

# Noise–mean relationship in mutated promoters

Gil Hornung,<sup>1</sup> Raz Bar-Ziv,<sup>1</sup> Dalia Rosin,<sup>1</sup> Nobuhiko Tokuriki,<sup>2</sup> Dan S. Tawfik,<sup>2</sup> Moshe Oren,<sup>3</sup> and Naama Barkai<sup>1,4</sup>

<sup>1</sup>Department of Molecular Genetics; <sup>2</sup>Department of Biological Chemistry; <sup>3</sup>Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Gene expression depends on the frequency of transcription events (burst frequency) and on the number of mRNA molecules made per event (burst size). Both processes are encoded in promoter sequence, yet their dependence on mutations is poorly understood. Theory suggests that burst size and frequency can be distinguished by monitoring the stochastic variation (noise) in gene expression: Increasing burst size will increase mean expression without changing noise, while increasing burst frequency will increase mean expression and decrease noise. To reveal principles by which promoter sequence regulates burst size and frequency, we randomly mutated 22 yeast promoters chosen to span a range of expression and noise levels, generating libraries of hundreds of sequence variants. In each library, mean expression ( $m$ ) and noise (coefficient of variation,  $\eta$ ) varied together, defining a scaling curve:  $\eta^2 = b/m + \eta_{\text{ext}}^2$ . This relation is expected if sequence mutations modulate burst frequency primarily. The estimated burst size ( $b$ ) differed between promoters, being higher in promoter containing a TATA box and lacking a nucleosome-free region. The rare variants that significantly decreased  $b$  were explained by mutations in TATA, or by an insertion of an out-of-frame translation start site. The decrease in burst size due to mutations in TATA was promoter-dependent, but independent of other mutations. These TATA box mutations also modulated the responsiveness of gene expression to changing conditions. Our results suggest that burst size is a promoter-specific property that is relatively robust to sequence mutations but is strongly dependent on the interaction between the TATA box and promoter nucleosomes.

[Supplemental material is available for this article.]

Biological processes are often executed by a small number of molecules per individual cell, leading to significant cell-to-cell variability (“noise”) in gene expression (Paulsson 2004; Maheshri and O’Shea 2007; Raj and van Oudenaarden 2008; Tawfik 2010). When analyzing gene expression noise, it is convenient to distinguish between intrinsic variations, resulting from stochastic production, and extrinsic variations propagating from global (e.g., ribosomes, polymerases, metabolites, etc.) or pathway-specific factors (Elowitz et al. 2002). Intrinsic noise is of particular interest, as it reflects on the transcription process itself (Paulsson 2004, 2005; Raser and O’Shea 2004; Raj and van Oudenaarden 2008; Rinott et al. 2011; So et al. 2011).

The prevailing model of gene expression noise assumes that genes transit stochastically between states that are permissive or nonpermissive to transcription (Paulsson 2004, 2005; Raser and O’Shea 2004; Friedman et al. 2006; Raj et al. 2006; Zenklusen et al. 2008; So et al. 2011). This two-state model predicts a scaling relationship between mean expression  $m$  and the coefficient of variation (noise, SD/mean)  $\eta$ :  $\eta^2 = b/m + \eta_{\text{ext}}^2$  where  $b$  is the typical number of protein molecules made during a single “on” state (“burst size”) and  $\eta_{\text{ext}}$  denotes the extrinsic noise (Paulsson 2004, 2005; Raser and O’Shea 2004; Bar-Even et al. 2006; Raj et al. 2006; Pedraza and Paulsson 2008; Tan and van Oudenaarden 2010; Taniguchi et al. 2010). Note that burst size,  $b$ , accounts for all the transcription-translation processes following the main stochastic event (burst initiation), integrating the number of mRNA molecules produced per burst and the number of protein molecules

made per each mRNA molecule. Upon a perturbation, the noise–mean relationship may change, depending on whether burst size or burst frequency were modulated (Pedraza and Paulsson 2008; Zenklusen et al. 2008; Tan and van Oudenaarden 2010).

Genome-wide analysis of the noise–mean relationship in yeast (Bar-Even et al. 2006; Newman et al. 2006) or *Escherichia coli* (Taniguchi et al. 2010; So et al. 2011) genes reported a general dependency that was well defined by the scaling relation  $\eta^2 = b/m + \eta_{\text{ext}}^2$ , suggesting a similar burst size for many genes. The expression of genes deviating from the scaling curve, displaying higher-than-expected noise (Bar-Even et al. 2006; Newman et al. 2006), was more responsive to changing conditions and also diverged more between related species (Tirosh and Barkai 2008; Choi and Kim 2009; Lehner 2010). Notably, high noise, responsiveness, and divergence were all correlated with the organization of gene promoters: All three measures were low in promoters lacking a TATA box and containing a nucleosome free region (NFR) proximal to the transcription start site (referred to as DPN promoters—depleted proximal nucleosome), and were high in TATA-containing promoters that lack NFR (OPN, occupied proximal nucleosome) (Field et al. 2008; Tirosh and Barkai 2008; Choi and Kim 2009). TATA box was also shown to increase noise in Pho5 expression (Raser and O’Shea 2004) and in synthetic promoters (Blake et al. 2006; Murphy et al. 2010).

The observation that genes with a characteristic promoter structure had a high noise (relative to that expected given their mean expression) is consistent with the idea that promoter sequence influences not only burst frequency but also burst size. Still, the principles by which promoter sequences regulate those two processes are not understood, primarily because most studies analyzing the interplay between promoter sequence variations and gene expression consider mean expression only (e.g., Yun et al. 2012).

#### <sup>4</sup>Corresponding author

E-mail Naama.barkai@weizmann.ac.il

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.139378.112>. Freely available online through the *Genome Research* Open Access option.

