Transcriptional Dynamics of the Eukaryotic Cell

Cory Batenchuk

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the MSc degree in Cellular and Molecular Medicine.

Department of Cellular Molecular Medicine

Faculty of Medicine

University of Ottawa

Abstract

Gene regulatory networks are dynamic and continuously remodelled in response to internal and external stimuli. To understand how these networks alter cellular phenotype in response towards specific challenges, my first project sought to develop a methodology to explore how the strength of genetic interactions changes according to environmental context. Defined as sensitivity-based epistasis, the results obtained using this methodology were compared to those generated under the conventional fitness-based approach. By integrating this information with gene expression profiles and physical interaction datasets, we demonstrate that sensitivity-based epistasis specifically highlights genetic interactions with a dynamic component.

Having investigated how an external stimulus regulates network dynamics, we next sought to understand of how genome positioning impacts transcription kinetics. This feat was accomplished by cloning two gene-reporter constructs, representing contrasting promoter architectures, across 128 loci along chromosome III in *S.cerevisiae*. By comparing expression and noise measurements for promoters with "covered" and "open" chromatin structures against a stochastic model for eukaryotic gene expression, we demonstrate that while promoter structure regulates burst frequency (the rate of promoter activation), positional effects in turn appear to primarily modulate burst size (the number of mRNA produced per gene activation event). By integrating these datasets with information describing global chromatin structure, we suggest that the acetylation state of chromatin regulates burst size across the genome. Interestingly, this hypothesis is further supported by nicotinamide-mediated inhibition of Sir2 which would appear to modulate burst size globally across the genome.

Foreward

Throughout my master's degree, I have had the opportunity to work on two distinct projects which are the object of my paper based thesis. The first, entitled: *Identification of response-modulated genetic interactions by sensitivity-based epistatic analysis*, extended the principle of epistasis to identify gene-gene interactions which are dynamically modulated according to the environmental-context. Defined as sensitivity-based epistatic analysis, we explicitly develop this principle under the assumption that genetic and environmental perturbations may be treated equivalently with respect to their impact on fitness. To validate this methodology, we contrast and compare our results to those generated by the conventional fitness-based approach. Specifically, by applying these methodologies to the study of both the transcriptional and proteomic response to the DNA damaging agent methyl-methanesulfonate, we demonstrate that this principle specifically highlights interactions with a dynamic component. By applying this principle systematically to generate sensitivity-based mini-array profiles (S-MAP), we demonstrate not only that we can identify many of the pathways and complexes involved in DNA repair, but that we can utilize this information to delineate complex hierarchy within the response. By identifying such regulatory events, this analysis provides an understanding of dynamic regulation within signal transduction networks. My contributions to this work involved all aspects including strain generation, strain validation, mating throughout the synthetic genetic array, establishment of the protocol for growth rate acquisition as well as data-analysis. Our technician, Lioudmila Tepliakova, in turn provided me with aid in strain generation and strain validation. The result of this work was presented via poster presentations at two conferences;

Batenchuk C., Teplikova M., and Kaern M., Dynamic Epistasis Among Transcriptional Regulators of DNA Damage Response Pathways. **Embo Practical course on Networks in Biology analysis, modeling and reverse engineering**, EuroMediterranean University, Bologna, Italy, 2009

Batenchuk C., Teplikova M., and Kaern M., Dynamic Epistasis Among Transcriptional Regulators of DNA Damage Response Pathways. **Progress in Systems Biology: the Brain and Mind,** Ottawa, Canada, 2009

and has been accepted under the title: "*Batenchuk, C., Tepliakova, L., Kærn, M. Identification of response-modulated genetic interactions by sensitivity-based epistatic analysis (2010) BMC Genomics, 11:493*". While this work has important implications for the epistatic and systems biology community, this project also aided me in an academic manner by developing my skills in Matlab coding, data integration and data analysis.

With these skills acquired during my first project, the second phase of my work sought to characterize the relationship between promoter kinetics and chromosomal positioning. In gene expression, two important attributes describe the expression levels driven from a given promoter. The first is the mean expression level and serves as a key determinant of downstream reaction kinetics by indicating the average concentration of protein products readily available within the cell. The second attribute pertains to the variance in the average gene expression across the population, alternatively described as gene expression noise, which provides insight into the regulation of a given locus. While the relationship between promoter kinetics and gene expression has been well characterized by combining experimental results with predictions established through theoretical physics (Ozbudak et al., 2002; Blake et al., 2003; Blake et al.,

2006), an understanding of the impact of genome positioning on this relationship remains to be identified. As an early attempt to identify these features, we investigate the relationship between mean expression and noise for two contrasting promoter architectures cloned across 128 loci along chromosome III in *S.cerevisiae*. The two promoters utilized in this analysis were selected to highlight the difference between promoters associated with an "open" or "covered" nucleosome structure. By comparing our results to those generated through stochastic simulations, we present a formal foundation for the principle that genome positioning potentiates its effects by regulating the burst size, defined as the number of mRNA produced per promoter activation event, rather than by regulating the frequency of promoter activation. By integrating this information with datasets describing histone acetylation and polymerase binding (Liu et al., 2005), and characterising the effects associated with nicotinamide perturbation, we suggest that the histone deacetylase Sir2 appears to serve as a regulator of the burst size at ~40% of all loci tested, consistent with the observation that Sir2 affects a rate limiting step post RNA polymerase initiation (Sekinger and Gross, 2001; Gao and Gross, 2008). While this study is of high importance to theoretical physicists, evolutionary biologist, as well as to the sirtuin community, this study has been submitted under the title: *Chromosomal position effects arise from Sir2-mediated variation in transcriptional bursting*. My contributions to this project involved all aspects other than strain generation, which was completed by Simon St.Pierre, Samyuktha Adiga, and Mila Teplikova, and the initial screens performed by Simon St.Pierre, Samyuktha Adiga, Anna Szuto, and Nazir Kabban, to establish that positional effects vary across promoter constructs; a result not discussed in this thesis.

Acknowledgements

I would like to thank Dr. Kristin Baetz, Leslie Mitchell and Simon St. Pierre for their help and guidance in yeast mating and discussions about results obtained through our analysis. Drs. Guri Giaever, Robert St. Onge and Trey Ideker for sharing experimental data to validate our methodology. Dr. David Bickel for discussions on statistical inference. Dr. Lionel Filion for help and guidance in flow cytometry. I would also like to give a special thanks to our technician, Lioudmila Tepliakova, who played a vital role in my projects by providing me with much appreciated aid in strain generation and validation, as well as to my mentor, Dr. Mads Kaern for taking the time to teach me valuable scientific skills ranging from data analysis to scientific writing. Without these two members, this work would not have been possible. This work was supported financially by the Canadian Institutes of Health Research, the Canadian Cancer Society, the National Science and Engineering Research Council, the Canada Research Chair program, and les Fonds Québécois de la Recherche sur la Nature et les Technologies.

Table of Contents

List of Figures and Tables

List of Abbreviations

CEN	Centromeric region
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
E-MAP	Epistatic mini-array profiling
ES-score	Environmental sensitivity score
Fig	Figure
F-score	Fitness-based epistasis score
H3K3	Histone 3 lysine 3
H3K14	Histone 3 lysine 14
H3K16	Histone 4 lysine 16
HC	High-confidence fitness-based epistatic interactions
HN	High noise
LE	Low expression
mRNA	Messenger ribonucleic acid
mM	Millimolar

MMS	Methyl-methanesulfonate
$\boldsymbol{P_{ADH1}}$	Promoter of *ADH1*
$\boldsymbol{P_{ACT1}}$	Promoter of *ACT1*
PIC	Pre-initiation complex
RC	Reduced-confidence fitness-based epistatic interactions
SGA	Synthetic genetic array
S-MAP	Sensitivity-based mini-array profiling
S-Score	Sensitivity-based epistasis score
TFs	Transcription factors
TEL	Telomeric regions
WT	Wildtype
yEGFP	yeast enhanced green fluorescent protein

List of Symbols

τ	Phenotypic trait
W	*F*itness
m	Growth rate
e^m	Exponential growth rate of the population
$\delta W(A,WT)$	Phenotypic impact of gene deletion A relative to wildtype
Δ	Deletion
ε_{Score}	Epistasis score
ε_{Fit}	Fitness based epistasis
F-Score	Fitness based epistasis score
ε_{Env}	Environmental epistasis
ε_{Sen}	Sensitivity based epistasis
S-Score	Sensitivity based epistasis score
k_{on}	Promoter activation rate
k_{off}	Promoter deactivation rate

k_{m}	mRNA synthesis rate
γ_{m}	mRNA decay rate
k_{p}	Protein translation rate
γ_{p}	Protein decay rate.
$\langle p \rangle$	Population-average expression level
η	Population-average gene expression noise
b_p	Translational burst parameter (or k_p/γ_M)
b	Transcription burst parameter (or k_m/ k_{off})

1. General Introduction

Sensitivity-Based Epistatic analysis

Origins of Epistasis

Epistasis is a fundamental principle which can be applied to indentify and interpret the hierarchical relationship among loci (Avery and Wasserman, 1992). According to the classical definition of epistasis, an epistatic interaction is observed between two loci when the phenotypic consequence of mutation (τ_A) is alleviated following mutation of a second locus (τ_{AB}). Alternatively described as alleviating epistasis (Onge et al., 2007), buffering epistasis (Segrè et al., 2005), or masking epistasis (Elena and Lenski, 1997), such interactions indicate an "AND-like" functional relationship between genes. For combinatorial gene deletions which enable complete rescue of the mutant phenotype relative to wildtype (i.e. $\tau_A < \tau_{AB} = \tau_B \leq \tau_{WT}$, see Figure I1a), it is typically inferred that both genes function antagonistically with one another. Classical examples include the interaction between Cdc25 and Wee1, whereby the G2/M cell cycle arrest induced by Cdc25 mutations is rescued following deletion of Wee1 in *Schizosaccharomyces pombe* (Fantes, 1979). A result which suggests that both regulators share opposite function in the regulation of the G2/M transition. For alleviating interactions which do not allow for the recovery of wildtype phenotype, but are rather less severe than anticipated since there is redundancy in the targets affected by single gene deletion, it is inferred that both genes function within a common pathway or complex (Onge et al ., 2007; Segrè et al ., 2005). Interestingly, pending equivalence between single- and double-mutant phenotypes, genes functioning within a common pathway can be distinguished from those involved in the same

protein complex. Co-equal relationships, in which $\tau_{AB} = \tau_A = \tau_B < \tau_{WT}$, are anticipated to reflect proteins functioning within the same complex (Onge et al., 2007), as deletion of any subunit impairs complex functionality (Figure I1b). Masking epistasis ($\tau_{AB} \leq \tau_A \leq \tau_B < \tau_{WT}$) is in turn observed when deletion of the upstream activator A results in loss of the downstream effector B, and/or its targets and is in turn associated with a greater phenotypic consequence (Figure I1c) (Aylor and Zeng, 2008; Onge et al., 2007).

While traditional epistatic analysis identifies functional interactions through alleviating relationships, the principle of epistasis has been extended to include the alternative scenario in which the phenotypic consequence is more severe than anticipated. Recent advances in genomics technologies have enabled large-scale identification of epistatic interactions in which gene deletion increases the phenotypic consequence typically associated with mutating the other. For example, synthetic genetic array (SGA) technology enables the rapid generation of large scale double-deletion yeast mutant libraries in which severe sickness or death highlights following double gene deletion highlight a genetic interaction between genes (Tong et al., 2001; Tong et al., 2004). In this analysis, the extreme phenotype of the double mutant arises from an aggravating epistatic interaction in which mutation in one causes the other to become essential for viability or mating. While synthetic lethal screens identify "extreme" phenotypes, aggravating epistasis can be referred to as any double deletion mutant whose phenotypic consequence is more severe than that anticipated following combination of single gene deletion (see Chapter 1.1.3). Such interactions typically refer to an "OR-like" scenario in which genes A or B must be present to prevent further loss of the trait (Figure I1d).

By integrating this information with alleviating interactions, epistatic analysis allows for near complete recovery of network architecture regulating the trait in question (Segrè et al .,

2005). Indeed, epistatic mini-array profiling (E-MAP) is an emerging technology which expands upon this principle to suggest that similarities in epistatic interaction profile across multiple loci is suggestive of shared function (Schuldiner et al., 2005). A principle which serves as a powerful tool to delineate pathway organization (Schuldiner et al., 2005), and reconstruct functional protein complexes (Collins et al., 2007) involved in active regulation of the trait.

Phenotypes Utilized in Epistatic Analysis

Early studies of epistasis were performed by measuring allele frequencies in the population (for a review see Carlborg and Haley, 2004) or by observing the phenotypic consequence following the mating of genotypes associated with a distinct trait (for a review see Avery and Wasserman, 1992; Boone et al., 2007). These studies were in turn limited by the requirement that each single allelic variant had to be associated with a distinct phenotype to follow its frequency or impact within the population. With the emergence of gene silencing procedures (Longtine et al., 1998; Elbashir et al., 2001), epistatic analysis can now includes mutants which are not associated with a distinct phenotypic consequence on their own. As such, these studies can in turn focus on a wide variety of traits ranging from gene expression (Fillingham et al., 2009) to cell size (Jorgensen et al., 2002). Among phenotypes, fitness is typically utilized to evaluate the presence or absence of epistasis among genes as it may be easily quantified in a systematic manner. Fitness phenotypes are characterized directly by measuring growth in liquid (Onge et al., 2007) or solid culture (Tong et al., 2001). Alternatively, indirect measurements of fitness may be provided through competition assays measuring genotype frequencies by flow cytometry (Breslow et al., 2008) or microarray based technologies (Pan et al., 2007).

Quantification of Epistasis

According to the principle of epistasis, the null hypothesis of independence between two genes is violated when the presence of a second gene deletion alters the phenotypic consequence typically associated with mutating the other (Boone et al., 2007). While four distinct models (multiplicative, additive, log or min) may be applied to reject the null hypothesis (Mani et al., 2008), the multiplicative neutrality function can serve as the basis for all alternative definitions (Mani et al., 2008), and predominates within recent large-scale studies of epistasis (Onge et al ., 2007; Segrè et al., 2005; Jasnos and Korona, 2007). Under this model, the strength of an epistatic interaction can be quantified as departure from the neutrality between fitness phenotypes where:

$$W_A \times W_B = W_{AB} \times W_{WT}.$$

To formalize this expectation, it is noted that the relative impact of deleting gene A should remain independent pending the presence ($\delta W_{(A,WT)}=W_A/W_{WT}$) or absence ($\delta W_{AB,B)}=W_{A,B}/W_B$) of gene B when the two are independent of one another. To identify if loss of gene B modulates the fitness impact typically associated with deletion of gene A, equality between these terms is set, resulting in the above multiplicative neutrality function.

To quantify the departure from this neutrality, the above function may be rearranged into an epistasis score (e-score) whereby:

$$\varepsilon_{Score} = \frac{W_{AB} \times W_{WT}}{W_A \times W_B} - 1.$$

As such, when the e-score is significantly greater than 0, an alleviating interaction occurs between the two genes. An observation typically associated with genes functioning antagonistically to one another (Figure I1a), or within a common complex (Figure I1b) or pathway (Figure I1c) . Aggravating interactions are in turn associated with the scenario in which the e-score is significantly lower than 0. Such interactions typically reflect genes which operate within parallel pathways (Figure I1d).

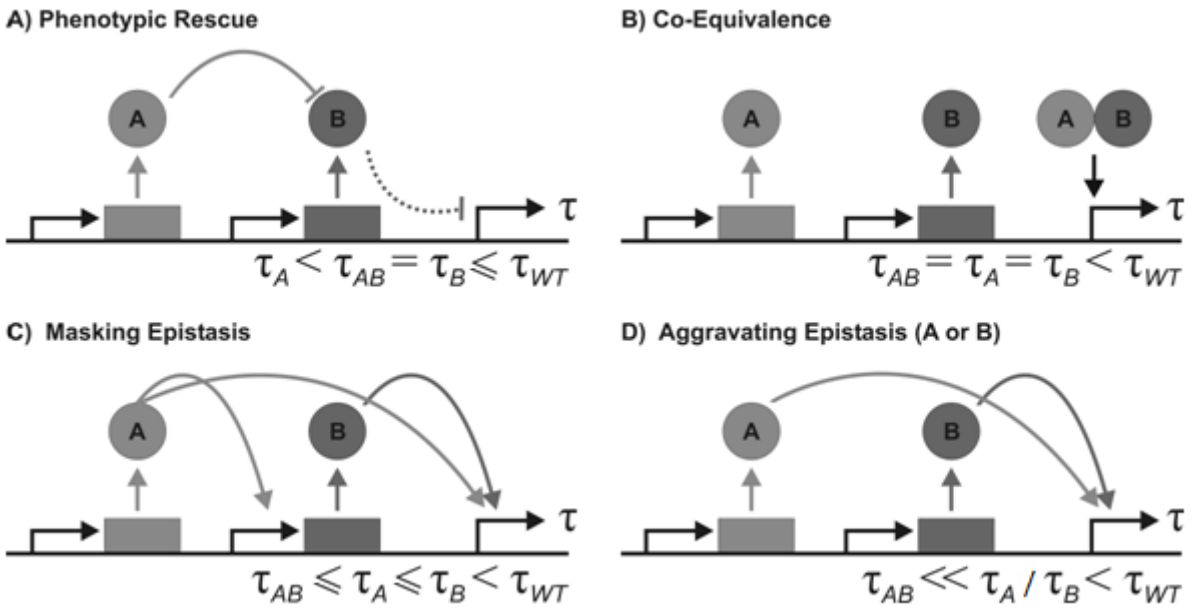


Figure I1. Sub-Classification of epistatic interactions pending regulation of the trait (τ).

Alleviating epistatic interactions are identified between genes A and B when the phenotypic impact of the double-deletion mutant is less severe than anticipated. Under the multiplicative definition of epistasis (Mani et al., 2008), this implies that $\tau_A \times \tau_B < \tau_{AB} \times \tau_{WT}$ and $\varepsilon_{Score} > 0$. Pending equivalence between wildtype, single- and double-deletion traits (τ), alleviating interactions can be sub-classified between **A)** phenotypic rescue, **B)** co-equivalence and **C)** masking relationships. **D)** Aggravating epistatic interactions in turn occur when the phenotypic consequence is greater than anticipated whereby $\tau_A \times \tau_B > \tau_{AB} \times \tau_{WT}$ and $\varepsilon_{Score} < 0$. Dotted lines represent interactions inhibited prior to gene deletion, while full lines represent active interactions. In the above figures, trait is defined as expression of a given locus.

Dynamic Epistasis

The response towards environmental perturbations is accompanied by rewiring of the core network architecture which inevitably alters the epistatic relationship among genes. Accurate pathway regulation is a factor required for orchestrating the specificity of the cellular response towards the environmental challenge. This process is accomplished through activation of the appropriate regulators which modulate cross talk between the different signal transduction pathways (McClean et al., 2007). Loss of such regulators or pathways would in turn result in severe phenotypic consequence, but only under specific conditions. For example, while deletion of ~80% of the loci in the genome in *S.cerevisiae* have little to no consequence on fitness in the absence of cellular stress, ~97% of these loci will have a severe impact in at least one of a thousand conditions tested (Hillenmeyer et al., 2008). It is in turn not surprising that both the sign and strength of an epistatic interaction are well documented to vary according to environmental contexts (Onge et al., 2007; Aylor and Zeng, 2008; Remold and Lenski, 2004; Musso et al., 2008). Such changes are in turn anticipated to reflect two primary mechanisms. Either 1) it reflects emergence of novel functional interactions among regulators (Harbison et al., 2004), or 2) the phenotypic impact of an interaction is apparent only under certain conditions. For example, the phenotypic consequence of deleting two parallel DNA repair modules on the fitness of the strain may have little effect in the absence of DNA damage, but become more pronounced once the organism is transferred to a DNA damaging environment.

Providing an understanding of the dynamics determining the presence or absence of epistasis between genes across environmental conditions represents an important feat in epistatic

analysis (Fiedler et al., 2009). As stated above, the impact of gene deletion on the fitness of the organism is typically utilized to evaluate the presence or absence epistasis between genes in single-celled organisms. However, as fitness phenotypes are defined according to the viability in a single environmental condition, identification of dynamically modulated genetic interactions using the above trait has remained enigmatic. Foremost, it is not because an interaction is identified in a single environment that it is inherently exclusive to that environmental condition. For example, a certain degree of variation in the probability of the interaction is likely to occur across environmental conditions and may cause the probability to be below threshold in one environment (e.g pV=0.049) and above the other (pV=0.051). A phenomenon inherently not associated with the environmental dependency of the interaction but to the identification process. Conversely, it is not because an interaction is identified across environmental conditions that it does not have an important role within the response towards the perturbation. This is the case in the above example between regulators of the DNA damage response which may have a slight impact in the absence of DNA damage, but inevitably have a greater phenotypic consequence in the presence of DNA damaging agents.

