nature genetics

Noise in protein expression scales with natural protein abundance

Arren Bar-Even[1], Johan Paulsson[2,3], Narendra Maheshri[4], Miri Carmi[1], Erin O'Shea[4], Yitzhak Pilpel[1] & Naama Barkai[1,5]

Noise in gene expression is generated at multiple levels, such as transcription and translation, chromatin remodeling and pathway-specific regulation. Studies of individual promoters have suggested different dominating noise sources, raising the question of whether a general trend exists across a large number of genes and conditions. We examined the variation in the expression levels of 43 *Saccharomyces cerevisiae* proteins, in cells grown under 11 experimental conditions. For all classes of genes and under all conditions, the expression variance was approximately proportional to the mean; the same scaling was observed at steady state and during the transient responses to the perturbations. Theoretical analysis suggests that this scaling behavior reflects variability in mRNA copy number, resulting from random 'birth and death' of mRNA molecules or from promoter fluctuations. Deviation of coexpressed genes from this general trend, including high noise in stress-related genes and low noise in proteasomal genes, may indicate fluctuations in pathway-specific regulators or a differential activation pattern of the underlying gene promoters.

Genetically identical cells growing under the same conditions can still vary greatly in their internal protein concentrations[1–14]. This inherent variability poses particular challenges for information processing in cells, which often need to generate precise and reliable computation in randomly changing environments. Thus, understanding the factors that affect noise in protein abundance, and their relative contributions, is of interest.

Several methods have been used to track the origin of fluctuations in protein levels. One strategy is to measure the variance in protein abundance as a function of transcription and translation rates[15–19]. Fluctuations that are due to random births and deaths of individual molecules show a distinct scaling behavior in which the variance is proportional to the mean (an effect loosely connected to the law of large numbers). This property can be used to identify which chemical species contribute to the fluctuations. For example, protein noise that scales this way in response to changes in both transcription and translation rates may result from randomness in the birth and death of proteins. In contrast, protein noise that shows this scaling in response to changes in the rate of transcription, but not in the rate of translation, is likely to reflect random birth and death of individual mRNAs[7,15,16]. The same principle can be used to study fluctuations that result from noise in an upstream component and propagate through the reaction network to increase the variance in protein levels. Applying this method to *Bacillus subtilis* suggested that most noise came from low-copy mRNA fluctuations[7,15,16]. Investigations in *S. cerevisiae* suggested instead that the dominant noise sources are pre-transcriptional and relate to low-abundance regulators[19] or a combination of slow promoter kinetics and pathway-specific factors[20].

Another strategy is to measure the simultaneous expression of two reporter genes[17,18,21–24]. The idea is that each of the two proteins has its own set of genes and mRNAs but shares global and pathway-specific factors with the other protein. This elegant method makes interpretation less model dependent: as long as the two reporters are identically regulated and do not affect each other, the noise contribution from the shared environment should equal the covariance between the reporters. An application to *Escherichia coli* showed both a noise that was specific to each reporter (termed intrinsic) and a shared noise (termed extrinsic), which was largely explained by a pathway-specific repressor[21]. Other studies in the same organism suggested that global factors can contribute significantly to the overall noise and emphasized the contribution of cell-cycle asynchrony to the observed variations[17,23]. Finally, a similar study in *S. cerevisiae* suggested that much of the noise originates from global factors rather than from pathway- or gene-specific factors[18,22].

By nature of the analysis, most studies have focused on a limited number of genes, many of which are highly expressed. This is appropriate when analyzing any particular mechanism in detail, but because the results have varied from study to study and from

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Systems Biology, Harvard University, Boston, Massachusetts 02115, USA. [3]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK. [4]Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [5]Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. Correspondence should be addressed to N.B. (naama.barkai@weizmann.ac.il) or Y.P. (pilpel@weizmann.ac.il).

Received 5 December 2005; accepted 25 April 2006; published online 21 May 2006; doi:10.1038/ng1807