To distinguish between the effects of promoter sequence on burst size and burst frequency, we generated large libraries of sequence-mutated promoters. Specifically, we chose 22 yeast promoters that span a range of expression and noise levels. Using mutagenic PCR, we generated hundreds of sequence variants of each promoter. Each variant was fused to a fluorescent reporter, and the associated mean expression and noise (coefficient of variation) in a population of identical cells were measured. We found that sequence variants in each of the libraries defined a scaling curve  $\eta^2 = b/m + \eta_{\text{ext}}^2$ , with a constant estimated burst size  $b$ , that was largely promoter-specific and was particularly large for OPN promoters containing a TATA box. A small fraction of sequence mutations leading to a large change in burst size was identified in the OPN-type promoters containing a TATA box. These changes were biased toward reducing burst size, and were almost fully explained by elimination of a TATA box or insertion of a new out-of-frame translation start site. Interestingly, mutations that deleted a TATA box in low-noise DPN-type promoters did not reduce burst size. Our results suggest that burst size is a promoter-specific property that is insensitive to most sequence mutations but is largely influenced by the interaction between TATA box and promoter nucleosomes.

## Results

### Noise in mutated promoters scales with mean abundance

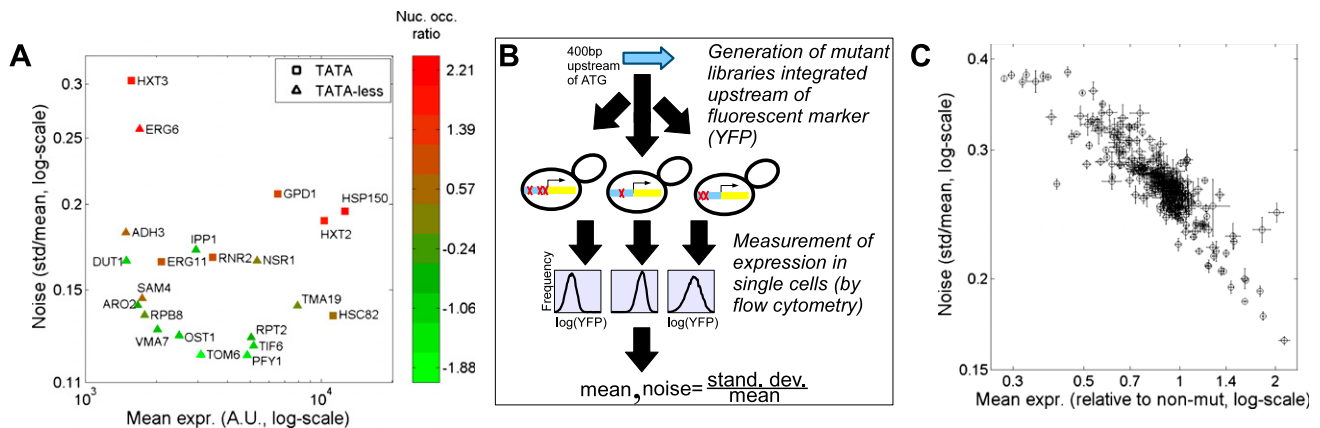
We chose 22 *Saccharomyces cerevisiae* promoters spanning a range of mid-to-high expression levels. Promoters chosen differed in their nucleosome organization and in the presence (or absence) of a TATA box (Fig. 1A; Supplemental Table S1). To randomly mutate promoter sequences, we performed mutagenic PCR amplification on a 400-bp region upstream of the ATG, containing the promoter region and the 5' UTR. Amplified DNA was cloned into a bacterial plasmid and chromosomally integrated into the same (neutral) locus upstream of a YFP reporter. This way, we created for each promoter a library of ~300 strains, each containing a different promoter variant (Fig. 1B).

We quantified the expression levels of each strain in the library using a flow cytometer (Fig. 1B; Methods) and summarized

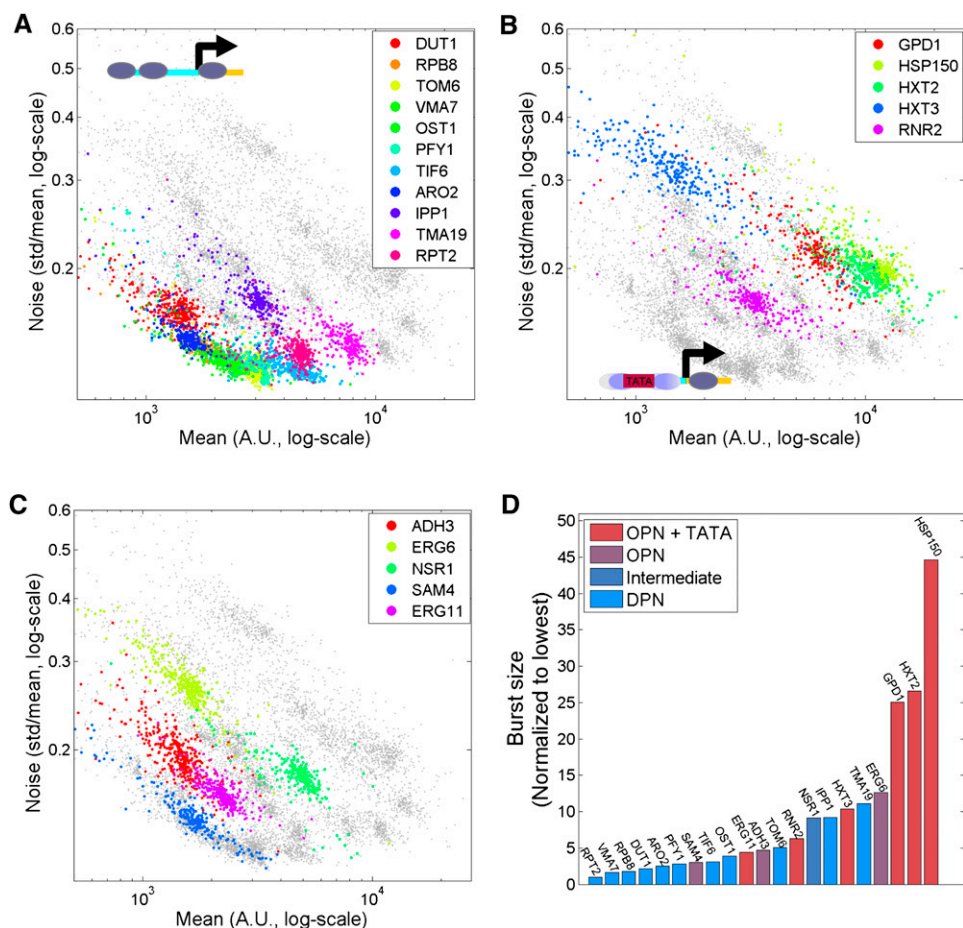
the distribution of single-cell fluorescence levels by its mean,  $m$ , and coefficient of variation (noise),  $\eta$  (raw results are provided as Supplemental Data). The noise levels of the nonmutated promoters varied, with OPN-type promoters containing a TATA box showing higher noise levels, as expected (Fig. 1A). Promoter mutations generally reduced mean expression, although we did observe some cases of increased expression.