Development of Sensitivity-Based Epistatic Analysis

To identify such dynamically regulated epistatic interactions, we build upon the classical definition of epistasis by explicitly incorporating the phenotypic effects caused by changing environment; an assumption frequently employed in chemogenomics (Hillenmeyer et al., 2008; Lehár et al., 2007; Parsons et al., 2004). The resulting neutrality function is in turn referred to as sensitivity-based epistatic analysis since fitness phenotypes are expressed as a ratio across the

two environmental conditions. To validate this methodology, we first analysed the results of a dataset acquired during my undergraduate degree (Cory Batenchuk, 2008). This screen characterised the response of 316 *Saccharomyces cerevisiae* mutant strains carrying combinatorial perturbations of 26 transcription factors (TFs) to the DNA-damaging agent methyl-methanesulfonate (MMS). We chose these TFs because comprehensive datasets describing the impact of MMS exposure on the TF-DNA binding patterns and the transcription profile following gene deletion has been made publically available (Workman et al., 2006). Following growth rate acquisition in liquid culture, we first conduct a conventional fitness-based epistatic analysis to highlight that this methodology does not capture the environmental-dependency of the interaction. We next implement our sensitivity-based epistatic analysis, to compare and contrast our results to those obtained using the fitness-based methodology. We in turn show that departure from sensitivity-based neutrality function enables the identification of context-dependent epistatic interactions and highlights novel functional relationships not captured by the fitness-based approach. Specifically, we show that only ~50% of the interactions identified in the fitness-based analysis are likely to have a dynamic component. By integrating this information with protein-DNA interactions and transcriptional profiles in which one TF binds or alters transcription of another, it appears that the majority of sensitivity-based epistatic relationships are owing to direct physical interactions which are altered by the presence of the DNA damaging agent. An enrichment not obtained using the fitness-based approach. To conclude our study, we apply our methodology to the dataset generated by St. Onge *et al.* (2007) to demonstrate that clustering of sensitivity mini-array profiles (S-Map) allows for an improved understanding of the DNA-damage response.

Chromosomal position effects

Gene expression noise within the population

Within an isogenic population of cells, considerable variation in protein levels are observed from one cell to the next. Defined as molecular noise, this variability originates from stochastic fluctuations in the general processes controlling gene expression (Kærn et al., 2005). While extrinsic factors such as the local cellular environment (Volfson et al., 2006), cell cycle phase (Kar et al., 2009), and cell volume (Halter et al., 2009) represent the primary contribution to eukaryotic noise (Raser and O'Shea, 2004), variation owing to fluctuation in the "finite number" of transcripts and molecular events involved in gene transcription have been shown to play an important role in defining the gene expression landscape across cells (Kærn et al., 2005; Becskei et al., 2005; Swain, 2004). For example, a key difference between the high-expression locus Pdr1 (~42000 proteins per cell; Ghaemmaghami et al., 2003) and the low expression locus Pol1 (~1050 proteins per cell; Ghaemmaghami et al., 2003) is the number of mRNA molecules per cell which passes from ~10 to ~3, respectfully (Zenklusen et al., 2008).

From both theoretical physics and experimental perspectives (Kærn et al., 2005; Becskei et al., 2005; Swain, 2004), discrete fluctuations in the kinetics involved in mRNA processing will have immense impact on the final protein levels within the cell. These processes can be deconvoluted into transcriptional and translational components where both play important roles in modulating molecular noise levels within the cell (Raser and O'Shea, 2004;Ozbudak et al., 2002; Blake et al., 2003; Blake et al., 2006). Gene expression noise associated with the translational process is assumed to primarily arise from static factors encoded within the nucleic

acid sequence such as codon usage (Blake et al., 2003) or ribosome binding affinity (Ozbudak et al., 2002). Transcriptional noise is in turn a property which is assumed to be primarily determined by discrete fluctuations in the processes involved in mRNA production. Notable targets include polymerase pausing, chromatin remodeling as well as transcription factors recruitment within the promoter region.

Transcriptional bursting in eukaryotic gene expression

In eukaryotes, rather than having mRNA synthesized in a continuous process, gene transcription can occur through short transient bursts of active mRNA production; a phenomenon now defined as “Transcriptional Bursting". According a theoretical model for eukaryotic gene expression, average noise and protein levels observed across the population of cells may be defined according to a subset of birth-death processes as illustrated in the schematic below.

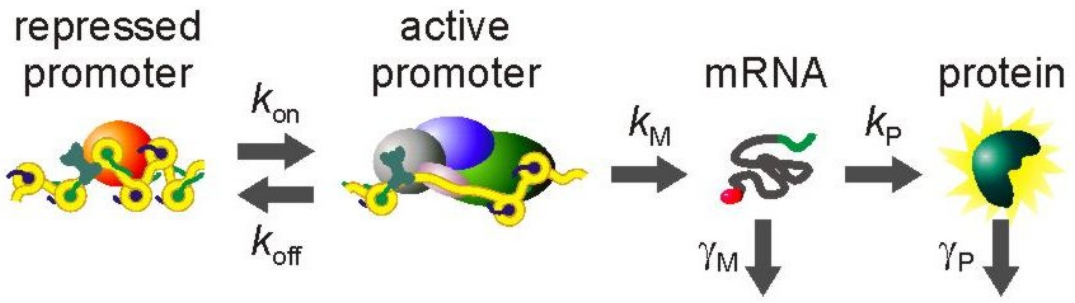


Figure I2. Schematic of eukaryotic gene expression.

Here, the frequency of gene activation (k_{on}) is a property defined by the number of rate limiting steps prior to activation of the burst of gene transcription. These include all factors leading to formation of the pre-initiation complex (PIC) including, but not limited to, transcription factor

binding, regulation of chromatin unwinding, and recruitment of the PIC (Kærn et al., 2005). The number of functional mRNA produced during the intermittent phase of high-gene transcription is determined by the rate of polymerase passage (k_m) as well as the stability of the open promoter complex (k_{off}). The number of proteins generated per mRNA species are defined by the rate of protein synthesis (k_p) as well as the rate of protein decay (γ_m), and are inevitably counterbalanced by the rate of mRNA degradation (γ_p).

Under the above schematic, expression levels within the system may be modeled according to a standard rate equation whereby:

$$\frac{d[mRNA]}{dt} = \frac{k_{on}}{k_{on} + k_{off}} \times k_m - \gamma_m \times [mRNA]$$

$$\frac{d[Protein]}{dt} = k_p \times [mRNA] - \gamma_p \times [Protein]$$

By solving both equations at steady state (i.e. where dxy/dt=0), average protein <*p*> and mRNA <*m*> levels observed across the population of cells may be defined as:

$$< m >= \frac{k_m}{\gamma_m}\left(1 - \frac{k_{off}}{k_{off} + k_{on}}\right)$$

$$< p >= \frac{k_m \times k_p}{\gamma_m \times \gamma_p}\left(1 - \frac{k_{off}}{k_{off} + k_{on}}\right).$$

Variance associated with the each of these processes is subsequently estimated through a poisson distribution as both equations are governed by birth-death processes. From steady state expression levels, Raser and O'Shea, 2004 provide an eloquent solution for the master equation of gene expression noise (η) whereby:

$$\eta^2 = \frac{k_P}{\gamma_M + \gamma_P}\frac{1}{\langle p \rangle} + \frac{\gamma_M \gamma_P k_{off}\left(\gamma_M + \gamma_P + k_{off} + k_{on}\right)}{k_{on}\left(\gamma_M + \gamma_P\right)\left(\gamma_M + k_{off} + k_{on}\right)\left(\gamma_P + k_{off} + k_{on}\right)}.$$

By fitting such theoretical models to experimental measurements, this study and others (Blake et al., 2003) suggest that infrequent promoter activation is required to recapitulate the relationship between transcriptional efficiency and η under certain circumstances. This theoretical observation would advocate that mRNA production may be generated in a pulsatile manner rather than as a continuous process for certain promoters. A theoretical concept latter confirmed by single-RNA counting in eukaryotic cells (Golding et al., 2005; Zenklusen et al., 2008).

Relationship between promoter architecture and noise

For promoters which are covered by tightly-bound nucleosomes, regulation of transcriptional bursting can be achieved by regulating the opening and closing of the local chromatin structure. These "covered" promoters are inherently associated with a high level of transcriptional noise as they are sensitive to nucleosome remodeling within the promoter region (Cairns, 2009). This is in contrast to "open" promoters which are anticipated to have lower noise levels, as the region surrounding their transcription start site is constitutively depleted of nucleosomes (Cairns, 2009). Characteristic of essential loci (Cairns, 2009), constitutive promoters encompass specific DNA sequence elements, such as Poly(dA::dT) tracts or binding sites for DNA bending proteins, which instigate a DNA structure unfavorable to nucleosome deposition (Segal and Widom, 2009; Angermayr et al., 2003). Such structures are anticipated to prevent transcriptional bursting and favor constitutive levels of gene expression within the cell

(Zenklusen et al., 2008). A factor essential to prevent the lethality associated with the down-regulation of essential loci (Fraser et al., 2004).

Modulation of gene expression noise across chromosomal positions

From observations that promoter regulation is dependent on genome localization, chromosomal positioning has been demonstrated to play an important role in defining expression heterogeneity (Becskei et al., 2005; Singh et al., 2010). Regulators anticipated to be involved include histone deacetylases which are well documented to regulate the spreading of silenced chromatin in yeast (Robyr et al., 2002; Rusché et al., 2002), and are well documented to vary in activity across the genome (Tsankov et al., 2006; Braunstein et al., 1993). Notable positional effects associated with such regulators include gene silencing within the telomeres and mating type loci. By restricting DNA accessibility, these regulators have the potential to impact both the frequency of promoter activation as well as the number of mRNA produced per activation event. Indeed, while positional effects have been suggested to modulate the frequency of gene activation (Becskei et al., 2005), histone acetylation has been shown to modulate a rate limiting step post PIC-formation or polymerase recruitment (Sekinger and Gross, 2001; Gao and Gross, 2008) in turn suggesting that positional effects may equally modulate burst size.

Such subtle differences in the regulation of chromosomal positioning effects have important contributions towards the evolution of chromosomal organization. If chromosomal positioning modulates its effects primarily by regulating the frequency of gene activation, then it is anticipated that promoters with a “covered” promoter architecture will vary significantly

across regions of the genome. In contrast, "open" promoters would be anticipated to resist such effects through the presence DNA sequence elements which promote nucleosome depletion to ensure constitutive gene expression from the promoter. If on the other hand chromosomal positioning modulates burst size, then it is anticipated that both promoters may be sensitive towards positional effects. This notable distinction would have important implications towards genome organization. Chiefly, if positioning primarily modulates burst frequency, open promoters would have a greater plasticity in their positioning in contrast to covered promoters which would require migration towards specific regions to achieve desired levels of gene expression. However, if both promoters are equally affected, then it is assumed that open promoters may migrate towards specific chromosomal regions to achieve desired expression levels by modulating a step post promoter activation.

Chromosomal positioning analysis

To understand the relationship between chromosomal positioning and promoter architecture, the promoters of the Act1 and Adh1 loci were cloned across 122 chromosomal positions along chromosome III in *Saccharomyces cerevisiae*. The open promoter of Act1 was selected as it contains a Reb1 site flanked by two AT rich elements to resist nucleosome deposition (McLean et al., 1995). The covered promoter of Adh1 was in turn selected as differentially regulated across environmental conditions (Bird et al., 2006; Denis et al., 1983), and is documented to be regulated by recruitment of the SAGA chromatin-remodeling complex (Bhaumik and Green, 2002; Shukla et al., 2006) to likely displace the tightly bound nucleosomes documented to cover its TATA-box region (Krogan et al., 2004). A phenomenon associated with

transcriptional bursting mechanisms (Zenklusen et al., 2008). Constructs driving yEGFP expression were next integrated at each of the 122 loci, and analyzed by flow cytometry. Through this experimental design, we hope to shed insight into three core aspects of chromosomal positioning effects. First to determine if chromosomal impacts gene expression kinetics from both open and covered promoter constructs. Second, to identify if burst frequency or size is most affected by chromosomal positioning. Third, to attribute these effects to known regulators of chromatin structure.

Following measuring how mean expression and transcriptional noise fluctuate along chromosome III in *Saccharomyces cerevisiae*, we reveal that the DNA sequence elements which resist nucleosome deposition in "open" promoters do not buffer against positional effects. This would suggest that burst size is indeed modulated across chromosomal position. In further support, the correlation between expression and noise for both promoters can be captured under a stochastic model for gene expression where burst size rather than burst frequency is assumed to vary across chromosomal positions. Following integration with high-throughput datasets describing histone acetylation and polymerase binding, and subsequent confirmation, we suggest that the histone deacetylase Sir2 appears to serve as a global regulator of burst size across the genome.

2. Identification of response-modulated genetic interactions by sensitivity-based epistatic analysis

Cory Batenchuk[1,2], Lioudmila Tepliakova[1,2], Mads Kærn[1,2,3*]

[1] Ottawa Institute of Systems Biology, University of Ottawa, [2] Department of Cellular and Molecular Medicine, University of Ottawa, [3] Department of Physics, University of Ottawa

Status: Accepted

Journal: BMC genomics

Format: Article

Abstract

Genetic networks are dynamic and continuously remodeled in response to environmental challenges. Here, we develop a method from first principles to identify genetic interactions that are modulated during an environmental response. Using the DNA damage response in budding yeast as a test case, we demonstrate that the approach can improve the inference of regulatory interactions from high-throughput data, and correctly associate genes with physical complexes and DNA-repair pathways known to be important in the recovery from drug-induced DNA damage.

Background

The principle of epistasis has been an important tool in functional genomics and genetics research for more than a century (Boone et al., 2007; Phillips, 2008). According to this principle, genes may be defined as epistatic to one another when the phenotypic impact associated with a given mutation is altered by the presence of a second gene mutation. By measuring epistasis scores, which quantify departure from a given neutrality model marking the absence of epistasis (reviewed by Mani et al., 2008), it is possible infer regulatory hierarchies, identify functional complexes, and infer genes functioning within common or parallel pathways (Avery and Wasserman, 1992; Onge et al., 2007; Segrè et al., 2005; Aylor and Zeng, 2008; Schuldiner et al., 2005; Collins et al., 2007; Fiedler et al., 2009). For example, aggravating interactions, which occur when the phenotypic impact of the double deletion is greater than predicted by neutrality,

may result from the loss of genes in compensatory pathways. Alternatively, alleviating interactions, which occur when the phenotypic impact is less than expected, may indicate that genes function within a common pathway or complex.

While epistasis reflects the structure of genetic networks in a given environment, the sign and strength of these interactions are expected to change in accordance to the substantial changes in physical interactions observed in response to external perturbations (see e.g. Harbison et al., 2004; Workman et al., 2006). Such changes are anticipated to reflect the activation or inactivation of different pathways across environments. Indeed, it is has been well established that epistasis depends on both genetic and environmental contexts (Onge et al., 2007; Aylor and Zeng, 2008; Remold and Lenski, 2004; Musso et al., 2008). Interestingly, while the phenotypic impact of a changing environment is extensively analyzed in studies of gene-drug and drug-drug interactions (see e.g., Hillenmeyer et al., 2008; Parsons et al., 2004; Lehner et al., 2006), the environmental modulation of epistasis between genes has received much less attention. Importantly, the analysis of fitness phenotypes may not enable a focus on pathways responding to specific environmental perturbations if the mutant strains involved have fitness defects in both the presence and absence of the perturbation (Onge et al., 2007). To address this issue, we have developed a method from first principles to specifically identify pair-wise genetic interactions that change dynamically between environments. This analysis of gene-gene-environment interactions is similar to the generalization of epistasis in terms of three-dimensional genotopes, (Beerenwinkel et al., 2007). This is achieved by explicitly incorporating environmental effects into the neutrality function used to identify epistatic relationships. It turns out that the derived neutrality function can be expressed in terms of sensitivity phenotypes. The method may thus be viewed not only as an identification scheme, but also as providing a formal basis for the sub-

classification of fitness-based genetic interactions using sensitivity recently proposed by St. Onge et al., 2007.

To explore the utility of sensitivity-based epistatic analysis, we examined two comprehensive *Saccharomyces cerevisiae* datasets describing the phenotypic impact of single and double gene-deletion in the presence and absence of the DNA-damaging agent methyl methanesulfonate (MMS). For the purpose of inferring transcriptional regulatory networks, we generated and analyzed 342 mutant strains carrying single and double deletions of 26 transcription factor (TF) genes. These TFs were selected due to the availability of comprehensive datasets describing the impact of MMS on their binding to downstream genes, as well as the genome-wide changes in MMS-induced differential gene expression following TF deletion (Workman et al., 2006). As a preamble, we derive the classical multiplicative neutrality function and perform a conventional fitness-based epistatic analysis to identify genetic interactions in the both presence and absence of MMS. We also discuss in more detail why the results of a fitness-based epistatic analysis should not be used on its own to determine if a genetic interaction plays a role in a given cellular response. We then derive the sensitivity-based neutrality function by adopting the common assumptions that genetic and environmental perturbations can be treated equivalently (Hillenmeyer et al., 2008; Parsons et al., 2004; Lehner et al., 2006), and that gene-environment interactions should remain invariant across genotypes in the absence of context-dependent epistasis.

Using the data obtained for single and double TF deletion mutants, we show that sensitivity-based epistatic analysis implicates a set of genetic interactions in the MMS-induced DNA damage response that is significantly different from that obtained using fitness phenotypes. Noticeably, only ~50% of the interactions identified using fitness phenotypes are also among

those identified using sensitivity. A direct quantitative comparison of the two sets confirms that the sensitivity-based analysis specifically identifies interactions that change between environments. To explore this further, we compare sets of sensitivity- and fitness-based genetic interactions with datasets generated by Workman *et al*, 2006 describing MMS-induced differential gene expression and protein-DNA interactions in the presence of MMS. This comparison demonstrates that sensitivity-based epistatic analysis can improve the identification of environmental-dependant regulatory relationships within transcriptional regulatory networks.

To evaluate the utility of sensitivity-based epistatic analysis for the identification of functional relationships among DNA repair genes, we analyzed a dataset generated by St. Onge et al, 2007. This dataset describes the phenotypic impact of MMS treatment on 349 single and double mutants carrying deletions of 26 genes conferring resistance to MMS. We demonstrate that hierarchical clustering of sensitivity-based epistasis signatures captures the composition and order of complexes and pathways with known roles in the DNA damage response. We also show that a sensitivity-based approach performs better than a fitness-based analysis for the identification of multicomponent protein complexes with known functions in drug-induced DNA damage repair.

Taken together, our results suggest that sensitivity-based epistatic analysis may provide a useful tool to map how environmental perturbations modulate the architecture of genetic networks and reveal new insight into the regulatory networks and pathways mediating cellular responses to changing environments.

Results and Discussion

Fitness-based epistatic analysis

The identification of genetic interactions using fitness phenotypes is typically based on the expectation that the absence of epistasis is marked by the equality:

$$W(X,Y) \times W(wt) = W(X) \times W(Y), \tag{1}$$

where $W(wt)$, $W(X)$, $W(Y)$ and $W(X,Y)$ are the fitness of the reference strain (wildtype, *wt*) and its single- and double-deletion derivatives, respectively. This relationship, which is attributed to Fisher (Boone et al., 2007), can be derived by comparing fitness defects caused by deleting gene *X* in the wildtype strain, defined by $\delta W(X,wt) = W(X)/W(wt)$, and a strain in which gene *Y* is also deleted, defined by $\delta W(X,Y) = W(X,Y)/W(Y)$. The equality in Eq. (1) is then obtained by assuming that the fitness defect caused by the deletion of *X* is independent of the presence or absence of gene *Y*, i.e., by setting $\delta W(X,wt) = \delta W(X,Y)$. Defining fitness in terms of relative growth rates, Eq. (1) predicts that the growth rate $m(X,Y)$ of the double mutant strain in the absence of epistasis is given by:

$$m(X,Y)_{\text{exp}} = \frac{m(X) \times m(Y)}{m(wt)}, \tag{2}$$

where $m(wt)$, $m(X)$ and $m(Y)$ are the growth rates of the wildtype and single mutant strains, respectively. The strength of an epistatic interaction can correspondingly be defined as the relative difference between the observed and expected double mutant growth phenotype:

$$\varepsilon_{\text{fit}} = \frac{m(X,Y) - m(X,Y)_{\text{exp}}}{m(X,Y)_{\text{exp}}} = \frac{m(X,Y) \times m(wt)}{m(X) \times m(Y)} - 1. \tag{3}$$

We refer to Eq. (3) as the fitness-based epistasis score (F-score) since it measures the deviation from neutrality, and since relative growth rate fitness and growth rates can be used interchangeably.

To conduct a fitness-based epistatic analysis, we measured the growth rates of 342 single- and double-deletion TF mutants in the absence and presence of MMS (Fig. 1a and Methods). Detailed results are provided in **Additional File 1**. Among the 26 single mutant strains, 15 had growth rates significantly different from that of the wildtype strain (Fig. 1b, T-Test; $P < 0.05$). Eleven of 14 TF mutants identified as MMS sensitive in the study performed by Workman *et al.*, 2006 are also identified in our screen. The three mutants "missing" from our set (*ecm22Δ*, *gcn4Δ*, and *yap1Δ*) all have *P*-values just above threshold ($P = 0.056$, 0.055 and 0.080, respectively) and could be classified as having growth defects in MMS using a higher significance threshold. Despite using conditions and methods that are significantly different, the overlap is comparable to that between the Workman *et al*, 2006 study and one by Begley *at al*, 2004 where 12 of 17 strains were identified in both studies using the same approach.