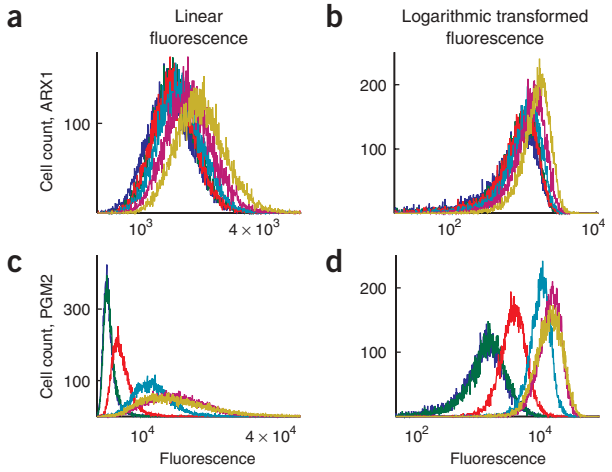

Figure 1 Single-cell distributions of fluorescence levels. (**a**,**b**) Cells expressing the high-abundance protein PGM2 were shifted from synthetic complete medium to medium containing 3% ethanol. (**c**,**d**) Cells expressing the low-abundance protein ARX1 were diluted from stationary phase. The cells were subjected to flow cytometry analysis at different time points after the transfer. Fluorescence distributions are shown on linear (**a**,**c**) and on logarithmic (**b**,**d**) scales. Blue, green, red, turquoise, magenta and yellow lines correspond to fluorescence distributions after 0, 30, 60, 90, 120 and 150 min from perturbation start, respectively. For low-abundance proteins, the fluorescence values appeared to follow normal distributions. By contrast, at high abundances, we more often observed a deviation from normality, with overrepresentation of high fluorescence values (see **Supplementary Note** and **Supplementary Fig. 6** online).

condition to condition, the generalizability of the conclusions remains to be seen. Moreover, most studies have been done under optimal growth conditions but do not consider how environmental conditions affect the noise. Comparing noise levels between conditions may be of particular interest because it may explain the enhanced phenotypic diversity often observed when cells are stressed.

In this study, we measured the cell-to-cell variation of 43 *S. cerevisiae* genes expressed as fusion proteins in this organism and grown under 11 different environmental conditions. Motivated by previous analyses, we attempted to correlate noise and mean expression levels. However, whereas previous studies have changed expression levels by explicitly tuning transcription and translation rates, we have explored the situations in which the levels of expression have been differently determined by the cells themselves. Differences in protein abundance between the proteins, and across different conditions, potentially could be achieved in numerous different ways, with very different effects on the noise. Despite this potential complexity, we found a strong correlation between cell-to-cell variability and mean expression level. We were struck that, in all gene classes examined and over a broad range of expression levels, the variance is roughly proportional to the mean. We suggest how this effect can be explained kinetically and how it points toward specific biological mechanisms both for setting averages and for generating fluctuations. By analyzing the deviation from this general trend, we find that genes of similar biological functions show similar noise features, indicating a contribution from pathway-specific components or mechanisms. In particular, most of the stress genes examined are significantly noisier than other genes tested, whereas components of the proteasome are the least noisy.

RESULTS

Noise versus mean protein abundance

We considered 43 proteins associated with four coexpressed transcription modules (stress, proteasome, ergosterol and rRNA processing; see Methods). We subjected cells expressing a GFP-fused version of each protein[25] to 11 different environmental conditions, and we used flow cytometry analysis to measure the single-cell fluorescence distributions at six subsequent time points ranging from 0–150 min after the transfer to each of the environments (see Methods and **Fig. 1**).

We quantified the noise by the (squared) coefficient of variation $\eta_p^2 = \sigma_p^2/\langle p\rangle^2$; that is, the variance σ_p^2 of protein abundance normalized by the square mean $\langle p\rangle^2$. Fluorescence values were normalized to actual protein numbers using known protein abundances[26] (Methods). **Figure 2** shows the noise as a function of mean abundance for all genes, under all conditions and at all time points.

At low abundances, the GFP concentration is barely detectable over the autofluorescence background. In this region, the noise seems to decrease as $1/\langle p\rangle^2$, as expected for a constant (autofluorescence) variance. At very high abundance, the noise is almost uncorrelated with the mean, reaching a minimum consistent with that reported previously[27] (**Supplementary Methods** online).

Perhaps the most notable behavior is observed at intermediate abundances, spanning about an order of magnitude and including most data points. In this region, the noise is strongly correlated with the mean, following $\eta_p^2 = C/\langle p\rangle$, with $C \approx 1{,}200$ as a proportionality constant. Protein molecules that are made and degraded with constant probabilities per second are expected to show a similar dependency, with $\eta_p^2 = 1/\langle p\rangle$. However, the high proportionality factor observed in our experiment rules out poissonian statistics: a protein that is present at 100,000 copies in our data set shows relative fluctuations of 10%, compared with 0.1% predicted by poissonian statistics.