As discussed above, sequence variants that change burst size are expected to change only the mean expression while leaving noise intact. In contrast, sequence variants that modulate burst frequency are expected to coordinately change noise and mean expression according to the relation  $\eta^2 = b/m + \eta_{\text{ext}}^2$  (Bar-Even et al. 2006; Friedman et al. 2006; Raj et al. 2006; Tan and van Oudenaarden 2010). Examining the noise–mean relation in all variants in our libraries, we found that mean expression and noise were strongly coupled (Figs. 1C, 2A–C; Supplemental Fig. S1): Mutations that increased mean expression reduced noise and vice versa. When plotted on a log-log plot, noise was roughly proportional to the (inverse) mean abundance ( $\eta^2 \sim b/m + \eta_{\text{ext}}^2$ ) (Figs. 1C, 2A–C). Within this model burst size  $b$  is given by the intercept of the curve, while  $\eta_{\text{ext}}$  is estimated by the asymptotic noise value at high mean. Only two promoters (*HXT3* and *PFY1*) showed a better fit to an alternative power-law model  $\eta \sim \alpha m^\beta$  (Supplemental Table S3), based on an F-test for the sum of squared residuals (Jaqaman and Danuser 2006). *HSC82* displayed a unique behavior, which will be discussed below (Fig. 5B, see below).

For each promoter, we estimated the burst size  $b$  and extrinsic noise  $\eta_{\text{ext}}$ , using linear regression (Methods; Supplemental Table S3). For most promoters, our estimates of  $\eta_{\text{ext}}$  were distributed around values that were reported previously for yeast ( $\eta_{\text{ext}} \sim 0.13$ ). The estimated burst size  $b$  largely varied, with each promoter defining a unique curve. Most of our DPN-type promoters traced a similar low-noise line, consistent with a low burst size (Fig. 2A,D). In contrast, each of the OPN-type TATA-containing promoters generated its own specific curve, consistent with a high burst size (Fig. 2B,D). Promoters with OPN architecture but without a TATA box, or promoters that had intermediate levels of nucleosome signal near the TSS, usually appeared on curves with intermediate noise levels (Fig. 2C). The difference in the estimated burst size of



**Figure 1.** Screen for mutations that affect noise. (A) Promoters used to generate the library. Shown is the noise of each nonmutated promoter as a function of its mean expression. The ratio of nucleosome occupancy close to the TSS and further upstream (Methods) is depicted in color code. Promoters with a low occupancy ratio are DPN type, whereas promoters with a high ratio are OPN. (Squares) Promoters with a conserved TATA box based on Basehoar et al. (2004). (B) Screen design: 400 bp upstream of the ATG of each yeast promoter were cloned, mutated using random PCR, and placed upstream of a yellow fluorescent protein. Expression and noise from single colonies were measured using flow cytometry. (C) Mean expression and noise for all variants in *ERG6* promoter library. (Error bars) Standard error of three biological repeats.



**Figure 2.** Scaling of noise with mean abundance. (A) Noise properties of NFR-containing, TATA-less promoters. Noise versus mean expression is shown for 11 promoters with a clear NFR and no TATA box. Different colors correspond to different promoters, as described. (B) Noise properties of TATA-containing promoters without NFR. Same as A for promoters that lack NFR and that contain a TATA box. (C) Noise properties of “intermediate” promoters. Same as A for promoters with intermediate nucleosome occupancy or occupied promoters with no TATA. In plots (A–C) the gray points in the background are pooled from all libraries. The values for mean and noise are provided as Supplemental Data. (D) Burst size for different promoters. The burst size was estimated by fitting the data to  $\eta^2 = b/m + \eta_{\text{ext}}^2$  (Methods), and normalized to the lowest burst size (RPT2). Promoters are ordered based on increase in burst size and colored based on nucleosome occupancy (OPN/DPN) and existence of TATA box.

the OPN versus DPN groups was significant ( $P = 0.02$ ;  $t$ -test) and resulted mostly from the high estimated burst size of the TATA-containing promoters.

### Mutations that deviate from the noise–mean scaling eliminate a TATA box or introduce an out-of-frame ATG

We searched for mutations that decouple the change in noise from that of mean expression. Such mutations are of particular interest because they may allow selection for high or low noise. For each promoter we calculated its best-fit curve (which in the majority of cases corresponded to  $\eta^2 = b/m + \eta_{\text{ext}}^2$ ), and examined all variants that deviated from the curve beyond a stringent threshold (see Methods). Only a small fraction of such variants were identified (Table 1), and most of them tended to reduce noise to a greater extent than predicted given their mean, implying a reduction in burst size (35 deviated downward vs. 10 upward, Table 1; see Supplemental Fig. S2A for other thresholds). Thus, random point mutations resulting in a large change in the noise versus mean relation are rare, and typically act to reduce the relative burst size.

Significant deviations from the scaling curve were found mostly in three promoters (*HXT2* in Fig. 3A, *GPD1* and *HSP150* in Supplemental Fig. S3). These three promoters are of the OPN type, contain a TATA box, and have the highest estimated burst size (Figs. 1A, 2D; Supplemental Table S1). To identify the mutations leading to the shift in the noise–mean relationship, we sequenced all of these variants. Notably, we find that they typically contain one of two mutation types: elimination of the TATA box or addition of an out-of-frame ATG start codon (Table 1). In the latter case, the upstream ATG probably reduces burst size via nonproductive translation, although some YFP is generated from the original ATG due to “slippage” in translation initiation or reinitiation (Meijer and Thomas 2002; Sachs and Geballe 2006). The upstream ATG may also lower transcript levels (Yun et al. 2012). The two types of mutations, TATA and upstream ATGs, also appeared in deviating variants from other promoters (Table 1).