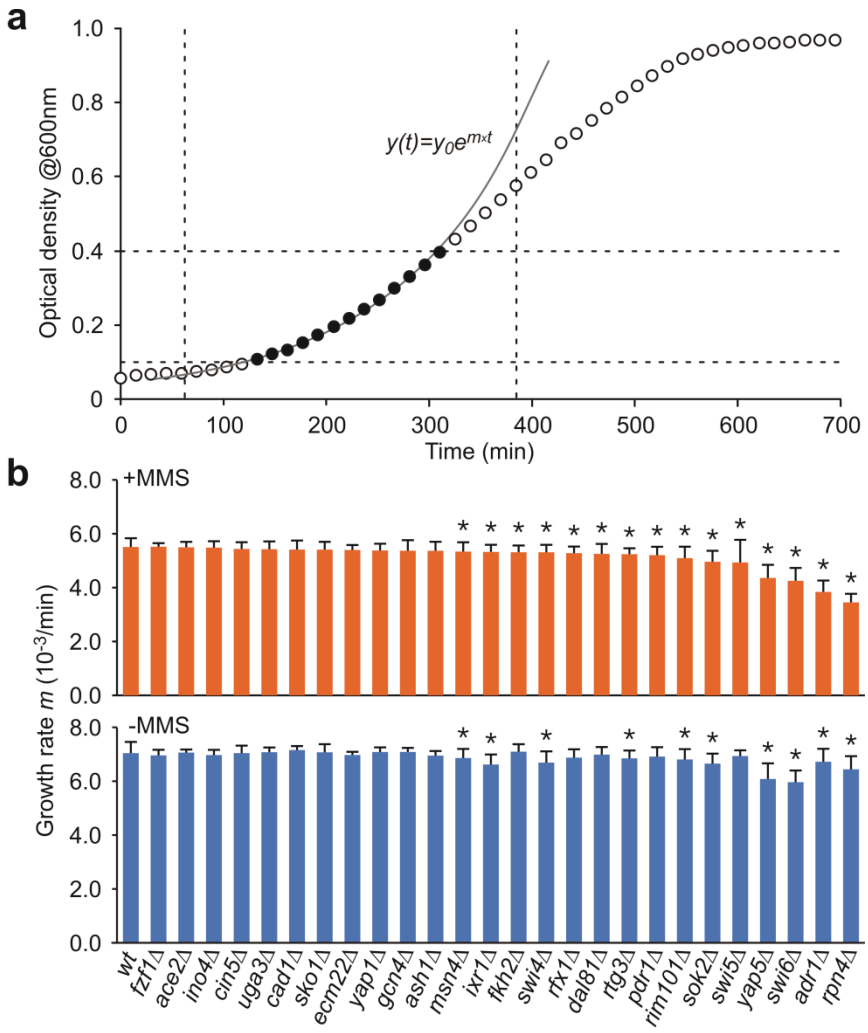


Figure 1. Growth rate measurements.

(**a**) Representative optical density time course (open circles) illustrating the data used to estimate the logarithmic growth rate (filled circles). (**b**) Growth rates of the single TF deletion mutants in the presence and absence of MMS. Error bars indicate standard deviation. Asterisks indicate strains with altered growth rates compared to the wildtype (T-Test; $P < 0.05$).

Fitness-based epistatic analysis can be performed using the measured single mutant growth rates directly (Onge et al., 2007), or by estimating the expected phenotypic outcome of double gene deletion from pooled fitness measurements (Collins et al., 2006). Both approaches have their advantages and disadvantages. While the former is associated with uncertainty arising from alterations in growth phenotypes during the strain generation procedure (Jasnos and Korona, 2007), the latter requires a low frequency of growth defects and genetic interactions. Since the frequency of statistically significant growth defects is high within the TF single mutant library, we employ a variant of the pooling method in which growth rates of the single mutant strains is estimated from the median double mutant growth rate corrected for the phenotypic impact of the second deletion (see Methods). In most cases, the estimated single mutant growth rates obtained using this method is consistent with their directly measured values (Fig. 2a). However, certain strains (*yap5*Δ, *sok2*Δ and *adr1*Δ) had deviations greater than 5%. This deviation could indicate a high number of epistatic interactions, or that a systematic bias was introduced during the generation of the double mutants. For example, the *yap5*Δ single mutant grew consistently slower than its double mutant progeny, suggesting that the mutant might carry a secondary mutation that is lost following mating. To mediate the impact of such experimental uncertainties, we used estimated growth rates for the *yap5*Δ, *sok2*Δ and *adr1*Δ strains in our subsequent analyses.

The results of the fitness-based epistatic analysis are summarized in Figs. 2b and 2c. Detailed results regarding F-scores obtained in both the presence and absence of MMS, as well as their associated *P* values, are provided in **Additional File 2**. Following strain generation, 316/325 of the possible double deletion strains were obtained for analysis. Figure 2b shows the

histograms of F-scores for the corresponding TF-TF pairs in the presence and absence of MMS, as well as 45 interactions identified when the criteria $P < 0.01$ is used to reject the null hypothesis that epistasis is absent (see Methods). A previous study using data similar to ours estimated a conservative false-discovery rate of ~20% (Onge et al., 2007). As expected, the F-score distributions are centred at zero in both environments (the median ε_{fit} is 0.007 and 0.0012 in the absence and presence of MMS, respectively), and scores associated with identified interactions are located in the tails of these distributions. Of the 45 interactions identified, a significant fraction (27/45) is identified in the presence of MMS while the remaining interactions are identified only in its absence.

The association of genetic interactions with specific environmental conditions using F-scores does not necessarily support correct interpretations about their environmental dependency. For example, the identification of an interaction in both the presence and absence of MMS does not inherently indicate an MMS-independent relationship. While the interaction may be conserved across most environments, it could be of particular importance in a specific environment. For example, genetic interactions important for the maintenance of chromosome integrity in all environments may be critical for the repair of MMS-induced DNA damage. Conversely, it should not be concluded that an interaction is important for MMS-induced response based on its identification exclusively in the presence of MMS. Several non-biological factors can contribute to a differential identification across different environments. For example, the true variance may by chance be over- or underestimated in one of the two environments. This may in turn cause the *P* value to be above its critical value in one environment and below it in the other. Within our dataset, we found that the variance among replicates is increased in the presence of MMS (data not shown), which inevitably introduce a bias towards identifying

interactions in its absence. For these reasons, it is not possible to conclude if a given genetic interaction plays a role in pathways responding to specific environmental perturbations based solely on the measurement of fitness in the presence of the perturbation.

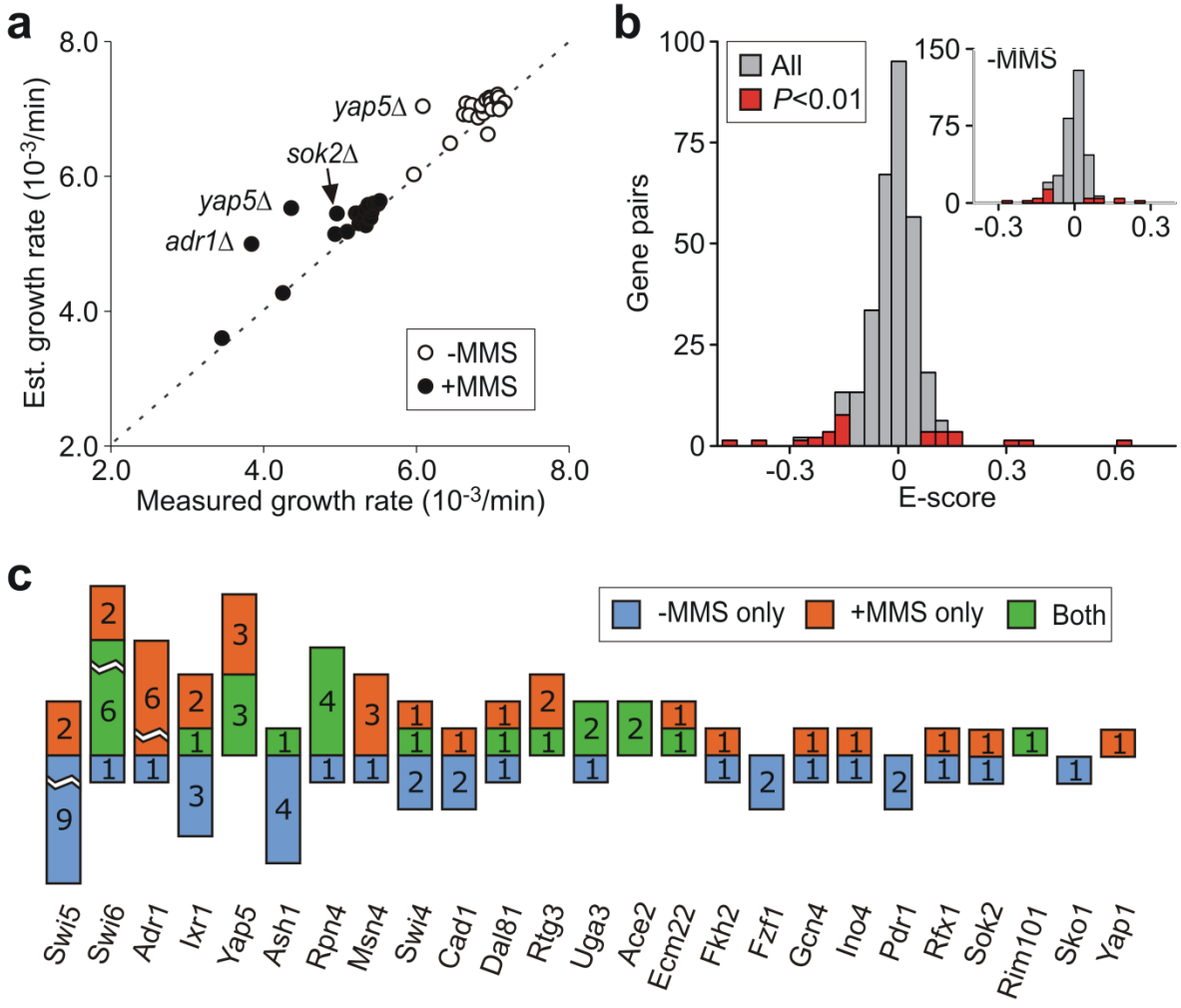


Figure 2. Fitness-based epistatic analysis.

(**a**) Correlation between measured and estimated single mutant growth rates. (**b**) Histograms of fitness-based epistasis scores (F-Score) calculated for all TF-TF pairs (grey) and those associated with high confidence genetic interactions (red) in the presence of MMS. Insert displays the corresponding histograms in the absence of MMS. (**c**) The number of epistatic interactions identified for each TF, categorized according to the environment where the interaction was

identified. 18 interactions are identified exclusively in absence of MMS, 15 are identified only in the presence of MMS and 12 are identified in both environments.

Quantifying gene-environment interactions

To derive a neutrality function that incorporates environmental effects, it is noted that the phenotypic impact of changing the environment should be independent of a gene deletion when the mutated gene is not involved in the cellular response to this change. The principle of epistasis can thus be extended to gene-environment interactions when it is assumed that genetic and environmental perturbations can be modelled equivalently with respect to their impact on fitness, an assumption frequently employed in chemical biology (see e.g., Hillenmeyer et al., 2008; Parsons et al., 2004; Lehner et al., 2006). To quantify the strength of gene-environment interactions analogously to that of genetic interactions, let the fitness defect caused by changing environment from $E1$ to $E2$ be given by $\delta W(wt,\Delta E) = m(wt,E2)/m(wt,E1)$ in the presence of gene X and by $\delta W(X,\Delta E) = m(X,E2)/m(X,E1)$ in its absence. When mutating gene X has no impact on the environmental response, i.e. $\delta W(wt,\Delta E) = \delta W(X,\Delta E)$, the absence of a gene-environment interactions is marked by the equality:

$$W(X,E2) \times W(wt,E1) = W(X,E1) \times W(wt,E2). \tag{4}$$

Equation (4) describes a neutrality function parallel to Eq. (1) in which a genetic perturbation has been substituted by an environmental perturbation to identify an interaction between a gene and the environmental condition rather than between genes. Using relative growth rate fitness, the expected growth rate of the mutant strain is in turn given by:

$$m(X,E2)_{\exp} = \frac{m(X,E1) \times m(wt,E2)}{m(wt,E1)}. \tag{5}$$

Defining sensitivity as the ratio of growth rates in the two environments, $S = m(E1)/m(E2)$, the relative difference between the observed and expected growth rate of the doubly perturbed strain can be written as:

$$\varepsilon_{\text{env}} = \frac{m(X,E2) \times m(wt,E1)}{m(X,E1) \times m(wt,E2)} - 1 = \frac{S(wt)}{S(X)} - 1. \quad (6)$$

We refer to Eq. (6) as the environmental sensitivity score (ES-score) since it quantifies the relative change in sensitivity to a new environment caused by a single genetic perturbation. Neutrality between gene *X* and the environmental change is inferred when deleting the gene has no impact on sensitivity and $\varepsilon_{\text{env}}(X) = 0$. Conversely, a non-zero ES-score implicates the gene in the cellular response to the environmental perturbation.

To identify which of our TFs are involved within the cellular response to MMS, we calculated ES-scores for the 26 single mutant strains (Fig. 3a). Seven of these mutants have *P*-values indicating a significant interaction ($P < 0.05$), including *rpn4Δ*, which displayed the greatest effect (ε_{env} = -0.31), and *adr1Δ*, *dal81Δ*, *fkh2Δ*, *swi5Δ*, *swi6Δ* and *pdr1Δ*, which displayed mild effects (ε_{env} between -0.04 and -0.10). Noticeably, all displayed fitness defects in the presence of MMS (Fig. 1). Conversely, not all strain associated with a fitness defect in the presence of MMS are accompanied by a high ES-score. Since the ratio of sensitivities in Eq. (6) may be expressed as a ratio of fitness values between the two environments, the relative impact of the mutation must be different across the two environments for the ES-score to assume a significant value. Consistent with this interpretation, with the exception of *rfx1Δ*, the eight mutants that display fitness defects in the presence of MMS but have low ES-scores also display fitness defects also in the absence of MMS (see Fig. 1).

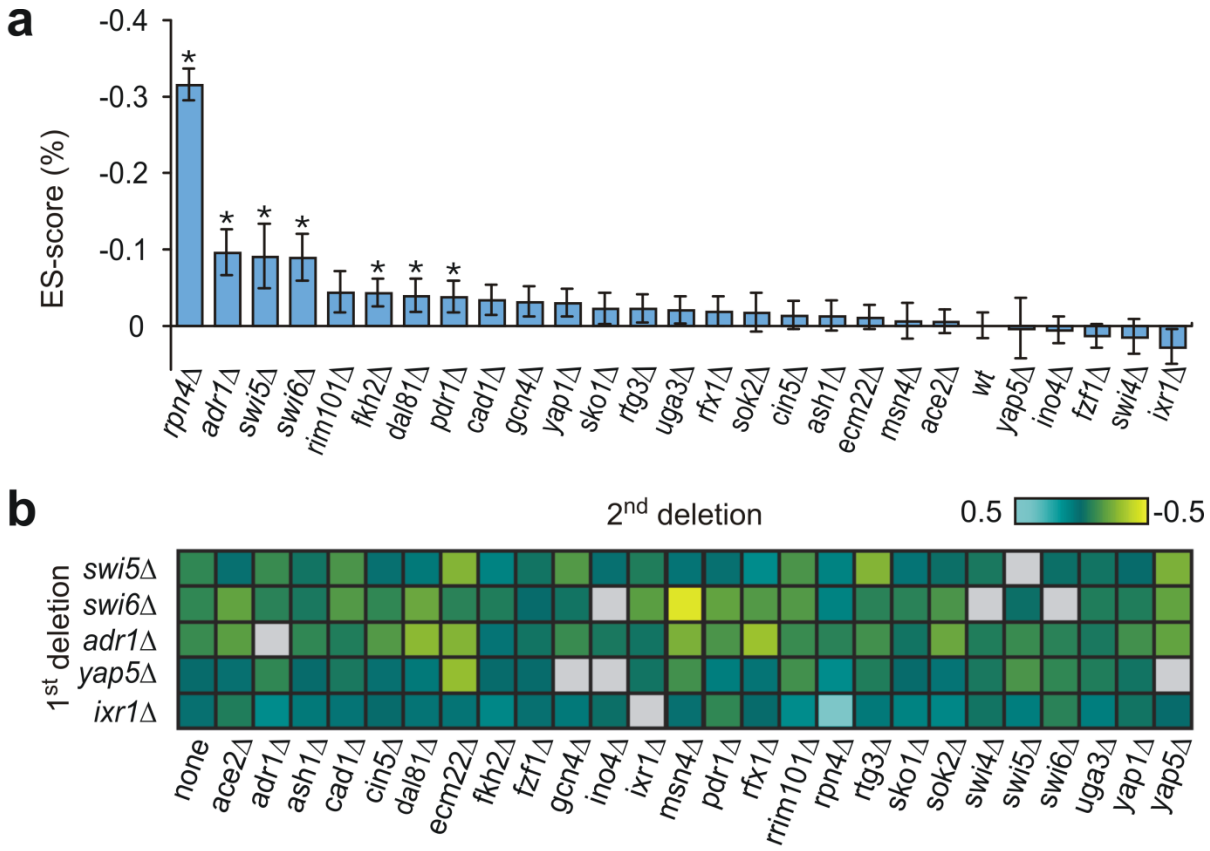


Figure 3. Analysis of environmental sensitivity scores.

(**a**) Mean ES-scores for the wildtype (*wt*) and the 26 single deletion mutants. Asterisks indicate strains with statistically significant environmental sensitivity ($P < 0.05$). (**b**) Examples of variation in ES-scores for selected deletion mutants following the introduction of a second TF deletion. Grey squares indicate double mutants not assayed.

Sensitivity-based epistatic analysis

Extending the definition of the ES-score in Eq. (6) to genetic backgrounds other than wildtype enables the identification of genetic interactions that change dynamically between environments. To demonstrate this, we note that the ES-score associated with deletion of gene *X*, in a strain that lacks gene *Y* is given by:

$$\varepsilon_{\text{env}} = \frac{m(X,Y,E2) \times m(Y,E1)}{m(X,Y,E1) \times m(Y,E2)} - 1 = \frac{S(Y)}{S(X,Y)} - 1. \quad (7)$$

In Fig. 3b, we illustrate that the ES-scores associated with specific TF deletions can vary considerably in the presence of a second TF deletion. In the plot, we include only the TFs with a high number of fitness-based epistatic interactions to specifically highlight the variation of environmental sensitivity across different genetic backgrounds.

To derive a neutrality function that incorporates environmental effects, we impose the definition of epistasis by assuming that mutating gene *Y* should not affect the phenotypic impact of mutating gene *X* when the two genes act independently. Considering the impact on sensitivity following deletion of gene *X* as the phenotype preserved across different genotypes, it immediately follows from the equality $\varepsilon_{\text{env}}(X) = \varepsilon_{\text{env}}(X,Y)$ that the absence of epistasis is marked by a sensitivity-based neutrality function where:

$$S(X,Y) \times S(wt) = S(X) \times S(Y). \quad (8)$$

Equation (8) is a direct analogue of the fitness-based neutrality function in Eq. (1) and the strength of the interaction between genes *X* and *Y* can correspondingly be quantified by the sensitivity-based epistasis score (S-Score):

$$\varepsilon_{\text{sen}} = \frac{S(X,Y) - S(X,Y)_{\text{exp}}}{S(X,Y)_{\text{exp}}} = \frac{S(X,Y) \times S(wt)}{S(X) \times S(Y)} - 1, \quad (9)$$

where $S(X,Y)_{\text{exp}}$ is the sensitivity satisfying Eq. (8) expected under the null hypothesis that epistasis is absent.

To compare and contrast the fitness- and sensitivity-based approaches, we identified the 45 most likely epistatic interactions using F- and S-scores, respectively. The results of sensitivity-based analysis are summarized in Fig. 4a, which displays the histograms of S-scores for all TF-TF pairs and the 45 interactions with the lowest *P* values. As in the fitness-based calculation, the S-score distribution is centred at zero (median ε_{sen}= -0.007) and S-scores associated with high-confidence interactions are located in the extreme tails of this distribution. Of the 45 interactions, 37 have *P* values below 0.05, while the remaining eight have *P* values between 0.05 and 0.07. The additional interactions are included only to allow for a comparison of interaction sets of equal size. Interestingly, only half of the fitness-based epistatic interactions (24/45) are among those also identified using sensitivity phenotypes (Fig. 4b). When a *P* value of 0.05 is used as the significance threshold, 16 of the 37 interactions are identified exclusively by the sensitivity-based method. Sensitivity-based epistatic analysis thus provides a perspective on TF-TF interactions in the DNA-damage response that is significantly different from that provided by fitness-based analysis.

To further explore the differences between the two methods, we plot in Fig. 4c the correlation between the absolute value of ε_{sen} and the absolute change in ε_{fit} across the two environments. This plot demonstrates that the strength of F-scores associated with interactions not identified by sensitivity vary little between the two environments. One example is the interaction between the cell cycle regulators *SWI6* and *ASH1*, which have a strongly alleviating interaction in both environments (ε_{fit}= 0.25 and 0.33, respectively), but has a low S-score (ε_{sen}= -0.06). In contrast to this, interactions identified solely by the sensitivity-based method involve an apparent change in the epistatic relationship between the two genes following MMS treatment. An example includes the interaction between the homologues, *ACE2* and *SWI5*, which have well-documented overlapping functions in cell cycle regulation (Voth et al., 2007). All interactions highlighted by the sensitivity-based method involve a marked change in F-scores between the two environments. For example, the four *SWI6* interactions identified by both methods have F-scores that are high in one environment and low in the other. This includes the interaction between *SWI6* and *MSN4* interaction, which is weak in the absence of MMS (ε_{fit}=-0.03) and strongly aggravating in its presence (ε_{fit}=-0.33), resulting in a high S-score (ε_{sen}= 0.6). Thus, sensitivity-based epistatic analysis allows for an assessment of the dynamic change in epistasis following an environmental perturbation. This may improve the identification of context-dependent regulatory relationships among genes, as well as the association of proteins to physical complexes and pathways involved in the response to environmental change.