To better understand the origin of the scaling behavior, we applied a theoretical analysis of noise propagation in reaction networks (**Fig. 3**). Models of noise propagation

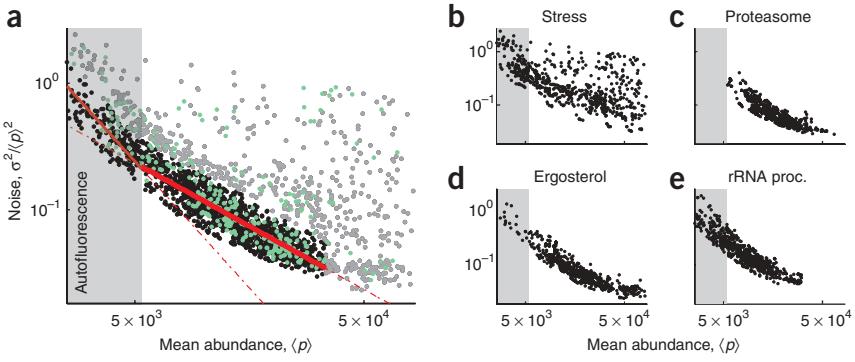

Figure 2 Scaling of noise with mean protein abundance. (**a**) Noise as a function of mean protein abundance. All genes in all conditions and time points are shown. Thick curve corresponds to $\log(\eta_p^2) = 1175 - \log(\langle p\rangle)$. Green filled circles represent initial, steady-state time points of stress perturbations ($t = 0$). Gray points were excluded from the fitting process (see Methods). The fit to the autofluorescence region (thin curve) corresponds to $\log(\eta_p^2) = 9.9 \cdot 10^5 - 2 \cdot \log(\langle p\rangle)$. (**b**–**e**) Noise versus mean protein abundance shown separately for genes belonging to a common module.

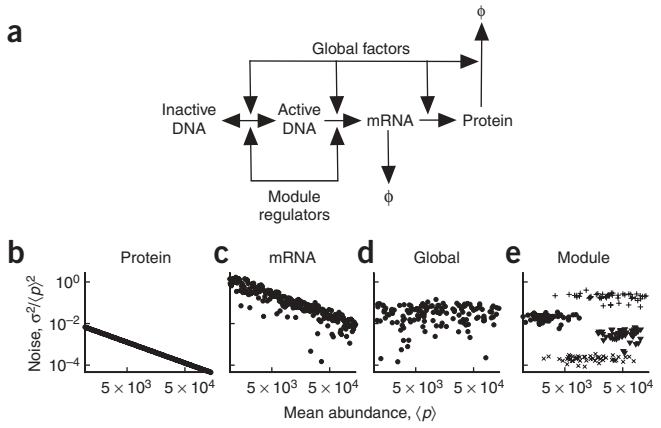

Figure 3 Theoretical analysis of noise propagation. (**a**) A schematic illustration of different upstream factors, creating fluctuations in the protein abundance. (**b**–**e**) Cartoon illustration of expected dependency of noise on mean protein abundance for different noise sources as shown (see Methods for simulation details). (**b**) Random synthesis and degradation of protein molecules. (**c**) Random synthesis and degradation of mRNA molecules. (**d**) Fluctuations in global factors. (**e**) Pathway-specific noise. Here we assumed that decay and synthesis rates are independent of $\langle p \rangle$.

typically assume a system that is in steady state, whereas in our case, most of the data reflects the behavior of the system after some perturbation. Notably, however, we obtained essentially the same scaling when considering all points and when confining the analysis to only the steady-state conditions (**Fig. 2a**; Methods). Therefore, we focus the theoretical analysis on steady-state fluctuations. Transient behavior is discussed below.

Is noise driven by global or pathway-specific factors?

We base our analysis on a general phenomenological model, described in Methods, that captures the essence of noise propagation. According to this model (see equation (3)), the protein noise that comes from the upstream component is determined by three factors: the noise in the upstream component itself (η_x^2), the steady-state susceptibility to upstream noise (H parameter) and time-averaging of the upstream noise (H and τ parameters).

In our study, we measured the expression variance for different genes under the same conditions. When x represents a global factor, its variance η_x^2 is the same for all cell cultures measured under a particular condition (**Fig. 3d**). Similarly, when x is specific to a given pathway (for example, a pathway-specific transcription factor), its variance η_x^2 will be the same for all pathway genes (**Fig. 3e**). By contrast, to explain the experimentally observed scaling (**Fig. 2a**), the magnitude of the upstream noise η_x^2 would have to decrease in proportion to the mean abundance of the measured genes. Moreover, as we also observed the same $\eta_p^2 = C/\langle p \rangle$ dependency within distinct groups of coexpressed genes (modules, **Fig. 2b–e**), the same argument rules out a changing pathway-specific noise. Notably, when we also considered the contribution of the susceptibility or the averaging times, this did not seem to alter this conclusion (Methods).

