS-v4TCS;
Ssk1P^0 = 1.88 x 10^{-3} μM;	
Ssk2P^0 = 1.394 x 10^{-3} μM;	dy(4) = v1MAP_plus-v1MAP_minus;
Pbs2P^0 = 0.0101 μM;	dy(5) = v2MAP_plus-v2MAP_minus;
Hog1P^0 = 0.088 μM;	dy(6) = v3MAP_plus-v3MAP_minus;
	dy = dy';

Parameter values are obtained from Klipp *et al.*

Table S2 | *YPD1* alleles and their measured in-vitro biochemical kinetic constants (Janiak-Spens et al., 2005)

YPD1	$k_{Sln1P->Ypd1}$ (s^{-1})	K_d (μM)
WT	29 ± 3	1.4 ± 0.3
K67A	33 ± 4	4.2 ± 1.5
R90A	11 ± 1	1.4 ± 0.6
Q86A	1.7 ± 0.3	1.4 ± 0.8
G68Q	0.003	~2

Table S3 | All unique HOG pathway mutations found in 45 evolved strains across 9 evolution experiments

Evolution experiment (a)									Gene	Chro	Genome position	Ancestral nucleotide	Evolved nucleotide	Gene position	Codon change	Amino acid change	Protein region	Impact (b)
1	2	3	4	5	6	7	8	9										
									SSK1	12	162383	A	T	1510 (504)	ATT -> TTT	Ile->Phe	Close to response regulator receiver domain	DEL 1.52 86
									SSK1	12	161959	A	G	1934 (645)	GAG -> GGG	Glu->Gly	Response regulator receiver domain	DEL 1.52 82
									SSK2	14	681882	T	C	1205 (402)	GTG -> GCG	Val->Ala	Ssk1 binding domain, essential for Ssk2 activation	DEL 1.93 73
									SSK2	14	684156	G	C	1281 (427)	TGG -> TGC	Trp->Cys	Essential for Ssk2 activation	DEL 1.93 73
									SSK2	14	681922	G	A	3515 (1172)	TGC -> TAC	Cys->Tyr	Unknown	DEL 1.99 70
									SSK2	14	681923	C	T	3516 (1172)	TGC -> TGT	Cys->Cys		
									SSK2	14	681259	C	T	4178 (1393)	CCC -> CTC	Pro->Leu	Kinase domain	DEL 1.92 92
									SSK2	14	680766	G	C	4671 (1557)	TGG -> TGC	Trp->Cys	Kinase domain	DEL 1.94 89
									SSK2	14	681025	G	T	4412 (1471)	GGA -> GTA	Gly->Val	Kinase domain	DEL 1.94 89
									SSK2	14	681040	C	T	4397 (1466)	CCA -> CTA	Pro->Leu	Kinase domain	DEL 1.92 91
									SSK2	14	684238	A	T	1199 (400)	GAT -> GTT	Asp->Val	Ssk1 binding domain, essential for Ssk2 activation	DEL 1.91 78
									SSK2	14	682852	G	A	2585 (862)	TGT -> TAT	Cys->Tyr	Unknown	TOL 2.22 24
									SSK2	14	681620	G	C	3817 (1273)	GGT -> CGT	Gly->Arg	Kinase domain	DEL 1.90 92
									SSK2	14	681049	A	G	4388 (1463)	TAC -> TGC	Tyr->Cys	Kinase domain	DEL 1.91 90
									SSK2	14	680990	G	T	4447 (1483)	GTT -> TTT	Val->Phe	Kinase domain	DEL 1.91 90
									HOG1	12	372203	G	T	583 (195)	GAC -> TAC	Asp->Tyr	Kinase domain	DEL 3.43 98
									HOG1	12	372103	C	G	483 (161)	TGC -> TGG	Cys->Trp	Kinase domain	DEL 3.34 98
									PBS2	10	179030	T	C	1070 (357)	TTG -> TCG	Leu->Ser	Close to kinase domain	DEL 1.97 95
									PBS2	10	179942	G	A	158 (53)	CGT -> CAT	Arg->His	Docking site for Ssk2	DEL 2.45 18
									PBS2	10	179973	T	G	127 (43)	TAC -> GAC	Tyr->Asp	Docking site for Ssk2	TOL 2.33 17
									PBS2	10	179918	G	T	182 (61)	CGT -> CTT	Arg->Leu	Docking site for Ssk2	DEL 2.38 16
									PBS2	10	178832	G	A	1268 (423)	GGT -> GAT	Gly->Asp	Kinase domain	DEL 1.94 97
									PBS2	10	178575	G	A	1525 (509)	GGT -> AGT	Gly->Ser	Kinase domain	DEL 1.94 97
									PBS2	10	178523	T	G	1577 (526)	ATG -> AGG	Met->Arg	Kinase domain	DEL 1.94 97
									PBS2	10	178182	C	T	1918 (640)	CGA -> TGA	Arg->Stop	NLS?	-

(a) Number of occurrences in which the mutation was observed in the five randomly selected evolved colonies from each of the nine experiments.

(b) The impact of the mutation on protein function was predicted using the SIFT software (Ng and Henikoff, 2001). The numbers represent the median sequence conservation and the number of sequences sampled at the amino acid position respectively.

Table S4 | List of yeast strains and plasmids used

Strain or plasmid	Genotype	Reference or source
Strain		
DMY017	*BY4741; MATa YER118c::kanMX4 HOG1::HOG1-YFP-HIS3 NRD1::NRD1-mRFP1-natR*	Muzzey et al.
DMY027	*DMY017; MYO2::P_{MYO2}-rtTA-LEU2 SSK22::URA3 P_{PBS2}::P_{TETO7}-CFP-kanMX-P_{TET07}*	
DMY028	*DMY017; MYO2::P_{MYO2}-rtTA-LEU2 SSK22::URA3 P_{YPD1}::P_{TETO7}-CFP-kanMX-P_{TET07}*	
DMY028-ev1	*DMY028; SSK1::SSK1^{I504F}*	
DMY028-ev2	*DMY028; SSK1::SSK1^{E645G}*	
DMY028-ev3	*DMY028; SSK2::SSK2^{V402A}*	
DMY028-ev4	*DMY028; SSK2::SSK2^{W427C}*	
DMY028-ev5	*DMY028; SSK2::SSK2^{C1172W}*	
DMY028-ev6	*DMY028; SSK2::SSK2^{P1393L}*	
DMY028-ev7	*DMY028; SSK2::SSK2^{W1557C}*	
DMY028-ev8	*DMY028; SSK2::SSK2^{G1471V}*	
DMY028-ev9	*DMY028; SSK2::SSK2^{P1466L}*	
DMY028-ev10	*DMY028; SSK2::SSK2^{D400V}*	
DMY028-ev11	*DMY028; SSK2::SSK2^{C862Y}*	
DMY028-ev12	*DMY028; SSK2::SSK2^{G1273R}*	
DMY028-ev13	*DMY028; SSK2::SSK2^{Y1463C}*	
DMY028-ev14	*DMY028; SSK2::SSK2^{V1483F}*	
DMY028-ev15	*DMY028; HOG1::HOG^{D195Y}*	
DMY028-ev16	*DMY028; HOG1::HOG^{C161W}*	
DMY028-ev17	*DMY028; PBS2::PBS2^{L357S}*	
DMY028-ev18	*DMY028; PBS2::PBS2^{R53H}*	
DMY028-ev19	*DMY028; PBS2::PBS2^{Y43D}*	
DMY028-ev20	*DMY028; PBS2::PBS2^{R61L}*	
DMY028-ev21	*DMY028; PBS2::PBS2^{G423D}*	
DMY028-ev22	*DMY028; PBS2::PBS2^{G509S}*	
DMY028-ev23	*DMY028; PBS2::PBS2^{M526R}*	
DMY028-ev24	*DMY028; PBS2::PBS2^{R640STOP}*	
DMY017-sm1	*DMY017; S.cer-SLN1::C.gla-SLN1*	
DMY017-sm2	*DMY017; S.cer-SLN1::C.alb-SLN1*	
DMY017-sm3	*DMY017; S.cer-YPD1::C.gla-YPD1*	
DMY017-sm4	*DMY017; S.cer-YPD1::C.alb-YPD1*	
DMY017-sm5	*DMY017; S.cer-SSK1::C.gla-SSK1*	
DMY017-sm6	*DMY017; S.cer-SSK1::C.alb-SSK1*	
DMY017-sm7	*DMY017; S.cer-SSK2::C.gla-SSK2*	

DMY017-sm8	*DMY017; S.cer-SSK2::C.alb-SSK2*	
DMY017-sm9	*DMY017; S.cer-PBS2::C.gla-PBS2*	
DMY017-sm10	*DMY017; S.cer-PBS2::C.alb-PBS2*	
DMY017-sm11	*DMY017; S.cer-YPD1::A.gos-YPD1*	
DMY017-sm12	*DMY017; S.cer-YPD1::K.lac-YPD1*	
DMY017-sm13	*DMY017; S.cer-YPD1::D.han-YPD1*	
DMY017-sm14	*DMY017; S.cer-PBS2::A.gos-PBS2*	
DMY017-sm15	*DMY017; S.cer-PBS2::K.lac-PBS2*	
DMY017-sm16	*DMY017; S.cer-PBS2::N.cra-PBS2*	
BG2 (C. glabrata)	*Clinical isolate*	Wurgler-Murphy et al.
Plasmid		
pYPD1^{K67A}	*YPD1*^{K67A}	Janiak-Spens et al.
pYPD1^{R90A}	*YPD1*^{R90A}	Janiak-Spens et al.
pYPD1^{Q86A}	*YPD1*^{Q86A}	Janiak-Spens et al.
pYPD1^{G68Q}	*YPD1*^{G68Q}	Janiak-Spens et al.
pGPBD21	*URA3* 2mm *PGAL1-PBS2DD* (*PBS2* with Ser514-Asp and Thr518-Asp mutations)	Wurgler-Murphy et al.
pGSS21	*URA3* 2mm *PGAL1-SSK2*DN (contains *SSK2* from Met1173 to Asp1579)	Maeda et al.

REFERENCES

Acar, M., Mettetal, J. T. and van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. Nat. Genet. *40*, 471-475.

Alon, U., Surette, M. G., Barkai, N. and Leibler, S. (1999) Robustness in bacteria chemotaxis. Nature *397*, 168-171.

Bahn, Y., Geunes-Boyer, S., Heitman, J. (2007) Ssk2 mitogen-activated protein kinase kinase kinase governs divergent patterns of the stress-activated Hog1 signaling pathway in *Cryptococcus neoformans*. Eukaryot. Cell *6*, 2278-2289.

Batchelor, E. and Goulian, M. (2002) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. Proc. Natl. Acad. Sci. USA *100*, 691-696.

Barkai, N. and Shilo, B. Z. (2007) Variability and robustness in biomolecular systems. Mol. Cell *28*, 755-760.

Barkai, N. and Leibler S. (1997) Robustness in simple biochemical networks. Nature *387*, 913-917.

Brown, J. M. and Firtel, R. A. (1998) Phosphorelay signaling: New tricks for an ancient pathway. Current Biology *8*, 662-665.

Ferrigno, P., Posas, F., Koepp, D., Saito, H., and Silver, P. (1998) Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. EMBO J. *17*, 5606-5614.

Hohmann, S. (2009) Control of high osmolarity signaling in the yeast *Saccharomyces cerevisiae*. FEBS Lett. *583*, 4025.

Janiak-Spens, F., Cook, P. F. and West, A. H. (2005) Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. Biochemistry *44*, 377-386.

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

Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., and Hohmann, S. (2005) Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. *23*, 975-982.

Krantz, M., Ahmadpour, D., Ottosson, L. G., Warringer, J., Waltermann, C., *et al.* (2009) Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway. Mol. Syst. Biol. *5*, 281.

Krantz, M., Becit, E. and Hohmann, S. (2006) Comparative genomics of the HOG-signaling system in fungi. Curr. Genet. *49*, 137-151.

Kollmann, M., Løvdok, L., Bartholomé, K., Timmer, J., Sourjik, V. (2005) Design principles of a bacterial signaling network. Nature *438*, 504-507.

Lestas, I., Vinnicombe, G. and Paulsson, J. (2010) Fundamental limits on the suppression of molecular fluctuations. Nature *467*, 174-178.

Little, J. W., Shepley, D. P. and Wert, D. W. (1999) Robustness of a gene regulatory circuit. EMBO J *18*, 4299-4307.

Ng, P. C. and Henikoff, S. (2001) Predicting deleterious amino acid substitutions. Genome Res. *11*, 863–874.

Maeda, T., Takekawa, M. and Saito, H. (1995) Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. Science *269*, 554-558.

Mettetal, J., Muzzey, D., Gomez-Uribe, C., and van Oudenaarden, A. (2008) The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. Science *319*, 482-484.

Moriya, H., Shimizu-Yoshida, Y., Kitano, H. (2006) In vivo robustness analysis of cell division cycle genes in Saccharomyces cerevisiae. PLoS Genet. *2 (7)*, e111.

Muzzey D., Gomez-Uribe, C, Mettetal, J. and van Oudenaarden, A. (2009) A systems-level analysis of perfect adaptation in yeast osmoregulation. Cell *138*, 160-171.

O'Rourke, S., and Herskowitz, I. (2004) Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. Mol. Biol. Cell *15*, 532-542.

Posas, F., Wurgler-Murphy, S. M., Witten, E. A., Thai, T. C. and Saito, H. (1996) Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1–YPD1–SSK1 "two-component" osmosensor. Cell *86*, 865-875.

Shinar, G., Milo, R., Martinez, R. and Alon, U. (2007) Input-output robustness in simple bacteria signaling systems. Proc. Natl. Acad. Sci. USA *104*, 19931-19935.

Shinar, G. and Feinberg, M. (2010) Structural sources of robustness in biochemical reaction networks. Science *327*, 1389-1391.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, III F. J., Doyle, J. (2004) Robustness of cellular functions. Cell *118*, 675-685.

Stephenson, K. and J.A. Hoch. (2002) Histidine kinase-mediated signal transduction systems of pathogenic microorganisms as targets for therapeutic intervention. Curr. Drug Targets - Infectious Disorders, *2 (3),* 235-24.

Stephenson, K. and J.A. Hoch. (2004) Developing Inhibitors to Selectively Target Two-Component and Phosphorelay Signal Transduction Systems of Pathogenic Microorganisms. Curr. Med. Chem. *11(6)*, 765-73.

Thomason, P. and Kay, R. (2000) Eukaryotic signal transduction via histidine-aspartate phosphorelay. J. Cell. Sci. *113*, 3141-3150.

Wagner, A. (2005) Robustness and Evolvability in Living Systems. (Princeton Univ. Press, Princeton, NJ).

Wurgler-Murphy, S. M., Maeda, T., Witten, E. A., Saito, H. (1997) Regulation of the *Saccharomyces cerevisiae* HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. Mol. Cell. Biol. *17*, 1289-1297.

Chapter 3

SINGLE CELL ANALYSIS OF SPLICING DYNAMICS AT SINGLE MOLECULE RESOLUTION

ABSTRACT | Splicing serves a unique regulatory role in the gene expression pathway where it can not only control the diversity of gene products, but it can also shape their mean expression and noise properties. Despite this, a single cell analysis of splicing is lacking. Here, we explore how *HAC1* RNA splicing contributes to heterogeneity in the unfolded protein response (UPR) in yeast by using single molecule imaging to count endogenous spliced and unspliced *HAC1* RNA in single cells. We find that different UPR-inducing stresses can alter the mean splicing kinetics from highly efficient to limiting. Furthermore, we observed that the cell-to-cell variability in the degree of splicing is differentially regulated under these conditions. By combining these measurements with stochastic gene expression models, we find that the increased variability can be explained by increased fluctuations in the splicing efficiency. Together, these results suggest that splicing (mis)regulation *in trans* can generate substantial variability in splicing outcomes, which might be advantageous for the cell population under stress.

INTRODUCTION

Single-cell measurements of mRNAs have revealed that the expression of genes can vary, sometimes dramatically, from cell to cell, and the biological role of these differences can be greatly amplified when the transcripts are regulatory molecules such as transcription factors (McAdams and Arkins, 1997; Suel et al., 2007; Maamar et al., 2007; Raj et al., 2010; Chubb et al., 2006; Paré et al., 2009). These measurements have also paved the way for a quantitative assessment of the different stochastic mechanisms of transcription, which led to the finding that many genes in eukaryotes exhibit transcriptional bursting (Zenklusen et al., 2008; Raj et al., 2008; Suteret et al., 2011; Chubb et al., 2006; Paré et al., 2009). While most previous studies have centered on transcription, a similar stochastic view of splicing is lacking. Splicing acts at the post-transcriptional level of gene regulation where, akin to transcription, expression can be controlled quantitatively. In a few exceptional cases, splicing overtakes the role of transcriptional regulation to become the dominant gene expression control for constitutively transcribed genes (Bingham et al., 1988). Given these, the abilities to quantitatively scrutinize splicing in single cells, and to analyze the transcription-coupled-to-splicing system's gene expression noise properties are desirable in furthering our understanding of gene expression regulation.

Current techniques for measuring splicing include real-time RT-PCR, northern blotting, and on a genome-wide scale, microarrays, which provide an ensemble average of the levels of spliced and unspliced RNA species in a cell population (Clark et al., 2002). Single-cell imaging methods based on fluorescence *in situ* hybridization and

RNA-binding proteins tagged with fluorescent proteins i.e. MS2-labeled or U1A-labeled RNAs have also been used to track the localization of RNAs during splicing (Zhang et al., 1994; Aragon et al., 2009; Brodsky and Silver, 2000). These visualization techniques, however, are limited in their abilities to monitor and quantify splicing directly, and often require genetic modifications and over-expression of the gene, thus making a quantitative analysis of splicing in single cells not possible. To circumvent these limitations, we devise a colocalization strategy using single molecule RNA fluorescence *in situ* hybridization (Raj et al., 2008) for tracking individual endogenous spliced and unspliced RNA transcripts. We present a framework for understanding how the relative balance between the kinetic steps (i.e. transcription, RNA turnover, splicing) involved in the synthesis of spliced transcripts contributes to the regulation of the mean gene expression levels and to cell-to-cell variations (Fig. 1a). Here, variability is quantified by the coefficient of variation, *CV*, which is defined as the standard deviation δr divided by the mean $\langle r \rangle$ of the RNA copy number. *CV* represents how "noisy" the spliced RNA production is relative to a Poisson process where $CV^2 = 1/\langle r \rangle$.

RESULTS

Single-cell visualization and quantification of spliced and unspliced transcripts

We devised an approach to adapt the single-molecule FISH technique to visualize and count endogenous spliced and unspliced RNAs in single cells with single transcript resolution. Our strategy involved targeting the exon sequences with probes labeled with a single Cy 5 fluorophore each and the intron sequences with Alexa 594 (Fig. 1b). Briefly,

we hybridized the fixed cells with the probes labeled with the two colors and imaged them using epifluorescence microscopy. Under this scheme, we identified the unspliced transcripts as spots that colocalize in both channels, while the individual singletons spots represented either the spliced species or the intron.

We applied this method to measure the *HAC1* RNA in yeast cells, which is found to be present in the unspliced form in the cytoplasm, and is thus untranslated under normal growth conditions (Bernales et al., 2006). But, once endoplasmic reticulum (ER) stress is triggered by protein misfolding, the unspliced *HAC1* RNA undergoes splicing, and the spliced RNA is then translated into proteins, which is essential for activating the unfolded protein response (Bernales et al., 2006). ER-stress-activated splicing of *HAC1* RNA does not require the canonical eukaryotic spliceosomal machinery. Instead, it requires only two components i.e. the ER-localized Rnase Ire1 protein and the tRNA ligase Rlg1 protein (Bernales et al., 2006). This system is well suited for our purpose because splicing can be easily controlled by stimuli that induce protein-misfolding, and both the unspliced and spliced products are spaced out in the cytoplasm where splicing occurs, which facilitates quantification of the two RNA species at the single-molecule level.