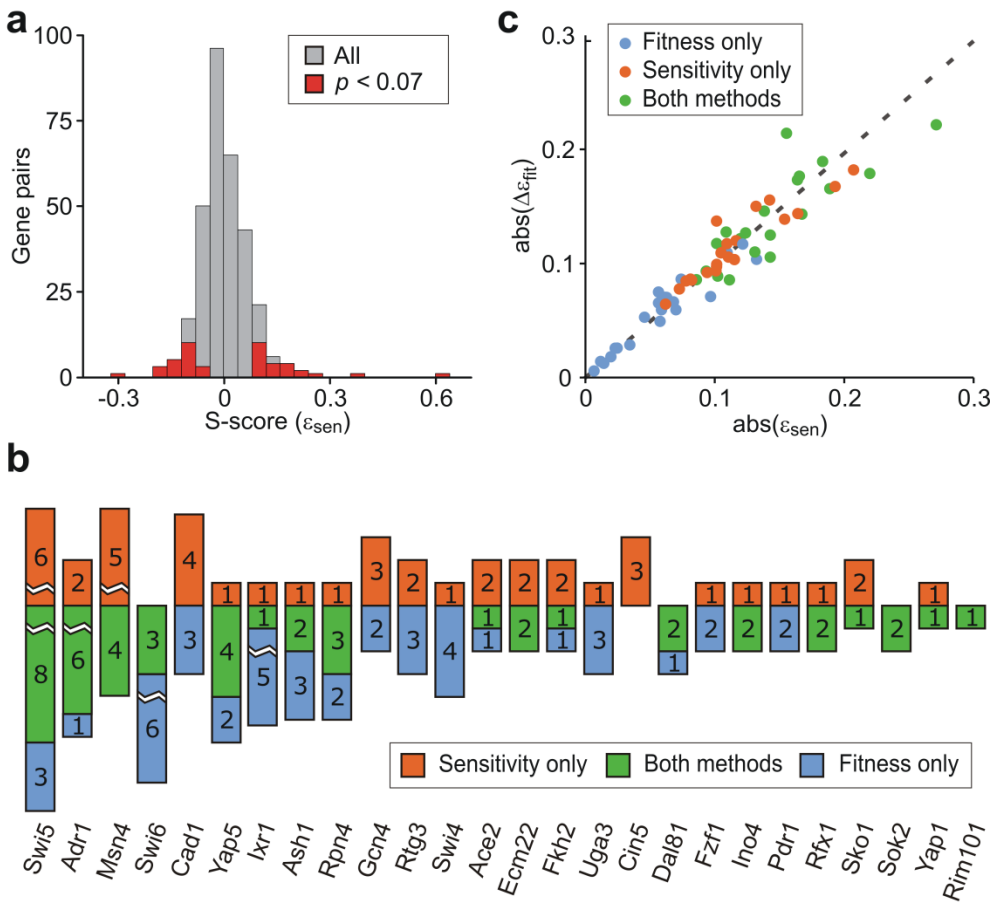


Figure 4. Comparison of fitness- and sensitivity-based epistatic analysis.

(**a**) Histograms of sensitivity-based epistasis scores (S-Score) for all TF pairs (grey) and the 45 most likely epistatic interactions (red). (**b**) The number of epistatic interactions identified for each TF categorized according to the methodology by which the interaction was identified. (**c**) Correlation between the absolute S-score and the absolute difference in F-scores in the presence and absence of MMS.

Inferring regulatory relationships

To evaluate the utility of sensitivity-based epistatic analysis in identifying putative MMS-dependent regulatory relationships, we compared sets of interactions identified by fitness- and sensitivity-based epistatic analysis to datasets generated by Workman *et al*, 2006 describing the loss of MMS-induced differential gene expression following TF deletion, referred to as genetic buffering (Workman et al., 2006) or regulatory epistasis (Tan et al., 2008), as well as protein-DNA interactions in the presence and absence of MMS. To ensure a fair comparison, we used a set of sensitivity-based interactions identified with $P < 0.05$ and two sets of fitness-based interactions identified in the presence of MMS. The first containing 27 high-confidence (HC) interactions with $P < 0.01$, and the second containing 62 reduced-confidence (RC) interactions with $P < 0.05$. The results of this analysis are summarized in Fig. 5a.

We first evaluated if the three sets of genetic interactions are enriched in direct genetic buffering whereby the deletion of one TF causes the loss of MMS-induced differential expression of another. Within the buffering dataset, there is evidence for genetic buffering interactions between 26 of the 316 TF pairs tested ($P < 0.05$). About one-third (9/26) of these direct buffering events are also identified by the sensitivity-based analysis corresponding to a significant 3.0 fold enrichment ($P = 0.001$, hypergeometric test). By contrast, the set of interactions identified using fitness phenotypes displays no significant enrichment over a random model (0.9 fold, $P = 0.68$ or 0.8 fold, $P = 0.79$ for the HC and RC sets, respectfully). In other words, if genetic buffering of one TF by another is viewed as evidence for a putative regulatory

relationship, the sensitivity-based analysis clearly outperforms a fitness-based model in identifying such interactions.

To determine which genetic interactions are supported by physical interaction data, we analyzed the Workman protein-DNA interaction dataset focusing on genes that are differentially expressed following MMS treatment. We evaluated two scenarios where TF-DNA binding might manifest as a genetic interaction in the presence of MMS – the direct binding of one TF to another and the co-binding of two TFs to a common downstream gene. The analysis of direct binding to differentially expressed TF genes (identified using $P < 0.05$) provides evidence for putative regulatory relationships among 10 TF pairs. Four of these interactions are also identified in sensitivity-based set of interactions corresponding to a significant 3.4 fold enrichment (P =0.02, hypergeometric test). By contrast, neither of the fitness-based sets displays enrichment.

To investigate the second scenario, we performed a two-step analysis. First, for each TF pair identified by sensitivity-based analysis, we perform a hypergeometric test by counting the number of differentially expressed genes bound by each TF and the number of genes bound by both. Here, the identification of differentially expressed genes uses a lower P-value ($P < 0.01$) to reduce the false positive rate. Within the set of interactions identified from sensitivity analysis, seven TF pairs display a significant enrichment in co-binding using a stringent cut-off of $P < 0.01$. To evaluate if this number of interactions is greater than expected from a random model, we counted the number of genes bound by any combination of TF pairs using the same criteria. This identified 46 TF pairs that are significantly enriched in co-binding among the 316 pairs tested. A hypergeometric test of these frequencies indicates no significant enrichment (1.3 fold, $P = 0.28$). Similar values are obtained for the fitness-based sets.

The most compelling evidence for the improved identification is offered by sensitivity-based epistatic analysis is obtained by considering the totality of the Workman data. When direct buffering, direct binding and shared target binding are all considered evidence for a putative regulatory relationship among TFs, nearly 50% of the interactions identified by using sensitivity phenotypes are supported by at least one line of evidence (18/37 interactions, 2.0 fold enrichment, P =0.001). By contrast, the sets of fitness-based interactions show no significant enrichment (Fig. 5a). Some of the identified interactions are well established in the literature. One example is the interaction between *FKH2* and *SWI5*, which, according to the *Saccharomyces* Genome Database, share a number of genetic interactions with genes involved in cell cycle progression. *FKH2* is essential for the correct cell cycle periodicity of *SWI5* transcription (Pic et al., 2004) and has been reported to prevent Swi5-specific activation of the cell cycle gene *CTS1* (Voth et al., 2007). Another notable example is the interaction between *SWI6* and *RPN4*, which co-localize to several common genes and both buffers the mitochondrial DNA repair gene *DIN7* (Workman et al., 2006). The TFs also share 26 of 61 genetic interactions with genes that have MMS-specific phenotypes and documented roles spanning numerous DNA repair modules, including homologous recombination and post replication repair (Pan et al., 2006). Existing genetic interaction data thus suggests that *SWI6* and *RPN4* are functionally linked in the MMS-response, in agreement with our observation of dynamic MMS-dependent genetic interaction between these genes.

To further compare the two methodologies, we calculated the true- and false-positive rates at varying P value thresholds when direct buffering, direct binding and shared target binding are all considered as evidence for a putative regulatory relationship. The results, which are displayed in Fig. 5b, indicate that sensitivity-based analysis can improve the identification of

regulatory relationships among TFs. Specifically, the sensitivity-based method identifies a higher number of true positives than the fitness-based method at any false-positive rate. This improvement becomes more evident when the predictive value, defined as the fraction of correctly identified interactions, is plotted for *P* values usually considered to imply statistical significance (Fig. 5c). While the sensitivity-based method achieves a success rate of about 50% for *P* values between 0.01 and 0.05, the success rate associated with the fitness-based method at best is in the 25-35% range.

a

Category	Sensitivity ($P < 0.05$)				Fitness ($P < 0.01$)				Fitness ($P < 0.05$)			
	hits	*N*	*R*	*P*	hits	*N*	*R*	*P*	hits	*N*	*R*	*P*
Genetic buffering	9	37	**3.0**	**0.001**	2	27	0.9	0.68	4	62	0.8	0.79
Direct binding	4	37	**3.4**	**0.020**	1	27	1.2	0.60	4	62	2.0	0.11
Shared binding	7	37	1.3	0.28	5	27	1.3	0.35	6	62	0.7	0.93
Combined	18	37	**2.0**	**0.001**	8	27	1.2	0.32	14	62	0.9	0.70

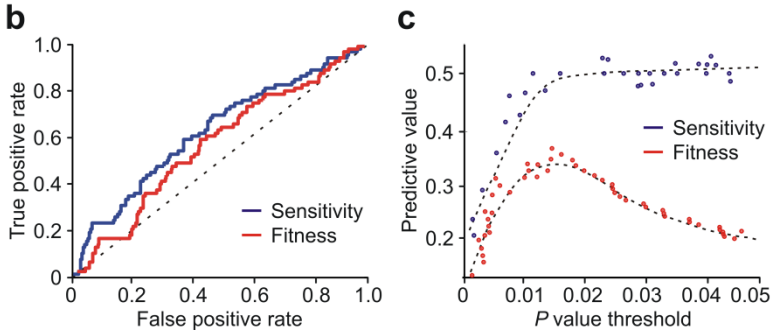


Figure 5. Statistical comparison of genetic interaction sets.

(a) Genetic buffering and genome-wide TF-DNA binding data (Workman dataset) is taken for evidence of putative regulatory relationships among TFs. *R* gives the ratio of frequencies and *P* the probability of observing this or a greater ratio by chance. Boldface is used to indicate statistically significant enrichments ($P < 0.05$) within the genetic interaction sets relative to the frequency (hits/*N*) of interactions within the Workman data. Enrichment is analyzed in four categories: (1) Genetic buffering of one TF by another, (2) Direct binding of one TF to the gene encoding another, (3) TF pairs that bind to the same target gene(s), (4) interactions supported by any of the categories (1)-(3). The number of hits in the Workman dataset among the 316 pairs tested for each category are 26, 10, 46 and 77, respectively. **(b)** Comparison of true- and false-positive rates associated with each methodology. **(c)** The predictive value of the fitness- and sensitivity-based methods at different *P* value thresholds. The predictive value is defined as the fraction of correctly identified interactions.

Inferring functional complexes and pathways

To explore if sensitivity-based epistatic analysis can be used to identify functional complexes and pathways, we conducted hierarchical clustering of S-score profiles (see Methods). Clustering of S-scores calculated for the TF dataset did not yield meaningful results (data not shown) presumably due to the diverse and only partially overlapping roles of the different TFs in the MMS response. As an alternative, we analyzed a dataset generated by St. Onge et al., 2007 describing the fitness of 349 mutants carrying single- and double-deletions of 26 genes displaying fitness defects in MMS.

The hierarchical clustering of S-scores, displayed in Fig. 6a, yields a grouping of the 26 genes that is consistent with known functional modules within the DNA damage response. These include members of the Rad6 epistasis group (*RAD5*, *RAD18* and *HPR5*), which functions within the post-replication repair (PRR) pathway (Friedl et al., 2001; Pfander et al., 2005), the Shu complex (*SHU1*, *PSY3*, *CSM2* and *SHU2*) involved in promoting the formation of homologous recombination repair (HRR) intermediates (Mankouri et al., 2007), the Rad52 epistasis group (*RAD54*, *RAD51*, *RAD57*, *RAD55*, *RAD2* and *RAD59*) involved in homologous recombination (Symington, 2002), as well as the Rtt101-Mms1 ubiquitin ligase (Zaidi et al., 2008) and the Mus81-Mms4 recombination factor (Ehmsen and Heyer, 2008). As expected, the genes within these clusters have MMS-enhanced alleviating interactions with one another (negative S-score).

Interestingly, the sensitivity-based clustering places *SGS1* within the Rad52 epistasis group in agreement with previous findings (Baldwin et al., 2005), but also appears to separate this group into two different components – one comprising *RAD54*, *RAD51*, *RAD57* and *RAD55* and the other comprising *RAD52* and *RAD59*. The former group are members of the Rad51-

dependent HRR pathway and function in parallel with members of the Shu complex to generate HRR intermediates processed by *SGS1* (Mankouri et al., 2007). The latter group is known to have additional functions in single stranded annealing not shared by the other members of the group (Symington, 2002). It may therefore not be surprising that *RAD52* and *RAD59* cluster farther from the Shu genes and that SGS1 clusters together with *RAD54*, *RAD51*, *RAD57* and *RAD55*. It is, however, interesting that hierarchical clustering of S-scores may be able to resolve how the different groups of genes act together within the MMS response. It is also interesting that the group comprising *RTT107*, which is grouped with Mms22-dependent repair in agreement with previous findings (Baldwin et al., 2005) *SLX4*, *CLA4*, and *MAG1* display strong aggravating interactions with members of the group comprising SGS1, the Shu complex genes and the genes the in Rad51-dependent and independent HRR. This suggests that *RTT107*, *SLX7*, *CLA4* and *MAG1* may function in parallel to the main HRR pathway. In the case of *SLX4*, this is consistent with the finding that the Slx4 and Sgs1 are part of functionally redundant endonuclease complexes (Fricke and Brill, 2003).

The clusters obtained using S-scores differ from those obtained by St. Onge et al, 2007 using fitness-based epistasis scores. Both analyses correctly identify the functional relationships between *RAD5* and *RAD18*, all members of the Shu complex, four members of the Rad52 epistasis group (*RAD52*, *RAD51*, *RAD55* and *RAD57*), the linkages between *RTT107*, *RTT101* and *MMS22*, as well as those between *MMS4* and *MUS81*. However, the fitness-based clustering failed to reveal the functional relationship between *HPR5* and members of the Rad6 epistasis group (*RAD5* and *RAD18*), the involvement of *RAD54* and *RAD59* with other members of the Rad52 epistasis group, as well as the upstream role of Sgs1 in processing HRR intermediates generated by Shu complex and the Rad51-dependent HRR pathway. Notably, while analysis of

fitness phenotypes performed by St.Onge et al, 2007 identified an alleviating interaction between *HPR5* and both *RAD5* and *RAD18*, the use of sensitivity-based clustering appears to better capture the interplay of *HPR5*, with both the Rad6 pathway (see Fig. 6a), consistent with the observation of direct physical interactions between *HPR5* and both *RAD5* and *RAD18* (Pfander et al., 2005).

To further compare the two methods, we evaluated their ability to correctly recover interactions among genes encoding multi-component protein complexes. Within a positively regulated pathway where *X* acts upstream of *Y*, deleting the upstream gene is expected to mask the phenotypic effect of deleting downstream gene (Avery and Wasserman, 1992). This phenotypic masking can be detected if $W_{XY} = W_X$ or $S_{XY} = S_X$. In terms of epistasis scores, this corresponds to $\varepsilon_{fit} = W_{wt}/W_Y$-1 when fitness phenotypes are used, and to $\varepsilon_{sen} = \varepsilon_{env}(Y) = S_{wt}/S_Y$-1 when sensitivity phenotypes are used. For encoding different components of a physical complex, it is further expected that $\varepsilon_{fit} = W_{wt}/W_X$-1 and $\varepsilon_{sen} = S_{wt}/S_X$-1, corresponding to co-equivalence among mutant phenotypes.

Figure 6b compares the fitness- and sensitivity-based methods in recovering phenotypic masking among protein complex genes. We focussed on three putative multi-component protein complexes involving members of the Rad6 epistasis group (*RAD5*/*RAD18*/*HPR5*), the Shu complex (*SHU1*, *PSY3*, *CSM2* and *SHU2*) and 3 members of the Rad51 HRR pathway (*RAD51*/*RAD57*/*RAD55*). The genes within each complex are annotated as interacting physically with one another according to the BioGRID database (Stark et al., 2006) and define a set of 24 directional interactions displaying phenotypic masking; two for each of the 12 gene pairs. Full data is provided in **Additional File 3**. We defined phenotypic masking as an alleviating interaction ($P < 0.05$) where the difference between respective single- and double- deletion

mutants for fitness ($\Delta_{thr\ x}= |\varepsilon_{fit} - W_{wt}/W_Y+1|$ and $\Delta_{thr\ y}= |\varepsilon_{fit} - W_{wt}/W_X+1|$) or sensitivity-based ($\Delta_{thr\ x}= |\varepsilon_{sen} - \varepsilon_{env}(Y)|$ and $\Delta_{thr\ y}= |\varepsilon_{sen} - \varepsilon_{env}(X)|$) measurements are below a certain threshold (Δ_{thr}). By applying this approach to test for phenotypic masking between all 636 directional interactions, the sensitivity-based identification outperforms that based on fitness phenotypes (Fig. 6b). This is more clearly demonstrated in Fig. 6c, which shows the fraction of masking relationship recovered when Δ_{thr} is less than 10%. Indeed, for $\Delta_{thr} = 0.1$, the sensitivity-based approach recovers 92% (22/24) of the predicted interactions, including those between Rad5 and Rad18, all the members of the Shu complex, as well as the putative complex involving Rad51, Rad55 and Rad57. By contrast, the fitness-based approach recovers only the interactions among members of the Shu complex, which accounts for less than 60% of the predicted relationships.

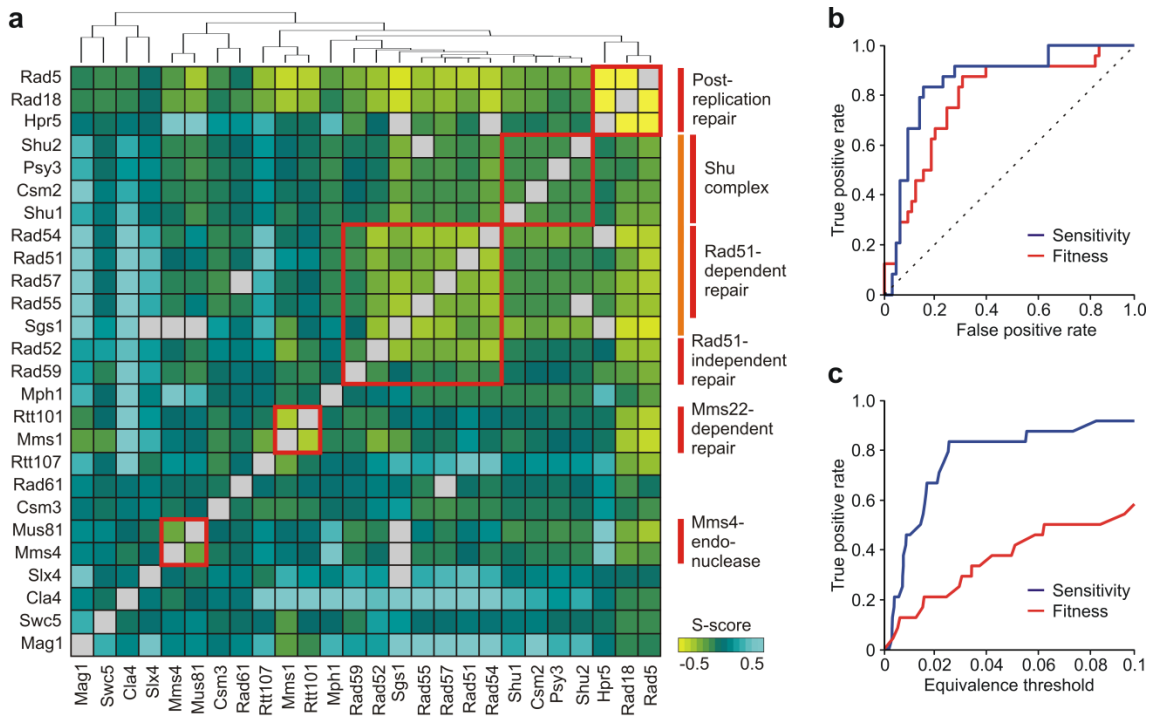


Figure 6. Inference of functional modules and complexes relationships.

(a) Hierarchical clustering of S-score profiles. Red bars indicate genes in the Rad5 epistasis group, members of the Shu complex, the Rad51-dependent and independent branches of the Rad52 epistasis group, members of the Mms22-dependent pathway and the Mms4/Mus81 endonuclease complex. The orange bar highlights the clustering of genes in the Shu complex with the Rad51-dependent pathway and Sgs1 (see text for details). Grey squares indicate double mutants not assayed. **(b)** Comparison of true positive and false positive rates for fitness- and sensitivity-based identification obtained by varying the threshold (Δ_{thr}) used to establish equivalence between single- and double-deletion phenotypes (see text). **(c)** Comparison of true positive rates for fitness- and sensitivity-based identification at different values of the equivalence threshold.

Conclusion

We have presented a method that extends conventional fitness-based epistatic analysis to specifically identify genetic interactions that are dynamically modulated in response to an environmental perturbation. The identification of such interactions may provide several advantages by allowing a focus on pathways responding specifically to a given environmental perturbation (Onge et al., 2007). Noticeably, within the TF dataset analyzed, as few as ~50% of the interactions identified using fitness phenotypes are also identified using sensitivity. These interactions represent linkages among transcriptional regulators that change in a response-specific manner. Thus, combining the two approaches may enable the segregation of genetic interactions within pathways involved in specific cellular responses, and interactions associated with core processes preserved across environments. This conclusion is supported by the analysis of genome-wide profiling of MMS-induced changes in transcription and protein-DNA interaction data. This analysis demonstrates a clear enrichment in putative regulatory relationships among TF pairs identified by sensitivity-based epistatic analysis, a result not afforded by the analysis of fitness phenotypes. Moreover, our analysis of epistasis within known DNA damage repair pathways confirms that quantifying the environmental dependency of genetic interactions can be used to associate genes with different functional groups, physical complexes and pathways. By applying this principle across a larger dataset encompassing additional environmental conditions, we anticipate this methodology could aid in deciphering the dynamics of gene networks.