Taken together, these results suggest that noise in global or pathway-specific factors may plausibly explain the region of high abundance, in which the noise level is practically independent of the mean abundance (**Fig. 2**). In contrast, neither noise type accounts well for the large region of intermediate abundance, in which noise is inversely proportional to the mean protein abundance.

Table 1 Measured noise in the dual-reporter assay

		ACS2	PWP1	TPS1	PRE9
Microscope	Mean fluorescence (AU/cell)	5.3×10^5	2.2×10^5	4.9×10^5	2.4×10^5
	Total noise	0.065	0.070	0.061	0.095
	Intrinsic noise	0.015	0.026	0.025	0.041
	Extrinsic noise	0.050	0.044	0.036	0.054
Flow cytometry	Mean abundance (protein/cell)	5.7×10^5	1.4×10^5	2.6×10^5	1.9×10^5
	Total noise	0.036	0.088	0.107	0.054

Flow cytometry values correspond to the data in **Figure 2a** for GFP fusion proteins in haploid cells. See text for details. AU, arbitrary units.

Is noise driven by mRNA fluctuations?

Early theoretical analyses of stochastic gene expression suggested that protein fluctuations are driven by underlying mRNA fluctuations[28,29]. In contrast to global factors, mRNA levels vary from gene to gene, making them possible candidates for the observed effects. Simple analysis (see Methods) predicts that the contribution of spontaneous mRNA fluctuations to the protein noise will scale as

$$\eta_p^2 = C/\langle p \rangle \qquad (1)$$

where $C \approx \lambda_p \tau_{\mathrm{mRNA}}$ is the average number of proteins made per transcript, λ_p is the translation rate and τ_{mRNA} is the mRNA average lifetime. Equation (1) thus predicts a scaling behavior between noise and mean protein abundance (**Fig. 3c**). To be consistent with the experimental relationship observed in our measurements, two conditions need to be satisfied: first, the average number of proteins made per transcript (C) needs to be similar for all genes and conditions in our data. Second, the differences in protein abundances observed in our data needs to be primarily due to changes in transcription levels rather than in post-transcriptional processes such as translation or protein degradation.

To estimate whether the observed level of C ($\sim$1,200) is consistent with biological parameters, we used genome-wide measurements of protein and mRNA abundances and mRNA degradation rates (Methods). The experimental errors for individual genes are too large to test whether C is well approximated as a constant. However, the average value over the 43 genes gives $C \approx 1{,}300$, markedly similar to the measured coefficient (Methods and **Supplementary Methods**). We note that because protein lifetimes are not known for most genes, we assumed that the effective protein degradation is dominated by dilution, with an effective average lifetime of 100 min. If the protein degradation rates instead were four times higher, with effective lifetimes of about 25 min, one would obtain $C \approx 220$. This would be significantly lower than the observed proportionality constant, but still in the right region. Considering all these results together, it seems that low-copy mRNA fluctuations, coupled with differential transcription rates as a major source of protein copy number variation, could explain the scaling behavior observed in our data.

Is noise driven by promoter fluctuations?

Random activation and inactivation of the gene promoter, resulting from changes in chromatin structure or from the stochastic binding and unbinding of transcription factors, may also contribute to protein noise. To examine whether this process can account for the observed

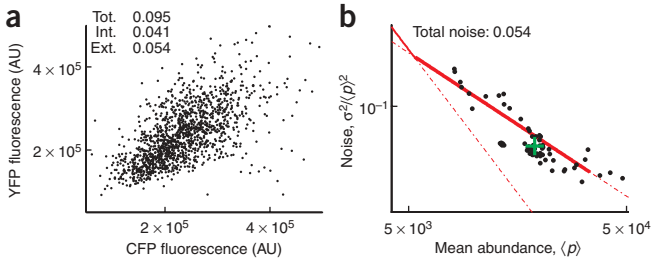

Figure 4 Noise pattern of PRE9. The noise patterns of PRE9 in the dual-reporter microscope assay (**a**) and in the GFP flow cytometry experiments (**b**). (**a**) Correlation between the CFP and the YFP fluorescence. The calculated total, intrinsic and extrinsic noises are at the top of the plot. (**b**) Red lines correspond to the trend lines shown in **Figure 2a**, black dots correspond to the measured mean abundance and noise of PRE9 at all the time points of all the perturbations, and the green plus signs represent the average mean abundance and noise (at the top of the plot) of the steady-state time points.

experimental trend, we consider the following model[30]: assume that each gene switches on and off with some constant probabilities, and that a protein is made (with constant probability) when the gene is in the 'on' state. The resulting protein noise can be calculated directly from equation (3) (**Supplementary Note** online). In the general case, no clear correlation between noise and mean abundance is predicted. However, the observed noise trend is retrieved in the special case in which, first, genes switch on and off rapidly relative to the protein's lifetime, and, second, each gene spends most of its time in the inactive state. The protein noise that comes from operator fluctuations is then as follows:

$$\eta_p^2 = \frac{\lambda_E/b}{\langle p \rangle} \qquad (2)$$

where the ratio λ_E/b corresponds to the average number of proteins made in the time window in which the gene is turned on.