We detected that most pre-stressed cells had spots that were mostly colocalized in the two channels, which represented unspliced *HAC1* RNA, and these were localized in the cytoplasm (Fig. 1c), consistent with measurements using *in situ* hybridization. In single cells, the number of Cy 5 exon spots and Alexa 594 intron spots were highly correlated ($r = 0.87$, $p = 1.2 \times 10^{-93}$) (Fig. 1d). To quantify spatial correlation between signals in the two channels, we used an analysis method based on particle image cross

correlation spectroscopy (Semrau et al., 2011). Of the total spots from 300 cells, 80% of the intron spots colocalized with the exon spots, while 78% of the exon spots colocalized with the intron spots (Supplementary Fig. S1). This degree of colocalization was in good agreement with the percentage reported by Raj *et al.* when using coding-sequence-specific probes and 3' UTR-sequence-specific probes for the GFP sequence. In each cell, we detected on average a total of 22 *HAC1* RNA molecules, with ~80% being the unspliced transcripts (Fig. 1d).

Measuring splicing under different UPR-inducing stresses

We measured splicing of *HAC1* RNAs in cells under different conditions that trigger ER stress i.e. using dithiothreitol (DTT), a drug that reduces disulfide bonds and causes protein misfolding, and heat stress. Upon adding DTT, we detected on average a significant increase in the number of spliced *HAC1* RNA and a concomitant decrease in unspliced RNA in cells within five minutes (Figs. 2a-b). At steady-state and at the cell population level, the total *HAC1* RNA remained largely similar to the pre-stress level. Furthermore, the dynamics and degree of splicing were in good agreement with literature and the measurements obtained using q-PCR (Bernales et al., 2006; Pincus et al., 2010) (Figs. 2a-b). We also simultaneously detected the formation of Ire1p foci upon activation of the ER-stress, which is responsible for the splicing of *HAC1* RNAs (Aragon et al., 2009) (Supplementary Fig. S2). In single cells, the ratio of spliced products to total RNA is not correlated with the total *HAC1* RNA (Supplementary Fig. S3), indicating that transcription and splicing play independent roles in the generation of spliced *HAC1* molecules.

HAC1 exhibits bursting transcription

We next considered a general stochastic model for spliced RNA production by expanding the two-state-promoter model for RNA production (Raj et al., 2008; Peccoud and Ycart, 1995) to incorporate a first-order splicing reaction. Here, the promoter stochastically fluctuates between "off" and "on" states, likely due to chromatin modifications (Becskei et al., 2005), and RNA is produced only in the "on" state, which is then processed by the splicing apparatus to generate the spliced RNA (Fig. 3a). Accordingly, the production of spliced RNA is characterized by 4 kinetic parameters i.e. λ, the rate of promoter switching to the "on" state; γ, the rate of promoter switching to the "off" state; μ, the rate of transcription while in the "on" state; and μ_s, the rate of RNA splicing.

If RNA were synthesized at a fixed rate, one would expect the statistics of the number of RNA per cell to fit a Poisson distribution, where the mean and variance are equal. But, we observed that the mean total *HAC1* RNA molecules per cell was approximately 22, while the variance was roughly 62, consistent with a bursting transcription model. Using a maximum likelihood method, we were able to fit the total *HAC1* RNA distribution to the analytical solution for the two-state transcription module to find expressions for the gene activation rate, λ, and the average number of RNA produced during each burst, μ/γ (average burst size) (Fig. 3b and Supplementary information). The spliced and unspliced *HAC1* RNA half-lives are similar and are approximately 20 min (Pincus et al., 2010) and we assume that they are decreased under heat stress (Grigull et al., 2004; Lindquist, 1981). As a measure of transcriptional initiation, we scored for the presence or absence of active transcription site in the nucleus

in cells, which represents nascent transcripts that had not yet diffused away (Zenklusen et al., 2008). Approximately 25% of the cells had transcription sites consisting of up to 7 nascent transcripts (Fig. 3c, Supplementary Fig. S4 and Supplementary information). We found that the average RNA burst statistics i.e. burst size and burst frequency (number of transcriptional events per RNA lifetime) of (3.8 ± 0.5 and 7.1 ± 0.9) (Fig. 3d) sufficiently describes both the cytoplasmic RNA abundance data and the nascent transcript count.

Mean splicing efficiency is altered under distinct stresses

At steady-state, the dependence of the mean levels of unspliced and spliced RNA on splicing rate falls into two distinct regimes (Fig. 4a and Supplementary information). If the rate of splicing μ_s is greater than the turnover rate of the unspliced RNA δ, most unspliced RNA ends up being spliced and thus the spliced RNA abundance is at its maximum and is independent of μ_s, while unspliced RNA abundance will scale inversely with μ_s. Conversely, if $\mu_s \ll \delta$, most unspliced molecules are degraded and the unspliced RNA abundance is essentially independent of μ_s. The level of spliced RNA, however, will scale with μ_s. We find that cells exposed to the two conditions exhibit behaviors that fall into different regimes of spliced RNA production efficiency i.e. in DTT-treated cells, the kinetics of splicing is much faster than the turnover rate of unspliced *HAC1*, which resulted in the efficient near-maximal production of spliced *HAC1* ($\langle s \rangle / \langle s \rangle_{max} = 0.82$). In contrast, the less efficient spliced transcript production observed under heat stress ($\langle s \rangle / \langle s \rangle_{max} = 0.48$) can be explained by comparable splicing and turnover rates, where splicing becomes almost limiting.

Variability in the degree of splicing is tightly regulated under DTT, but is compromised under heat stress

We obtained the analytical intrinsic noise expressions for the spliced and unspliced species based on the set of reactions in Fig. 3a (Supplementary information), and determined the dependence of the cell-to-cell variability in RNA expression on splicing kinetics. We find that similar to the mean behavior of the RNA expression; its variability displays two distinct characteristics (Fig. 4b). Where splicing is limiting in the synthesis of spliced transcripts, increases in splicing rate reduces the variability of the spliced transcript levels, but not that of the unspliced transcript abundance. But where splicing is not limiting, the unspliced transcript noise increases with splicing rate while the variability of the spliced transcript level is insensitive to even large variations in splicing rate.

While we find that environmental stresses elicit distinct effects on the mean behavior of the transcript abundance, we next asked if cell-to-cell variability in the transcript abundance is differentially regulated under these conditions. We find that for both of the populations exposed to either DTT or heat shock, most of the cell-to-cell variations in transcript abundance can be accounted for by intrinsic noise ($CV_{dtt,\ measured}/CV_{dtt,\ predict} = 1.13$ and $CV_{heat,\ measured}/CV_{heat,\ predict} = 1.24$). Stochastic simulations of the model fit adequately well to the RNA distributions for spliced and unspliced *HAC1* in both cases (Fig. 4c-d and Supplementary information), supporting that most of the variation originates from random gene activation events.

Since every cell possesses different numbers of spliced and unspliced RNAs, their degrees of splicing can vary widely among one another. To analyze the variability in the

degree of splicing, we examined the spliced and total transcript levels in individual cells. We determined the best fit line to the data, and hence, deviation from this line represents variability. On average, the majority of the *HAC1* transcripts in cells exposed to DTT were present in the spliced from (best-fit slope = 0.82 ± 0.011), while only approximately half of the total transcripts in cells under heat stress consisted of spliced RNAs (best-fit slope = 0.47 ± 0.023). By computing the coefficient of variation of the root-mean-square deviation in expression across all cells, we found significantly increased variability in the degree of splicing (DOS) in cells subjected to heat stress as compared to cells exposed to DTT ($CV_{RMSE,\ heat}$ = 0.50, $CV^{observed}_{RMSE,heat}/CV^{predict}_{RMSE,heat}$=1.44 and $CV_{DTT,\ RMSE}$ = 0.14, $CV^{observed}_{RMSE,dtt}/CV^{predict}_{RMSE,dtt}$ = 1.13).

To quantitatively understand the greater DOS variability, we considered models including extrinsic noise at the levels of transcription and splicing, and determined analytically as well as via simulations their effects on the normalized covariance of spliced and unspliced transcripts in single cells (Paulsson, 2004). We find that any extrinsic noise at the transcriptional level, i.e. μ and δ, would only increase the spliced-unspliced-transcript covariance $\bar{\sigma}_{su}$, whereas heterogeneity in the splicing rate μ_s would reduce the covariance (Supplementary information). For the DTT-treated population, the observed normalized covariance can be attributed mostly to the two-state promoter fluctuations ($\bar{\sigma}_{su,predict}/\bar{\sigma}_{su,measured}$ = 1.3) (Fig. 4e). But, the covariance expected purely from these fluctuations is significantly higher than the measured covariance in the heat stress case, which turned out to be negative ($\bar{\sigma}_{su,predict}/\bar{\sigma}_{su,measured}$ = -4.8) (Fig. 4e). Based on these models, the poor covariance in the heat stress condition can be generated by increased fluctuations in splicing rate (Fig. 4f), which can arise from, for

example, increased variations in the expression levels of splicing factors. But, these fluctuations are minimal under DTT stress. Together, these results suggest that the variability in the efficiency of the splicing machinery might be differentially regulated under different stress conditions.

DISCUSSION

By counting single *HAC1* molecules and monitoring its transcriptional activity in single cells, we found that the gene exhibits a wide expression and the statistics of variation indicate that the gene is transcribed in bursts, rather than constitutively. Most of the recent noise studies have also found evidences of transcriptional bursting in genes across different eukaryotes (Zenklusen et al., 2008; Raj et al., 2008; Suteret et al., 2011; Chubb et al., 2006; Paré et al., 2009), suggesting the prevalence of this kinetic mode of transcription in eukaryotic gene expression in general. We investigated the kinetics of spliced *HAC1* expression in different environmental stresses using two-color colocalization to distinguish and to count individual spliced and unspliced RNAs in each cell. From these measurements, we were able to derive the splicing efficiencies using a stochastic splicing-coupled-to-transcription gene expression model, which sufficiently captures the variation in transcript expression. We found that the splicing rate is significantly increased in DTT than in heat stress, such that in the former case, splicing is not limiting in the production of spliced transcripts, and while in the latter case, splicing becomes limiting.

Furthermore, by simultaneously monitoring unspliced and spliced transcripts in single cells, we found that the variability in the degree of splicing differed greatly between these two UPR-induced stresses, and increased splicing rate fluctuations are necessary to explain the higher variability. Thus, it appears that not only can the mean splicing efficiency be altered in response to different environmental conditions, but the variability in the splicing kinetics can even be changed. The expression of the splicing factor Ire1 protein has been observed to be tightly regulated via negative autoregulation under normal conditions (Tirasophon et al., 1998). In support of this, overexpression of Ire1p results in constitutive activation of the URP and slow growth in yeast cells (Bernales et al., 2006). Thus, it seems likely that the splicing mis-regulation i.e. increased splicing rate fluctuations under heat shock is mostly, if not partly, a result of the compromised autoregulation of Ire1p expression. Almost 400 genes (5% of the yeast genome) have been identified to be transcriptionally controlled by the Ire1p-*HAC1*-mediated UPR in yeast (Travers et al., 2000). This greater variability would likely translate to greater population heterogeneity in *HAC1* protein expression, which can result in a highly variegated unfolded protein response.

We further determined the analytical solution of the general splicing-coupled-to-transcription model for the dependence of RNA expression and expression variability on splicing rate. The results from the model propose two classes of genes with distinctive spliced RNA expression regulation encoded *in cis*. One class of genes are excellent substrates of the splicing apparatus (very high splicing rates), where perturbations to the splicing reaction i.e. fluctuations in splicing factor levels has a minimal impact on the mean and cell-to-cell variability of their spliced transcript expression. These fluctuations,

however, can greatly affect the expression of their unspliced transcript counterparts. The other class are genes which are relatively poorer substrates (very low splicing rates), where their spliced transcript expression and expression noise are highly sensitive to fluctuations in the splicing reaction, and yet, the expression of their unspliced species are robust to these changes. These gene-specific differences in sensitivity of mean and variance of gene expression to splicing fluctuations may be correlated with its gene function.

These results demonstrate that *HAC1* RNA splicing (mis)regulation *in trans* can generate substantial gene expression variation, which can result in heterogeneous stress responses among cells, and also highlight the need for quantitative single-cell analyses in the understanding of cell-to-cell heterogeneity in probabilistic splicing outcomes.

MATERIALS AND METHODS

Ire1p-GFP strain construction and cell growth

To visualize both the Ire1 protein and *HAC1* RNA simultaneously, we transformed a construct with GFP inserted into the cytosolic portion of Ire1 adjacent to its transmembrane region (Aragon et al., 2009) (a gift from the Walter lab) into our yeast strain. All experiments involved growing cells in synthetic media at 30 °C to an optical density (at 600 nm) of 0.4, after which they were either treated with 10mM DTT (Sigma) or shifted to 37 °C for the ER stress induction experiments. DTT stocks were made fresh from powder stored at 4 °C for each experiment.

Fluorescence *in situ* hybridization and image acquisition

We performed FISH as described by Raj et al., 2008. Cells were fixed by adding 32% (v/v) formaldehyde to the yeast culture to a final concentration of 4% (v/v) for 45 minutes at room temperature. The cell wall was digested using lyticase (Sigma), and stored in 70% (v/v) ethanol at 4 °C. Prior to hybridization, cells were rehydrated with 10% (v/v) formamide, 2 x SSC for 10 minutes. We used a hybridization buffer consisting of 10% (v/v) formamide, 2 × SSC, 1 mg/ml BSA, 10 mM VRC, 0.5 mg/ml *Escherichia coli* tRNA and 0.1 g/ml dextran sulfate. All hybridizations were carried out in solution using *HAC1* probes to Cy5 (GE Amersham) and Alexa 594 (Invitrogen) at optimal concentrations determined empirically, and were carried out overnight at 37 °C. The cells were attached to coverslips coated with concanavalin A (Sigma). Images were acquired using a Nikon TE2000 microscope equipped with a Princeton Instruments camera, a 100 x oil immersion objective and custom filter sets capable of distinguishing between the different fluorophores used. Data was taken as stacks of images with a z direction step size of 0.2 μm using filters appropriate for DAPI, FITC, Alexa 594 and Cy5 (Chroma Technology). To minimize photobleaching during imaging, we used antifade reagents including Trolox and the oxygen-scavenging solution glucose oxidase (Raj et al., 2008). The probes used are listed in Supplementary Table 1.

Spot counting analysis

Using custom-written MATLAB (Mathworks) software ImageM, we reduced the stacked images to two-dimensional images by maximum projection, and counted the number of fluorescent spots, each of which corresponds to a single RNA, using the semi-automated

method outlined by Raj et al., 2008. We segmented the yeast cells manually and manually counted the number of cells with active transcription sites with the aid of custom-written scripts. Nuclear segmentation was done by thresholding using DAPI. Single transcript intensity was defined as the integrated intensity of the spot identified using the two-dimensional Gaussian mask algorithm. We determined the number of nascent transcripts at the site of transcription by dividing the spot intensity of the transcription site by the single transcript intensity, and rounding off to the nearest integer.

Colocalization analysis

We used custom-written MATLAB software to implement the particle image cross correlation spectroscopy analysis method described by Semrau et al., 2011.

Numerical modeling

The theoretical model for RNA abundance is based on that described by Raj et al., 2006. The analytical solution for the two-state model is:

$$\rho(m) = \frac{\Gamma\left(\frac{\lambda}{\delta}+m\right)}{\Gamma(m+1)\Gamma\left(\frac{\lambda}{\delta}+\frac{\gamma}{\delta}+m\right)}\frac{\Gamma\left(\frac{\lambda}{\delta}+\frac{\gamma}{\lambda}\right)}{\Gamma\left(\frac{\lambda}{\delta}\right)}\left(\frac{\mu}{\delta}\right)^{m} \quad {}_1F_1\left(\frac{\lambda}{\delta}+m,\frac{\lambda}{\delta}+\frac{\mu}{\delta}+m,-\frac{\mu}{\delta}\right)$$

where λ, μ and γ are defined as in the text, and δ is obtained from literature (Pincus et al., 2010). To simulate the nascent transcript distribution, we used another parameter τ to define the duration during which RNA II polymerase remains on the gene for the synthesis of a complete transcript. If the polymerase elongates at 2 kb min^{-1}, the amount

of time needed for the synthesis of *HAC1*, a 1 kb transcript, would require at least 30 seconds (Mason and Struhl, 2005). We fitted the total RNA abundance data to the equation above using the maximum likelihood method, and obtained well defined parameter values: rate of activation (λ/δ) = 7.1, inactivation (γ/δ) = 21.8 and transcription (μ/δ) = 82.2. Using these parameters and τ, we determined the nascent transcript distributions using Monte Carlo simulations (Zenklusen et al., 2008). Stochastic simulations of the spliced and unspliced RNA numbers were performed by implementing Gillespie's direct method (Gillespie, 1977) in Matlab. The parameters governing RNA production dynamics obtained earlier were used in the simulations. We performed simulated annealing to obtain the best-fit splicing rate μ_s that minimizes the difference between the CDF of the simulated and experimental spliced and unspliced RNA abundance data. To simulate the effects of splicing rate fluctuations, μ_s is assumed to be normally-distributed with mean $\langle \mu_s \rangle$ and $\sigma_{\mu s}$ is allowed to vary.

SUPPLEMENTARY INFORMATION

Steady-state model: dependence of mean spliced and unspliced levels on μ_s

To model the spliced and unspliced RNA dynamics, we extend the stochastic 2-state transcription model to include a splicing step which obeys 1st order kinetics. The set of chemical reactions are:

$$a \xrightarrow{\lambda(1-a)} a+1$$

$$a \xrightarrow{\gamma a} a-1$$

$$u \xrightarrow{\mu a} u+1 \qquad [1]$$

$$u \xrightarrow{\delta u} u - 1$$

$$\{u, s\} \xrightarrow{\mu_s u} \{u - 1, s + 1\}$$

$$s \xrightarrow{\delta_s s} s - 1$$

where λ and γ represent rates of promoter activation and inactivation, μ and μ_s are the unspliced and spliced RNA synthesis rates, and δ and δ_s are their corresponding degradation rates. We use A, U and S to represent the activated gene, the unspliced and spliced species, and a, u and s to represent their numbers of molecules in the cell.

The dynamics for the averages of the molecular species can be described by:

$$\frac{d\langle a \rangle}{dt} = -\gamma\langle a \rangle + \lambda(1 - \langle a \rangle)$$

$$\frac{d\langle u \rangle}{dt} = -(\delta + \mu_s)\langle u \rangle + \mu\langle a \rangle$$

$$\frac{d\langle s \rangle}{dt} = -\delta_p\langle s \rangle + \mu_s\langle u \rangle$$

At steady-state, the mean number of unspliced RNA for this system is given by:

$$\langle u \rangle = \frac{T}{\delta + \mu_s}$$

where $T = \mu\lambda/(\lambda + \gamma)$. And that of the spliced RNA is:

$$\langle s \rangle = \frac{\mu_s(\frac{1}{\delta + \mu_s})}{\delta_s} T$$

Expressing u and s as a function of μ_s/δ gives:

$$\langle u \rangle = \frac{1}{\delta} T \left[1 + (\frac{\mu_s}{\delta})\right]^{-1}$$

$$\langle s \rangle = \frac{1}{\delta_s} T \left(\frac{\mu_s}{\delta}\right) \left[1 + \left(\frac{\mu_s}{\delta}\right)\right]^{-1}$$

In the case of efficient splicing i.e. $\mu_s \gg \delta$:

$$\langle u \rangle \cong \frac{1}{\mu_s} T$$

$$\langle s \rangle_{max} \cong \frac{1}{\delta_s} T$$

The level of spliced RNA is maximal and is independent of the splicing rate, while the unspliced RNA amount scales inversely with it.