Integrating physical and phenotypic data into comprehensive and accurate models of regulatory networks and pathways remains a major challenge in systems biology (Beyer et al., 2007). The mapping of biomolecular interactions and transcriptional profiling provide fundamental insight into the substantial remodelling of gene regulatory networks that take place following environmental perturbations. However, it is not always clear if and how observed changes in the physical interaction network manifest at the physiological level. This can be clarified using the phenotypic information provided by sensitivity-based epistatic analysis since the dynamically modulated interactions identified by this method are likely to reflect the remodelling of network architecture in response to environmental cues. As such, the method may have important applications in the inference and analysis of biological networks.

Methods

Strains.

Double gene deletion mutants were generated as described (Tong and Boone, 2007). To construct a single-deletion "starter" strain library, the TF-encoding open reading frames were deleted using a PCR-based gene replacement strategy conferring uracil prototrophy or kanamycin resistance. Kanamycin-resistant single-deletion mutants derived from strain BY4741 (MATa *his3*Δ *leu2*Δ *met15*Δ *ura3*Δ) were obtained from Open BioSystems. Uracil prototrophic strains were derived from strain Y7092 (MATα can1Δ::STE2pr-LEU2 *lyp1*Δ *ura3*Δ0 *leu2*Δ0 *his3*Δ1 *met15*Δ0, kind gift of Dr. Kristin Baetz).

Growth Assays.

Glycerol stocks maintained at -20^{o}C were thawed at 4^{o}C and 20µl used to inoculate 380µl YPD media, containing 10g/l yeast extract (Wisent), 20g/l of Bactopeptone (Fisher), 20g/l of dextrose (Fisher) and 0.042 g/l adenine (Sigma), followed by incubation overnight at 30^{o}C under continuous shaking (250rpm). 20µl aliquots were subsequently diluted with 280µl YPD and the optical density at 600nm (OD) measured using a PerkinElmer Victor3V 1420 Multilabel Counter following incubation at 30^{o}C for 1.5 hours. The OD was then adjusted to ~0.16 by dilution with fresh YPD, and 35µl added to 35µl YPD or 35µl YPD supplemented with 0.015% MMS (Sigma) in a 384 well plate. Each well was overlaid with a 6µl layer of light mineral oil (Sigma) to minimize evaporation. Growth curves were estimated by measuring OD at ~15 minute intervals

for 10 hours at 30°C in no less than 19 and 4 replicates for single- and double-deletion strains, respectively. A custom Matlab script was used to calculate growth rates from OD values in the range from 0.1 to 0.4, and obtained between 60 and 360 minutes after inoculation by fitting to an exponential growth model. Following manual inspection, growth rate estimates were computed based no less than 10 (-MMS) or 12 (+MMS) data points. A decreased OD window was used in a few cases to allow for the analysis of strains with slow initial growth. Single mutant growth rate were estimated from double mutant data using the following procedure. For each TF, a set of growth rates $\mu(X) = \mu_1(X) \ldots \mu_N(X)$ was calculated from Eq. (1) under the hypothesis that epistasis with each of the other TFs is absent, i.e., $\mu_i(X) = m(X,Y_i) \times m(wt)/m(Y_i)$ where Y_i refers to the second TF deleted. The single mutant growth rate is then estimated by the median of $\mu(X)$.

Statistical analysis.

Statistical significance was assessed using parametric bootstrapping. Simulated data, consisting of random numbers drawn from distributions with the same mean and variance as the experimental data was used to estimate the probability of observing an epistasis score as extreme, or more extreme than the observed epistasis score by chance under the null hypothesis that epistasis is absent. The null hypothesis was imposed on the simulated data by drawing the appropriate double mutant growth rate from a distribution with a mean (m_0) given by the growth rate expected in the absence of epistasis and a variance given by $m_0^2 \times (cv_1^2 + cv_2^2)$ where cv_1 and cv_2 are the coefficients of variation associated with the measured double mutant growth rate and the median coefficients of variation of all double mutant growth rates, respectively. P-values for each epistasis score was computed based on 300000 trials.

The assignment of fitness-based epistasis to specific environments (*E1*, -MMS; *E2*, +MMS) was based on the *P* values in the two environments. Interactions were associated with the absence of MMS if $P_{\text{-MMS}} < 0.01$ and $P_{\text{+MMS}} > 0.01$, to the presence of MMS if $P_{\text{-MMS}} > 0.01$ and $P_{\text{+MMS}} < 0.01$, and to both environments if $P_{\text{-MMS}} < 0.01$ and $P_{\text{+MMS}} < 0.01$. The *P* values associated with protein-DNA interactions and loss of differential expression in TF deletion strains were provided by Dr. Trey Ideker and analyzed as described (Workman et al., 2006).

Hierarchical clustering.

The analysis was performed in the *R* programming language using the pvclust package with default parameters (correlation-based measure of dissimilarities between objects, and agglomeration based on averages) (Suzuki and Shimodaira, 2006).

Acknowledgements

We would like to thank Dr. Kristin Baetz, Leslie Mitchell and Simon St. Pierre for their help and guidance with the SGA methodology and laboratory procedures, Dr. David Bickel for insightful discussions on statistical inference, and Drs. Guri Giaever, Robert St. Onge and Trey Ideker for sharing experimental data. This work was supported financially by the Canadian Institutes of Health Research, the National Science and Engineering Research Council Canada Research Chair program, and le Fonds Québécois de la Recherche sur la Nature et les Technologies.

Abbreviations

TF: transcription factor; MMS: methyl methansulfonate; F-score: fitness-based epistasis score; ES-score: environmental sensitivity score; S-Score: sensitivity-based epistasis score; HC: high-confidence fitness-based epistatic interactions; RC: reduced-confidence fitness-based epistatic interactions.

3. Chromosomal position effects arise from Sir2-mediated variation in transcriptional bursting

Cory Batenchuk[1,2], Simon St-Pierre[1,2], Lioudmila Tepliakova[1,2], Samyuktha Adiga[1,3], Anna Szuto[1,3], Nazir Kabbani[1,3], John C. Bell[3,4], Kristin Baetz[1,3], Mads Kærn[1,2,5*]

1 Ottawa Institute of Systems Biology, University of Ottawa, 2 Department of Cellular and Molecular Medicine, University of Ottawa, 3 Department of Biochemistry, Immunology and Microbiology, University of Ottawa, 4 Ottawa Hospital Research Institute, 5 Department of Physics, University of Ottawa

Status: Submitted

Journal: Biophysical Journal

Format: Brief Communication

Brief communication

Gene expression noise varies with genomic position and is a driving force in the evolution of chromosome organization. Here, we performed a systematic analysis of chromosomal position effects by characterizing single-cell gene expression from euchromatic positions spanning across a eukaryotic chromosome. We demonstrate that position primarily modulates the size of transcriptional bursts rather than their frequency, and that the histone deacetylase Sir2 plays a role in this process.

Cells display considerable variability in gene expression due to fluctuations in the rates of gene activation, transcription and translation(Kærn et al., 2005; Mcadams and Arkin, 1997; Raser and O'Shea, 2005). In eukaryotes, slow promoter kinetics can result in transcriptional bursting and high cell-to-cell variability (noise) as multiple mRNAs are synthesized in rapid succession at irregular intervals(Blake et al., 2003; Zenklusen et al., 2008). This phenomenon is thought intimately linked to gene position(Becskei et al., 2005; Singh et al., 2010) through spatial variation in the recruitment and retention of transcription factors, nucleosomes and chromatin remodeling complexes. However, high-throughput studies of noise in the expression of endogenous yeast proteins (Newman et al., 2006; Bar-Even et al., 2006) failed to provide strong support for this hypothesis, presumably due to the masking of position effects by gene-specific variables. Consequently, it remains unclear to what degree gene expression is modulated across the chromosomal landscape, and what factors contribute to these effects.

To characterize the relationship between gene position and gene expression independently of gene-specific variables, we integrated reporter cassettes at 128 euchromatic loci along chromosome III in budding yeast *Saccharomyces cerevisiae* (Fig. 1a). We characterized reporter expression driven by promoters with two contrasting architectures(Cairns, 2009). The promoter of the *ADH1* gene (P_{ADH1}) was selected as an example of a "covered" promoter, which regulates gene expression through SAGA(Bhaumik and Green, 2002) and a consensus TATA-box occupied by nucleosomes(Krogan et al., 2004). Both of these features are hallmarks of high transcriptional noise (Zenklusen et al., 2008; Newman et al., 2006). The promoter of *ACT1* (P_{ACT1}) is an example of a contrasting "open" promoter architecture that maintains a structure unfavorable to nucleosome deposition through the presence of an Poly(dA::dT) tract and a binding site for the DNA bending protein Reb1(Angermayr et al., 2003; McLean et al., 1995). Structures which allow the promoter to be constitutively active, and anticipated to result in relatively low gene expression noise (Cairns, 2009).

We quantified reporter gene expression in single cells by flow cytometry. For each position, we calculated the population-average expression and gene expression noise from single-cell measurements (see Supplemental Methods). Interestingly, expression driven by P_{ADH1} was bimodal (Fig. 1b) at two positions located between the heterochromatic regions of the left telomere and the silenced HML mating type loci. A similar effect, attributed to the existence of two distinct expression states, was recently observed in an engineered system when two interacting silencing gradients flank a reporter expression cassette (Kelemen et al., 2010). However, for P_{ACT1} expression at the same positions (Fig. 1c), and in all other cases, the distributions of reporter fluorescence were unimodal and consistent with fluctuations around a unique expression state.

Average expression and expression noise are both subject to considerable variation across the chromosome for both open and covered promoters (Fig. 1d,e) with P_{ADH1} displaying the highest coefficient of variation. Nevertheless, expression noise is significantly correlated between the two promoters ($p = 3.9\times10^{-9}$, Supplementary Table 1), suggesting that the observed position effects are linked to variation of a common, promoter-independent factor.

To putatively identify such factors, we compared our data to polymerase II occupancy and histone modifications across chromosome III (Liu et al., 2005). We observed significant correlations between our data and euchromatic positions within regions depleted in polymerase II binding or acetylation of histone lysines H3K9, H3K14 and H4K16 targeted by the histone deacetylase Sir2 (Imai et al., 2000) (Supplementary Table 1). Notably, expression noise is negatively correlated with polymerase binding (Fig. 1f, $p = 5.1\times10^{-8}$ for P_{ACT1}, $p = 2.9\times10^{-4}$ for P_{ADH1}) and H4K16 acetylation (Fig. 1f, $p = 3.8\times10^{-15}$ for P_{ACT1}, $p = 2.4\times10^{-15}$ for P_{ADH1}). Additionally, regions with the lowest polymerase binding or the highest apparent Sir2 activity are robustly enriched in low expression or high noise positions for both promoters (Supplementary Fig. 1). These regions, which accounts for most of the observed correlations, are predominantly located adjacent to heterochromatic regions where it is well established that Sir2 has high activity.

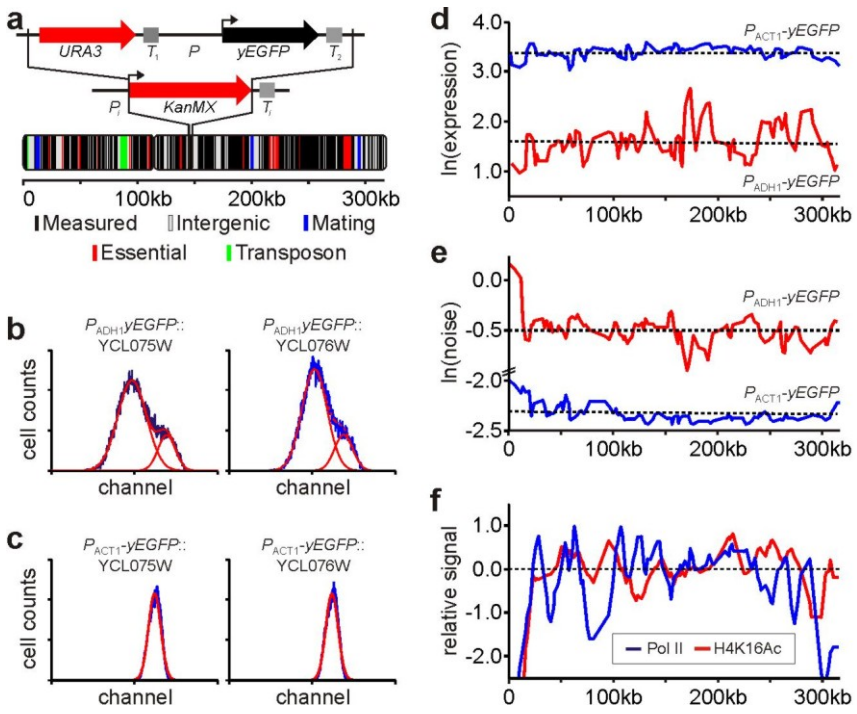


Figure 1. Variation in reporter expression and expression noise across chromosome III.

(**a**) Experimental design. A cassette containing a fluorescent reporter gene downstream from P_{ADH} or P_{ACT1} replaces the KanMX cassette used to delete non-essential chromosome III genes. (**b**) Bimodal fluorescence histograms for P_{ADH1} expression at positions near the left telomeric region. Red curves illustrate the de-convolution of the histogram into two underlying distributions. (**c**) Unimodal fluorescence histograms obtained for P_{ACT1} expression at the same positions. (**d**) Population-average P_{ACT1} expression (blue) and P_{ADH1} expression (red) as a function of chromosomal coordinate. (**e**) Noise in P_{ACT1} expression (blue) and in P_{ADH1} expression (red) as a function of chromosomal coordinate. (**f**) Relative polymerase II occupancy (blue) and relative acetylation of H4K16 (red) as a function of chromosomal coordinate (see Supplemental Methods).

To gain insight into the mechanistic origin of chromosomal position effects, we examined a stochastic model of stochastic gene expression (Fig. 2a) to determine which of four different scenarios best account for our experimental observations. In this model, transcriptional burst size is given by the average number of mRNA synthesized once the promoter is active and is defined as the ratio of the mRNA synthesis rate (k_M) and the promoter deactivation rate (k_{off}). The frequency of the bursts is determined by the promoter activation rate (k_{on}). The first scenario (Model A) corresponds to the absence of transcriptional bursting. In this model, expression noise is determined primarily by fluctuations in the number of mRNA molecules and is independent of how position modulates mRNA synthesis. The three additional scenarios correspond to position-dependent modulation of transcriptional bursting by different mechanisms. In models B and C, position is assumed to modulate burst size through k_M and k_{off}, respectively. In model D, burst frequency is modulated through k_{on}. These different scenarios can be distinguished since they predict different dependencies between noise and average expression (see Supplemental Information).

The measured dependency between noise and average expression can be fitted well to the model where the rate of mRNA synthesis varies across the chromosome ($r = 0.96$ for P_{ACT1} and $r = 0.91$ for P_{ADH1}, Fig. 2b). For P_{ADH1} expression (Fig. 2b, insert), cross-validation indicates that the models involving variation of burst size explain the data significantly better than that involving variation of burst frequency ($p = 5.8\times10^{-5}$ for Model B versus D, and $p = 0.018$ for Model C versus D). In fact, the burst frequency model performs no better than the model where transcriptional bursting is absent ($p = 0.51$ for Model D versus A). For P_{ACT1} expression, the four

models perform equally well (data not shown) suggesting that the low expression noise associated with this promoter is due to the absence of transcriptional bursting. A result anticipated for an open promoter construct.

To further investigate the hypothesis that variation in burst size contributes to chromosomal position effects, we quantified the impact of the Sir2 inhibitor, nicotinamide, across the chromosome (Fig. 2c), as well as the effect of Sir2 deletion on P_{ADH1} expression from five randomly chosen positions (Fig. 2d). Unsurprisingly, nicotinamide treatment had the greatest impact on positions associated with high Sir2 activity (Supplementary Fig. 2), in agreement with the hypothesis that Sir2 attenuate expression in regions adjacent to heterochromatin. Moreover, the effect of nicotinamide on noise measured across the chromosome is consistent with the modulation of burst size (Fig. 2c, Supplementary Figure 3). Specifically, for P_{ADH1} expression, modulation of k_M or k_{off} is significantly better at predicting the effect of nicotinamide treatment compared to a model where k_{on} is varied ($p = 2.7\times10^{-37}$ for Model B versus D, and $p = 3.7\times10^{-46}$ for Model B versus D, Fig. 2c, *insert*). This is further substantiated by the effect of Sir2 deletion on gene expression noise, which can be accurately predicted by variation in burst size, but not by variation in burst frequency (Fig. 2d). It is noted, however, that significant variations across loci persists after the deletion of Sir2 (Fig. 2d, *insert*), indicating that Sir2 is not the only factor contributing to position effects.

Interestingly, our finding that chromosomal position modulates burst size rather than frequency is consistent with the finding that the chromatin structure established by Sir2 is permissive to promoter activation, and suppresses mRNA synthesis by inhibiting the recruitment of factors necessary for transcript elongation(Sekinger and Gross, 2001). Moreover, while Sir2 is

typically viewed as restricted to heterochromatic regions, systematic analysis has documented widespread binding of Sir2 across eukaryotic regions, including *ACT1* and *ADH1* at their native loci, and many other highly transcribed genes(Tsankov et al., 2006). Therefore, it seems reasonable to hypothesize that the position effects we observe for the *ACT1* and *ADH1* promoters are due to variation in mRNA synthesis rates and are linked to variation in Sir2 activity across the chromosome. However, both the mechanism involved in Sir2-mediated transcriptional silencing and the extent of Sir2 activity outside heterochromatic regions remain controversial. Additional experiments, such as single mRNA counting (Zenklusen et al., 2008; Ghandi et al., 2010), are needed to clarify these issues.

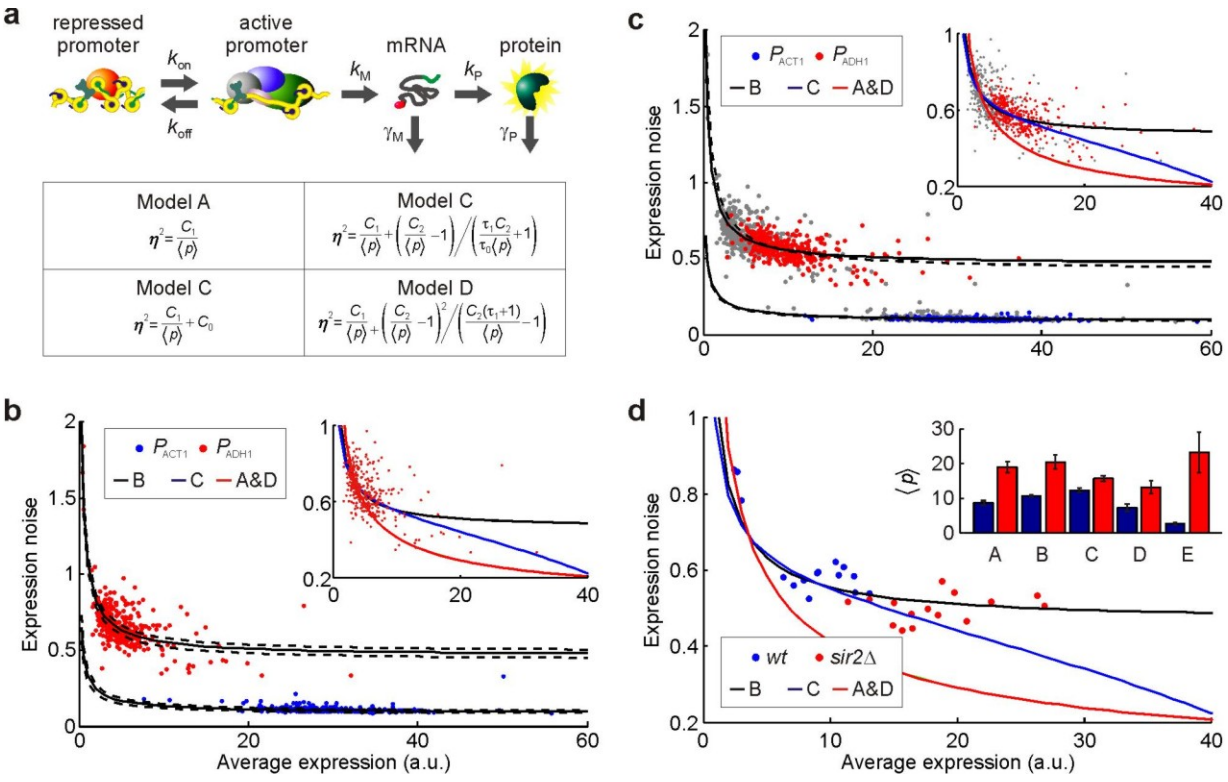


Figure 2. Chromosomal position effects are consistent with modulation of transcriptional burst size.

(**a**) Four different models predict different relationships between noise η and population-average expression $\langle p \rangle$. The constants k_{on} and k_{off} represents promoter activation and deactivation rates, k_M is the rate of mRNA synthesis, k_P is the rate of protein synthesis per mRNA, and γ_M and γ_P are mRNA and protein decay rates, respectively. The fitting constants are defined by: $C_1 = k_P/(\gamma_M + \gamma_P)$, $C_2 = k_M k_P/(\gamma_M \gamma_P)$, $\tau_0 = \gamma_P/\gamma_M$, $\tau_1 = k_{on}/\gamma_M$, $\tau_2 = k_{off}/\gamma_P$ (see Methods). (**b**) Full curves illustrate the fit to Model B. Broken curves indicate the 95% confidence interval. Insert displays comparison of model fits to P_{ADH1} data. Curves fitted to Model A and D overlap. (**c**) Full and broken curves illustrate the fit to Model B to data obtained in the absence and presence of the drug. Insert displays how different models fit the P_{ADH1} dataset. Curves fitted to Model A and D overlap. (**d**) The effect of Sir2 deletion is consistent with modulation of burst size but not its frequency. P_{ADH1} expression was characterized for five loci (A: YCR020C; B: YCL030C; C: YCL037C; D: YCR060W; E: YCL064C) in the presence (blue) and absence (red) of Sir2. Curves fitted to Model A and D overlap. Insert displays the effect of Sir2 deletion on the average expression. Position E near heterochromatin is most severely affected.