We observed a few variants that deviated from the scaling curve although their sequence did not identify a consistent mutation pattern (a pattern was not identified possibly due to their low numbers, Table 1). Importantly, the magnitude of their deviation was generally lower than in the TATA and upstream ATG

**Table 1.** Mutants that deviated substantially from the scaling curve

Promoter	Type	TATA status <sup>a</sup>	No. of colonies in library	Direction of deviation	No. of deviating colonies	Percent of colonies that deviate	TATA mutants <sup>b</sup>	ATG generation <sup>c</sup>
<i>ADH3</i>	OPN	+	374	Up	2	0.5%	0	0
			374	Down	1	0.3%	0	1
<i>ARO2</i>	DPN	–	223	Up	2	0.9%	—	0
<i>ERG6</i>	OPN	–	334	Down	1	0.3%	—	0
<i>GPD1</i>	OPN	++	243	Down	8	3.3%	6	2
<i>HSP150</i>	OPN	++	339	Up	1	0.3%	0	0
			339	Down	10	3.0%	5	5
<i>HXT2</i>	OPN	++	315	Down	5	1.6%	5	0
<i>HXT3</i>	OPN	++	284	Down	2	0.7%	1	0
<i>IPP1</i>	DPN	–	256	Down	3	1.2%	—	1
<i>NSR1</i>	Interm	–	287	Down	3	1.0%	—	0
<i>OST1</i>	DPN	–	229	Up	1	0.4%	—	0
<i>PFY1</i>	DPN	–	201	Up	1	0.5%	—	0
<i>RNR2</i>	OPN	++	324	Up	1	0.3%	0	0
			324	Down	1	0.3%	0	1
<i>RPT2</i>	DPN	–	319	Up	2	0.6%	—	0

Promoters that are not shown did not exhibit substantial deviations.

<sup>a</sup>Existence of a TATA box: + has TATA, ++ has a conserved TATA.

<sup>b</sup>Colonies with mutation in a TATA box.

<sup>c</sup>Colonies with an upstream out-of-frame ATG.

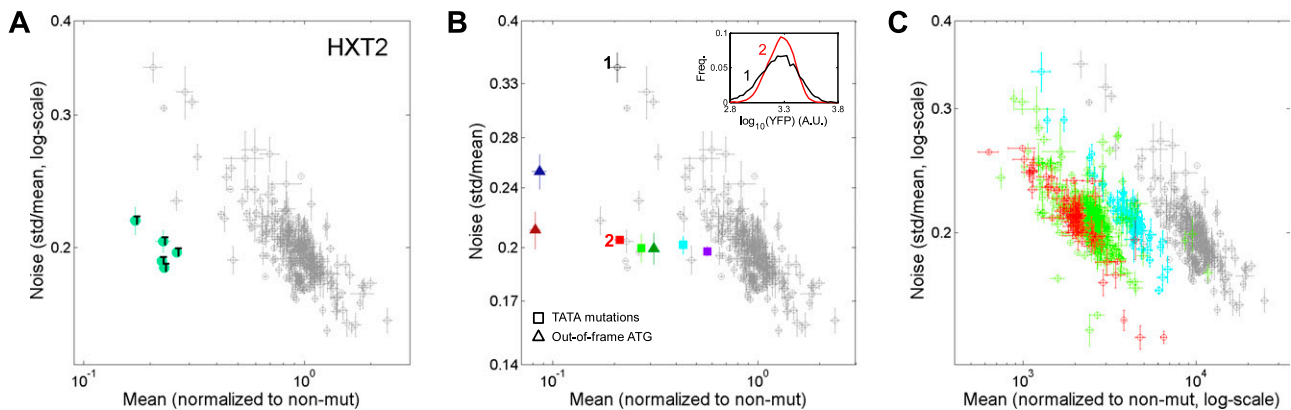
mutants of the high-noise promoters (Supplemental Fig. S2B). We conclude that most mutations that strongly modify the noise–mean scaling in our data set eliminated a TATA box or produced an out-of-frame ATG.

### The effect of the TATA box depends on promoter context

The observation that mutations in the TATA box reduce the estimated burst size (noise relative to mean expression) is consistent with previous studies (Raser and O’Shea 2004; Blake et al. 2006; Murphy et al. 2010). Yet, in addition to the three high-noise promoters where variants deviating from the noise–mean relationship were easily identifiable, our data set included additional promoters that contained a TATA sequence (a conserved TATA sequence in *HXT3*, *RNR2*, *ERG11*, and *HSC82*; unconserved TATA in *ADH3* and *DUT1*). In neither of those promoters did we identify a mutation

leading to a significant change in burst size. In order to detect more TATA mutations even if they do not have an effect on burst size, we sequenced >75 randomly selected colonies from 20 of our 22 libraries (details in Supplemental Table S2 and Supplemental Data; *ERG11* was not sequenced and *HSC82* is discussed in a separate section). Consistent with the above, for the three high-noise promoters *HXT2*, *GPD1*, and *HSP150*, nearly all TATA mutations significantly reduced burst size (9/10,  $P = 3 \times 10^{-11}$ ) (Table 2), with only one TATA box mutant retaining the scaling behavior (TATAAgAG in *HSP150*).

Interestingly, for the other promoters, mutations that eliminated the TATA did not lead to a significant reduction in burst size (Table 2). Therefore, in those promoters the TATA box is not likely to influence burst size. We conclude that the effect of the TATA box on the estimated burst size depends on the promoter context. A compelling hypothesis is that the TATA box interacts with nucle-



**Figure 3.** Mutations that deviate from the noise–mean scaling. (A) Noise versus mean expression of mutants of the *HXT2* library. (Green) Mutants that deviate from the noise–mean scaling and were sequenced. Those that had a mutated TATA are denoted by T. (B) Mutations that eliminate the TATA box or introduce an unproductive ATG reduce burst size. Mutations were directed into the *HXT2* promoter to eliminate the TATA box (squares: red, TATgAAAA; green, TATAgAAA; cyan, TATAAAcA; purple, TAcAAAA) or to introduce an out-of-frame ATG in different positions (triangles: red, 13; blue, 34; green, 23). (Inset) Single-cell distributions of expression from two mutants of similar expression but different noise (marked by 1 and 2 in the figure). (C) TATA box mutations modify the scaling law. The three TATA mutants shown in B were subjected to further random mutagenesis and the resulting mutants are shown. Color code as in B. (Error bars) Standard error of three biological repeats. Similar analysis is presented for *GPD1* and *HSP150* in Supplemental Figure S3.