Thus, if the genes are mostly off, and the average level of gene activity is modulated by tuning the 'on rate', one would obtain a dependency of the noise on the mean protein abundance as observed in our data. To be consistent with our data, two additional conditions need to be satisfied: first, the average number of proteins made per instance of gene activation (λ_E / b) needs to be similar for all genes and conditions in our data. Second, the differences in protein abundances observed in our data needs to be primarily due to tuning the rate of gene activation, rather then to downstream processes. Finally, we note that achieving $\lambda_E/b \approx 1{,}200$ is biologically conceivable, but the mechanisms of gene activation are not yet characterized in sufficient detail to provide a quantitative estimate.

Intrinsic noise measured in two-color experiment

Our analysis above concluded that the scaling behavior between the noise and mean abundance observed in our data (**Fig. 2a**) reflects stochastic fluctuations in mRNA levels. This result suggest that gene-specific fluctuation in mRNA or promoter activation adds a substantial, if not dominant, contribution to protein variability. In contrast, recent studies in *S. cerevisiae* have shown that promoter-specific fluctuations (intrinsic noise) are negligible compared with noise that is generated by shared factors (extrinsic noise)[18,22]. As most previous experiments have considered highly abundant reporter genes, we wished to directly examine the relative extent of intrinsic noise for the fusion proteins analyzed in our study.

We created dual-reporter diploid strains by fusing one copy to cyan fluorescent protein (CFP) and the other copy to yellow fluorescent protein (YFP). Since CFP is hard to visualize, we focused on four proteins of relatively high abundance (**Table 1**). Three of these proteins (PWP1, TPS1 and PRE9) were associated with the intermediate region, where scaling was observed, whereas the fourth (ACS2) corresponded to the region of high abundance, where no dependence between noise and mean abundance was observed (**Fig. 4** and **Supplementary Fig. 1** online). Following the experimental model introduced in ref. 21, we characterized both the intrinsic and the extrinsic contributions to protein noise. We used linear regression to correct for cell-size effect (which were gated in our flow cytometry experiment; see **Supplementary Methods**). After this correction, the total noise observed in the microscope was comparable to that observed by flow cytometry (**Table 1**). Notably, for PWP1, TPS1 and PRE9, proteins of intermediate abundance, the extent of intrinsic noise was of the same order as the extrinsic noise, whereas the highly abundant ACS2 was characterized by extrinsic noise that was more than three times higher than the intrinsic noise. Taken together, these results suggest that for proteins of intermediate abundance, internal noise contributes substantially to the total protein noise.

Scaling behavior during transient response to perturbations

The theoretical discussion above was limited to steady-state behavior. However, we observed essentially the same scaling relationship for the transient time points after the start of perturbations (**Fig. 2a** and Methods). Notably, most conditions that substantially changed the doubling time also altered the observed pre-factor C, as expected if protein degradation time was indeed controlled by dilution (**Supplementary Methods** and **Supplementary Fig. 2** online). To verify that the scaling reflected an actual response to perturbation and was not due to the basal differences of gene expression levels, we normalized both the mean and the variance with respect to their pre-stimulus values. Indeed, these normalized variables were also proportional (**Fig. 5**).

The transient response to perturbation has been studied theoretically within a single-gene model[15]. Similar to the steady-state discussion above, this model predicts a scaling behavior between the variance and the mean, provided that the protein noise is due to mRNA fluctuations and that the change in protein abundance after the perturbation is due to changes in mRNA levels. Our data agree qualitatively with this expected scaling, suggesting that changes in protein abundances are achieved primarily through modulation of mRNA levels and that intrinsic fluctuations in mRNA levels add a substantial contribution

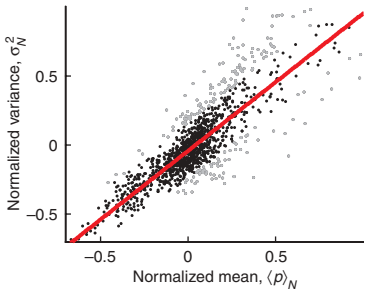

Figure 5 Transient response to perturbations. Normalized variance ($\sigma_N^2 = (\sigma_t^2 - \sigma_0^2)/\sigma_0^2$) versus normalized mean ($\langle p \rangle_N = (\langle p \rangle_t - \langle p \rangle_0)/\langle p \rangle_0$) of all genes, under all perturbations and at all time points. Gray points were excluded from the fitting process (see Methods). Curve corresponds to $\sigma_N^2 = 0.99\langle p \rangle_N - 0.04$.