While in the case of inefficient splicing i.e. $\mu_s \ll \delta$:

$$\langle u \rangle_{max} \cong \frac{1}{\delta} T$$

$$\langle s \rangle \cong \frac{1}{\delta_s} T \left(\frac{\mu_s}{\delta}\right)$$

the converse is true. Thus, the levels of spliced and unspliced RNA produced are determined by the relative values of μ_s and δ, rather than the absolute value of μ_s. And the ratio of spliced RNA to total RNA, which we defined as the degree of splicing, is dependent only on μ_s and δ_s.

$$\langle u \rangle + \langle s \rangle = \frac{\mu_s \left(\frac{1}{\delta + \mu_s}\right)}{\delta_s} T(1 + \delta_s/\mu_s)$$

$$\frac{\langle s \rangle}{\langle u \rangle + \langle s \rangle} = \frac{1}{1 + \delta_s/\mu_s}$$

Analytical linear gene expression noise approximations

Theoretical derivations of the gene expression noise for transcription and translation models have been rigorously developed over the years (Paulsson, 2004; Ozbudak et al., 2002; Raj et al., 2006; Zenklusen et al., 2008; Pedraza and Paulsson, 2008; Peccoud and Ycart, 1995). We adapt Paulsson's FDT approach to the case of a transcription-coupled-to-splicing system to determine the analytical expressions for the noise of the molecular species involved.

The equation for the covariance matrix is:

$$\frac{d}{dt}\sigma = A\sigma + \sigma A^T + \Omega B \qquad [2]$$

where A represents the Jacobian matrix, defined by

$$a_{ij} = \frac{\partial \frac{\partial \langle n_i \rangle}{\partial t}}{\partial \langle n_j \rangle}$$

and B describes the randomness of biochemical events defined as,

$$b_{ij} = \sum v_{jk} v_{ik} R_k$$

with v_{jk} representing the number of j molecules that will change in the k-th reaction at the macroscopic rate R_k, and Ω represents the average cell volume, and finally, σ represents a matrix of covariances between variables x, y and z. Here x, y and z represent the molecular species A, U and S in model [1] respectively.

From the model in [1], we get

$$A = \begin{bmatrix} -(\gamma+\lambda) & 0 & 0 \\ \mu & -(\delta+\mu_s) & 0 \\ 0 & \mu_s & -\delta_p \end{bmatrix} \equiv \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}$$

and

$$B = \begin{bmatrix} (\lambda+\gamma)x_1 + \lambda & 0 & 0 \\ 0 & (\delta+\mu_s)x_2 + \mu x_1 & -\mu_s x_2 \\ 0 & -\mu_s x_2 & \delta_p x_3 + \mu_s x_2 \end{bmatrix} \equiv \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ 0 & b_{32} & b_{33} \end{bmatrix}$$

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}$$

By linearizing [2] and setting Ω to unity such that:

$$A'\sigma' + B' = 0 \qquad [3]$$

where

$$A' = \begin{bmatrix} 2a_{11} & a_{12} & 0 & a_{12} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{11}+a_{22} & 0 & 0 & a_{12} & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{11}+a_{33} & 0 & 0 & a_{12} & 0 & 0 & 0 \\ a_{21} & 0 & 0 & a_{11}+a_{22} & a_{12} & 0 & 0 & 0 & 0 \\ 0 & a_{21} & 0 & a_{21} & 2a_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{21} & a_{31} & a_{32} & a_{22}+a_{33} & 0 & 0 & 0 \\ a_{31} & 0 & 0 & a_{32} & 0 & 0 & a_{11}+a_{33} & a_{12} & 0 \\ 0 & a_{31} & 0 & 0 & a_{32} & 0 & a_{21} & a_{22}+a_{33} & 0 \\ 0 & 0 & a_{31} & 0 & 0 & a_{32} & a_{31} & a_{32} & 2a_{33} \end{bmatrix}$$

$$B' = \begin{bmatrix} b_{11} \\ 0 \\ 0 \\ 0 \\ b_{22} \\ b_{23} \\ 0 \\ b_{32} \\ b_{33} \end{bmatrix}$$

$$\sigma' = \begin{bmatrix} \sigma_{xx} \\ \sigma_{xy} \\ \sigma_{xz} \\ \sigma_{yx} \\ \sigma_{yy} \\ \sigma_{yz} \\ \sigma_{zx} \\ \sigma_{zy} \\ \sigma_{zz} \end{bmatrix}$$

we solve for σ'. From the solution, we obtain the square of the noise η of the unspliced RNA as:

$$\eta_u^2 = \frac{\sigma_{yy}^2}{\langle y \rangle^2} = \frac{1}{\langle u \rangle} + \frac{\sigma_a^2}{\langle a \rangle^2} \frac{(\lambda + \gamma)^{-1}}{(\delta + \mu_s)^{-1} + (\lambda + \gamma)^{-1}}$$

where $\langle a \rangle = \lambda/(\lambda + \gamma)$ and $\sigma_a^2/\langle a \rangle^2 = 1/\langle a \rangle - 1$ since the number of active genes is binomial rather than poisson-distributed. The 1st term describes the noise from random births and deaths of unspliced RNA, and the 2nd term describes the noise obtained by time-averaging the random changes in gene activity.

And for the spliced RNA, we obtain:

$$\eta_s^2 = \frac{\sigma_{zz}^2}{\langle z \rangle^2} = \frac{1}{\langle s \rangle} + \frac{1 - \langle a \rangle}{\langle a \rangle} \frac{\delta_s(\mu_s + \delta)(\gamma + \lambda + \mu_s + \delta + \delta_s)}{(\gamma + \lambda + \delta_s)(\delta + \mu_s + \delta_s)(\gamma + \lambda + \delta + \mu_s)}$$

where the 1st term represents random births and deaths of spliced RNA and the 2nd term represents the noise from a two-step time-averaging of random changes in gene activity. Note that the noise contribution from the unspliced RNA is negligible unlike that expected from a transcription-translation system. This arises because one species is converted into another in the splicing reaction.

And we obtain the covariance, σ_{yz}, between spliced and unspliced RNA as:

$$\sigma_{yz}^2 = \langle a \rangle \frac{\mu \mu_s}{\delta_s(\delta + \mu_s)} \frac{\delta_s}{\delta_s + (\delta + \mu_s)} \left(2 + (1 - \langle a \rangle) \frac{\mu(\lambda + \gamma + \delta + \mu_s + \delta_s)}{(\lambda + \gamma + \delta + \mu_s)(\lambda + \gamma + \delta_s)}\right)$$

The normalized covariance takes the form:

$$\frac{\sigma_{yz}^2}{\langle y \rangle \langle z \rangle} = \frac{\sigma_{zz}^2}{\langle z \rangle^2} - \frac{1}{\langle s \rangle}$$

Effects of promoter state and splicing fluctuations on the covariance of the 2 RNA species

We assume that transcription and splicing fluctuations exhibit a characteristic period τ_μ and $\tau_{\mu s}$ respectively, and their autocorrelation function can be defined as:

$$F_\mu(\tau) = \langle \mu(t+\tau)\mu(t) \rangle - \langle \mu(t) \rangle^2 = \sigma_\mu^2 e^{-\tau/\tau_\mu}$$

$$F_{\mu_s}(\tau) = \langle \mu_s(t+\tau)\mu_s(t) \rangle - \langle \mu_s(t) \rangle^2 = \sigma_{\mu_s}^2 e^{-\tau/\tau_{\mu_s}}$$

where $\langle \cdot \rangle$ denotes averaging with respect to t. We include these factors to analyze their effects on the noise of the 2 RNA species.

Using FDT as before, we obtain:

$$\eta_u^2 = \frac{\sigma_{uu}^2}{\langle u \rangle^2} = \frac{1}{\langle u \rangle} + \frac{\sigma_\mu^2}{\langle \mu \rangle^2} \frac{\tau_\mu}{\tau_u + \tau_\mu}$$

$$\eta_s^2 = \frac{\sigma_{ss}^2}{\langle s \rangle^2} = \frac{1}{\langle s \rangle} + \frac{\sigma_\mu^2}{\langle \mu \rangle^2} \frac{\tau_\mu}{\tau_s + \tau_\mu} + \frac{\sigma_{\mu_s}^2}{\langle \mu_s \rangle^2} \frac{\tau_{\mu_s}}{\tau_s + \tau_{\mu_s}}$$

where $\tau_u = (\delta + \mu_s)^{-1}$ and $\tau_s = \delta_s^{-1}$.

The normalized covariance is given by:

$$\frac{\sigma_{us}^2}{\langle u \rangle \langle s \rangle} = \frac{\sigma_\mu^2}{\langle \mu \rangle^2} \left(\frac{\tau_\mu}{\tau_u + \tau_\mu} \frac{\tau_u}{\tau_u + \tau_s} + \frac{\tau_\mu}{\tau_s + \tau_\mu} \frac{\tau_s}{\tau_u + \tau_s} \right)$$

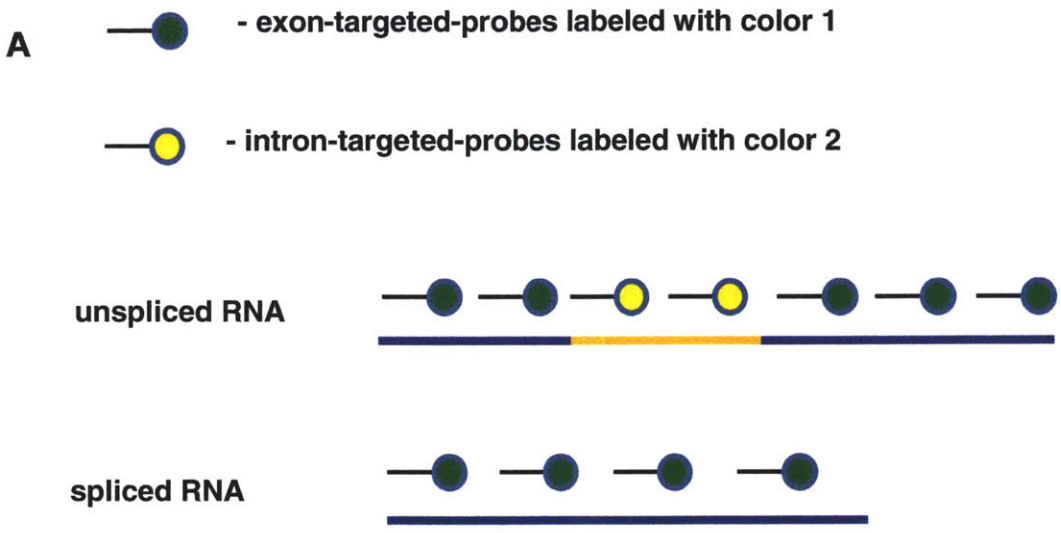
and from the expressions for η_u^2 and η_s^2, we obtain:

$$\frac{\sigma_{us}^2}{\langle u \rangle \langle s \rangle} = \frac{\tau_u}{\tau_u + \tau_s} \left(\eta_u^2 - \frac{1}{\langle u \rangle} \right) + \frac{\tau_s}{\tau_u + \tau_s} \left(\eta_s^2 - \frac{1}{\langle s \rangle} \right) - \frac{\tau_s}{\tau_u + \tau_s} \left(\frac{\sigma_{\mu_s}^2}{\langle \mu_s \rangle^2} \frac{\tau_{\mu_s}}{\tau_s + \tau_{\mu_s}} \right)$$

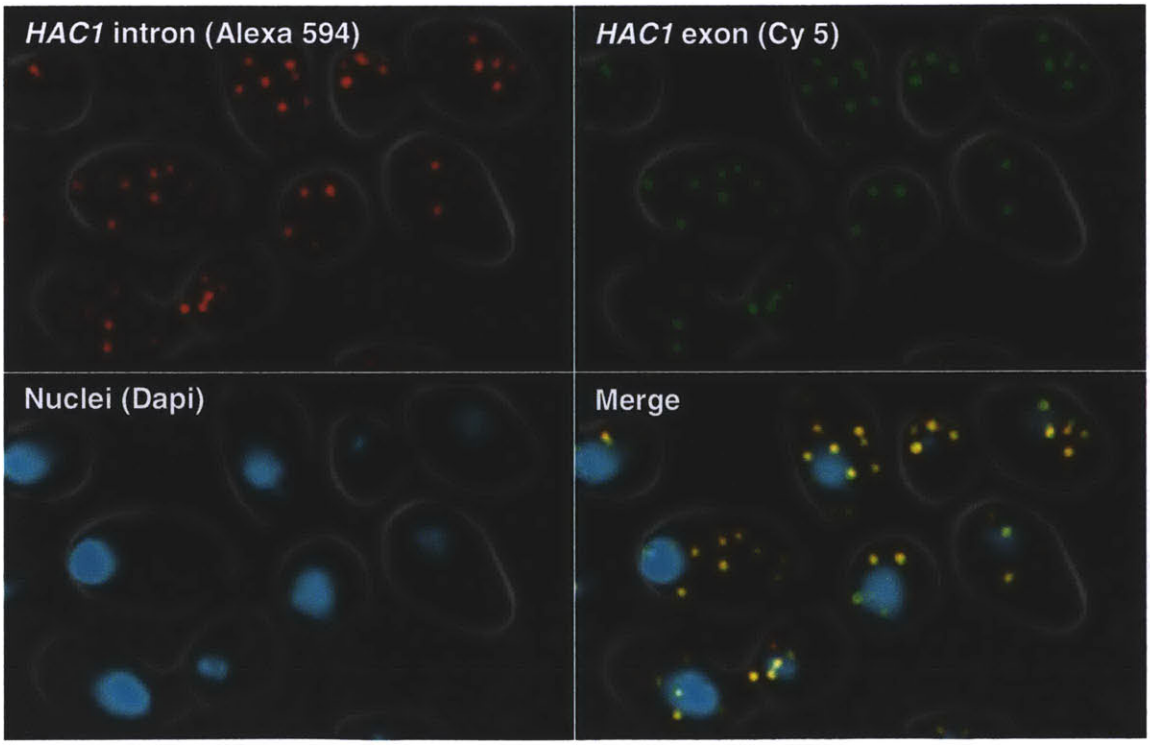
From this expression, we observe that transcription rate fluctuations increase the spliced-unspliced RNA covariance, whereas splicing rate fluctuations reduce the covariance.

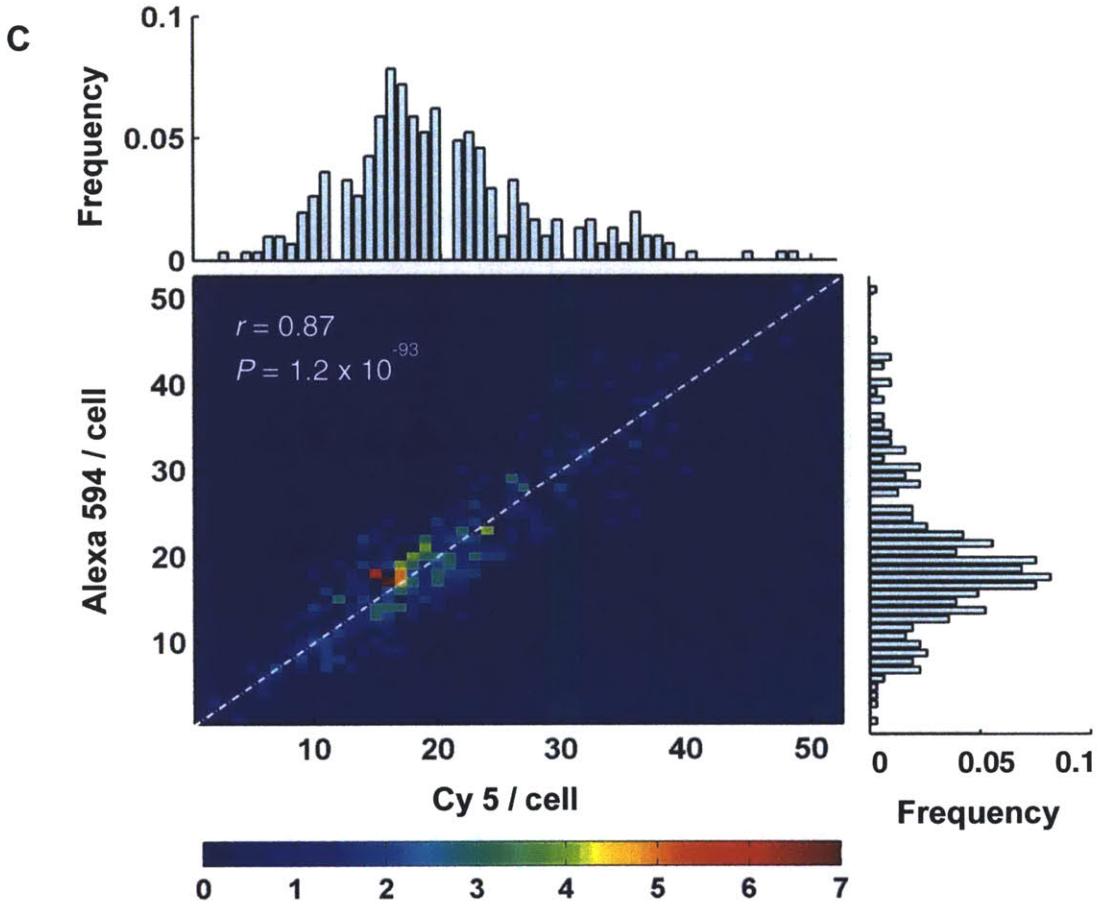
FIGURES AND TABLES

Fig 1



B




C
Frequency
0.1
0.05
0
r = 0.87
P = 1.2 x 10^-93
Alexa 594 / cell
50
40
30
20
10
10
20
30
40
50
Cy 5 / cell
0
0.05
0.1
Frequency
0
1
2
3
4
5
6
7

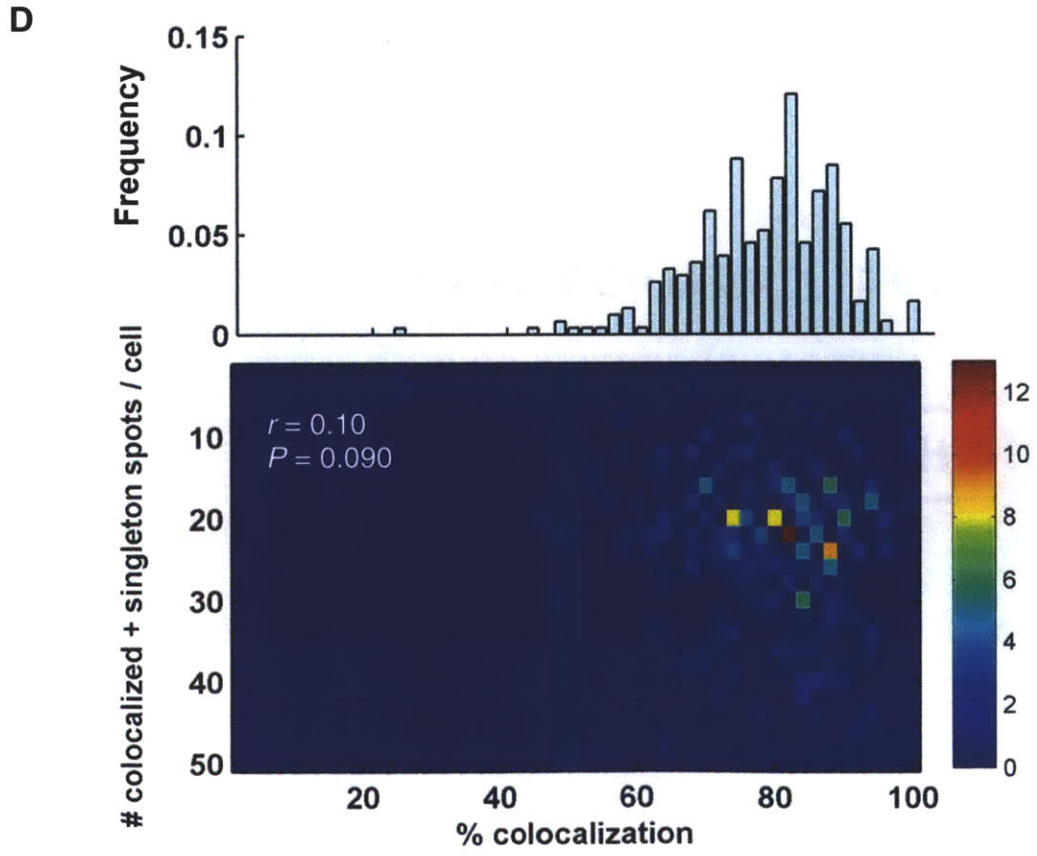
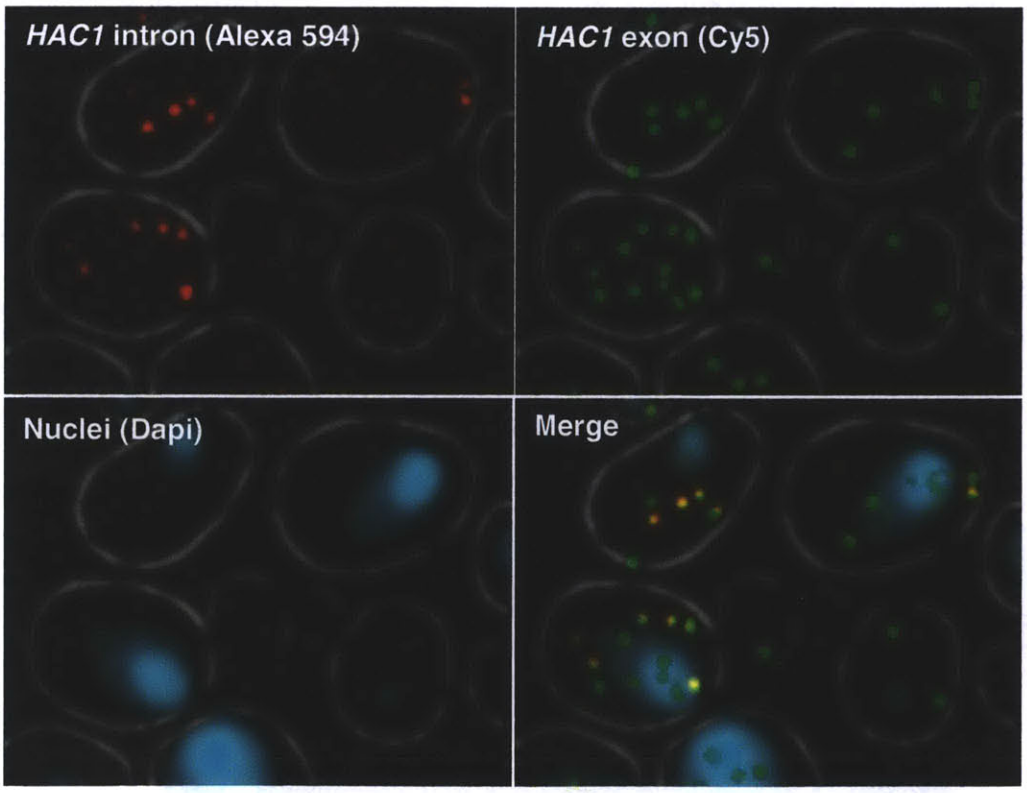