**Table 2.** Behavior of TATA mutants and new ATG mutants discovered by extensive unbiased sequencing

Promoter	TATA status <sup>a</sup>	TATA mutants that deviated	ATG mutants that deviated	TATA mutant, not deviated	ATG mutant, not deviated	Total deviated <sup>b</sup>	Total not deviated <sup>b</sup>	P-value for TATA <sup>c</sup>	P-value for ATG <sup>d</sup>
<i>ADH3</i>	+	0	0	2	0	0	54	NS	NS
<i>DUT1</i>	+	0	0	4	0	0	64	NS	NS
<i>ERG6</i>	–	0	0	0	0	1	210	NS	NS
<i>GPD1</i>	++	4	1	0	0	5	119	$5.5 \times 10^{-7}$	$4.2 \times 10^{-2}$
<i>HSP150</i>	++	1	0	1	0	1	44	$4.3 \times 10^{-2}$	NS
<i>HXT2</i>	++	4	0	0	0	4	101	$2.1 \times 10^{-7}$	NS
<i>HXT3</i>	++	1	0	2	0	1	68	$4.2 \times 10^{-2}$	NS
<i>IPP1</i>	–	0	1	0	2	2	53	NS	NS
<i>NSR1</i>	–	0	0	0	1	0	52	NS	NS
<i>OST1</i>	–	0	0	0	2	0	164	NS	NS
<i>PFY1</i>	–	0	0	0	2	0	67	NS	NS
<i>RNR2</i>	++	0	1	7	2	1	262	NS	$1.1 \times 10^{-2}$
<i>RPT2</i>	–	0	0	0	2	0	71	NS	NS
<i>SAM4</i>	–	0	0	0	4	0	57	NS	NS
<i>VMA7</i>	–	0	0	0	2	0	75	NS	NS

<sup>a</sup>Existence of a TATA box: + has TATA, ++ has a conserved TATA.

<sup>b</sup>Out of colonies that were selected at random for sequencing and had at least one mutation.

<sup>c</sup>Fisher's Exact Test for enrichment of TATA mutations in colonies that deviated from the scaling curve.

<sup>d</sup>Fisher's Exact Test for enrichment of new ATG mutations in colonies that deviated from the scaling curve.

osomes, as the three high-noise promoters (*HXT2*, *GPD1*, and *HSP150*) in which TATA had a distinctive effect on the estimated burst size are all characterized by a high OPN measure, defined by the ratio of nucleosome signal proximal to the TSS and further upstream (Fig. 1A; Supplemental Table S1).

We repeated this analysis for mutants with a new out-of-frame ATG, but due to low numbers it was difficult to reach significant conclusions (Table 2). However, we do expect this type of mutation to be context-dependent (Meijer and Thomas 2002; Sachs and Geballe 2006).

### TATA mutations or out-of-frame ATGs are sufficient to reduce burst size in the high-noise promoters

To verify that the decrease in estimated burst size of the three high-noise promoters was due to the mutations in the TATA box and not to other mutations present in the same promoters, we mutated the TATA box consensus sequence at four different bases in the three high-noise promoters. We further generated out-of-frame ATG codons at different positions to verify the consequence of generating such out-of-frame translation start sites (Supplemental Table S4). As expected, all these mutants decreased expression. The noise levels of these mutants was lower than the curve formed by most other mutants, signifying a change in expression due to decreased burst size (*HXT2* in Fig. 3B; *GPD1* and *HSP150* in Supplemental Fig. S3). We next subjected the TATA-mutated promoters to further PCR-dependent random mutagenesis, generating three libraries of hundreds of mutated variants each. Notably, those variants all fell on a new noise–mean scaling curve that was lower (reduced *b*) compared with the wild-type curve (*HXT2* in Fig. 3C; *GPD1* and *HSP150* in Supplemental Fig. S3). These results suggest that the effect of TATA mutations in reducing burst size of the three high-noise promoters is independent of other mutations.

### Unique effect of TATA box mutations on the response to environmental stress

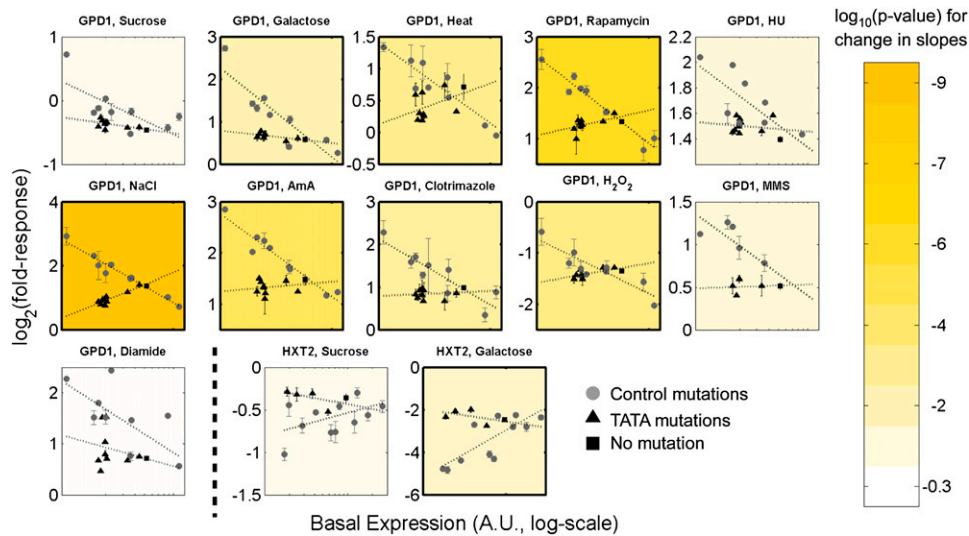
Our results suggest that maintaining high expression noise requires an intact TATA box. The TATA box was also implicated in high responsiveness to environmental perturbations (Basehoar

et al. 2004; Choi and Kim 2008, 2009; Tirosh and Barkai 2008; Lehner 2010). To examine whether eliminating the TATA box modulates the responsiveness to environmental changes, we measured the fold change in expression of different TATA mutant promoters following different environmental perturbations (Methods). As mutating the TATA box reduced the basal expression level, we selected, as control, mutants that changed expression to the same extent as the TATA mutations but were maintained on the scaling curve (thereby having a significantly higher noise level). The sequences of these promoter mutants are provided as Supplemental Data.