Acknowledgements

This work was funded by Early Researcher Awards from the Ontario Government for K.B. and M.K., a scholarship from le Fonds Québécois de la Recherche sur la Nature et les Technologies for C.B., the Canadian Cancer Society for K.B. (grant #020309), the Canadian Institute of Health Research for M.K. (grant #079486), and the National Science and Engineering Research Council for M.K. (grant #313172-2005). K.B. is a Canada Research Chair in Chemical and Functional Genomics. M.K. is a Canada Research Chair in Systems Biology.

Supplementary materials and methods

Strain generation.

Promoter-reporter expression cassettes were obtained by inserting a PCR-amplified *yEGFP* into the pRS406 vector backbone (Stratagene) using SacI and EcoRI. The endogenous P_{ACT1} and P_{ADH1} promoters were isolated from genomic DNA (strain BY4741) by PCR amplification of a 1kb region upstream of the *ACT1* or *ADH1* start codon, and inserted into the modified pRS406 vector using BamHI or XhoI and EcoRI. The resulting vector (**Supplementary Fig. 5**) was designed such that reporter expression following chromosomal integration is shielded from distal upstream effects by an *URA3* expression cassette and a *CYC1* transcriptional terminator positioned upstream of the reporter cassette. Promoter sequences were confirmed by sequencing.

Genomic integration of the reporter-expression cassette was achieved by PCR-mediated gene replacement of the KanMX cassette used to generate the deletion library in the parental strain BY4741 (Open Biosystems). Successful integrations were confirmed by testing uracil prototrophic strains for the loss of kanamycin resistance by replica plating onto YPD plates containing 2 mg/ml G418 . PCR confirmation of 15 randomly selected strains confirmed for loss of kanamycin resistance yielded an efficiency of 100%.

Loci selected for integration include all but one none-dubious euchromatic open reading frames (ORFs). The only ORF not analyzed is YCL073C, which encodes a protein of

unconfirmed function, since the corresponding deletion mutant is not available. We employed stringent criteria to mitigate gene deletion effects and experimental error (see below). Therefore, data was not available for all loci in triplicate. For P_{ACT1} expression no data was not available for the following ORFs: YCL073C, YCR003W, YCR020W-B, YCR028C and YCR053W. For P_{ADH1} expression no data was available for the following ORFs: YCL076W, YCL075W, YCL073C, YCR107W, YCR053W, YCR020W-B, YCR002C and YCL034W.

Media and growth conditions.

Strains were arrayed in a 96-well plate format and stored at -20 ^{o}C in YPD media supplemented with 15% w/vol glycerol. Prior to analysis, plates were thawed on ice for 1 hr and 20 µL aliquots inoculated into 400 µL YPD supplemented with 42 mg/l adenine. Following ~17 hours of growth at 30^{o}C and continuous shaking (250 rpm), optical density at 600 nm (OD_{600}) was measured using a Victor3V 1420 Multilabel Counter (PerkinElmer). Each sample was subsequently diluted to an OD_{600} of ~0.006 and incubated for an additional 6 hours. Samples were supplemented with 5 mM of nicotinamide (Sigma) prior to incubation when applicable. This concentration was previously shown in Ref. (Bitterman et al., 2002) to inhibit Sir2 activity in vivo.

Flow cytometry and data processing.

Flow cytometry was carried out on a 1024-channel Beckman Coulter FC 500 MPL System equipped with a custom 96-well plate loader. A 20 mW Argon laser provided excitation at 488 nm and fluorescence emission (FL1) was collected through a 525±15 nm band-pass filter.

Typically, 30000 events were recorded for each sample. Three replicates were analyzed for most samples.

Extrinsic variability was minimized by analyzing only events within a narrow forward-scatter (FS) and side-scatter (SS) gate capturing a small fraction of the total cell population (see e.g., Becskei et al., 2005; Newman et al., 2006). Correspondingly, the smaller the gate, the larger is the uncertainty in the estimate of the mean FL1 signal and its variance. Additionally, since FS/SS values varies across different genetic background, cells captured by narrow gate at a fixed location will not consistently correspond to the most representative sampling of the population.

To circumvent these issues, we calculated, the mean FL1 signal and its variance for events with identical FS and SS values acquired within an elliptical auto-gate capturing 80% of all events measured on the same day for the same promoter. This methodology, results in up to ~10^6 estimates for each point on a 1024-by-1024 grid of FS and SS values. The mean overall signal and its variance were subsequently calculated as the weighted average of individual estimates. This greatly reduces the extrinsic component of the noise (**Supplementary Fig. 6**). The impact of sample-specific effects (e.g., gene deletion effects or the strain generation errors) were mitigated by excluding samples with a low event count (<5000) or with less than 40% of events within the FS/SS autogate from further analysis.

To correct for autofluoresence (AF), we treated AF and *yEGFP* fluorescence as independent random variables. With this assumption, population-average FL1 signal attributable to the average protein abundance $\langle p \rangle$ is given by $\langle p \rangle = \langle p_{\mathrm{FL1}} \rangle - \langle p_{\mathrm{AF}} \rangle$ where subscripts FL1 and AF refer to the raw FL1 signal for *yEGFP* expressing and non-expressing strains, respectively.

The noise associated with variation in protein abundance among them is then given by the equation $\eta^2 = \left(\sigma^2_{\mathrm{FL1}} - \sigma^2_{\mathrm{AF}}\right)\langle p \rangle^2$ where σ^2 terms are the variances of the raw FL1 signal for *yEGFP* expressing and non-expressing strains. Noise estimates from samples with an average FL1 value less than three standard deviations above autofluorescence were excluded from further analysis.

Enrichment analyses.

Enrichment analysis was performed in Matlab. Loci within low expression (LE) and high noise (HN) regions were identified when the number of measurements within the 20th (expression) or the 80th percentile (noise) acquired across a three-loci moving window was greater than that expected under a random model (hypergeometric test, $p < 0.01$). The position of loci within identified regions are displayed in **Supplementary Fig. 1a** (low expression) and **Supplementary Fig. 1b** (high noise).

Regions de-enriched in polymerase occupancy or Sir2-mediated histone acetylation (H3K9, H3K14 and H4K16) were identified by analyzing a 12.5 kb window surrounding the native open reading frame. Loci were classified as de-enriched in polymerase activity or as enriched in Sir2 activity when the number of measurements within the 25th percentile was greater than that expected under a random model (hypergeometric test, $p < 0.005$). The identified regions are displayed in **Supplementary Fig. 1c** (low polymerase activity) and **Supplementary Fig. 1d** (high Sir2 activity). Data for this analysis was generated by Liu et al.(Liu et al., 2005) In all cases, low polymerase or high-Sir2 regions are associated with LE and HN across both promoters (hypergeometric test, $p< 0.05$).To ensure robustness of this analysis, we

systematically characterized the effect of varying the percentile cutoffs by ±5%, the size of the moving window by ±2.5kb and critical *p* values by a two-fold increase or decrease. Consistent fold enrichment was observed for both polymerase and Sir2 activity as illustrated in **Supplementary Fig. 1e** and **Supplementary Fig. 1f**, respectively.

The loci most affected by nicotinamide treatment were identified using a similar approach. In this case, however, the loci-specific effect was determined by evaluating the deviation from the average effect of nicotinamide treatment. First, the average effect of nicotinamide was fitted to a linear model (**Supplementary Fig. 2a**). Next, the deviation from this trend was determined by examining the residuals of this fit (**Supplementary Figs. 2c and 2d**). The positions within the 80th percentile of the residuals for average expression or the 20th percentile of the residuals for expression noise were classified as those most affected by the treatment. These positions are significantly enriched in positions that also have high Sir2 activity (**Supplementary Fig. 2e**).

Model discrimination.

We consider the steady state dependency of noise η on average protein abundance $\langle p \rangle$ for the model in Fig. 2a. The model predicts that the average protein abundance is given by:

$$\langle p \rangle = \frac{k_M k_P}{\gamma_M \gamma_P}\left(1 - \frac{k_{off}}{k_{off} + k_{on}}\right), \qquad \textbf{Eq. (S1)}$$

where k_M is the rate of mRNA synthesis, k_P is the rate of protein synthesis per mRNA, k_{on} is the promoter activation rate, k_{off} is the promoter deactivation rate, and γ_M and γ_P are the rates of mRNA and protein decay, respectively. The maximal average protein abundance is given by $p_{\max} = k_M k_P / \gamma_M / \gamma_P$.

The intrinsic noise is given by (see Raser and O'Shea, 2004):

$$\eta^2 = \frac{k_P}{\gamma_M + \gamma_P}\frac{1}{\langle p \rangle} + \frac{\gamma_M \gamma_P k_{off}\left(\gamma_M + \gamma_P + k_{off} + k_{on}\right)}{k_{on}\left(\gamma_M + \gamma_P\right)\left(\gamma_M + k_{off} + k_{on}\right)\left(\gamma_P + k_{off} + k_{on}\right)}. \qquad \textbf{Eq. (S2)}$$

In the following, we consider four different variants of the dependency in Eq. (S2): Fast promoter kinetics (Model A), and modulation of k_M (Model B), k_{off} (Model C) or k_{on} (Model D).

For Model A, fast promoter kinetics are achieved by assuming that $k_{off} + k_{on} >> 1$. This assumption yields a noise dependency that is independent of how the promoter kinetic constants are modulated whereby:

$$\eta_A^2 = \frac{k_P}{\gamma_M + \gamma_P}\frac{1}{\langle p \rangle} = \frac{C_1}{\langle p \rangle}, \qquad \textbf{Eq. (S3)}$$

and $C_1 = k_P / (\gamma_M + \gamma_P)$.

For Model B, varying k_M only impacts the noise through variation of the average protein abundance. Correspondingly, the noise given by Eq. (S2) assumes that k_{off} and k_{on} are constant across chromosomal positions. This dependency can be written as:

$$\eta_B^2 = \frac{C_1}{\langle p \rangle} + \frac{\tau_0^2 \tau_2}{\tau_1 (1+\tau_0)} \frac{\tau_0 + \tau_1 + \tau_0 \tau_2 + 1}{(\tau_1 + \tau_0 \tau_2 + 1)(\tau_0 + \tau_1 + \tau_0 \tau_2)} = \frac{C_1}{\langle p \rangle} + C_0. \qquad \textbf{Eq. (S4)}$$

where the three time-scale parameters are given by $\tau_0 = \gamma_P/\gamma_M$, $\tau_1 = k_{on}/\gamma_M$ and $\tau_2 = k_{off}/\gamma_P$, and C_0 is a constant defined by these parameters.

For Model C, the noise dependency can be obtained by first expressing k_{off} in terms of the average protein abundance in Eq. (S1), then inserting the resulting function into Eq. (S2). The result is given by:

$$\eta_C^2 = \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{\tau_1}{1+\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)\left(\frac{C_2}{\langle p \rangle} - 1\right)}{\left(\frac{\tau_1}{\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)\left(\tau_1 \frac{C_2}{\langle p \rangle} + 1\right)} \approx \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)}{\left(\frac{\tau_1}{\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)}, \qquad \text{Eq. (S5)}$$

where $C_2 = p_{\max}$ and the approximation is valid when $\tau_0 << 1$.

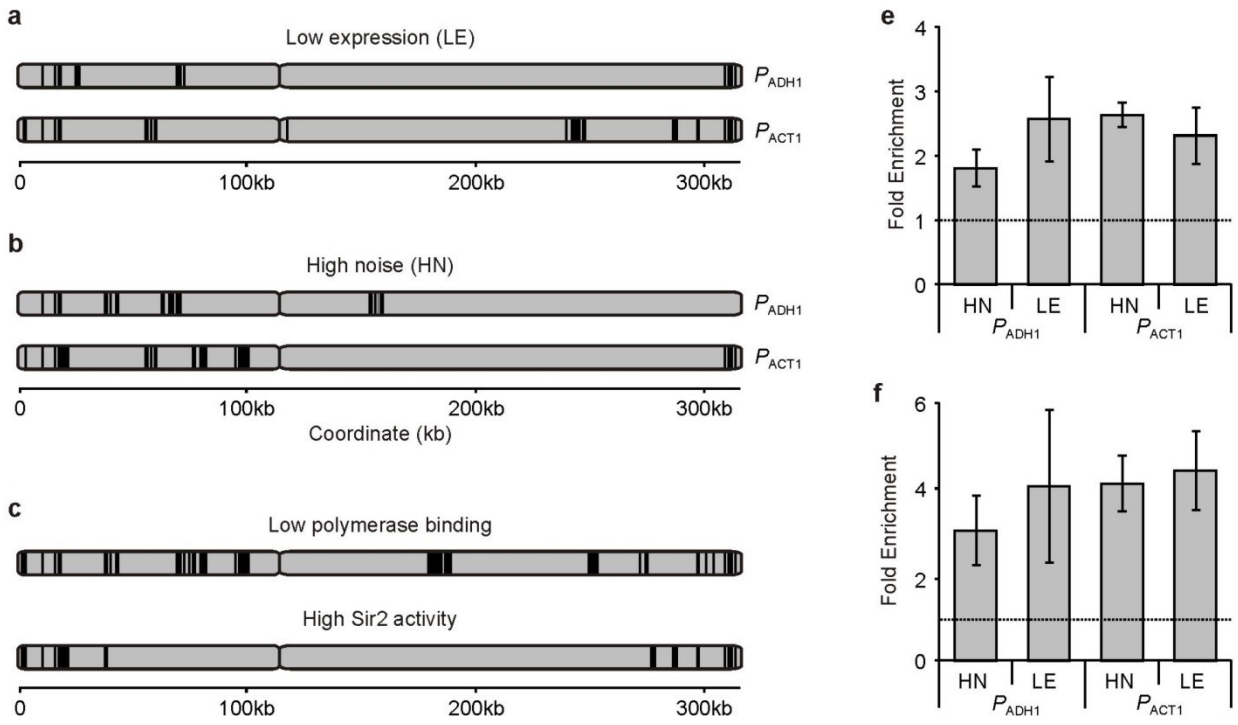
For Model D, where k_{on} impacts gene expression noise, a similar approach yields a noise dependency given by:

$$\eta_D^2 = \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)^2 \left(\frac{C_2}{\langle p \rangle}\left(\frac{\tau_0 \tau_2}{1+\tau_0} + 1\right) - 1\right)}{\left(\frac{C_2(\tau_0 \tau_2 + 1)}{\langle p \rangle} - 1\right)\left(\frac{C_2(\tau_2 + 1)}{\langle p \rangle} - 1\right)} \approx \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)^2}{\left(\frac{C_2(\tau_2 + 1)}{\langle p \rangle} - 1\right)}. \qquad \text{Eq. (S6)}$$

where $C_2 = p_{\max}$ and the approximation is valid when $\tau_0 << 1$.

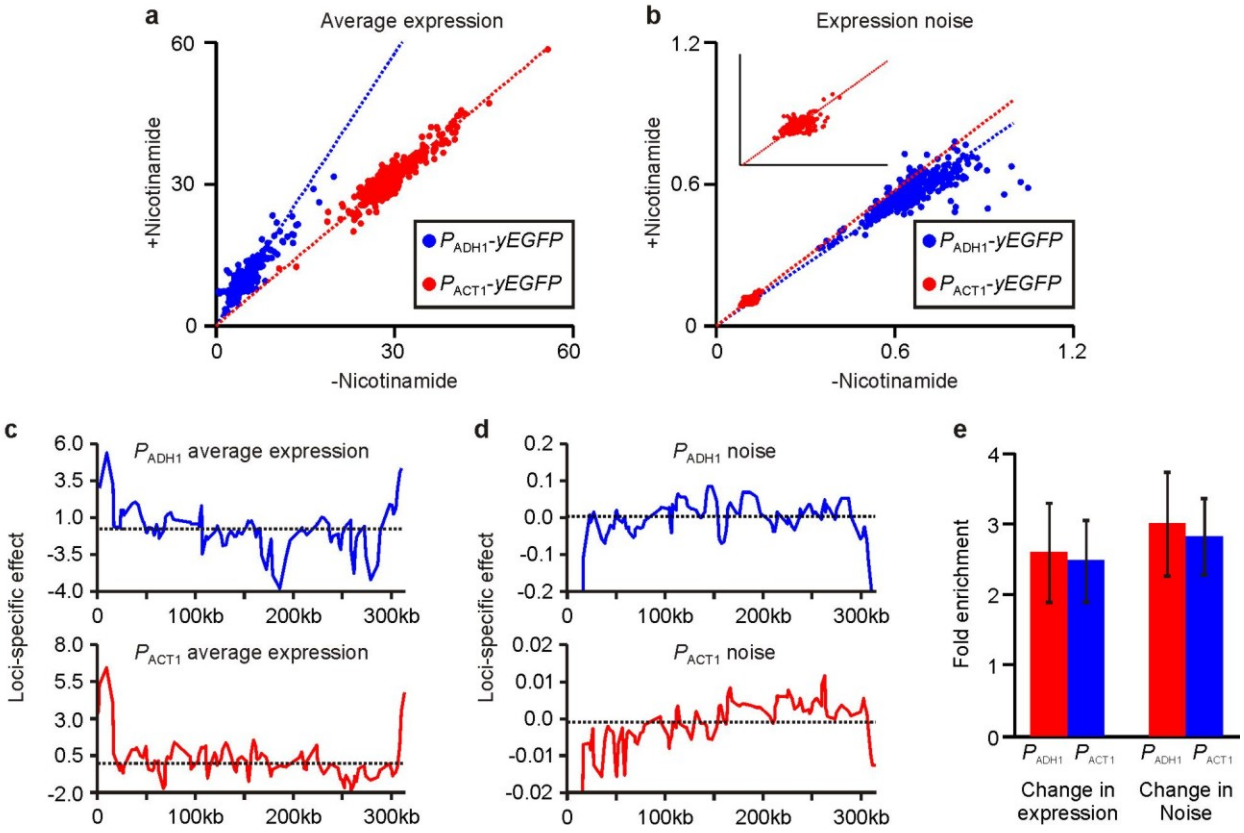
Experimental data were fitted to the noise dependencies predicted for the four different models using robust nonlinear least squares implemented using Matlab "fit" function. The assumption that $\tau_0 << 1$ has minimal impact on the goodness of fit (data not shown). Estimated parameter values are given in **Supplementary Table 2**.

We discriminated among different models by examining the logarithmic error ratios (LER) defined for two models *X* and *Y* whereby LER = $\log(SSE_Y/SSE_X)$ and SSE is the sum squared error. The *p* values associated with the null hypothesis that the two models perform equally well was calculated by "50x2" cross-validation for the chromosomal-wide datasets. Briefly, the loci were split randomly into two sets of equal size, one for training and one for validation. For each split, the four models were fitted to the training loci data and the SSE calculated from the validation loci data. A second SSE was calculated by swapping the training and validation loci, and the process repeated 50 times to obtain a distribution of LER and its standard deviation. An LER that is consistently positive throughout this process indicates that model *Y* performs better than model *X*. The resulting *p* values were calculated with the Matlab ztest function using estimated values of the mean LER and its standard deviation (**Supplementary Table 3**). The mean LER and its standard deviation were calculated similar for the Sir2 deletion data using "leave-one-out" cross validation.



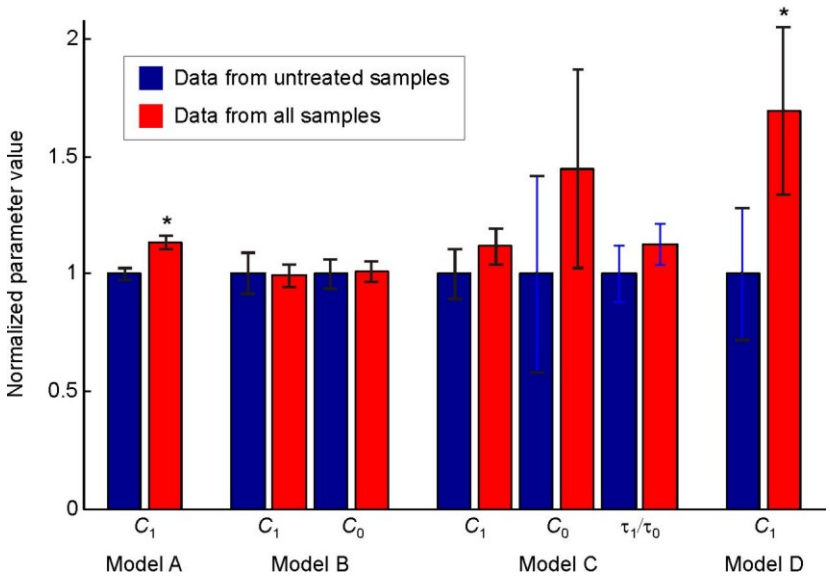
Supplementary Figure 1: Analysis of low expression and high noise regions

(**a**) The position of loci within regions identified as low expression (LE) for the two promoters. (**b**) The position of loci within regions identified as high noise (HN) for the two promoters. (**c**) The positions of loci identified as enriched in depletion of polymerase binding or depletion of Sir2-mediated histone acetylation (H3K9, H3K14 and H4K16). (**e**) Fold enrichment of LE and HN loci within regions of low polymerase binding. (**f**) Fold enrichment of LE and HN loci within regions of high Sir2 activity. Error-bars represent standard deviations calculated by systematic variation of the parameters used to classify loci. The details of the enrichment analysis are provided in Methods.