D
Frequency
0.15
0.1
0.05
0
r = 0.10
P = 0.090
# colocalized + singleton spots / cell
10
20
30
40
50
20
40
60
80
100
% colocalization
12
10
8
6
4
2
0

Figure 1 | Single-cell visualization and quantification of spliced and unspliced RNA. **A**, Schematic of design of probes targeted exclusively to the exon or intron sequences and were labeled with two different colors i.e. Alexa 594 and Cy 5. The unspliced RNA is detected as a colocalized spot in both channels. **B,** A single optical slice image of *HAC1* RNA in normal growing yeast cells hybridized to Cy 5-labeled exon probes and Alexa 594-labeled intron probes, and costained with dapi (cyan). Unspliced *HAC1* are detected as colocalized spots in both channels (yellow). **C**, Scatterplot of the number of Alexa 594 spots and Cy 5 spots in single cells grown under normal conditions. Dotted line denotes $y = x$. Marginal histograms indicate the distribution of Alexa 594 and Cy 5 spots per cell. **D**, Scatterplot of the total number of colocalized and singleton spots and the degree of colocalization in single cells. Marginal histogram indicates the distribution of colocalization percentage.

Fig 2

A



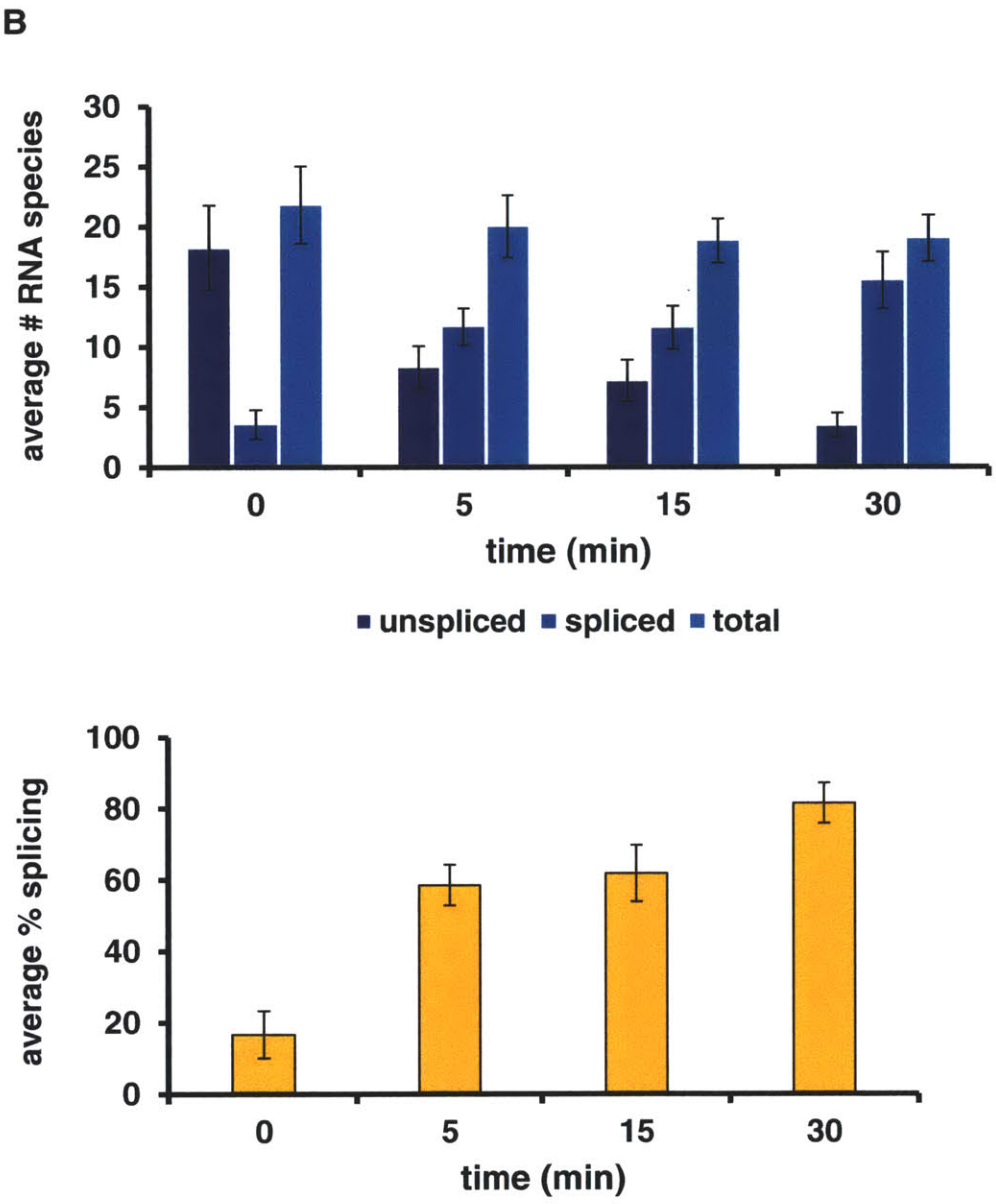
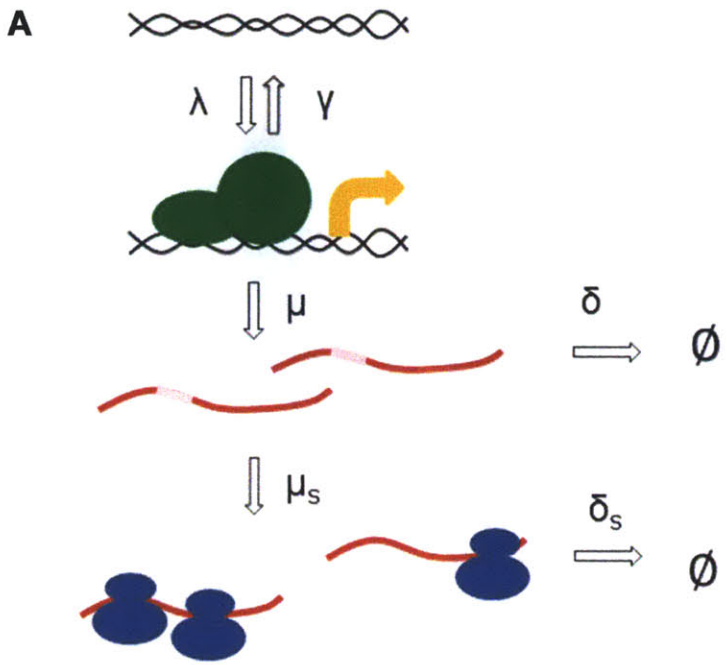
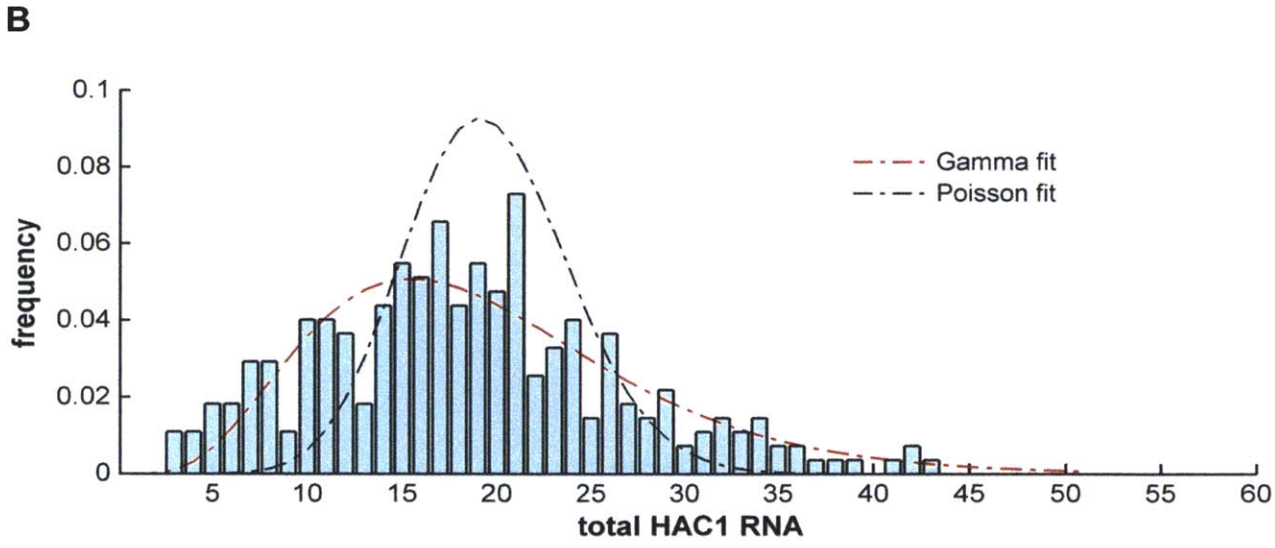


Figure 2 | Measuring splicing under UPR-inducing stresses. A, A single optical slice image of *HAC1* RNA in DTT-treated yeast cells for 15 minutes, and hybridized to Cy 5-labeled exon probes and Alexa 594-labeled intron probes, and costained with dapi (cyan). **B**, (Top) Dynamics of mean levels of spliced, unspliced and total *HAC1* in cell populations treated with DTT. (Bottom) Dynamics of mean percentage of splicing.

Fig 3


A
λ
γ
μ
δ
Ø
μs
δs
Ø


B
0.1
0.08
0.06
0.04
0.02
0
frequency
5
10
15
20
25
30
35
40
45
50
55
60
total HAC1 RNA
Gamma fit
Poisson fit

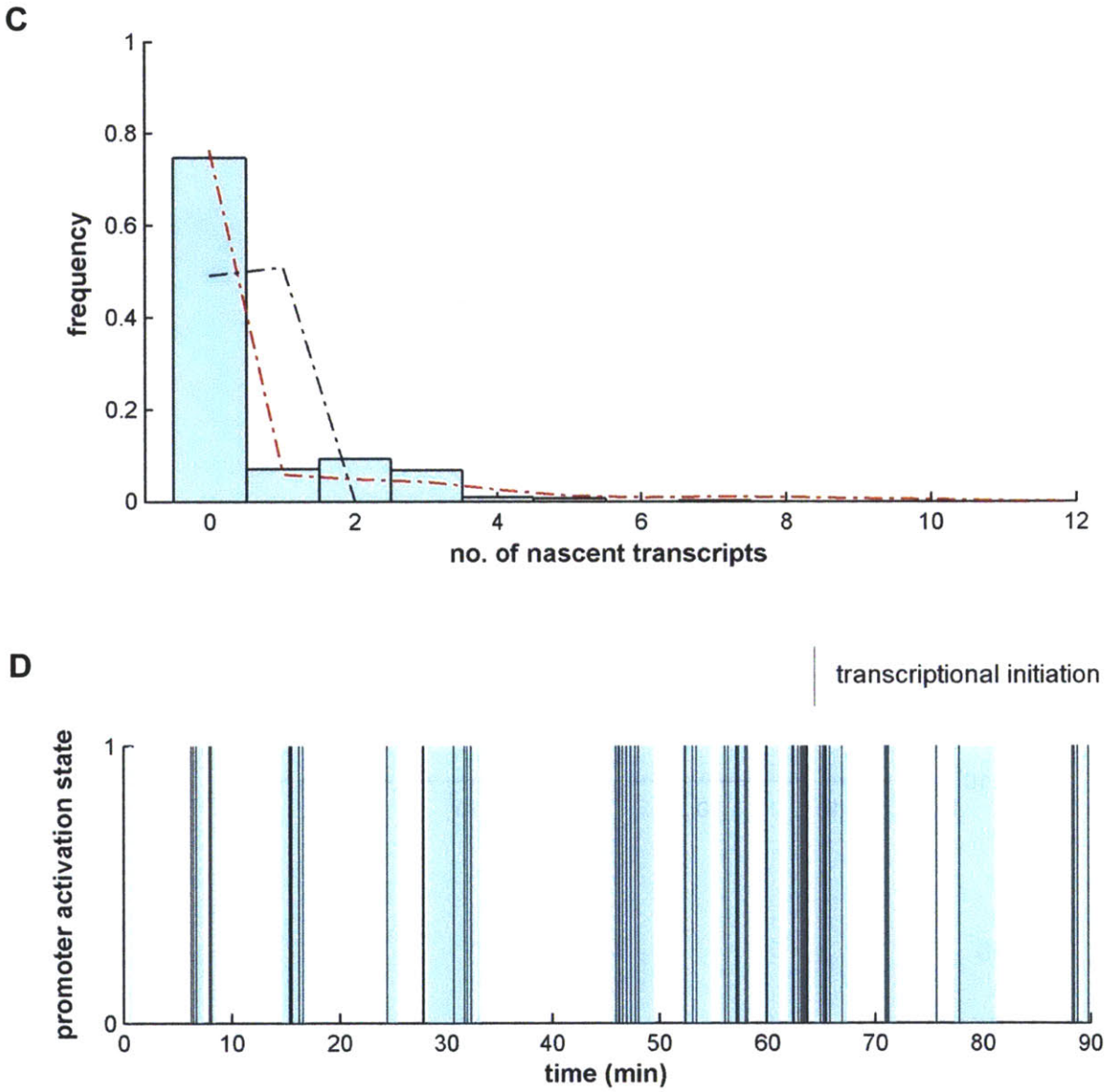
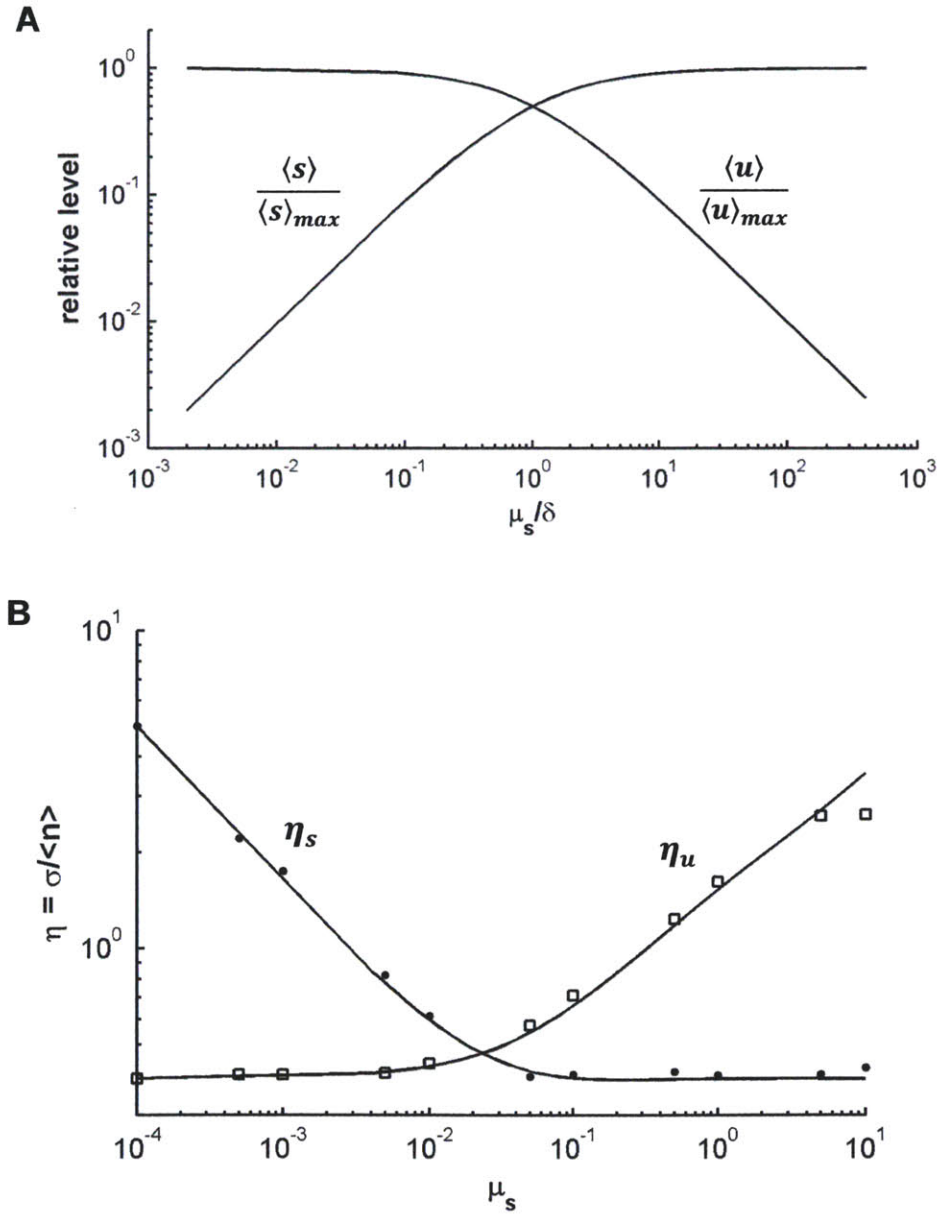