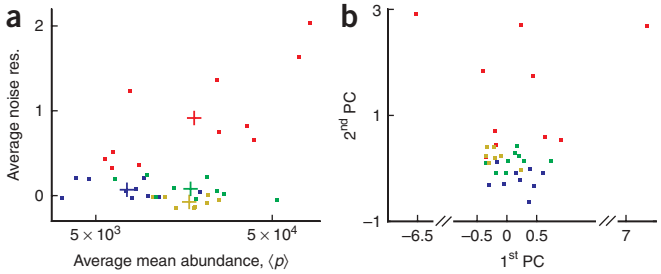

Figure 6 Noise residuals. Genes are colored according to their module affiliation: red, stress; yellow, proteasome; green, ergosterol; and blue, rRNA processing. (**a**) Noise residual versus mean abundance. Each point represents a gene, with both parameters (residuals and means) averaged over all perturbations and time points. Modules are represented by plus signs, with both parameters (residuals and means) averaged over all genes belonging to each module. *P* values, obtained by rank-sum tests, of the hypothesis that the medians of the noise residuals of two modules are equal, are as follows: stress, proteasome: 10^{-140}; stress, ergosterol: 10^{-118}; stress, rRNA processing: 10^{-119}; proteasome, ergosterol: 10^{-24}; proteasome, rRNA processing: 10^{-16}; and ergosterol, rRNA processing: 0.34. (**b**) PCA of noise residuals, normalized to the first time point of each condition ($R_N = R_t - R_0$). All conditions and time points were considered. The first two principal components capture 79% of the overall variance. Modules are ordered linearly, on the second component axis, according to induction by the stress conditions; increasing values correspond to increased induction.

to the protein noise. Quantitatively, the proportionality factor we find is somewhat smaller then that predicted theoretically (which is $1 + d$, with $0 < d < 1$, and d decreasing with time as the system approaches steady state)[15], suggesting that the protein levels might be maintained at a quasi–steady state (**Supplementary Note**). Because the underlying mechanisms of the adjustment are unknown, it is difficult to speculate further on this potential discrepancy.

The dependence of noise residuals on module affiliation

The analysis above indicates that a large noise contribution to our data comes from stochastic fluctuations in mRNA levels, resulting either from random birth and death processes of individual mRNA molecules or from stochastic gene activation. However, some genes also show large deviations from this trend, reflecting either additional noise sources specific to individual proteins, or variations in kinetic parameters such as translation rate. These deviations are captured by 'noise residuals', defined as the difference between the actual noise level and its expected value, given its mean expression level (Methods). It is interesting to note that noise residuals are predominantly positive. Negative residuals could, in principle, result from buffering mechanisms that would function to reduce the extent of intrinsic noise. The lack of such residuals may indicate that noise buffering, if it exists, occurs primarily at the higher level of cellular processing and not at the level of protein expression itself.

The noise residuals show clear, module-specific characteristics (**Fig. 6a**). In particular, for a given mean abundance, stress genes are very noisy, whereas the proteasome genes show very little variation. When projected on the first two principal components using principle component analysis (PCA), genes belonging to different modules clearly cluster together (**Fig. 6b**).

DISCUSSION

The relationship between the noise in gene expression and the mean protein abundance has been examined in several recent studies[16–20]. Expression levels were then tuned either by mutation or by varying the amounts of activators and repressors, sometimes using genes that are synthetically engineered. That is an appropriate and useful method for identifying the origin of fluctuations for the particular genes under study, but it does not indicate the general relation between noise and mean abundance in nature. Here we tested this relationship by examining proteins that differ naturally in their expression level and by analyzing expression levels in a variety of environments. Our data indicate that under most natural conditions, the variance is approximately proportional to the mean expression. Notably, we observed the same scaling behavior for all groups of genes, conditions and time points examined. Similar scaling behavior has been observed in the analysis of complex networks in physics and engineering[31,32].