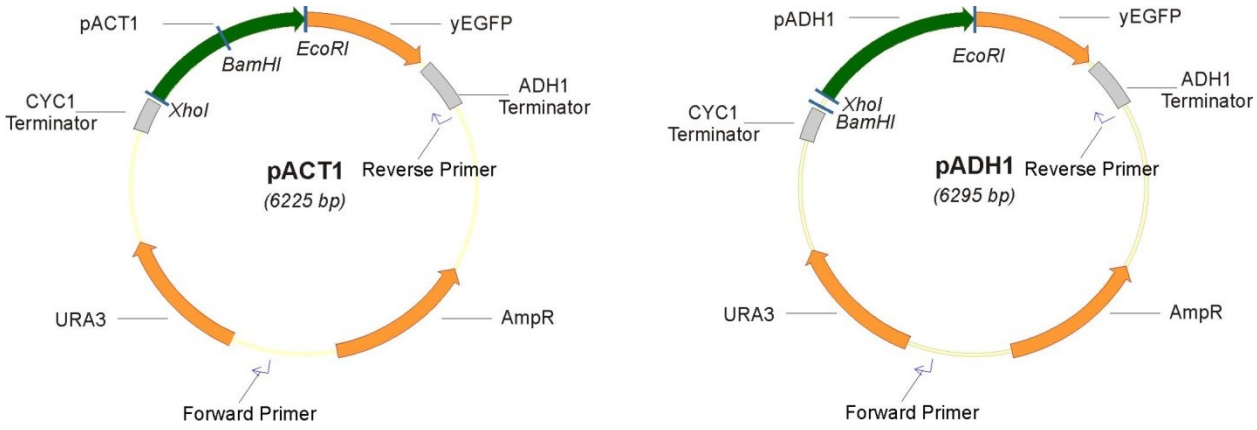
Supplementary Figure 2: Effects of nicotinamide treatment

(**a**) Correlation between population-averaged expression in the presence and absence of 5mM nicotinamide. The slopes are 1.92 and 1.05 for P_{ADH1} and P_{ACT1} population-averaged expression, respectively. (**b**) Correlation between gene expression noise in the presence and absence of 5mM nicotinamide. The insert is an enlargement of the P_{ACT1} data. The slopes are 0.86 and 0.96 for P_{ADH1} and P_{ACT1} noise, respectively. (**c**) Loci-specific effects of nicotinamide on average expression (see Methods). (**d**) Loci-specific effects of nicotinamide on gene expression noise. (**e**) Enrichment of loci most affected by nicotinamide treatment within regions of high Sir2 activity. Error bars represent standard deviations calculated by systematic variation of the parameters used in the statistical analysis.



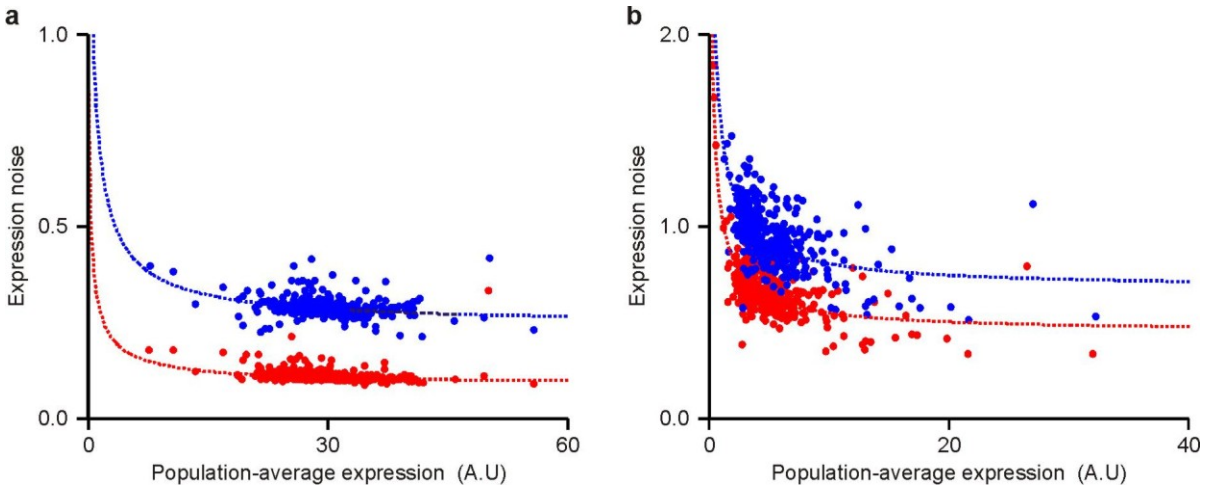
Supplementary Figure 3: Impact of nicotinamide on fitted model parameters for P_{ADH1}.

The four models were fitted to data obtained from untreated samples only (blue) and all samples (red). The error-bars indicate 95% confidence intervals. Asterisk indicates a statistically significant change in the parameter value (t-test,p<0.05). The absence of a statistically significant difference in the fitted parameters indicates that the model obtained using untreated samples can account for the effect of nicotinamide treatment. All parameters were normalized to unity for the untreated data. Actual parameter values are given in Supplementary Table 2. Only the parameter C_1 is shown for Model D (see Supplementary Table 2 for details).



Supplementary Figure 4: Maps of constructed plasmids

The Forward Primer and Reverse Primer indicate the positions used for genomic integration.



Supplementary Figure 5: Extrinsic noise reduction

Reduction of extrinsic noise for (**a**) P_{ACT1} and (**b**) P_{ADH1} expression. Blue points are noise measurements without forward- and side-scatter gating. Red points are noise measurements within a fixed 80% autogate when cells are grouped by their forward- and side-scatter channel values and average expression and noise are obtained as the weighted average of the mean and the noise within these groups (see Methods). Broken curves are fits to Model B.

a **P_{ACT1}/P_{ADH1} correlations**

	r	*p* value
Average	0.29	1.50E-03
Noise	0.51	3.90E-09

b **P_{ACT1} noise correlations**

	r	*p* value
PolII	-0.48	1.50E-08
H3K14Ac	-0.52	4.10E-10
H3K9Ac	-0.6	1.20E-13
H4K16Ac	-0.63	3.80E-15

c **P_{ADH1} noise correlations**

	r	*p* value
PolII	-0.32	2.90E-04
H3K14Ac	-0.35	7.70E-05
H3K9Ac	-0.45	2.50E-07
H4K16Ac	-0.64	2.40E-15

Supplemental Table 1: Correlation coefficients and p values

Correlation coefficients *r* and *p* values associated with the null hypothesis of no correlation for different experimental datasets. (**a**) Correlations between average expression and expression noise for the two promoters. (**b**) Anti correlations between RNA polymerase II binding and histone acetylation for noise in P_{ACT1} and P_{ADH1} expression respectively.

a P_{ADH1} data untreated samples

	C_1	CI	C_0 or C_2	CI	τ_1/τ_0	CI	τ_2	CI
Model A	1.70	±0.05	-	-	-	-	-	-
Model B	0.21	±0.02	0.95	±0.05	-	-	-	-
Model C	0.64	±0.07	46.11	±19.28	3.04	±0.36	-	-
Model D	1.70	±1.22	40.00	NA	-	-	3.9E+04	±2.0E+07

b P_{ADH1} data all samples

	C_1	CI	C_0 or C_2	CI	τ_1/τ_0	CI	τ_2	CI
Model A	1.93	±0.04	-	-	-	-	-	-
Model B	0.21	±0.01	0.94	±0.05	-	-	-	-
Model C	0.72	±0.05	66.73	±19.63	3.42	±0.27	-	-
Model D	2.88	±0.60	40.00	NA	-	-	4.7E+04	±3.8E+07

Supplemental Table 2: Fitted model parameters for P_{ADH1} expression

CI is the confidence interval. The confidence interval for C_2 is not applicable (NA) since the parameter is fixed at the lower bound for the maximal average protein abundance. Note that the value of τ_2 is unconstrained.

a **P_{ACT1} data untreated samples**

	Slow versus fast kinetics			Size versus frequency		K_{off} versus
	Model B	Model C	K_m	Model B	Model C	k_{off}
Mean	4.1E-02	-5.5E-03	-6.9E-04	4.2E-02	-4.8E-03	-4.7E-02
Stdev	1.0E-01	7.0E-02	1.8E-03	1.0E-01	7.0E-02	6.2E-02
p-value	6.6E-01	4.7E-01	3.5E-01	1.0E-01	4.7E-01	7.8E-01

b **P_{ADH1} data untreated samples**

	Slow versus fast kinetics			Size versus frequency		K_{off} versus
	Model B	Model C	K_m	Model B	Model C	K_m
Mean	5.5E-01	4.5E-01	-2.5E-05	5.5E-01	4.5E-01	-9.4E-02
Stdev	1.4E-01	2.2E-01	8.8E-04	1.4E-01	2.2E-01	1.3E-01
p-value	5.8E-05	1.7E-02	5.1E-01	5.8E-05	1.8E-02	2.4E-01

c **P_{ADH1} data prediction of treated samples**

	Slow versus fast kinetics			Size versus frequency		K_{off} versus
	Model B	Model C	Model D	Model B	Model C	K_m
Mean	1.7E+00	1.6E+00	5.0E-05	1.7E+00	1.6E+00	-5.4E-02
Stdev	1.3E-01	1.2E-01	2.0E-05	1.3E-01	1.2E-01	8.4E-02
p-value	2.7E-37	3.7E-46	6.0E-03	2.7E-37	3.7E-46	2.4E-01

Supplemental Table 3: Model discrimination using chromosome-wide data

The mean and standard deviation (stdev) associated with the log error ratio (LER) associated with different model comparisons (see Methods). The *p*-value is associated with the null hypothesis that the mean LER deviated from zero and was obtained using a one-sided z-test. Slow versus fast kinetics refers to the testing of Model A against Model B, Model C and Model D. Size versus frequency refers to testing Model D against Model B and Model C. k_M versus k_{off} refers to testing of Model C versus Model B. (**a, b**) Comparison of how well different models fits to data obtained for untreated data for P_{ACT1} and P_{ADH1} expression, respectively. (**c**) Comparison of how well different models are able to predict the effect of nicotinamide treatment.

4. General Discussion

Sensitivity-Based Epistatic analysis

Throughout our epistatic analysis of the DNA damage response in *S.cerevisiae*, we have developed a method that explicitly indentifies epistatic interactions with a dynamic component. We have demonstrated that such effects can be captured by extending the neutrality function between a gene and its environment (Hillenmeyer et al., 2008; Lehár et al., 2007; Parsons et al., 2004) to incorporate additional genetic backgrounds. This results in a sensitivity-based neutrality function ($S_{XY} \times S_{wt} = S_X \times S_Y$) from which departure is indicative of a genetic interaction that changes in response to an environmental perturbation. The identification of genetic interactions with dynamic components may differ significantly from those identified in a single environment. Noticeably, in our dataset only ~50% of the interactions identified by the conventional fitness-based approach in the presence of MMS are also identified using sensitivity. These interactions represent dynamic functional relationships that are likely to be modulated in a response-specific manner. Thus, combining the two approaches may enable the segregation of genetic interactions representing response-specific linkages between regulatory functional pathways, and genetic interactions associated with core processes preserved across most environments.

While certain of the interactions gained through this analysis are associated with well documented regulatory interactions, such as those between Swi5 and Fkh2 (Pic et al., 2000) and Swi5 and Ace2 (Voth et al., 2007), others represent novel findings which require further investigation into their role within the DNA damage response. This includes the interaction

between Msn4 and Swi6 which has little to no impact in the absence of DNA damage, but was severely aggravating in its presence. While Msn4 activity is intimately linked with the chromatin regulator NuA4 (Mitchell et al., 2008), and Swi6 has equally been documented to physically interact with many of its subunits according to the Biogrid Database (Stark et al., 2006) including Eaf1, Eaf3, Eaf6, Eaf7, Epl1, Swc4, Yaf9, and Yng2, we believe this interaction may reflect loss of function of NuA4. We believe that this aggravating interaction may reflect one of two NuA4 dependant mechanisms. Either 1) presence of either of the two TFs is sufficient to maintain NuA4 functionality within the DNA damage response or 2) NuA4 functionality buffers for the loss of one of the two transcriptional regulators. Hypotheses which should be investigated in future studies.

Integrating information on response-specific functional relationships among cellular regulators may provide clues to the dynamic modulation of genetic networks in response to external stimuli. In fact, the transcriptional regulators investigated in this study were chosen because of the comprehensive dataset made available by Workman et al., 2006. This dataset includes information on genome-scale protein-DNA interactions in the presence and absence of MMS, as well as information on the loss of MMS-induced differential gene expression following TF deletion, referred to as genetic buffering or regulatory epistasis (Tan et al., 2008). From these experiments, direct TF-TF interactions in which one TF binds to or regulates expression of another are anticipated to result in a "dynamically" regulated epistatic interaction. Within our transcriptional network direct TF-TF interactions inferred from both protein-DNA or genetic buffering experiments serve as an accurate predictor of a sensitivity-based interaction. A result not shared by fitness-based epistatic analysis. From the Workman datasets, TF co-binding to differentially expressed targets should equally highlight epistatic interactions with a dynamic

component. While no enrichment in such co-binding events was observed within the fitness-based dataset, only a slight enrichment was observed for sensitivity-based interactions. The lack of a strong overlap between these two datasets may in part reflect the observation that few of the targets differentially bound across environmental condition are differentially expressed throughout the DNA damage response. Indeed, when we plot the number genes differentially bound by a given TF against the number of genes it buffers (Workman et al., 2006), no correlation is observed between the two datasets (Figure D1). As such, it is assumed that only a small portion of the interactions inferred from co-binding to differentially expressed targets manifest a functional consequence on the expression profile. While co-binding information on its own does not provide strong support for sensitivity-based epistatic interactions, integrating this information with direct TF-TF interaction accounts for ~50% of the interactions. An enrichment not obtained when the fitness-based approach is analyzed were only ~25% are supported by one on the lines of physical evidence. This observation would suggest that while co-binding to differentially expressed targets is not the primary determinant of a dynamically regulation, such information provides support an important fraction of our interactions.

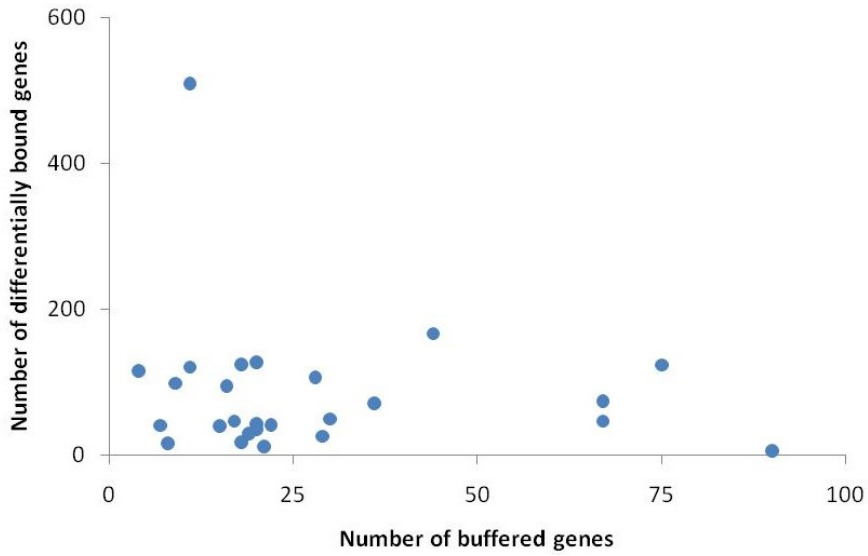


Figure D1. Number of genes differentially bound by a given transcription factor as a function of the number of genes it buffers.

Genes differentially bound refers to the sum of genes identified as only expressed in one of the two environmental conditions. Data illustrated represents the response for our 26 TFs in response to MMS as defined by Workman et al., 2006.

While our TF dataset did not encompass physical complexes specifically identified as involved within the DNA damage response, we next applied our analysis to a previously published dataset generated by St.Onge et al, 2007. This dataset comprises a set of 26 core regulators involved within the DNA-damage response for which extensive protein-protein interaction datasets are publically available (Cherry et al., 1998). To perform this analysis, we utilized both sensitivity and fitness phenotypes to identified physical complexes using two distinct methodologies. The first methodology, defined as epistatic score profiling, was implemented under the assumption that genes involved in the same physical complex should share a similar pattern of epistatic interactions. Interestingly, hierarchical clustering of sensitivity-based epistatic profiles allowed recapturing near complete network hierarchy within the DNA damage response. The sensitivity-based methodology improved complex identification over its fitness counterpart not only by identifying a greater number of physical complexes, but also by improving coverage within these complexes as a greater number of subunits were appropriately sub-classified. This includes *HPR5* involvement with members of the Rad6 epistasis group (*RAD5* and *RAD18*), *RAD54* and *RAD59* function within the Rad52 epistasis group, as well as the upstream role of Sgs1 in the processing HRR intermediates, as well as identification of the MMS1-RTT101 functional complex. The second methodology, defined as co-equivalence testing, is the more classical approach. Through this methodology, physical complexes are inferred when the phenotype of the double mutant is equivalent to that of the single mutant. Through this analysis, we demonstrated that sensitivity-based analysis provides a certain improvement in that it greatly reduces the equivalence threshold utilized to identify functional complexes.

From these analyses, we have demonstrated that sensitivity-based epistatic analysis certainly enables the identification of dynamically regulated genetic interactions. However, certain caveats are associated with this methodology. Noticeably, propagation of error is a major limitation of our methodology. While the traditional fitness-based methodology is associated with the measurement of four parameters ($W_A, W_B, W_{A,B}$ and W_{WT}), the sensitivity-based methodology inherently doubles the number of parameters as measurements must be performed across environmental conditions. This in turn decreases the statistical significance of a given interaction and impedes interaction identification. A second issue associated with this methodology is that the sign of the interaction is not associated with a distinct network state, contrary to the more traditional fitness based analysis. For example, genes which operate within the same pathway or complex can be associated with an aggravating or alleviating sensitivity-based interaction depending on which environmental condition the complex is active. If the complex is active only in the presence of the environmental perturbation, an aggravating sensitivity-based interaction is inferred. If the complex is active only in its absence, an alleviating sensitivity-based interaction is inferred.

Integrating physical and phenotypic data into comprehensive and accurate models of cellular regulation remains a major challenge in systems biology (Beyer et al., 2007). The mapping of biomolecular interactions and transcriptional profiling provide fundamental insight into the substantial remodeling of gene regulatory networks that take place following environmental perturbations. However, it is not always clear if and how observed changes in the physical interaction network manifest at the physiological level. The phenotypic information provided by sensitivity-based epistatic analysis may prove useful for the analysis of complex gene regulatory

networks. Having demonstrated that this methodology provides the mapping of dynamically changing genetic interactions, this methodology may generate important clues that can assist in the identification of pathways and modules more likely to be active within a given network model; a feat not accomplished by physical interaction datasets nor fitness-based genetic interactions. As such, we anticipate future studies will implement this methodology to analyze a wide range of phenotypes raging from fitness to gene expression studies.

Chromosomal position effects

Having completed this first project by the end of the first year of my master's degree, I next sought to finalize a second project which investigated how chromosomal positioning modulates transcription kinetics. By cloning two contrasting promoter architectures across an entire chromosome, we have systematically mapped how chromosomal positioning modulates gene expression. The first important conclusion was that while a promoter dynamically regulated by nucleosome structure (P_{ADH1}) varied significantly across positions, the promoter associated with a constitutive open chromatin structure (P_{ACT1}) was also associated with significant positional effects. A result unanticipated under the hypothesis that the elements which resist nucleosome positioning in open promoters should in principle buffer against positional effects if genome positioning potentiates its effects by regulating the frequency of gene activation. This observation would in turn suggest that the factors which mediate positional effects may impact a step in gene transcription more-or-less independent of promoter activation.

To further develop this notion, we fit the relationship between mean expression and noise for the two promoters to theoretical models for eukaryotic gene expression. A key difference in the correlation between mean expression and noise for the two promoter is the asymptotic limit for

noise. To investigate which parameters could accurately recapture this difference, we derive theoretical models describing the absence of transcriptional bursting, as well as three models which simulate the effects anticipated when chromosomal position modulates transcriptional bursting through variation in the rate of transcription (k_m), promoter activation (k_{on}), or promoter inactivation (k_{off}). All four models accurately recapture the trend for P_{ACT1}. A result which suggests that transcriptional bursting is absent for this promoter. In contrast, models where the rate of transcription (k_m) or promoter inactivation (k_{off}) accounts for chromosomal positioning effects improve the trend established for P_{ADH1}. While the number of mRNA produced per burst is determined by the ratio between these two parameters, a theoretical approach in turn supports the notion that transcriptional burst size rather than burst frequency, is primarily targeted by chromosomal positioning effects.

To investigate the factors which contribute towards such effects, we integrated our analysis with high-throughput information describing histone post-translational modifications. Interestingly, genes with low expression and high noise appear to be enriched in low levels of H3K3, H3K14 and H4K16 acetylation. Events which are associated with the function of Sir2, a histone deacetylase. A result somewhat anticipated as low expression high noise regions are primarily located adjacent to the HML and telomeric domains; well documented targets of the Sir2 histone deacetylase (Sauve et al., 2006). It is however surprising that Sir2 has the potential to impact gene expression from both covered (P_{ADH1}) and open (P_{ACT1}) promoter constructs. An observation which suggests that that Sir2 may bypass the elements in the open promoter construct which prevent nucleosome deposition. Consistent with this notion, Sir2 activity establishes a chromatin structure that is permissive to PIC formation and polymerase recruitment but prevents full length mRNA synthesis (Sekinger and Gross, 2001; Gao and Gross, 2008). An

experimental result which in our simulations would be associated with a variation in Km, and would impact a step independent of the nucleosome free region generated by the DNA sequence elements present in the Act1 promoter.