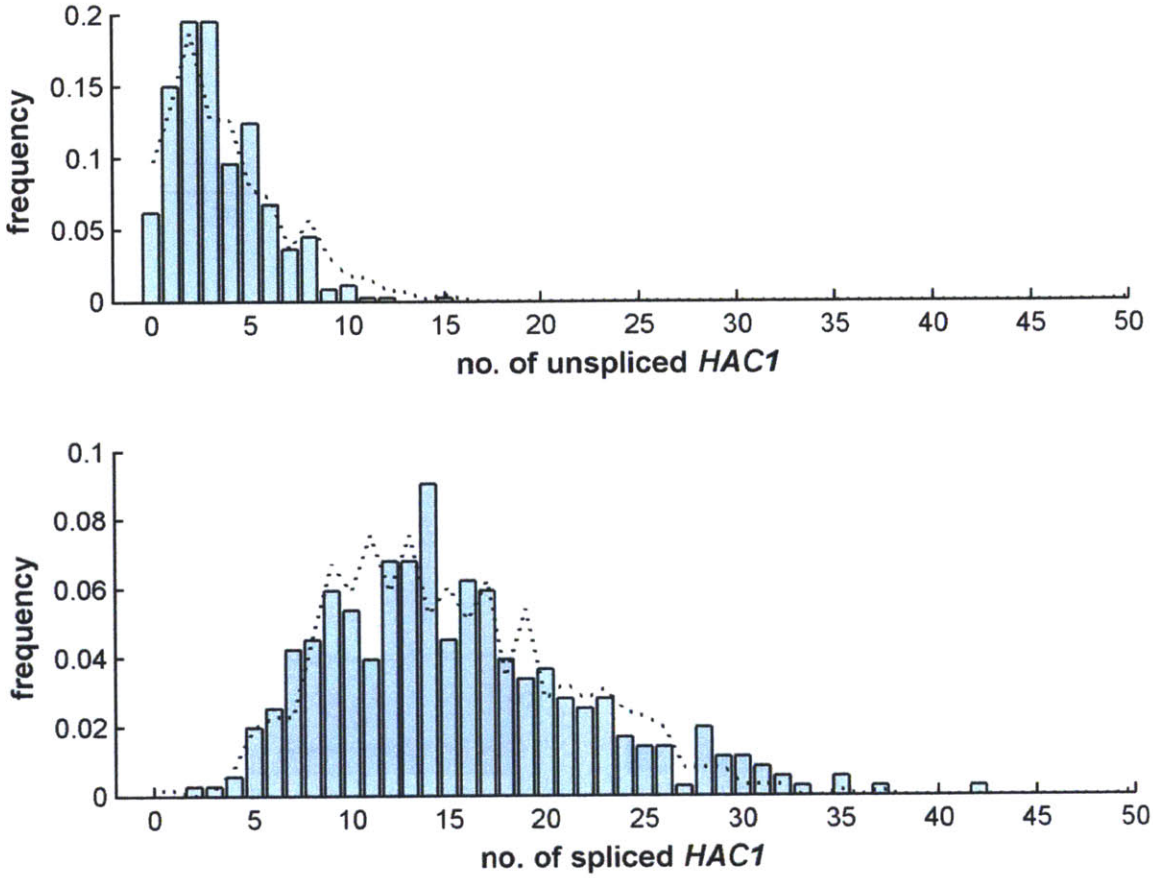
Figure 3 | *HAC1* exhibits transcriptional bursting. A, Schematic of a two-state promoter-transcription-coupled-to-splicing model. λ and γ represent promoter activation, inactivation rates, and μ and μ_s represent transcriptional and splicing rates respectively. δ and δ_s represent turnover rates of the unspliced and spliced RNA respectively. **B**, Distribution of total *HAC1* RNA in cells under normal conditions, and fitted separately to a gene-activation-inactivation model (Gamma distribution) and a Poisson distribution for comparison. **C**, Distribution of number of nascent transcripts, and fitted to simulations assuming either a gene-activation-inactivation model or a Poissonian transcription model.

D, Gene-activation-inactivation model used to simulate the *HAC1* expression kinetics. Cyan shadings indicate periods during which the promoter is active per cell generation time. The promoter switches between on and off states depicted by 1 and 0 on the y-axis. Initiation of a single transcript is denoted by a vertical line.

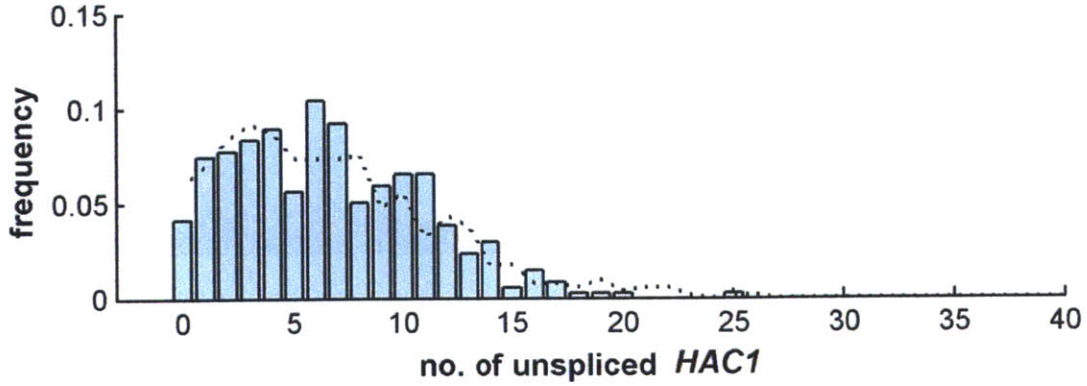
Fig 4

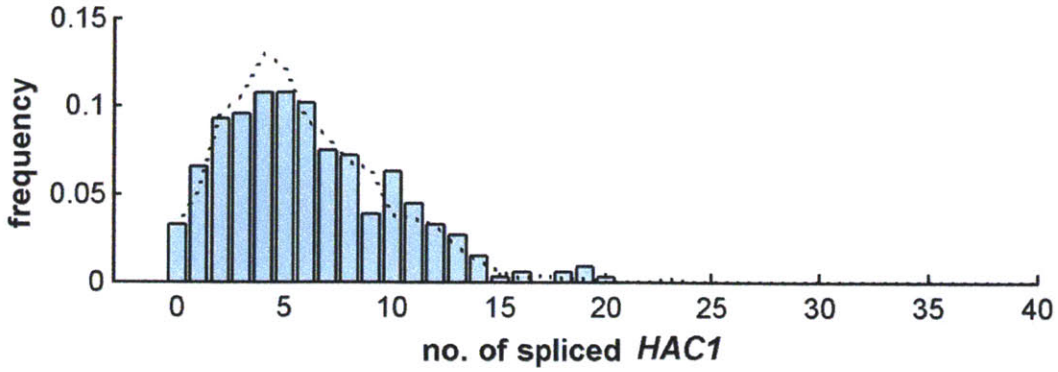


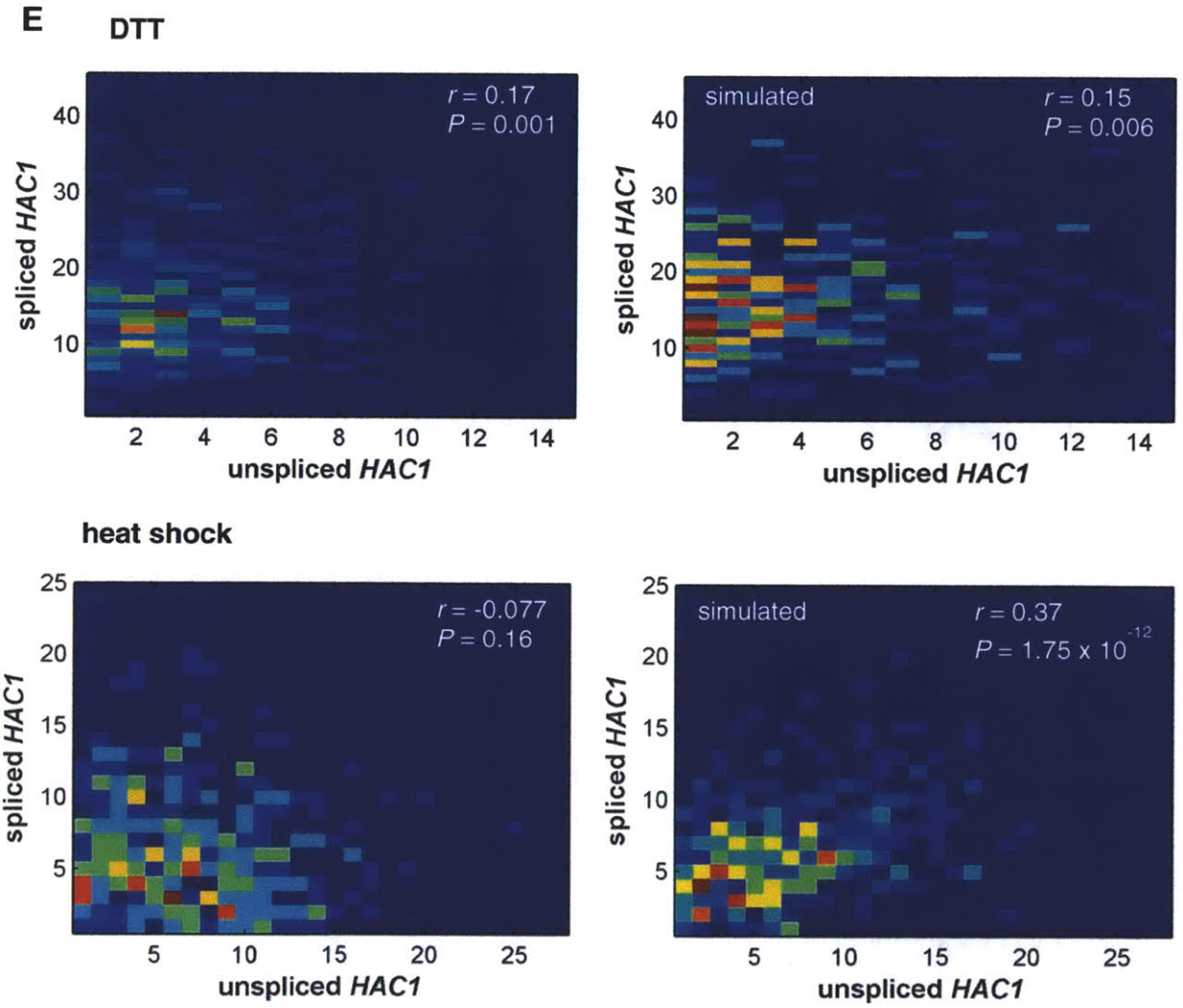
C

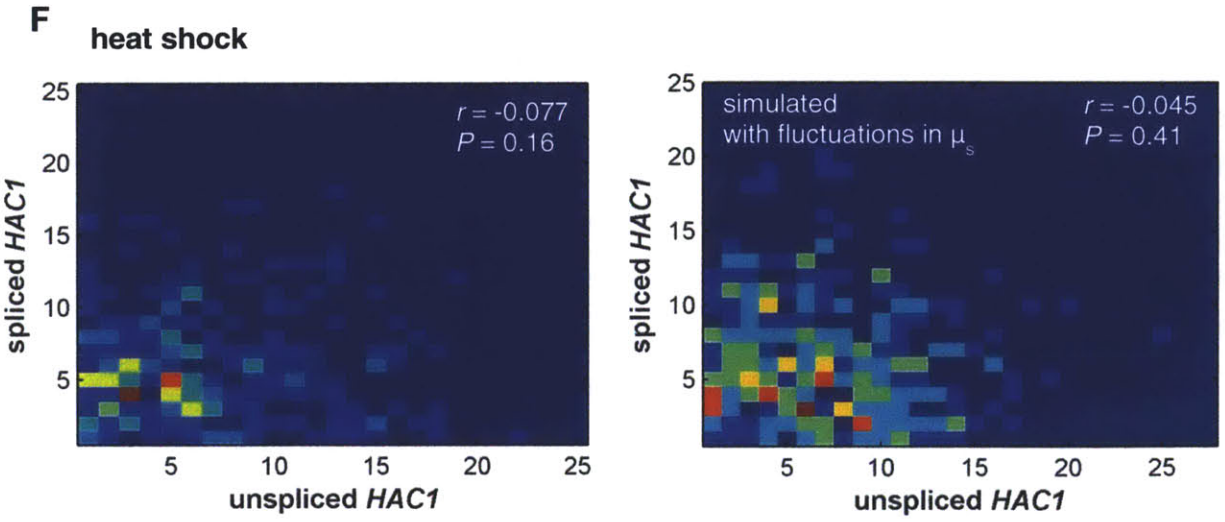

frequency
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30 35 40 45 50
no. of unspliced *HAC1*
frequency
0.1
0.08
0.06
0.04
0.02
0
0 5 10 15 20 25 30 35 40 45 50
no. of spliced *HAC1*

D


frequency
0.15
0.1
0.05
0
0 5 10 15 20 25 30 35 40
no. of unspliced *HAC1*


0.15
0.1
0.05
0
frequency
0
5
10
15
20
25
30
35
40
no. of spliced HAC1


E
DTT
r = 0.17
P = 0.001
simulated
r = 0.15
P = 0.006
spliced HAC1
unspliced HAC1
heat shock
r = -0.077
P = 0.16
simulated
r = 0.37
P = 1.75 x 10-12



$$\frac{\sigma_{us}^2}{\langle u\rangle\langle s\rangle} = \underbrace{\frac{\tau_u}{\tau_u+\tau_s}\left(\eta_u^2-\frac{1}{\langle u\rangle}\right)+\frac{\tau_s}{\tau_u+\tau_s}\left(\eta_s^2-\frac{1}{\langle s\rangle}\right)}_{\text{synthesis rate fluctuations}} - \underbrace{\frac{\tau_s}{\tau_u+\tau_s}\left(\frac{\sigma_{\mu_s}^2}{\langle\mu_s\rangle^2}\frac{\tau_{\mu_s}}{\tau_s+\tau_{\mu_s}}\right)}_{\text{splicing rate fluctuations}}$$

Figure 4 | Considerably higher variability in the degree of splicing is observed under heat stress, which can be explained by increased fluctuations in splicing efficiency. **A**, Steady-state dependence of relative mean unspliced and spliced RNA abundance on normalized splicing rate μ_s/δ. $\langle\cdot\rangle$ denotes averaged values. **B**, Dependence of η or CV (standard deviation divided by the mean) on splicing rate. The solid lines represent analytical results confirmed by predictions from simulations (denoted by markers). The parameters fitted to the *HAC1* RNA abundance data were used in the simulations. Details of the analytical expressions are given in the Supplementary information. **C**, Distributions of unspliced and spliced *HAC1* RNA in cells treated with DTT for 30 minutes, and fitted to stochastic simulations of the model in Fig. 3a using the fit parameters for the transcription module (dotted line). **D**, Distributions of unspliced and spliced *HAC1* RNA in cells subjected to heat shock for 30 minutes, and fitted to simulations. The transcriptional dynamics used were the same, except that the unspliced and spliced RNA degradation rates were increased. And fluctuations in the mean splicing

rate were simulated to obtain the fit in the plot. Details of the simulation procedures and parameters used are given in the Materials and Methods section. **E**, Scatterplots of spliced and unspliced *HAC1* RNA counts. Experimental data is shown on the right, while results from simulations are on the left. Simulations were performed as described in B and C above assuming no extrinsic noise in transcript abundance for the DTT- and heat-shock-treated population data. **F**, Fluctuations in the mean splicing rate were included to the model to generate the plot for the heat shock data (right). The expression for normalized covariance of the spliced and unspliced RNA is shown at the bottom.

SUPPLEMENTARY FIGURES AND TABLES

Table S1. List of probe sequences for *HAC1*

Exon probe number	Sequence	Intron probe number	Sequence
1	gttcaaaatcagtcatttcc	1	tcatcgtaatcacggctgga
2	gttcgattgcgaattactag	2	cttgttcactgtagtttcct
3	ttgaagttggtagggatagc	3	agcaaaagctggggctagtg
4	ttttccttggaggcagagtc	4	aaaaaaaagaaaaaaagcag
5	ttttcctcttttgtcttggc	5	catcagagaaccacgactaa
6	acgctcgatccttcgctgtt	6	gtactttaaccggctcctcc
7	cagctcttctgtttctcaaa	7	accctgcattctgcttttga
8	tttttctctctgctctggtg	8	gaaaaaaagaaagcttccaa
9	gagatactgcagatgtagtc	9	gttcaggaaaaactagcata
10	ccaaaagagaacattttctc	10	taagaaaagaatggctctat
11	ggttgacgctgttcagtaaa	11	caagccgtccatttcttagt
12	tcgtggtcagccagtttttc		
13	gtggctgcaagtcaacgcgt		
14	caagagaagcaacaaaagcg		
15	ctctggaaatccctgtactc		

16	tgtccagtgaagcgcccctc		
17	agacgacgagtgcgaactgg		
18	gaggtgaaggtgtgaacgta		
19	gcaggctccattgtacagtt		
20	catactcttgggcgacaaag		
21	tcttggtccgacgcggaatc		
22	acatctgcagctcccatgaa		
23	actctggtacattttccgtc		
24	tctacggcaggtagcgtcgt		
25	ccgcatcaaacaaattgttg		
26	tgggtctgccaacggcgagg		
27	ttcccgctatatcgtcgcag		
28	gaattgtcaaagggtagact		
29	acgccaattgtcaagatcaa		
30	actgcgcttcggacagtaca		
31	aattcaaatgaattcaaacc		
32	tgaagtgatgaagaaatcat		

Fig S1

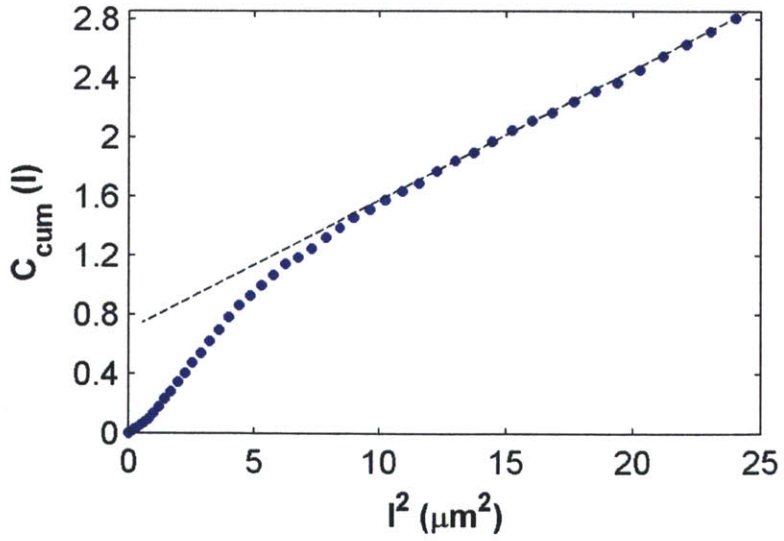
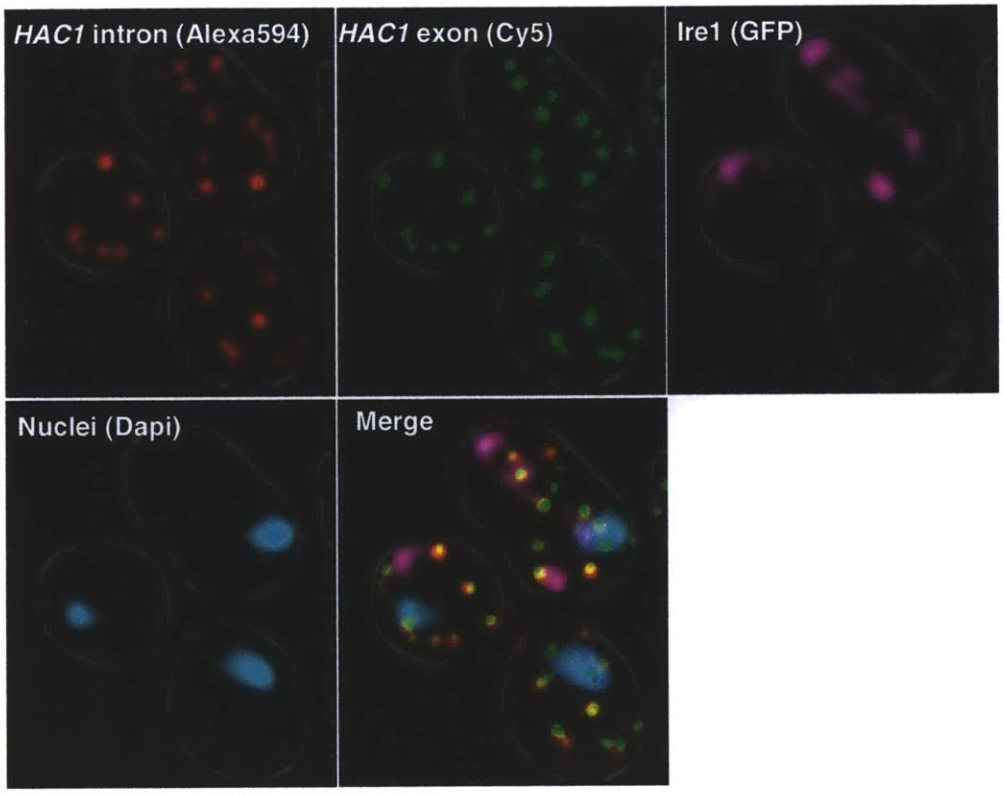


Figure S1 | Colocalization analysis using particle image cross correlation spectroscopy method. The cumulative correlation function C_{cum} plotted as a function of the radius l of a circle. The dotted line is fitted to the linear portion of the curve. The offset of the dotted line is equal to the correlation factor (or percentage colocalization).