Figure 4 presents the fold response of *GPD1* and *HXT2* promoter mutants across different conditions (11 and two conditions, respectively). We observe a striking difference in the behavior of promoters with TATA box mutations versus those with control mutations. The fold response of TATA-mutated promoters was similar to that of the original promoter, regardless of the effect that the mutation had on basal expression level. In contrast, the response of the control mutants strongly increased with decreasing basal expression. We quantified the dependence between the (log–) fold response and the (log–) basal expression level through linear regression, and compared the resulting slopes of the TATA mutants and control mutants via ANCOVA (values for slopes are provided in Supplemental Table S5, and values for the fold response of each promoter are provided as Supplemental Data). We found a significant difference in the slopes between the TATA mutant group and the control, under most conditions. The combined *P*-values (Fisher's method) over all conditions for *GPD1* was  $<10^{-16}$  and 0.036 for *HXT2*. Thus, in the tested promoters, the effect of TATA mutations on the responsiveness of expression is distinct from that of other mutations.

Interestingly, the dependence of the fold response of promoter mutants on basal expression level, and the deviation of the TATA mutants from this behavior, resemble the scaling behavior of noise versus mean expression. Indeed, the average fold change over all conditions is correlated to the noise levels (Supplemental Fig. S4). This result, although for two promoters only, highlights another aspect of the relationship between noise and responsiveness (Kaneko 2007; Tirosh and Barkai 2008; Choi and Kim 2009; Lehner 2010).





**Figure 4.** The responsiveness of promoter mutants to environmental perturbations. The fold change in expression of *GPD1* and *HXT2* promoter mutants to different conditions, plotted against the basal expression level. (Black triangles) TATA mutants; (gray circles) control mutations; (black square) the original promoter (see Supplemental Table S4 and other Supplemental Data for details on mutants). (Dotted lines) Linear regression curves. The background of each figure is a color-coded scale for the significance of the difference between the regressed slopes of the TATA mutants and control mutants (based on ANCOVA). The figure is outlined if the  $P$ -value is  $<0.05$ . The conditions for *GPD1*, from left to right: sucrose, galactose, heat shock, rapamycin, hydroxyurea, NaCl, Antimycin A, Clotrimazole,  $H_2O_2$ , MMS, and diamide. For *HXT2*: sucrose and galactose (see Supplemental Methods for details on conditions). Note that for *HXT2* the observed response is a decrease in expression. Values for slopes are provided in Supplemental Table S5, and values for the fold response of each promoter are provided as Supplemental Data.

A previous study on synthetic promoters showed that the TATA contributes to a constant fold change in expression, regardless of *cis*-regulatory site identity and arrangement (Mogno et al. 2010). Because changes in environmental conditions are probably reflected in the binding of transcription factors to the promoter, our results, together with the previous study, support a unique role for TATA in defining gene expression level, independently of other interactions on the promoter.

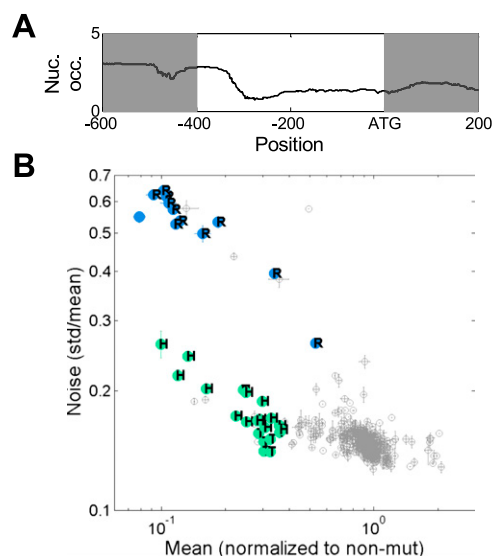
#### *HSC82* promoter: Two branches in the noise–mean plane

The *HSC82* promoter presents an interesting exception to the general scaling behavior we described. This promoter is unique as it has a TATA box, characteristic of high-noise promoters, although its nucleosome arrangement is of intermediate nature and closer to that of the DPN architecture that characterizes low-noise promoters (Fig. 5A; Erkin et al. 1995a,b, 1996). The combination of a DPN promoter with a conserved TATA box is rare, observed in only  $\sim 30$  yeast promoters ( $\sim 0.5\%$ ). The depletion of nucleosomes in the *HSC82* promoter likely results from binding of Reb1, a nucleosome-evicting factor (Erkin et al. 1995a,b, 1996). Expression of the wild-type *HSC82* promoter was high and of a rather low noise (Fig. 1A; Supplemental Table S1). Most *HSC82* promoter variants had only a minor effect on the mean abundance or noise level. Yet, a small fraction of mutations strongly reduced expression, forming two branches on the mean–noise plane: one of low noise and one of high noise (Fig. 5B).

Sequencing revealed that all mutations in the low-noise branch targeted either the TATA box or the binding site for HSF1, a transcription factor known to regulate *HSC82*. In contrast, all mutations in the high-noise branch targeted the Reb1 binding site. Thus, Reb1 binding to the *HSC82* promoter contributes to expression levels but reduces expression noise.

## Discussion

Gene expression is encoded in the gene promoter sequence. Understanding how promoter sequence mutations influence gene expression is therefore of obvious interest. Most studies, however, characterize promoter mutations by measuring mean expression



**Figure 5.** Unique behavior of the *HSC82* promoter mutants. (A) Nucleosome occupancy profile of *HSC82* promoter: note the pronounced NFR. (B) Noise versus mean expression of mutants of the *HSC82* library. Mutants form two branches on the noise–mean plane. Mutants from the upper and lower branch were sequenced (blue and green, respectively). (R) Mutation in Reb1 binding; (T) mutation in the TATA box; (H) mutation in the HSF1 binding site. (Error bars) Standard error of three biological repeats.

only (e.g., Yun et al. 2012), and therefore do not distinguish between mutations affecting burst frequency versus those that modulate burst size. In this study, we relied on the theoretical idea of using gene expression noise to distinguish between those two processes in order to examine the sensitivity of burst size and burst frequency to mutations in promoter sequence. Our main finding is that each promoter has a characteristic burst size that is largely insensitive to most promoter mutations. This burst size is small in DPN-type promoters and high in OPN-type, TATA-containing promoters. In the latter promoter class, but not in the former, high burst size was strongly dependent on a functional TATA box.