To validate the hypothesis that Sir2 modulates chromosomal positioning effects by varying transcriptional burst size, we characterized the effects of nicotinamide treatment, a well documented inhibitor of Sir2 functionality (Sauve et al., 2006). Addition of nicotinamide had widespread effect on noise and gene expression driven by both promoters constructs, affecting more than 40% of loci tested for each promoter. The resulting correlation between gene expression and noise mimics the trend established through our positional analysis. A result which suggest that Sir2 may function as a global regulator of transcriptional burst size across the genome, consistent with the notion that Sir2 is polymerase recruitment but prevents full length mRNA synthesis (Sekinger and Gross, 2001; Gao and Gross, 2008).

Such widespread effects differ from the typical description that Sir2 functions predominantly within the mating type locus (HML/HMR), telomeric regions and rDNA in yeast (Sauve et al. 2006). While loci surrounding these regions were most affected by nicotinamide treatment (other than HMR which is not silenced by Sir2 in our genetic background), many other regions were significantly affected. Consistent with this observation, ChIP-Chip analysis performed on Sir2 demonstrated its widespread binding throughout the genome, including at approximately a quarter (244/1013) of the most frequently transcribed genes (Tsankov et al., 2006). Noticeably, both Act1 and Adh1 appear to be bound by Sir2 according to the Chip-ChIP study (Tsankov et al., 2006), whereby the promoter of Adh1 is equally bound by both Rap1 and Abf1 (Bird et al., 2006; Hae Yong Yoo et al., 1995), well documented regulators which recruit Sir2 proteins to silenced loci (Huang, 2002). Interestingly, early studies assessing the impact of Sir2 over-

expression within the cell noted a global decrease in acetylation levels of histones H2B, H3 and H4 (Braunstein et al., 1993). While histone acetylation is primarily associated with euchromatic domains, this study suggests that over-expression of Sir2 may have additional functionality within the cell consistent with its widespread binding across the genome.

To confirm that our observation is not attributed to pleiotropic effects associated with nicotinamide treatment, we next sought to validate our observation through Sir2 deletion. This analysis was performed by monitoring the impact of Sir2 deletion on a subset of loci where the P_{ADH1} construct cloned. This analysis in turn sought to provide a direct link between Sir2 activity and its potential to modulate transcriptional burst size across the genome. While the construct located adjacent to the HML locus (YCL064C) was associated with the greatest effect, all strains were associated with an increase in gene expression and a decrease in transcriptional noise following Sir2 deletion. Once again, fitting to theoretical models demonstrate that this response is best captured by models where it is assumed that transcription burst size is regulated by Sir2. Through this analysis, we have not eliminated the possibility that the trends established are associated with indirect effects inherent to Sir2 deletion such as an increase in extrinsic noise due to cell cycle progression. Although extrinsic noise should in principle impact the asymptotic limit for noise, real-time measurement of single cell mRNA counts in the presence and absence of Sir2 may be required to further substantiate our hypothesis.

The observation that transcriptional burst size, rather than burst frequency, is modulated across chromosomal positions has important implication in the forces driving genome organization. In previous analysis of chromosomal organization, it was observed that essential gene clusters are associated with nucleosome depletion (Batada and Hurst, 2007). The authors in turn suggest that nucleosome depletion promotes the formation of essential gene clusters, as their

migration towards regions dynamically regulated by chromatin structure may have a negative impact cell viability. However, as many of these promoters contain elements which resist nucleosome deposition, this observation may merely reflect the consequence of such clustering, rather a selective force driving migration towards such regions. Through our analysis, we observe that chromosomal positioning effects effectively bypass the elements which induce an open promoter structure. As such, it seems plausible, if not likely, that essential genes selectively migrate towards open chromatin regions to ensure an adequate burst size. Consequently, covered promoters may migrate towards specific regions to modulate gene expression in a process bypassing promoter structure thereby allowing for a greater complexity in gene regulation across environmental conditions.

5. References

Angermayr, M., U. Oechsner, and W. Bandlow. 2003. Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. *J.Biol.Chem.* 278:17918-17926.

Avery, L., and S. Wasserman. 1992. Ordering gene function: The interpretation of epistasis in regulatory hierarchies. *Trends Genet.* 8:312-316.

Aylor, D.L., and Z.-. Zeng. 2008. From classical genetics to quantitative genetics to systems biology: Modeling epistasis. *PLoS Genet.* 4.

Baldwin, E.L., A.C. Berger, A.H. Corbett, and N. Osheroff. 2005. Mms22p protects Saccharomyces cerevisiae from DNA damage induced by topoisomerase II. *Nucleic Acids Res.* 33:1021-1030.

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

Batada, N.N., and L.D. Hurst. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat.Genet.* 39:945-949.

Becskei, A., B.B. Kaufmann, and A. Van Oudenaarden. 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat.Genet.* 37:937-944.

Beerenwinkel, N., L. Pachter, B. Sturmfels, S.F. Elena, and R.E. Lenski. 2007. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol. Biol.* 7.

Begley T.J., Rosenbach A.S., Ideker T., Samson L.D. 2004. Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping *Mol. Cell*. 16:**1**17-125.

Beyer, A., S. Bandyopadhyay, and T. Ideker. 2007. Integrating physical and genetic maps: From genomes to interaction networks. *Nat. Rev. Genet.* 8:699-710.

Bhaumik, S.R., and M.R. Green. 2002. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo. *Mol.Cell.Biol.* 22:7365-7371.

Bird, A.J., M. Gordon, D.J. Eide, and D.R. Winge. 2006. Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J.* 25:5726-5734.

Bitterman, K.J., R.M. Anderson, H.Y. Cohen, M. Latorre-Esteves, and D.A. Sinclair. 2002. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast Sir2 and human SIRT1. *J.Biol.Chem.* 277:45099-45107.

Blake, W.J., G. Balázsi, M.A. Kohanski, F.J. Isaacs, K.F. Murphy, Y. Kuang, C.R. Cantor, D.R. Walt, and J.J. Collins. 2006. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Mol.Cell.* 24:853-865.

Blake, W.J., M. Kærn, C.R. Cantor, and J.J. Collins. 2003. Noise in eukaryotic gene expression. *Nature.* 422:633-637.

Boone, C., H. Bussey, and B.J. Andrews. 2007. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 8:437-449.

Braunstein, M., A.B. Rose, S.G. Holmes, C.D. Allis, and J.R. Broach. 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes and Development.* 7:592-604.

Breslow, D.K., D.M. Cameron, S.R. Collins, M. Schuldiner, J. Stewart-Ornstein, H.W. Newman, S. Braun, H.D. Madhani, N.J. Krogan, and J.S. Weissman. 2008. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature Methods.* 5:711-718.

Cairns, B.R. 2009. The logic of chromatin architecture and remodelling at promoters. *Nature.* 461:193-198.

Carlborg, Ö., and C.S. Haley. 2004. Epistasis: Too often neglected in complex trait studies? *Nature Reviews Genetics.* 5:618-625.

Cherry, J.M., C. Adler, C. Ball, S.A. Chervitz, S.S. Dwight, E.T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, and D. Botstein. 1998. SGD: Saccharomyces genome database. *Nucleic Acids Res.* 26:73-79.

Collins, S.R., K.M. Miller, N.L. Maas, A. Roguev, J. Fillingham, C.S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S.L. Berger, P. Hieter, Z. Zhang, G.W. Brown, C.J. Ingles, A. Emili, C.D. Allis, D.P. Toczyski, J.S. Weissman, J.F. Greenblatt, and N.J. Krogan. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature.* 446:806-810.

Collins, S.R., M. Schuldiner, N.J. Krogan, and J.S. Weissman. 2006. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 7.

Cory Batenchuk. 2008. Quantifying the Impact on DNA Damage Recovery of Transcription Factor Deletion and Transcriptional Regulatory Network Topology. :32.

Denis, C.L., J. Ferguson, and E.T. Young. 1983. mRNA levels for the fermentative alcohol dehydrogenase of Saccharomyces cerevisiae decrease. *J.Biol.Chem.* 258:1165-1171.

Ehmsen, K.T., and W.-. Heyer. 2008. Saccharomyces cerevisiae Mus81-Mms4 is a catalytic, DNA structure-selective endonuclease. *Nucleic Acids Res.* 36:2182-2195.

Elbashir, S.M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* 411:494-498.

Elena, S.F., and R.E. Lenski. 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature.* 390:395-398.

Fantes, P. 1979. Epistatic gene interactions in the control of division in fission yeast. *Nature.* 279:428-430.

Fiedler, D., H. Braberg, M. Mehta, G. Chechik, G. Cagney, P. Mukherjee, A.C. Silva, M. Shales, S.R. Collins, S. van Wageningen, P. Kemmeren, F.C.P. Holstege, J.S. Weissman, M.-. Keogh, D. Koller, K.M. Shokat, and N.J. Krogan. 2009. Functional Organization of the S. cerevisiae Phosphorylation Network. *Cell.* 136:952-963.

Fillingham, J., P. Kainth, J.-. Lambert, H. van Bakel, K. Tsui, L. Peña-Castillo, C. Nislow, D. Figeys, T.R. Hughes, J. Greenblatt, and B.J. Andrews. 2009. Two-Color Cell Array Screen Reveals Interdependent Roles for Histone Chaperones and a Chromatin Boundary Regulator in Histone Gene Repression. *Mol.Cell.* 35:340-351.

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

Fricke, W.M., and S.J. Brill. 2003. Slx1 - Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1 - Top3. *Genes and Development.* 17:1768-1778.

Friedl, A.A., B. Liefshitz, R. Steinlauf, and M. Kupiec. 2001. Deletion of the SRS2 gene suppresses elevated recombination and DNA damage sensitivity in rad5 and rad18 mutants of Saccharomyces cerevisiae. *Mutation Research - DNA Repair.* 486:137-146.

Gao, L., and D.S. Gross. 2008. Sir2 silences gene transcription by targeting the transition between RNA polymerase II initiation and elongation. *Mol.Cell.Biol.* 28:3979-3994.

Ghaemmaghami, S., W.-. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, and J.S. Weissman. 2003. Global analysis of protein expression in yeast. *Nature.* 425:737-741.

Ghandi, S.J., D. Zenklusen, T. Lionnet, and R.H. Singer. 2010. Transcription of functionally related constitutive genes is not coordinated. *Nature Structural and Molecular Biology.*

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

Hae Yong Yoo, So Young Jung, Young Ho Kim, J. Kim, G. Jung, and Hyune Mo Rho. 1995. Transcriptional control of the Saccharomyces cerevisiae ADH1 gene by autonomously replicating sequence binding factor 1. *Curr.Microbiol.* 31:163-168.

Halter, M., J.T. Elliott, J.B. Hubbard, A. Tona, and A.L. Plant. 2009. Cell volume distributions reveal cell growth rates and division times. *J.Theor.Biol.* 257:124-130.

Harbison, C.T., D.B. Gordon, T.I. Lee, N.J. Rinaldl, K.D. Macisaac, T.W. Danford, N.M. Hannett, J.-. Tagne, D.B. Reynolds, J. Yoo, E.G. Jennings, J. Zeitlinger, D.K. Pokholok, M. Kellis, P.A. Rolfe, K.T. Takusagawa, E.S. Lander, D.K. Gilford, E. Fraenkel, and R.A. Young. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature.* 430:99-104.

Hillenmeyer, M.E., E. Fung, J. Wildenhain, S.E. Pierce, S. Hoon, W. Lee, M. Proctor, R.P. St.Onge, M. Tyers, D. Koller, R.B. Altman, R.W. Davis, C. Nislow, and G. Giaever. 2008. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science.* 320:362-365.

Huang, Y. 2002. Transcriptional silencing in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Nucleic Acids Res.* 30:1465-1482.

Imai, S.-., C.M. Armstrong, M. Kaeberlein, and L. Guarente. 2000. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature.* 403:795-800.

Jasnos, L., and R. Korona. 2007. Epistatic buffering of fitness loss in yeast double deletion strains. *Nat. Genet.* 39:550-554.

Jorgensen, P., J.L. Nishikawa, B.-. Breitkreutz, and M. Tyers. 2002. Systematic identification of pathways that couple cell growth and division in yeast. *Science.* 297:395-400.

Kærn, M., T.C. Elston, W.J. Blake, and J.J. Collins. 2005. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics.* 6:451-464.

Kar, S., W.T. Baumann, M.R. Paul, and J.J. Tyson. 2009. Exploring the roles of noise in the eukaryotic cell cycle. *Proc.Natl.Acad.Sci.U.S.A.* 106:6471-6476.

Kelemen, J.Z., P. Ratna, S. Scherrer, and A. Becskei. 2010. Spatial epigenetic control of mono- and bistable gene expression. *PLoS Biology.* 8.

Krogan, N.J., K. Baetz, M.-. Keogh, N. Datta, C. Sawa, T.C.Y. Kwok, N.J. Thompson, M.G. Davey, J. Pootoolal, T.R. Hughes, A. Emili, S. Buratowski, P. Hieter, and J.F. Greenblatt. 2004. Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. *Proc.Natl.Acad.Sci.U.S.A.* 101:13513-13518.

Lehár, J., G.R. Zimmermann, A.S. Krueger, R.A. Molnar, J.T. Ledell, A.M. Heilbut, G.F. Short III, L.C. Giusti, G.P. Nolan, O.A. Magid, M.S. Lee, A.A. Borisy, B.R. Stockwell, and C.T. Keith. 2007. Chemical combination effects predict connectivity in biological systems. *Molecular Systems Biology.* 3.

Lehner, B., C. Crombie, J. Tischler, A. Fortunato, and A.G. Fraser. 2006. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat. Genet.* 38:896-903.

Liu, C.L., T. Kaplan, M. Kim, S. Buratowski, S.L. Schreiber, N. Friedman, and O.J. Rando. 2005. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biology.* 3.

Longtine, M.S., A. McKenzie III, D.J. Demarini, N.G. Shah, A. Wach, A. Brachat, P. Philippsen, and J.R. Pringle. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast.* 14:953-961.

Mani, R., R.P. St. Onge, J.L. Hartman IV, G. Giaever, and F.P. Roth. 2008. Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.* 105:3461-3466.

Mankouri, H.W., H.-. Ngo, and I.D. Hickson. 2007. Shu proteins promote the formation of homologous recombination intermediates that are processed by Sgs1-Rmi1-Top3. *Mol.Biol.Cell.* 18:4062-4073.

Mcadams, H.H., and A. Arkin. 1997. Stochastic mechanisms in gene expression. *Proc.Natl.Acad.Sci.U.S.A.* 94:814-819.

McClean, M.N., A. Mody, J.R. Broach, and S. Ramanathan. 2007. Cross-talk and decision making in MAP kinase pathways. *Nat.Genet.* 39:409-414.

McLean, M., A.V. Hubberstey, D.J. Bouman, N. Pece, P. Mastrangelo, and A.G. Wildeman. 1995. Organization of the Saccharomyces cerevisiae actin gene UAS: Functional significance of reiterated REB1 binding sites and AT-rich elements. *Mol.Microbiol.* 18:605-614.

Mitchell, L., J.-. Lambert, M. Gerdes, A.S. Al-Madhoun, I.S. Skerjanc, D. Figeys, and K. Baetz. 2008. Functional dissection of the NuA4 histone acetyltransferase reveals its role as a genetic hub and that Eaf1 is essential for complex integrity. *Mol.Cell.Biol.* 28:2244-2256.

Musso, G., M. Costanzo, M. Huangfu, A.M. Smith, J. Paw, B.J. San Luis, C. Boone, G. Giaever, C. Nislow, A. Emili, and Z. Zhang. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res.* 18:1092-1099.

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

Onge, R.P.S., R. Mani, J. Oh, M. Proctor, E. Fung, R.W. Davis, C. Nislow, F.P. Roth, and G. Giaever. 2007. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* 39:199-206.

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

Pan, X., P. Ye, D.S. Yuan, X. Wang, J.S. Bader, and J.D. Boeke. 2006. A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell.* 124:1069-1081.

Pan, X., D.S. Yuan, S.-. Ooi, X. Wang, S. Sookhai-Mahadeo, P. Meluh, and J.D. Boeke. 2007. dSLAM analysis of genome-wide genetic interactions in Saccharomyces cerevisiae. *Methods.* 41:206-221.

Parsons, A.B., R.L. Brost, H. Ding, Z. Li, C. Zhang, B. Sheikh, G.W. Brown, P.M. Kane, T.R. Hughes, and C. Boone. 2004. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways.. *Nat. Biotechnol.* 22:62-69.

Pfander, B., G.-. Moldovan, M. Sacher, C. Hoege, and S. Jentsch. 2005. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature.* 436:428-433.

Phillips, P.C. 2008. Epistasis - The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9:855-867.

Pic, A., F.-. Lim, S.J. Ross, E.A. Veal, A.L. Johnson, M.R.A. Sultan, A.G. West, L.H. Johnston, A.D. Sharrocks, and B.A. Morgan. 2000. The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. *EMBO J.* 19:3750-3761.

Raser, J.M., and E.K. O'Shea. 2005. Molecular biology - Noise in gene expression: Origins, consequences, and control. *Science.* 309:2010-2013.

Raser, J.M., and E.K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science.* 304:1811-1814.

Remold, S.K., and R.E. Lenski. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in Escherichia coli. *Nat. Genet.* 36:423-426.

Robyr, D., Y. Suka, I. Xenarios, S.K. Kurdistani, A. Wang, N. Suka, and M. Grunstein. 2002. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell.* 109:437-446.

Rusché, L.N., A.L. Kirchmaier, and J. Rine. 2002. Ordered nucleation and spreading of silenced chromatin in Saccharomyces cerevisiae. *Mol.Biol.Cell.* 13:2207-2222.

Sauve, A.A., C. Wolberger, V.L. Schramm, and J.D. Boeke. 2006. The biochemistry of sirtuins. *Annual Review of Biochemistry.* 75:435-465.

Schuldiner, M., S.R. Collins, N.J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J.F. Greenblatt, J.S. Weissman, and N.J. Krogan. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell.* 123:507-519.

Segal, E., and J. Widom. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr.Opin.Struct.Biol.* 19:65-71.

Segrè, D., A. DeLuna, G.M. Church, and R. Kishony. 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* 37:77-83.

Sekinger, E.A., and D.S. Gross. 2001. Silenced chromatin is permissive to activator binding and PIC recruitment. *Cell.* 105:403-414.

Shukla, A., P. Bajwa, and S.R. Bhaumik. 2006. SAGA-associated Sgf73p facilitates formation of the preinitiation complex assembly at the promoters either in a HAT-dependent or independent manner in vivo. *Nucleic Acids Res.* 34:6225-6232.

Singh, A., B. Razooky, C.D. Cox, M.L. Simpson, and L.S. Weinberger. 2010. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys.J.* 98:L32-L34.

Stark, C., B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research.* 34:D535-539.

Suzuki, R., and H. Shimodaira. 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics.* 22:1540-1542.

Swain, P.S. 2004. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J.Mol.Biol.* 344:965-976.

Symington, L.S. 2002. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiology and Molecular Biology Reviews.* 66:630-670.

Tan, K., H. Feizi, C. Luo, S.H. Fan, T. Ravasi, and T.G. Ideker. 2008. A systems approach to delineate functions of paralogous transcription factors: Role of the Yap family in the DNA damage response. *Proc. Natl. Acad. Sci. U.S.A.* 105:2934-2939.

Tong, A.H.Y., and C. Boone. 2007. High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in Saccharomyces cerevisiae. *Methods in Microbiology.* 36:369-386,706-707.

Tong, A.H.Y., M. Evangelista, A.B. Parsons, H. Xu, G.D. Bader, N. Pagé, M. Robinson, S. Raghibizadeh, C.W.V. Hogue, H. Bussey, B. Andrews, M. Tyers, and C. Boone. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 294:2364-2368.

Tong, A.H.Y., G. Lesage, G.D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G.F. Berriz, R.L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D.S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J.N. Levinson, H. Lu, P. Ménard, C. Munyana, A.B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.-. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S.L. Wong, L.V. Zhang, H. Zhu, C.G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F.P. Roth, G.W. Brown, B. Andrews, H. Bussey, and C. Boone. 2004. Global Mapping of the Yeast Genetic Interaction Network. *Science.* 303:808-813.

Tsankov, A.M., C.R. Brown, M.C. Yu, M.Z. Win, P.A. Silver, and J.M. Casolari. 2006. Communication between levels of transcriptional control improves robustness and adaptivity. *Molecular Systems Biology.* 2.

Volfson, D., J. Marciniak, W.J. Blake, N. Ostroff, L.S. Tsimring, and J. Hasty. 2006. Origins of extrinsic variability in eukaryotic gene expression. *Nature.* 439:861-864.

Voth, W.P., Y. Yu, S. Takahata, K.L. Kretschmann, J.D. Lieb, R.L. Parker, B. Milash, and D.J. Stillman. 2007. Forkhead proteins control the outcome of transcription factor binding by antiactivation. *EMBO J.* 26:4324-4334.

Workman, C.T., H.C. Mak, S. McCuine, J.-. Tagne, M. Agarwal, O. Ozier, T.J. Begley, L.D. Samson, and T. Ideker. 2006. A systems approach to mapping DNA damage response pathways. *Science.* 312:1054-1059.

Zaidi, I.W., G. Rabut, A. Poveda, K. Hofmann, J. Malmström, H. Ulrich, K. Hofmann, P. Pasero, M. Peter, and B. Luke. 2008. Rtt101 and Mms1 in budding yeast form a CUL4DDB1-like ubiquitin ligase that promotes replication through damaged DNA. *EMBO Rep.* 9:1034-1040.

Zenklusen, D., D.R. Larson, and R.H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural and Molecular Biology.* 15:1263-1271.