Fig S2

A



B

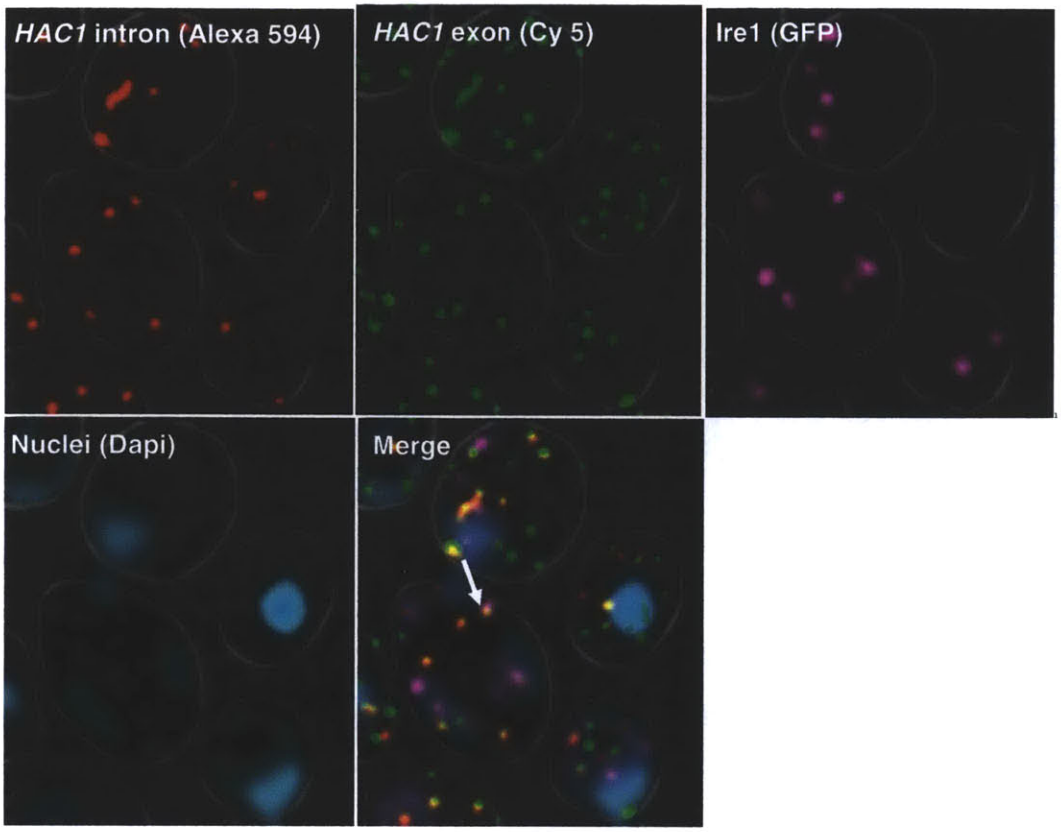


Figure S2 | Detection of clustering of activated splicing factor Ire1p responsible for the splicing of *HAC1* under UPR-induced stress. A, A single optical slice image of *HAC1* RNA and Ire1p-fused-with-GFP in normal growing cells hybridized to Cy 5-labeled exon probes and Alexa 594-labeled intron probes, and costained with dapi (cyan). Unspliced *HAC1* are detected as colocalized spots in both channels (yellow). Ire1p is diffusely localized before UPR stress. **B**, A single optical slice image of *HAC1* RNA and Ire1p-fused-with-GFP in DTT-treated cells for 15 minutes fixed and hybridized to Cy 5-labeled exon probes and Alexa 594-labeled intron probes, and costained with dapi (cyan). Ire1p cluster detected colocalizing with an unspliced *HAC1* RNA, indicated by the white arrow.

Fig S3

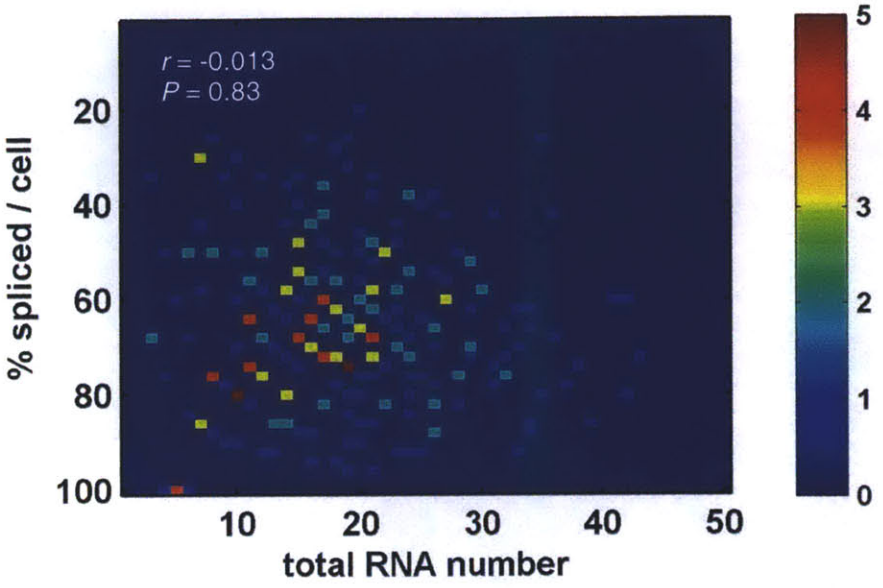
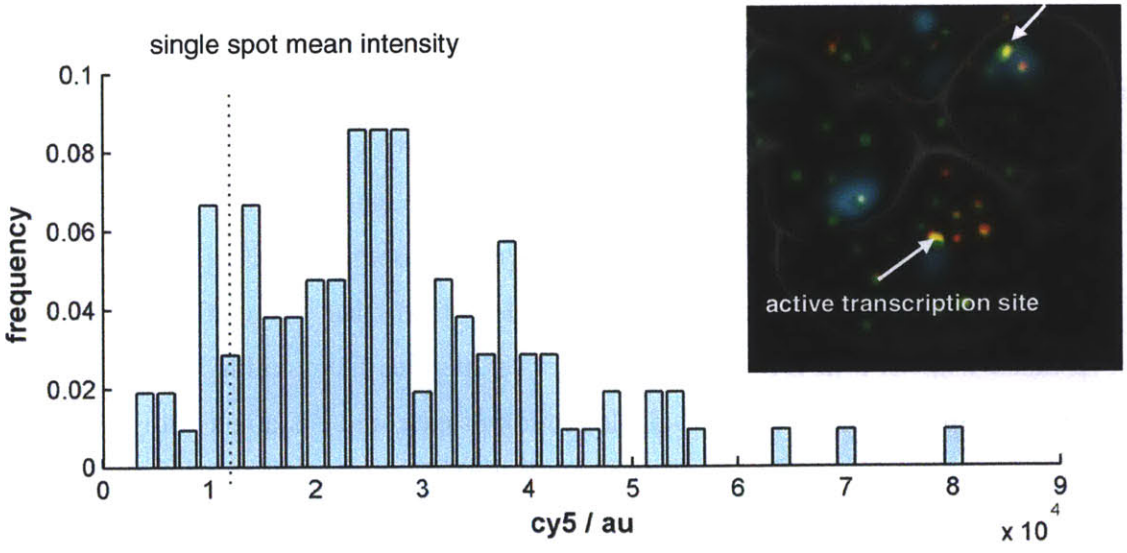


Figure S3 | The degree of splicing and the total *HAC1* count are uncorrelated in single cells. Scatterplot of the degree of splicing (spliced RNA count divided by total RNA count) and total *HAC1* RNA count.

Fig S4



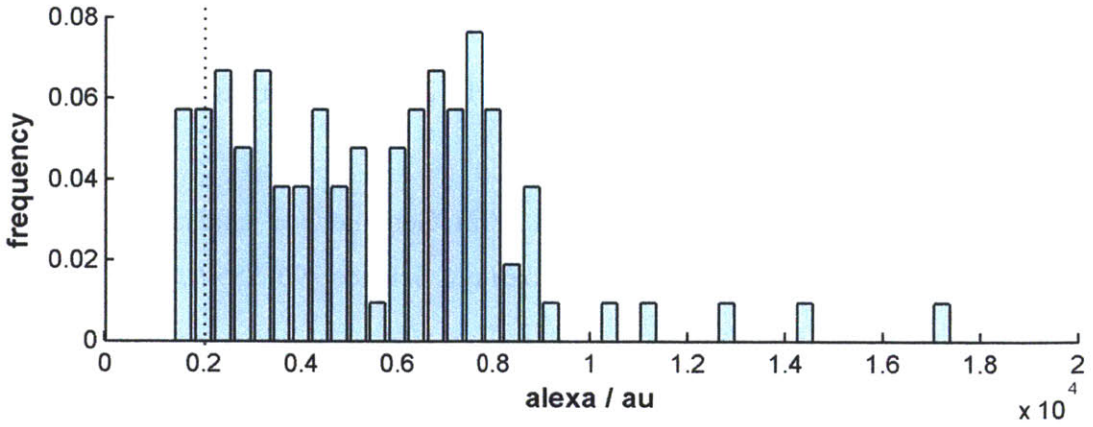


Figure S4 | Determination of nascent transcript counts. Histogram of nuclear transcription site intensities (integrated). The cytoplasmic RNA intensities representing a single RNA can be fit to a Gaussian distribution and the mean is depicted by the dotted line. The nuclear signal intensities detected by Cy5 (top) and Alexa 594 (bottom) can be fitted to integer multiples of the single-peak intensity. The inset shows the active transcription sites (in yellow, and indicated by white arrow).

REFERENCES

Aragon, T., E. van Anken, D. Pincus, I.M. Serafimova, A.V. Korennykh, C.A. Rubio, P. Walter. (2009) Messenger RNA targeting to endoplasmic reticulum stress signaling sites. Nature *457*, 736–740.

Becskei, A., Kaufmann, B.B. and van Oudenaarden, A. (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat. Genet. *37*, 937–944.

Bernales, S., Papa, F.R., Walter, P. (2006) Intracellular signaling by the unfolded protein response. Annu Rev Cell Dev Biol *22*, 487–508.

Bingham, P. M., Chou, T. B., Mims, I., Zachar, Z. (1988) On/off regulation of gene expression at the level of splicing. Trends Genet., May *4(5),* 134-8.

Brodsky A.S., Silver, P.A. (2000) Pre-mRNA processing factors are required for nuclear export. RNA Dec; *6(12)*, 1737-49.

Chubb, J. R., Trcek, T., Shenoy, S. M., Singer, R. H. (2006) Transcriptional pulsing of a developmental gene. Curr. Biol. *16*, 1018-1025.

Clark, T. A., Sugnet, C. W., Ares, M. Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science *296(5569)*, 907–910.

Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem *81*, 2340-2361.

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D., Hughes, T.R. (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals post-transcriptional control of ribosome biogenesis factors. Mol. Cell. Biol. *24*, 5534–5547.

Lindquist, S. (1981) Regulation of protein synthesis during heat shock. Nature *293*, 311-314.

Maamar, H., Raj, A., and Dubnau, D. (2007) Noise in gene expression determines cell fate in Bacillus subtilis. Science *317*, 526–529.

Mason, P. B. and Struhl, K. (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. Mol. Cell *17*, 831–840.

McAdams, H.H., and Arkin, A. (1997) Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA *94*, 814–819.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. and van Oudenaarden, A. (2002) Regulation of noise in the expression of a single gene. Nat. Genet. *31*, 69-73.

Paré, A. et al. (2009) Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr. Biol. *19*, 2037-2042.

Paulsson, J. (2004) Summing up the noise in gene networks. Nature *427*: 415-418.

Peccoud, J. & Ycart, B. (1995) Markovian modeling of gene-product synthesis. Theor. Popul. Biol. *48*, 222–234.

Pincus D, et al. (2010) BiP binding to the ER-stress sensor Ire1 tunes the homeostatic behavior of the unfolded protein response. PLoS Biol *8*, e1000415.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., Rifkin, S. A., Andersen, E., van Oudenaarden, A. (2010) Variability in gene expression underlies incomplete penetrance. Nature *463*, 913-918.

Raj, A., Van Den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–79.

Semrau, S. Holtzer, L., González-Gaitán, M., Schmidt, T. (2011) Quantification of biological interactions with particle image cross-correlation spectroscopy (PICCS). Biophys. J., *100(7)*, 1810-1818.

Smith, C.W., Patton, J.G. and Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. Annu. Rev. Genet. *23*, 527-577.

Suel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B. (2007) Tunability and noise dependence in differentiation dynamics. Science *315*, 1716–1719.

Suteret, D. M., Molina, N., Gatfield, K., Schneider, U., and Naef, F. (2011) Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472-474.

Tirasophon, W., Welihinda, A. A., Kaufman, R. J. (1998) The endoribonuclease activity of mammalian IRE1 autoregulates its mRNA and is required for the unfolded protein response. Genes Dev.*12*, 1812–24.

Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S., Walter, P. (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101(3)*, 249–58.

Zenklusen, D., Larson, D.R., Singer, R.H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol *15*, 1263–1271.

Zhang, G., Taneja, K.L., Singer, R.H., Green, M.R. (1994) Localization of pre-mRNA splicing in mammalian nuclei. Nature *372*, 809–812.

Chapter 4

DEVELOPING A SINGLE CELL ASSAY FOR TRANSLATION

ABSTRACT | Translation is one of the key post-transcriptional mechanisms for regulating gene expression in eukaryotes. Ribosomal profiling via RNA sequencing and microarrays are powerful methods for monitoring translational activity on a genome-wide level. However, these techniques do not allow translational profiling of single cells. Here, we present an approach using single-molecule fluorescence *in situ* hybridization (sm-FISH) and protein synthesis inhibitors to measure ribosome numbers of individual transcripts in single cells. We describe preliminary results where we applied our method to monitor translation in exponentially growing yeast, and to explore the changes in translational regulation upon switching the cells from nutrient rich to starvation conditions.

INTRODUCTION

Protein synthesis is one of the most important processes in the cell, given that most cellular functions are mediated by proteins. The development of tools for monitoring translation has led to the discovery of translational gene control that goes beyond the traditional view that all cellular messages are readily translated into proteins. Specifically, genome-wide studies of translation in yeast using ribosomal profiling techniques have found genes that are subjected to extensive translational regulation under normal growth conditions and in response to stress and environmental stimuli (Arava et al., 2003; Kuhn et al., 2001; Preiss et al., 2003; Ingolia et al., 2009). In mammalian cells, the fraction of genes with incongruent mRNA-protein expression is considerably higher, suggesting that the protein abundance of most genes is strongly controlled at the translational level (Schwanhäusser et al., 2011).

Despite these advances in the study of translation, no current technique allows translational analysis in single cells, although such information is important for understanding fundamental questions on gene expression. It has long been appreciated that cellular processes driven by small numbers of molecules are subjected to stochastic variation, which contributes to variability within a cell population and can have phenotypic consequences (Arkin et al., 1998; Chang et al., 2008; Spencer et al., 2009; Weinberger et al., 2008). Such translational noise can generate phenotypic heterogeneity that might be advantageous in fluctuating environments, or might hinder the cells' ability to respond appropriately under conditions where a precise response is critical. Here, we present an approach to use single-molecule RNA fluorescence *in situ* hybridization (sm-

FISH) (Raj et al., 2008) to quantify translation in single cells. The aim is to use sm-FISH and protein synthesis inhibitors which dissociate ribosomes from mRNAs i.e. puromycin, for detecting ribosomal occupancies of each mRNA in a cell population with single-cell resolution. We describe preliminary results where we applied the method to monitor translation in exponentially growing budding yeast, and to study the translational control of genes under different growth conditions.

RESULTS

Using sm-FISH and protein synthesis inhibitors to infer translational status

The degree of association of mRNAs with ribosomes is a direct measure of translational activity. Hence, we designed our approach to determine this in single cells. A typical sm-FISH experimental procedure involves designing oligonucleotide probes tiling the entire coding sequence of the mRNA of interest, with the assumption that the mRNA is fully naked and entirely accessible to probes (Fig. 1a). However, in reality, a more accurate depiction of the mRNA would most likely include secondary structures and its association with proteins such as ribosomes and RNA-binding proteins, thus limiting the number of FISH probes that can bind to the transcript (Fig. 1a). Taking advantage of protein synthesis inhibitors i.e. puromycin, which is a small molecule that mimics acyl–transfer RNA and serves as a polypeptide chain terminator (Blobel and Sabatini, 1971; Rodriguez et al., 2006), we surmise that live cells treated with puromycin prior to fixation for FISH would have a greater number of probes that hybridize to the transcript for actively translating genes, as compared to cells without the puromycin

treatment (Fig. 1b). Notably, puromycin affects only elongating ribosomes (Maroney et al., 2006; Nottrott et al., 2006), and thus, this approach detects mRNAs undergoing active translation.

Using this approach, the difference in the mean number of probes bound to the mRNA in the two conditions will reflect its degree of association with ribosomes. To achieve this, our strategy will involve (i) measuring the intensity of spots in cells hybridized with a single fluorescent probe; (ii) measuring and gating the intensities of mRNA spots in cells hybridized with a tiling array of fluorescent probes detected using sm-FISH (Fig. 2); (iii) repeating procedure (i) and (ii) for cells with and without puromycin treatment.

Mesauring translation in exponentially growing yeast

To test the validity of our approach, we applied it to assess translation in rapidly growing diploid yeast cells. Four of the five genes in our preliminary study showed the shift in the mRNA intensity distribution towards higher values in the puromycin condition as expected (Fig. 3). No appreciable change in intensity was observed for *GCN4* transcripts consistent with ribosomal profiling measurements (Hinnebusch et al., 2005; Arava et al., 2003; Ingolia et al., 2009). *GCN4* is a well-studied example of a translationally regulated gene (Hinnebusch et al., 2005). During log-phase growth, *GCN4* mRNA is highly expressed but its protein counterparts are absent due to translational repression by uORFs in its 5'UTR. This repression is relieved under amino-acid starvation, where *GCN4* proteins are critical for activating the amino-acid stress response.

These results support that the shift in mRNA intensities with puromycin is specific to the effects of ribosomal release from mRNAs.

Using the intensity of a single fluorescent probe, we can translate mRNA spot intensity into the number of probes bound to an mRNA for the two distributions and accordingly, derive the mean number of probes possibly occluded from binding to the mRNA due to ribosomes (Fig. 4a-b). We confirmed experimentally that the fluorescence intensity scales linearly with the number of probes (Fig. 5), indicating minimal fluorophore quenching effects.

Each ribosome has a footprint of approximately 30 nucleotides (Wolin and Walter, 1988), while we designed our probes to be 20 nucleotides each. We sought to determine the effective average number of probes disrupted from binding to an mRNA due to a single ribosome. To do this, we first determined the minimum probe length necessary for the probe to bind to its target sequence. We hybridized cells to probe sets of varying probe lengths i.e. 11, 14, 17 and 20 nucleotides, and measured their corresponding spot signal intensities (Fig. 6). From fitting the data to a sigmoidal function as expected from a thermodynamic probe affinity model (Yilmaz and Noguera, 2004), we find that the critical length required to achieve half-maximal intensity is roughly 16.5 nucleotides. We then derived analytically as well as simulated kinetic models of translation to determine the relationship between the minimum length L_s required for probe binding and the corresponding number of probes occluded per ribosome (Fig. 7) (Supplementary information). We estimate that at the minimum probe length of 16.5 nucleotides, a single ribosome prevents on average 2.1 probes from

binding. In our subsequent analyses, we used this scaling factor to derive the number of bound ribosomes from the probe count measurements.