Our results were consistent amongst 21 of the 22 yeast promoters tested. Those promoters were chosen to span a wide range of expression values and noise levels. Their promoter sequences also differed in terms of both nucleosome organizations (DPN vs. OPN types) and the presence of a TATA box. The associated gene functions were diverse, with no significant enrichment to any particular functional category. Therefore, the scaling behavior we observe is likely to be a general property of yeast promoters.

Interestingly, all mutations causing large changes in burst size were mapped to just two types of mutations: elimination of a TATA box or out-of-frame ATG sequences. Hence, the presence or absence of a TATA box plays a major role in defining burst size, consistent with previous reports (Raser and O’Shea 2004; Blake et al. 2006; Murphy et al. 2010). Yet, having a TATA box is not necessary and sufficient for high burst size, because high burst size promoters can be TATA-less, and TATA-containing promoters can still have low burst size (see also ERG6 and ERG11 in our data set). Further, mutations in the TATA box did not reduce burst size in promoters of the DPN architecture. Together, we consider that burst size is determined by the interaction between TATA box and other properties such as the pattern of nucleosome occupancy.

Our study focused on mutations in promoter sequence. Such *cis*-mutations play an important role in the evolution of gene expression between related species. Still, they represent only one class of all possible mutations. It would therefore be interesting to extend the analysis to *trans*-mutations (Raser and O’Shea 2004; Rinott et al. 2011), as well as consider other mutational mechanisms such as recombination.

Despite this limitation, it is tempting to speculate on the implications of our study for the evolution of gene expression noise. It is widely assumed that noise in gene expression is the result of purifying evolutionary (Batada and Hurst 2007; Lehner 2008; Wang and Zhang 2011) or positive selection (Kussell and Leibler 2005; Blake et al. 2006; Acar et al. 2008; Çağatay et al. 2009; Zhang et al. 2009). In our data set that included a large number of mutations, we observed a clear bias: Very few mutations increased burst size (higher noise beyond the expected dependency on the mean), compared with mutations that decreased it. This may suggest that low burst size promoters are stable against random *cis*-mutations and that selection is required to protect high burst size promoters (specifically TATA-containing promoters) from becoming low burst size ones. That selection indeed acts to maintain the TATA box in promoters is best indicated by the conservation of TATA box between species (Basehoar et al. 2004).

Clearly, such selection may not act directly on noise, but could be a byproduct of selection for other features (Kaneko 2007; Lehner 2010; Tawfik 2010). For example, as suggested by our results, the TATA box may have a role in increasing expression level while maintaining the fold response to perturbations. Further-

more, selection might act to remove alternative ATGs because they reduce translational efficiency and produce undesirable polypeptides. Nevertheless, it remains an intriguing possibility that, at least under certain circumstances, selection for high noise may act directly to shape the structure of gene promoters.

## Methods

### Generation of mutant libraries

Promoter fragments (400 bp from ATG) were cloned from the yeast genome and subjected to random mutagenesis using the GeneMorph II Random Mutagenesis kit (Stratagene). PCR fragments were cloned into a specifically designed plasmid. The plasmid contained a selection marker and sequences that are homologous to genomic regions in a yeast strain with an integrated YFP. Plasmid was linearized and transformed into the yeast. Single colonies were collected and kept for further analysis. Details of plasmids, yeast strains, PCR, and cloning procedure are found in the Supplemental Material.

### Measurements of expression and noise using flow cytometry

Full details of measurement procedure are given in the Supplemental Material. Briefly, yeast libraries were replicated into 150  $\mu$ L YPD in U-bottom 96-well plates and grown overnight at 30°C. Saturated cultures were resuspended and diluted  $\sim$ 1:150 into prewarmed 96-well plates with 130–150  $\mu$ L SC-HIS media. Diluted cells were grown for  $\sim$ 5.5–7 h, at 30°C with vigorous shaking, and fluorescence was measured by flow cytometry on the BD LSRII system (BD Biosciences) with a High Throughput Sampler extension (HTS). Excitation wavelength was 488 nm and emission was collected via 525/50 filter.

Expression and noise were calculated based on the measured fluorescence distributions. To minimize effects of cell size and cell cycle on the measured noise, we used two procedures: First we took for our analysis only a subpopulation of cells which had low forward and side scatter values, and is enriched in G1 cells. Second, we used linear regression to correct the variation in fluorescence for effects of cell size. Exact details of analysis procedures and other normalization procedures are provided in the Supplemental Material. Raw results are provided as Supplemental Data.

### Fitting a curve to the noise–mean data and defining deviating mutants

We fitted the following three models to the noise versus mean curves of our promoters:

$$\eta^2 = \frac{b}{m} \quad (1)$$

$$\eta^2 = \eta_{\text{ext}}^2 + \frac{b}{m} \quad (2)$$

$$\eta = \alpha \cdot m^{-\beta} \quad (3)$$

The fitting was performed using iteratively reweighted least squares with a bisquare weighting function, implemented in Matlab’s `robustfit` function with default parameters. In order to avoid fitting artifacts from points that do not change expression level, values of mean expression with  $|\log_2(m/m_{\text{wt}})| < 0.35$  were averaged into 200 equally spaced bins.

The results of the fit are presented in Supplemental Table S3 and Supplemental Figure S1. Models number 2 and 3 outperformed the simple model 1. In two cases, *HXT3* and *PFY1*, model 3 was better than model 2 ( $P$ -value < 0.01 based on F-test for the sum of squared residuals; Jaqaman and Danuser 2006).