Although this scaling mimics the noise from random birth and death of individual protein molecules (which is substantial only at low protein numbers), the variances are three orders of magnitude too high to be explained by such low-copy effects. Rather, our theoretical analysis suggests that this scaling reflects fluctuations in mRNA levels that arise from the random birth and death of individual mRNA molecules (in other words, low-copy mRNA noise) or from the stochastic activation of gene promoters. Several recent studies have suggested that noise is dominated by cell-to-cell variations in global factors[17,18,21–23], which are substantially larger than any gene-specific effects. It seems that most results from these studies came from the region of high abundance. Our study suggests that for intermediate abundances, the gene-specific noise is not overshadowed, in fact, by the contribution of global factors, but instead it is comparable, if not dominant. Indeed, previous analysis of low-abundance genes identified the rare gene activation events as a major noise source[19].

Comparing the noise residual levels of genes from different mRNA expression modules, we have observed module-dependent noise residuals. In particular, we find that the stress genes are highly noisy. This noise may be due to variations in the expression level of some common regulator. For example, the transcription factor MSN4 is present in only a few dozen copies, but it is implicated in the regulation of hundreds of stress-related genes. Stochastic transitions of MSN4 between promoters of different genes may increase gene expression noise. Another source of variability may be the gene activation pattern. For example, previous studies[22] have suggested that the presence of a TATA box enhances noise at a given mean abundance, presumably by increasing the number of transcripts synthesized when the gene is activated[22]. Indeed, the promoter of most stress genes contains a TATA box, whereas this sequence is absent in other genes of similar expression levels.

It has also been suggested that the level of gene expression noise is influenced by general chromatin effects[19,20,22]. We have noticed, however, that genes that are present at close proximity still show very different noise properties. For example, although the distance from *HXK1* to the proteasome genes *RPN12* and *PRE4* is only 263 and 5183 bp, respectively, HXK1 is one of the highly fluctuating proteins, whereas RPN12 and PRE4 have very small noise residuals. At least in this particular example, potential effects of chromatin seem to either be short-ranged or overshadowed by other factors.

It is conceivable that noise may be detrimental for some genes, whereas in others it may be tolerated or even advantageous. For example, it has been hypothesized that essential genes and genes that work in a complex will be characterized by low noise levels[33]. In support of that, we have observed that the essential genes coding for the proteasome subunits have the smallest noise residuals, whereas the stress genes, which are all dispensable, show the highest noise residuals. Moreover, the four dispensable ergosterol genes have significantly higher noise residuals than the five essential ergosterol genes

($P < 10^{-100}$, rank sum test). However, the difference between the residuals of the six dispensable genes and the three essential genes of the rRNA processing module is not significant ($P = 0.51$, rank sum test). An intriguing possibility is that for some genes and under some conditions, enhanced noise may even be beneficial at the population level. It remains to be understood whether the enhanced noise observed here in stress genes could have been selected for by evolution or whether it is a mere result of lack of constraint on the variability in expression of such genes[34].

Note added in proof: the scaling behavior we report was observed independently by others[40].

METHODS

Strains and growth conditions. We used 43 strains from the yeast GFP clone collection, purchased from Invitrogen (divided by module affiliation). The fusion proteins used corresponded to the following genes: stress: *TPS2*, *HSP104*, *HSP78*, *SSE2*, *GSY2*, *SSA4*, *PGM2*, *HXK1*, *TPS1*, *HSP42*, *HSP26* and *HSP12*; proteasome: *RPN3*, *RPN7*, *RPN6*, *RPN8*, *PRE10*, *RPN12*, *PRE4*, *PRE9*, *PUP2* and *SCL1*; ergosterol: *ACS2*, *ERG5*, *ERG13*, *MVD1*, *ERG10*, *CYB5*, *ERG1*, *ERG11*, *ERG6* and *ERG3*; rRNA processing: *NOC2*, *ARX1*, *PWP1*, *URA7*, *DBP3*, *PRS1*, *CIC1*, *AAH1*, *PRS4*, *BRX1* and *APT1*. Only 38 of those 43 genes had sufficiently high fluorescence above the autofluorescence background. Those genes were chosen for the next steps of the analysis. The discarded genes were *RPN3*, *PRS1*, *RPN7*, *HSP26* and *CYB5*.

Experimental perturbations. Inoculation and overnight growth are described in **Supplementary Methods**. All experiments were conducted on cells after overnight growth in 10 ml medium, collected at an optical density (OD) of ~0.2, unless otherwise mentioned (see **Supplementary Methods**). For each experiment, flow cytometry measurements were taken every 30 min until the last time point of 150 min. Throughout the experiment, cells remained in the Unimax1010 Incubator-Shaker. As a control, each experiment included cells that lacked GFP but that were subject to the same treatment as well as cells that lacked GFP and were not subject to the perturbation. Protein localization does not seem to affect the fluorescence reading (**Supplementary Note**).