To assess the feasibility of our approach, we compared our results in Fig. 3 to data from a genome-wide polysomal profiling study performed similarly in rapidly growing yeast (Arava et al., 2004). For the 5 genes studied, we observed a good correlation between the mean ribosome numbers determined using our method and the data reported by Arava et al. (Fig. 8). We detected mRNAs that are minimally translating (0.2 ribosome i.e. *GCN4*) and those that are highly actively translating (5 ribosomes i.e. *CAR2*). However, we also observed that measurements from sm-FISH were systematically lower than Arava's data for all 5 genes.

Measuring translation in amino acid starvation conditions

Next, we tested our approach by monitoring the translational status of the 5 genes in yeast cells under starvation conditions. We applied the same procedure as described earlier, and determined the change in the number of bound ribosomes per mRNA in amino acid-starved cells as compared to cells in rich media. We compared the data obtained using sm-FISH with the results from ribosomal profiling using RNA-sequencing (Fig. 9). Similar to the findings reported by Ingolia et al., we detected that among the 5 genes studied, *SRO9* and *CAR2* were translationally repressed (fold change < 1) while *GCN4* and *TRR2* were translationally upregulated (fold change > 1) under starvation. *MET22*, on the other hand, remained as translationally active in both rich and starvation conditions (fold change = 1).

Distribution of ribosome numbers per mRNA

Other than the mean metric, we were interested in determining the full distribution of ribosome numbers for the cell population. To this end, we deconvolved the intensity distributions for *MET22* measured with and without puromycin using the single fluorescent probe intensity data and we obtained the distributions of occluded probe numbers under the two conditions. Unexpectedly, we found very narrow distributions in both cases (Fig. 10a). From these two distributions, we inferred the Fano factor (variance divided by mean) of the ribosome distribution to be 0.8 ± 0.2, which indicates Poissonian statistics (Fano factor of a Poisson process is 1). We similarly computed the Fano factor using Arava's data, and observed a 9-fold difference between our results (Fig. 10b).

DISCUSSION

Results so far suggest the feasibility of monitoring translation in budding yeast using sm-FISH and protein synthesis inhibitors. We showed that the mean ribosome numbers of 5 genes measured in yeast cells under rich and starvation conditions largely agree with the results from ribosomal profiling using microarrays and RNA-sequencing. The possible explanations for systematic deviations of our results from Arava's data could be that puromycin does not completely remove ribosomes on the RNA, or that we were over-estimating the single fluorescent probe intensity, or that these differences were strain-specific, and or that our assay detects only actively translating mRNAs while polysomal analysis captures both stalled and active ribosomes which are bound onto mRNAs. These could be systematically addressed by 1. titrating the concentration of

puromycin and measuring the resultant ribosome count. 2. measuring the intensities of single fluorescent probes attached onto glass slides and comparing them to measurements of single fluorescent probes hybridized in cells. 3. performing sucrose gradient and polysomal analysis in our strain and comparing the results to those obtained using sm-FISH.

In addition to reporting the mean, we were interested in determining the full ribosome distribution. We observed that the intensity distributions can be well described by a log-normal distribution, suggesting that the measured intensity fit the model of the product of an intrinsic intensity and extrinsic variable factors. The intensity distribution measured with puromycin can be approximately converted into the intensity distribution obtained without puromycin by the scaling factor I^-/I^+ and vice versa. However, when we deconvolved the intensity distributions to derive the corresponding probe count distributions, they turned out to be close to a Poisson distribution. We estimated the Fano factor of the *MET22* ribosome distribution derived using FISH to be roughly 0.8, which is inconsistent with Arava's data where the Fano factor is significantly greater i.e. 7. The greater inaccuracy intrinsic to single probe intensity measurements, which we used in the deconvolution, might have contributed to the stark discrepancy. Furthermore, our assumption of the underlying model for the measured intensities (whether measured intensity is the sum of the intensities of single fluorophores or whether it is the product of an intrinsic intensity and some variable factors) might have been incorrect.

Hence, while we showed proof of principle of our method, the appropriate interpretation of the intensity distributions and the validity of extracting the full ribosome distribution from these data remain to be worked out. Nonetheless, we have shown that

the *in situ* method, which eliminates the need for cell lysis and RNA extraction, provides an alternative assay for measuring the mean number of ribosomes per mRNA in yeast at the single transcript level. Future experiments using higher resolution imaging systems coupled with more advanced analysis methods would be expected to improve the technique to achieve the end objective of profiling translation in single cells.

MATERIALS AND METHODS

Yeast strain and media

GCN4 is the classic translationally regulated gene, however, it is highly expressed and we are unable to resolve individual transcripts using sm-FISH. To control its expression to obtain spatially well-separated mRNAs in yeast cells ideal for imaging, we transformed a plasmid with pTETO7 driving *GCN4* and pMYO2 driving rtTA into a homozygous *GCN4* knockout diploid strain in the BY4743 background (Open Bioystems). Thus, without doxycycline, *GCN4* is only basally-transcribed. This strain was subsequently used for all the experiments described in the text. All experiments involved growing cells in 5 ml of synthetic -URA media at 30 °C overnight (>16 hours), and diluting to an OD_{600} of 0.01 in 50 ml pre-warmed fresh media, and then allowed to grow to a final OD_{600} of 0.4. For the amino acid starvation experiment, we pelleted the cells for 5 minutes at 5000xg at 30 °C, and removed all media, followed by resuspending the cells in an equal volume of pre-warmed synthetic media without amino acids, and we repeated this 2x. Cells were returned to 30 °C with vigorous shaking for 20 minutes before fixing.

Puromycin treatment

For the experiments with puromycin, cells grown to semi-log phase were placed in fresh media containing a final concentration of 100 μg/ml puromycin (Sigma), and allowed to grow for roughly 3-5 minutes at 30 °C with vigorous shaking. After which, the cells were pelleted for 5 minutes at 5000xg at 30 °C, media was removed, and the pelleted cells were fixed.

Fluorescence *in situ* hybridization

We performed FISH as described by Raj et al., 2008. Pelleted cells were immediately fixed by adding 4% (v/v) formaldehyde diluted in PBS directly to the culture for 45 minutes at room temperature. The cell wall was digested using lyticase (Sigma), and stored in 70% (v/v) ethanol (for at least 3 hours) at 4 °C. Prior to hybridization, cells were rehydrated with 10% (v/v) formamide, 2 x SSC for 10 minutes. We used a hybridization buffer consisting of 10% (v/v) formamide, 2 × SSC, 1 mg/ml BSA, 10 mM VRC, 0.5 mg/ml *Escherichia coli* tRNA and 0.1 g/ml dextran sulfate. All hybridizations were carried out in solution using probes to Alexa 594 (Invitrogen) at optimal concentrations determined empirically, and were carried out overnight at 37 °C. The cells were attached to coverslips coated with concanavalin A (Sigma) prior to imaging. All probes used in this study were coupled to Alexa 594 (Invitrogen).

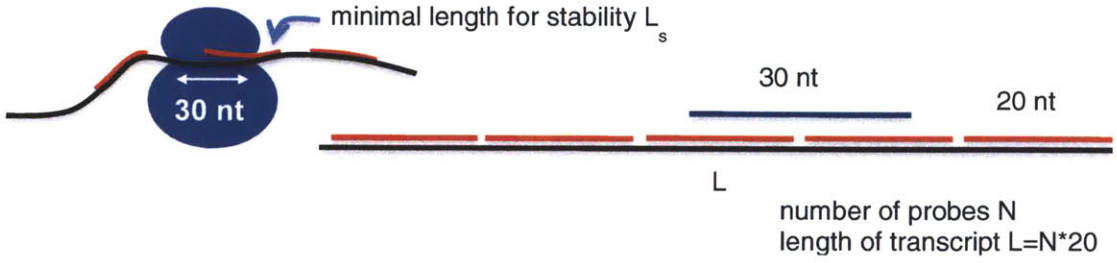
Image acquisition and analysis

Images were acquired using a Nikon TE2000 epifluorescence microscope equipped with a Princeton Instruments camera and a 100 x oil immersion objective. The microscope

settings used were slow camera readout (10kHz), 2x2 pixel binning and 1.5x magnification (1 pxl = 173 nm). Data was taken as stacks of images with a z direction step size of 0.2 μm using filters appropriate for Alexa 594 (Chroma Technology). To minimize photobleaching during imaging, we used antifade reagents including Trolox and the oxygen-scavenging solution glucose oxidase (Raj et al., 2008). Using custom-written MATLAB (Mathworks) software, we reduced the stacked images to two-dimensional images by maximum projection, and fitted the fluorescent spots to a 2D Gaussian as the model for the point spread function. Single transcript intensity was defined as the integrated intensity of the spot identified using the two-dimensional Gaussian mask algorithm.

SUPPLEMENTARY INFORMATION

Relationship between minimum length L_s required for probe binding and number of probes occluded per ribosome



Probability p for a ribosome excluding a particular probe $p = {}^{(10 + 2L_s)}/_{L}$

Average number of probes excluded per ribosome $f = Np = {}^{(10 + 2L_s)}/_{20}$

FIGURES AND TABLES

Fig 1

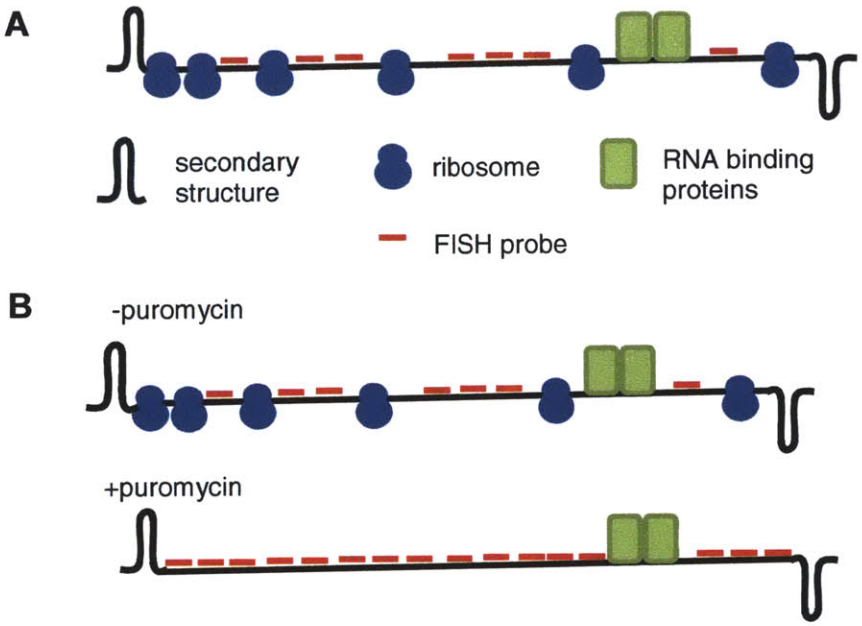


Figure 1 | Method for detecting ribosomes using sm-FISH and protein synthesis inhibitors. A, A typical depiction of the *trans* (ribosomes and RNA binding proteins) and *cis* (secondary structure) factors on a single mRNA transcript. **B**, Addition of the protein synthesis inhibitor puromycin is predicted to increase accessibility of the RNA strand to FISH probes.

Fig 2

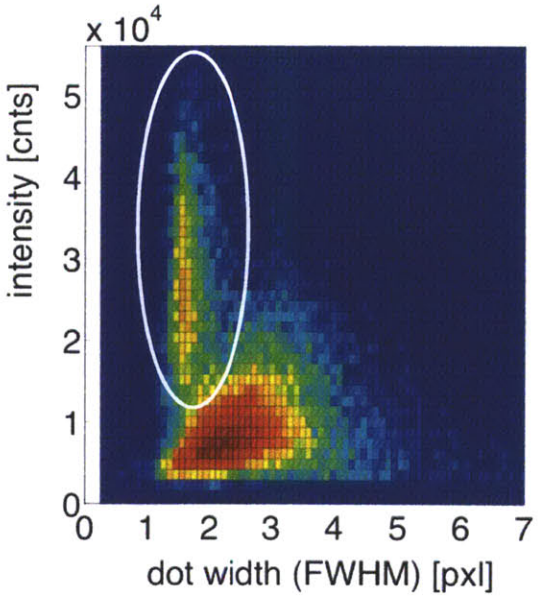


Figure 2 | Gating of spots based on spot intensity and width. Scatterplot of spot intensity and width for all mRNA spots in the cell population. The white circle represents the gated population with intensities greater than the minimum intensity (measured from a single fluorophore), and which fall within the expected dot width (~2 pixels) (See Materials and Methods).

Fig 3

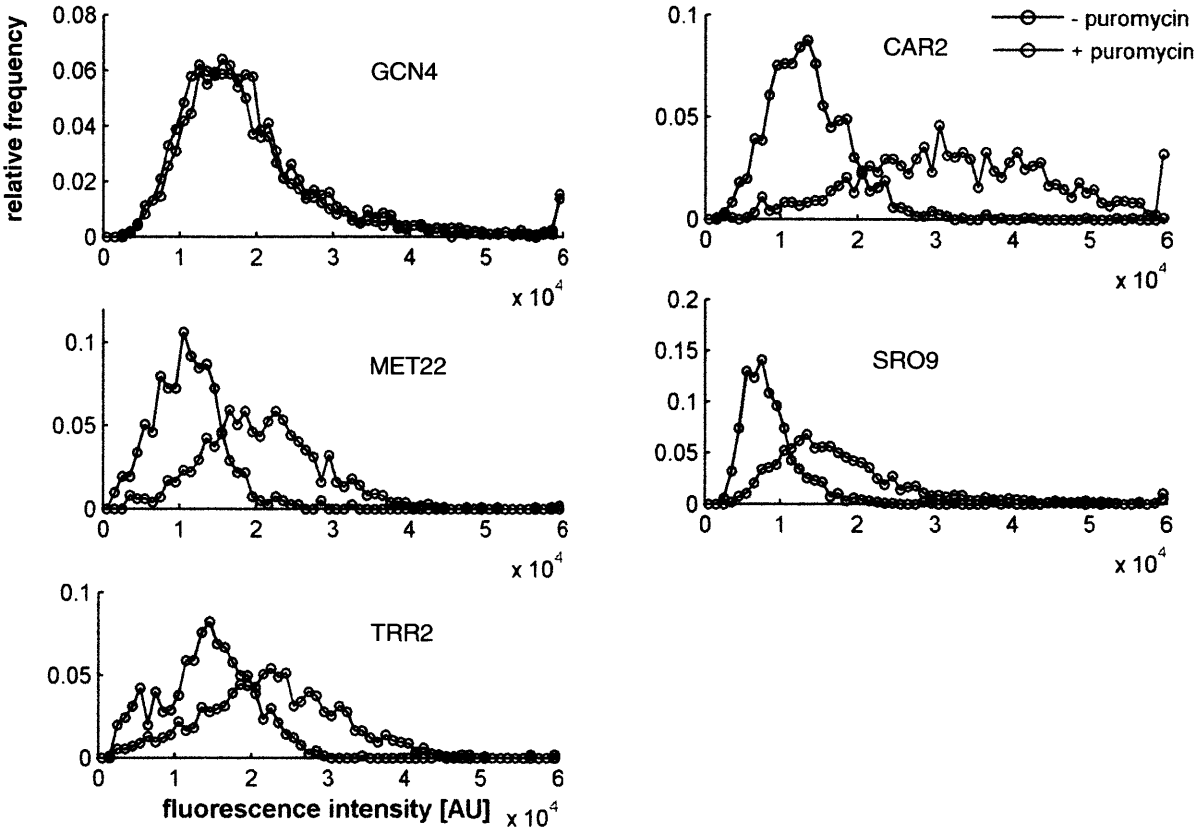


Figure 3 | Distributions of RNA spot intensities with and without puromycin. Histograms of intensities of FISH spots corresponding to the respective genes detected in cell populations treated separately with and without puromycin.

Fig 4

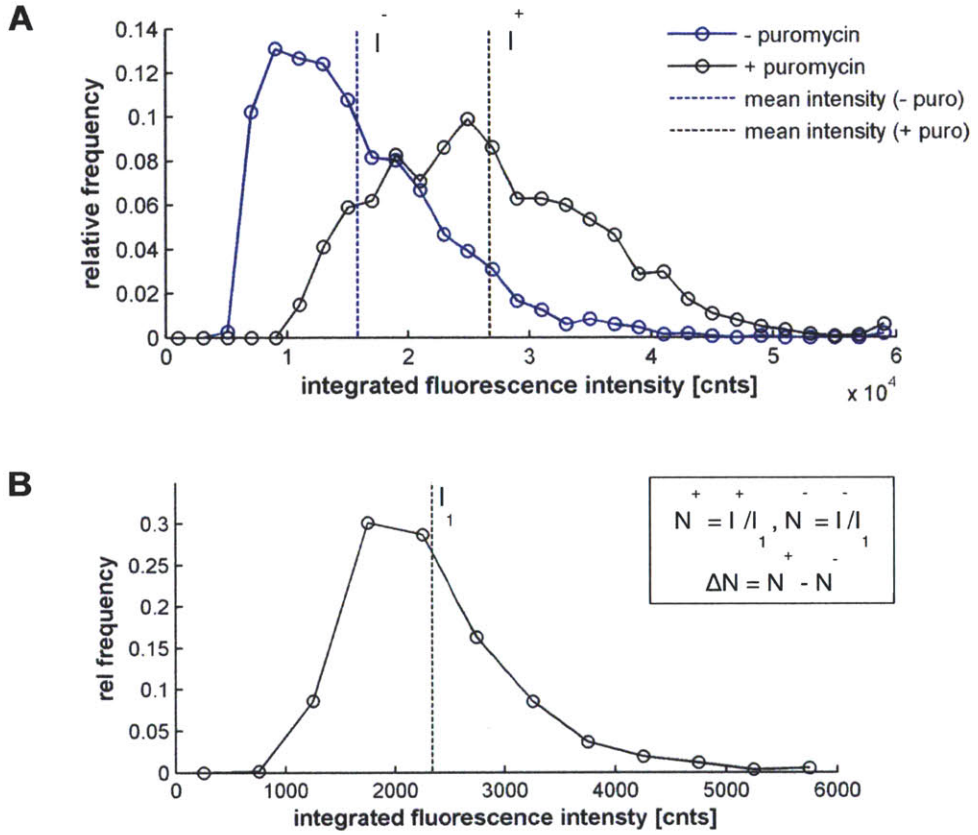


Figure 4 | Determining the mean number of probes disrupted by ribosomes. A, Representative histograms of intensities of spots detected in cells treated with or without puromycin. Dotted lines represent mean intensities of each distribution denoted by I^- and I^+. **B**, Representative histogram of intensity of spots detected in cells hybridized with a single fluorescently labeled probe. Dotted line represents the mean intensity denoted by I_1. N^+ and N^- represent the mean number of probes bound to each RNA transcript in each condition, and ΔN represents the mean number of probes possibly excluded from binding by ribosomes.

Fig 5

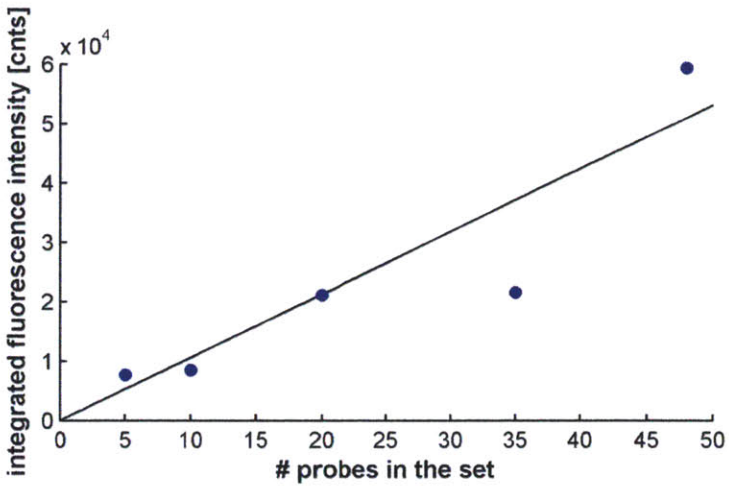


Figure 5 | Linear scaling of intensity with number of probes. Mean intensities of spots detected in cells hybridized with varying number of probes (denoted by blue spots), and are fitted to a linear line (n = 1 experiment).