To define the colonies that deviate from the curve we used the best-fit curve (model 2 or 3) and calculated the log ratio of the predicted noise based on the curve and the measured noise. Table 1 presents colonies that deviated from an abs(log-ratio) of 0.263, corresponding to a statistically significant deviation with a  $P$ -value of 0.05 after a Bonferroni correction. Different thresholds are presented in Supplemental Figure S2.

### Testing high-noise variants using colony isolation

We ascertained that yeast colonies which display higher-than-expected noise levels are not in fact composed of a mixture of two colonies. Such yeast colonies were streaked and single colonies were collected. The fluorescence of the single colonies was measured together with the original colony, to see if there are changes in the expression level and/or decrease in noise.

### Conditions used to perturb gene expression

The conditions used to perturb gene expression were rapamycin, sodium chloride, hydrogen peroxide, antimycin A, MMS, diamide, hydroxy-urea, clotrimazole, 37°C heat shock, and two carbon sources: galactose or sucrose. Exact concentrations and procedures are given in the Supplemental Material.

### Nucleosome occupancy ratio

The nucleosome occupancy ratio is the nucleosome signal proximal to the transcription start site relative to signal further upstream. Values shown here are mean from two studies (Tirosch and Barkai 2008; Tsui et al. 2011) that were based on different nucleosome occupancy data and slightly different definitions for proximity. Each data set was normalized to have mean 0 and standard deviation 1 before averaging.

### Acknowledgments

We thank Einat Dayan for technical assistance. We thank the lab of Eran Segal for yeast strain and protocols. Furthermore, we thank Itay Tirosch, Ilya Soifer, and all of our lab members for helpful discussions. G.H would like to express his gratitude to N. Broudo-Hornung for support and understanding. This work was supported by the NIH (P50GM068763), the ERC, and the Hellen and Martin Kimmel Award for Innovative Investigations.

### References

Acar M, Mettetal JT, van Oudenaarden A. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40**: 471–475.

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

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709.

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* **39**: 945–949.

Blake WJ, Balázs G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* **24**: 853–865.

Çağatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139**: 512–522.

Choi JK, Kim Y-J. 2008. Epigenetic regulation and the variability of gene expression. *Nat Genet* **40**: 141–147.

Choi JK, Kim Y-J. 2009. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat Genet* **41**: 498–503.

Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297**: 1183–1186.

Erkine AM, Adams CC, Gao M, Gross DS. 1995a. Multiple protein-DNA interactions over the yeast *HSC82* heat shock gene promoter. *Nucleic Acids Res* **23**: 1822–1829.

Erkine AM, Szent-Gyorgyi C, Simmons SF, Gross DS. 1995b. The upstream distribution of the *HSP82* and *HSC82* genes of *Saccharomyces cerevisiae*: Regulatory elements and nucleosome positioning motifs. *Yeast* **11**: 573–580.

Erkine AM, Adams CC, Diken T, Gross DS. 1996. Heat shock factor gains access to the yeast *HSC82* promoter independently of other sequence-specific factors and antagonizes nucleosomal repression of basal and induced transcription. *Mol Cell Biol* **16**: 7004–7017.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216. doi: 10.1371/journal.pcbi.1000216.

Friedman N, Cai L, Xie XS. 2006. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* **97**: 168302. doi: 10.1103/PhysRevLett.97.168302.

Jaqaman K, Danuser G. 2006. Linking data to models: Data regression. *Nat Rev Mol Cell Biol* **7**: 813–819.

Kaneko K. 2007. Evolution of robustness to noise and mutation in gene expression dynamics. *PLoS ONE* **2**: e434. doi: 10.1371/journal.pone.0000434.

Kussell E, Leibler S. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**: 2075–2078.

Lehner B. 2008. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* **4**: doi: 10.1038/msb.2008.11.

Lehner B. 2010. Conflict between noise and plasticity in yeast. *PLoS Genet* **6**: e1001185. doi: 10.1371/journal.pgen.1001185.

Maheshri N, O'Shea EK. 2007. Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* **36**: 413–434.

Meijer HA, Thomas AAM. 2002. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem J* **367**: 1–11.

Mogno I, Vallania F, Mitra RD, Cohen BA. 2010. TATA is a modular component of synthetic promoters. *Genome Res* **20**: 1391–1397.

Murphy KF, Adams RM, Wang X, Balázs G, Collins JJ. 2010. Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res* **38**: 2712–2726.

Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**: 840–846.

Paulsson J. 2004. Summing up the noise in gene networks. *Nature* **427**: 415–418.

Paulsson J. 2005. Models of stochastic gene expression. *Phys Life Rev* **2**: 157–175.

Pedraza JM, Paulsson J. 2008. Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**: 339–343.

Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135**: 216–226.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**: e309. doi: 10.1371/journal.pbio.0040309.

Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **304**: 1811–1814.

Rinott R, Jaimovich A, Friedman N. 2011. Exploring transcription regulation through cell-to-cell variability. *Proc Natl Acad Sci* **108**: 6329–6334.

Sachs MS, Geballe AP. 2006. Downstream control of upstream open reading frames. *Genes Dev* **20**: 915–921.

So L, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet* **43**: 554–560.

Tan RZ, van Oudenaarden A. 2010. Transcript counting in single cells reveals dynamics of rDNA transcription. *Mol Syst Biol* **6**: 358. doi: 10.1038/msb.2010.14.



- Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**: 533–538.
- Tawfik DS. 2010. Messy biology and the origins of evolutionary innovations. *Nat Chem Biol* **6**: 692–696.
- Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* **18**: 1084–1091.
- Tsui K, Dubuis S, Gebbia M, Morse RH, Barkai N, Tirosh I, Nislow C. 2011. Evolution of nucleosome occupancy: Conservation of global properties and divergence of gene-specific patterns. *Mol Cell Biol* **31**: 4348–4355.
- Wang Z, Zhang J. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci* **108**: E67–E76.
- Yun Y, Adesanya TMA, Mitra RD. 2012. A systematic study of gene expression variation at single-nucleotide resolution reveals widespread regulatory roles for uAUGs. *Genome Res* **22**: 1089–1097.
- Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**: 1263–1271.
- Zhang Z, Qian W, Zhang J. 2009. Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol* **5**: 299. doi: 10.1038/msb.2009.58.

*Received February 19, 2012; accepted in revised form July 5, 2012.*