Fig 6

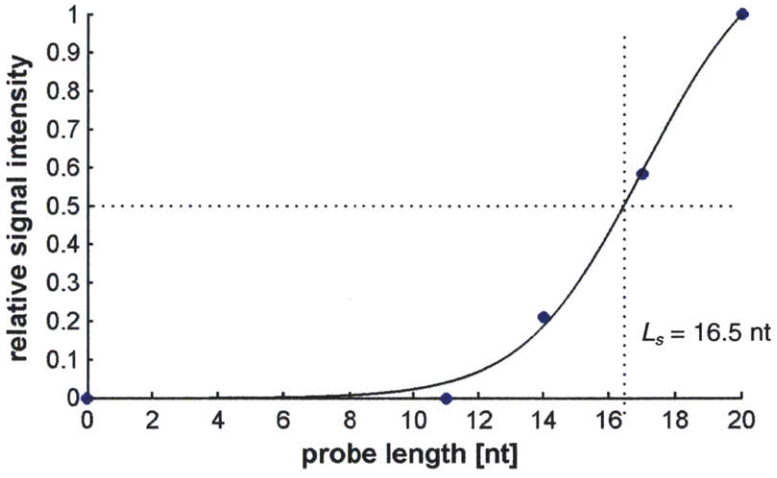


Figure 6 | Relative spot intensity as a function of probe length. Mean intensities of spots detected in cells hybridized with probes of varying length (11, 14, 17 and 20 nucleotides) and normalized to the maximum mean intensity (denoted by blue spots) (n =

1 experiment). The data is fitted to a sigmoidal curve. The minimum probe length L_s required to achieve half-maximal signal intensity is estimated to be 16.5 nucleotides.

Fig 7

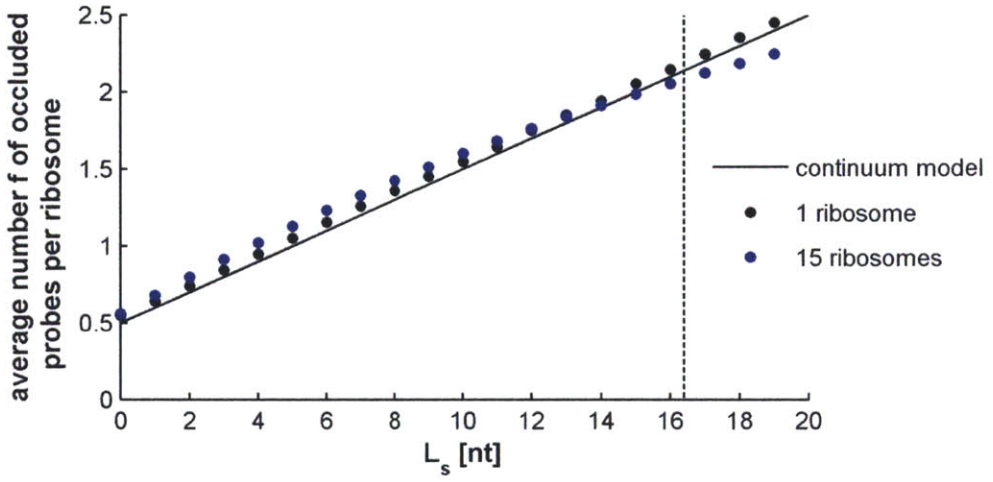


Figure 7 | The average number of occluded probes per ribosomes as a function of L_s. The analytical result is represented by the solid black line, while simulation results assuming different total number of ribosomes bound on a single transcript are indicated by black and blue spots. The dotted line denotes the minimum probe length L_s = 16.5 nt determined in Fig. 6, and roughly 2.1 probes on average are prevented from binding.

Fig 8

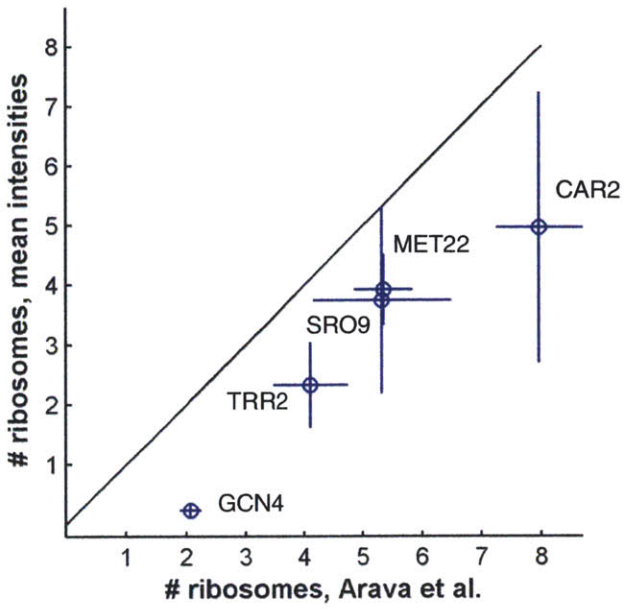


Figure 8 | Comparison of our results (y-axis) with results from the polysomal profiling method using microarrays reported by Arava *et al.* Performing the procedure described in the text using the distributions in Fig. 3 and the estimated scaling factor derived earlier, we computed the average number of bound ribosomes per transcript for the 5 genes, and compared our results to Arava's *et al.* Data denotes mean (n = 3) ± SEM.

Fig 9

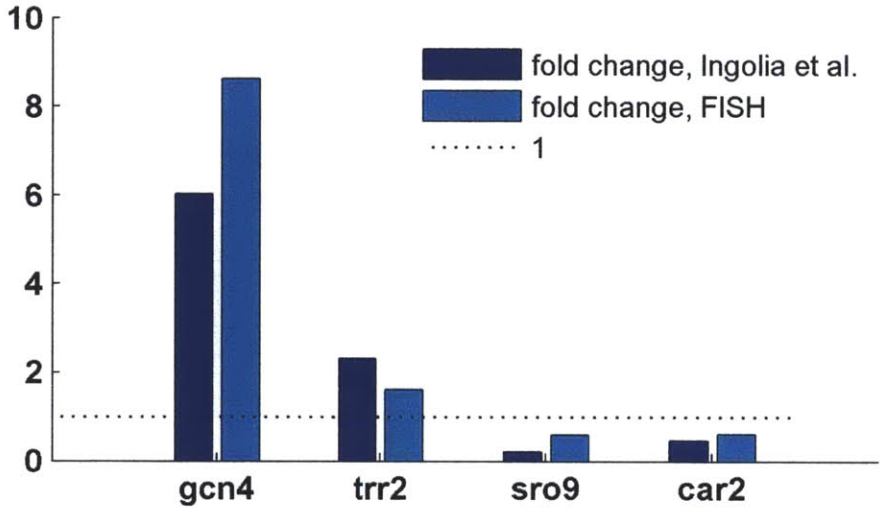
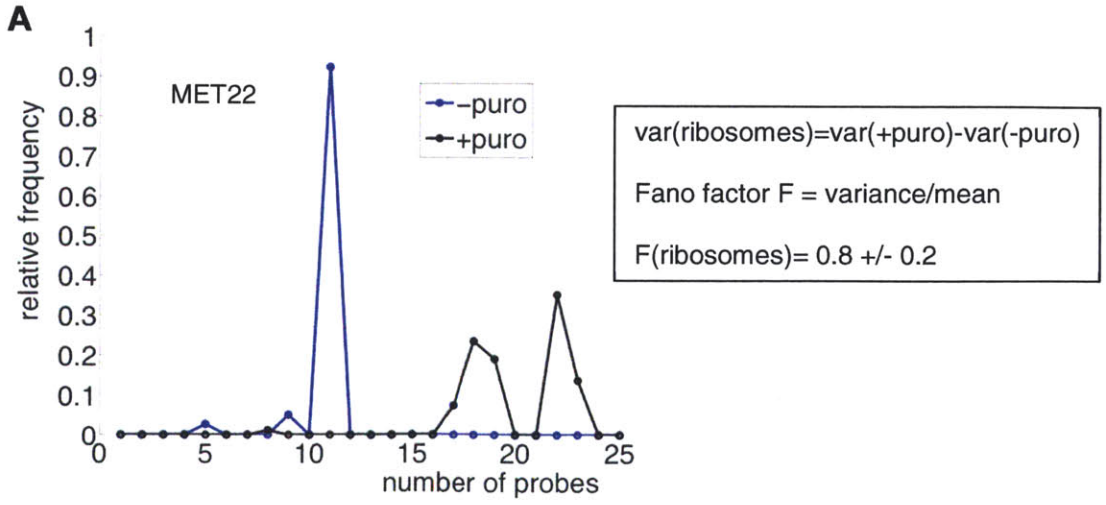


Figure 9 | Comparison of our results with results from the ribosomal profiling method using RNA-sequencing reported by Ingolia *et al.* Fold change denotes the mean number of bound ribosomes per mRNA in cells under amino acid starvation conditions normalized by the value obtained from cells in rich media. Dotted line represents a fold change of 1.

Fig 10



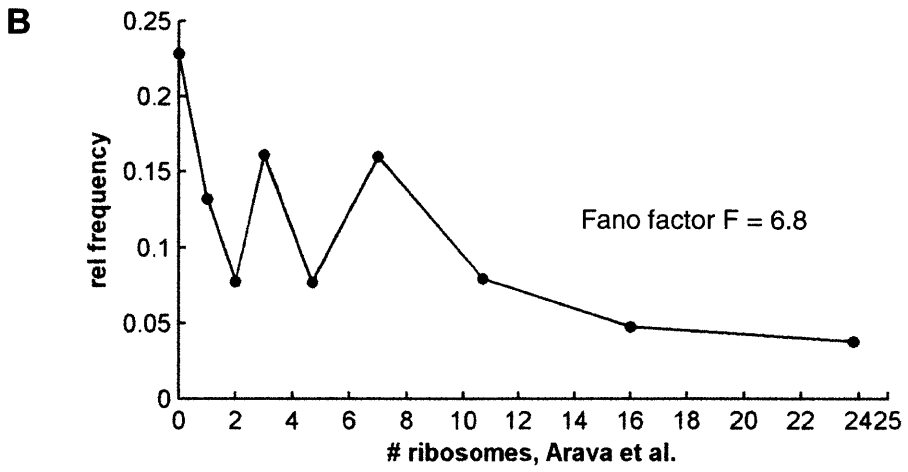


Figure 10 | Comparison of the ribosome distributions inferred from FISH with Arava's data. A, *MET22* intensity distributions in Fig. 3 were deconvolved with the single fluorescent probe data to obtain the probe count per transcript distributions with and without puromycin. The Fano factor of the ribosome distribution was computed using the expressions described in the box. **B,** *MET22* ribosome distribution obtained from the Arava *et al.* data.

REFERENCES

Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O. and Herschlag, D. (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *100*, 3889-3894.

Arkin A., Ross J., McAdams H. H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. Genetics *149*, 1633-1648.

Blobel, G. and Sabatini, D. (1971) Dissociation of mammalian polyribosomes into subunits by puromycin. Proc. Natl. Acad. Sci. USA *68*, 390-394.

Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., Huang, S. (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Hinnebusch, A. G. (2005) Translational regulation of *GCN4* and the general amino acid control of yeast. Annu. Rev. Microbiol. *59*, 407-450.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., Weissman, J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Kuhn, K. M., DeRisi, J. L., Brown, P. O. and Sarnow, P. (2001) Global and specific translational regulation in the genomic response of *Saccharomyces cerevisiae* to a rapid transfer from a fermentable to a nonfermentable carbon source. Mol. Cell. Biol. *21*, 916-927.

Maroney, P. A., Yu, Y., Fisher, J. and Nilsen, T. W. (2006) Evidence that microRNAs are associated with translating messenger RNAs in human cells. Nature Struct. Mol. Biol. *13*, 1102-1107.

Nottrott, S., Simard, M. J., Richter, J. D. (2006) Human *let-7a* miRNA blocks protein production on actively translating polyribosomes. Nature Struct. Mol. Biol. 13, 1108-1114.

Preiss, T., Baron-Benhamou, J., Ansorge, W. and Hentze, M.W. (2003) Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. Nat. Struct. Biol. *10*, 1039-1047.

Raj, A., Van Den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–79.

Rodriguez, A. J., Shenoy, S. M., Singer, R. H. and Condeelis, J. (2006) Visualization of mRNA translation in living cells. J Cell Biol. *175(1)*, 67-76.

Spencer, S. L., Gaudet, S., Albeck, J. G., Burke J. M., Sorger, P. K. (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Schwanhäusser, B., Busse, D., Li. N, Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M. (2011) Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Weinberger, L. S., Dar, R. D., Simpson, M. L. (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. Nat Genet *40*, 466-470.

Wolin, S. L. and Walter, P. (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA EMBO J, *7*, 3559-3569.

Yilmaz, L. S., Noguera, D. R. (2004) Mechanistic approach to the problem of hybridization efficiency in fluorescent *in situ* hybridization. Appl Environ Microbiol *70*, 7126–7139.

Chapter 5

DISCUSSION

SUMMARY OF FINDINGS

In this dissertation, I present the results from both systems- and single-cell level analyses of signaling and gene expression pathway processes in *Saccharomyces cerevisiae*. All three studies used analytical tools from physics and engineering to deduce important properties of biological systems. In the first investigation, we employed computational sensitivity analysis to predict the system behavior of the osmosensing signaling pathway to changes in kinetic rate constants. We demonstrated that the sensitivity of Hog1 activation dynamics to genetic perturbations can be predicted by a simple biochemical model. By decoupling the network into two separate modules, we showed that an input-output analysis of each of the modules can generate the measured disparity in their tolerance to kinetic parameter fluctuations. Our analysis suggests that the input-output relation of catalytic signaling pathways i.e. MAPK cascade are intrinsically sensitive to kinetic rate perturbations. By contrast, signaling governed by stoichiometric biochemical reactions, i.e. phosphorelay, exhibit robust input-output functions. Consistent with this model, we found that the input-output relation of Hog1 activation is severely impacted by protein coding sequence changes in the MAPK cascade genes, but not the phosphorelay genes.

In the second project, we analyzed how *HAC1* RNA splicing contributes to variability in splicing outcomes and heterogeneity in the unfolded protein response in *Saccharomyces cerevisiae*. We combined the single molecule FISH method with

colocalization analysis to count endogenous spliced and unspliced *HAC1* transcripts in single cells. From a statistical analysis of the distributions of cytoplasmic and nuclear *HAC1* RNA per cell, we showed that the statistics of variation are consistent with a transcriptional bursting model. We further used a stochastic gene-activation-inactivation model to determine the kinetic rates governing the transcriptional dynamics of *HAC1*. Based on a two-state-transcription-splicing model, we derived the analytical noise expressions for the spliced and unspliced RNA species. We showed that the cell-to-cell variability in the degree of splicing is tightly regulated in the presence of a UPR-inducing chemical agent, but it is compromised under heat stress. By considering models including extrinsic noise at the splicing or transcriptional level, we showed that the decrease in the covariance of the two RNA species (increased variability in the degree of splicing) under heat stress can be generated by increased fluctuations in the splicing rate.

In the third project, we embarked on the development of a single-cell translation assay. We combined the single molecule FISH technique with the use of protein synthesis inhibitors to measure ribosome numbers of each transcript in single cells. We showed that puromycin-treated cells exhibit higher mean mRNA spot intensities than untreated cells when detected using sm-FISH. We estimated that the minimum probe length required for hybridizing to its target sequence is roughly 16 nucleotides. From this minimum probe length and our probe design of 20 nucleotides each, we determined that one ribosome disrupts on average two probes from binding. We showed that our results of the mean ribosome numbers per mRNA for the five genes measured in exponentially-growing yeast correlate well with the results obtained using polysomal analysis via microarrays. We further showed that our method detected the (mean) change in the translational

regulation of these five genes upon switching the cells from rich to starvation conditions. Further experiments and analyses would be required to determine the full distribution of ribosome count per mRNA, and to ultimately extract single cell information from the data.

Cells are constantly faced with the enormous challenge of maintaining reliable execution of functions while being exposed to an onslaught of time-varying external (genetic or non-genetic) perturbations from the environment, and intrinsic fluctuations due to the stochasticity of discrete biochemical events happening within the cell. What are the ways in which cells can tackle this problem? And, how do these perturbations affect phenotypic heterogeneity? Our investigations in the first project identified the stoichiometric phosphoryl-transfer mechanism as a means for buffering genetic variation. Such stoichiometric systems are not only found in signaling pathways in bacteria and yeast, but they are widely used in metabolic networks which involve transfer of metabolites rather than phosphoryl groups. The basis of the robustness of stoichiometric systems arises from the understanding that their steady-state outputs are only reliant on the influx and efflux of phosphoryl through the system, and is independent of kinetic rate changes in internal system components. A similar mechanism has been used to describe the robustness of the two-component osmosensing signaling system in *Escherichia coli* to component concentration variations (Shinar et al., 2007; Shinar and Feinberg, 2010).

While this example showcased a system that can generate minimal cell-to-cell differences in output, we present in the second project a different system that has a less precise input-output function, and is capable of generating considerable variability in its output between individuals. Systems that make sharp decisions such as the switch-like

UPR-induced splicing and the ultrasensitive response of the MAPK cascade in the Hog1 (Huang and Ferrell, 1996) pathway tend to exhibit less robust input-output functions to perturbations. Hence, fluctuations in internal systems components i.e. fluctuations in splicing factor concentrations can compromise splicing regulation, and generate variable splicing outcomes leading to heterogeneous UPRs among clonally identical cells. In a similar fashion, we showed in the first project that, in contrast to the robustness of Hog1 activation to phosphorelay perturbations, changes made to the MAPK cascade often imposed serious (and variable) consequences on signaling.

Systems built with feedback circuits can reduce the effects of these intrinsic fluctuations; however, even with the most optimal feedback, there is a fundamental limit to noise reduction on the order of a fourth root of the number of regulatory molecules (Lestas et al., 2010). Building the circuit to perform time-averaging is another way to buffer noise (Raj et al., 2006). For example, the lifetime of proteins is usually longer than the average interval between protein production bursts. This results in protein accumulation over time, which averages out the variability generated by bursty expression. Another noise-reducing design is the implementation of successive rate limiting reactions with similar reaction rates, since the noise of the system output scales inversely with the square root of the number of intermediate steps (Pedraza and Paulsson, 2008). RNA maturation for example, requires precursor RNAs to go through a series of conversions into intermediate RNA species before the final mRNA product is being successfully synthesized.

The development of single cell experimental techniques has made it possible to directly examine the probabilistic biochemical events that occur in the cell, and to

compare the individuality of responses between cells (Raj and van Oudenaarden, 2008). We present in the third project our efforts to develop a method to quantify translation in single cells, with the objectives of deciphering the precision of its input-output function and determining its consequences on gene expression heterogeneity. These studies presented here provide examples that connect variations with the underlying biochemical network, and the biological functions these encode, and suggest ways in which cells can employ to suppress or utilize these unpredictable changes.

REFERENCES

Huang, C. Y. F and Ferrell, J. E. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA *93*, 10078-10083.

Lestas, I., Vinnicombe, G. and Paulsson, J. (2010) Fundamental limits on the suppression of molecular fluctuations. Nature *467*, 174-178.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Pedraza, J. M., Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. Science *319*, 339-343.

Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

Shinar, G., Milo, R., Martinez, R. and Alon, U. (2007) Input-output robustness in simple bacteria signaling systems. Proc. Natl. Acad. Sci. USA *104*, 19931-19935.

Shinar, G. and Feinberg, M. (2010) Structural sources of robustness in biochemical reaction networks. Science *327*, 1389-1391.