We used 11 conditions. Eight of them were stress conditions: diamide (1.5 mM), hydrogen peroxide (H_2O_2) (0.3 mM), methyl methane-sulfonate (0.04% wt/vol), heat shock (30 °C→37 °C), dithiothreitol (4 mM), clotrimazole (10 μM), rapamycin (65 ng/ml) and ethanol (3%). The other three conditions were stress relaxation conditions: nitrogen depletion relaxation, stationary phase relaxation and glycerol growth relaxation by glucose addition. See **Supplementary Methods** for experimental details.

Flow cytometry measurements. Flow cytometry experiments were conducted using the Becton-Dickinson FACSAria machine and standard protocols. To reduce cell size variability, cells were filtered based on FSC-W and FSC-A filters. See **Supplementary Methods** for details.

Noise and noise residuals. For each fluorescence distribution, we measured the mean abundance ($\langle p \rangle$) and the variance (σ^2). In each experiment, $\langle p \rangle$ was normalized by subtracting the mean fluorescence of a controlled culture not expressing GFP, measured in parallel, under the same perturbation.

The trend lines in **Figure 2** were fitted using MATLAB's curve fitting toolbox. First, we considered each condition and time point separately and fitted the slope best describing the dependency of $\log(\eta^2)$ as a function of $\log\langle p \rangle$ for all genes measured, in the middle region of **Figure 2a**. In all 66 cases (11×6), the slope ranged from –0.7 to –1.4, with an average of –1.09.

To obtain the best-fitted curve and define its prefactor, we combined all data points and used an iterative procedure that discards outliers (that is, points whose distance from the best-fitted line is >0.5). This analysis was done using a fixed slope (–2 or –1, for the low- and intermediate-abundance regions, respectively). The process converged in five iterations. For all the data points in the intermediate-abundance region (18% of the points were excluded) we got $\log(\eta_p^2) = 1175 - \log(\langle p \rangle)$. Using only the steady-state time points ($t = 0$ for each of the stress conditions) gives $\log(\eta_p^2) = 1189 - \log(\langle p \rangle)$. Using all data points without discarding outliers results in $\log(\eta_p^2) = 1416 - \log(\langle p \rangle)$.

The precise position of the line separating the $1/\langle p \rangle^2$ versus $1/\langle p \rangle$ regions was (arbitrarily) chosen so that the two fitted curves coincide. The right border of the $1/\langle p \rangle$ region was chosen by eyeballing.

The noise residual was defined as the vertical distance of a given point from the fitted line in the log-log plot.

General model of noise propagation. Consider a protein p that is synthesized and degraded at some probabilities per time unit. Further assume that these probabilities depend on the level of an independent upstream component x, which also fluctuates, providing a random environment for the protein dynamics. Mathematically, this model can be captured by a two-variable time-continuous Markov process. Assuming that a single molecule is made or degraded per reaction event, and that the process has a stable stationary state, it has been shown that the noise arising from any such processes can be approximated[7,35] by

$$\eta_p^2 \approx \frac{1}{H_{pp}} \times \frac{1}{\langle p \rangle} + \eta_x^2 \times \frac{H_{xp}^2}{H_{pp}^2} \times \frac{H_{pp}/\tau_p}{H_{pp}/\tau_p + H_{xx}/\tau_x} \qquad (3)$$

The τ-parameters denote average lifetimes, whereas the H-parameters are elasticities that measure the normalized sharpness of nonlinearities in the reaction network connecting x with p. The elasticities are defined by, for instance, $H_{px} = \partial \ln(R_p^-/R_p^+)/\partial \ln x$, where R_p^+ and R_p^- are synthesis and degradation rates. For example, in the case where $R_p^+ = \lambda x$ and $R_p^- = \beta x^3 p$, so that x proportionally stimulates the synthesis of p and cubically stimulates the degradation of p, the elasticity is $H_{px} = 3 - 1 = 2$. Parameter H_{pp}/τ_p is the effective adjustment rate of the protein, and vice versa for x. For more details and explanations, see refs. 7,35.

Equation 3 suggests that proteins p are subject to two types of randomness, coming from the random birth and death of individual protein molecules (spontaneous fluctuations, observed for a given value of x), and from fluctuations in x. The spontaneous fluctuations, captured by the first term of equation (3), display the same qualitative scaling behavior as observed in our experiments, $\eta_p^2 = C/\langle p \rangle$ (**Figs. 2a** and **3b**). However, a quantitative fit would require extraordinarily low elasticities ($C = 1/H_{pp} \approx 1{,}200$). Such elasticities correspond to processes where the steady-state fixed points are almost unstable: for example, autocatalytic protein synthesis just at the balance point of unlimited synthesis. To our knowledge, such an extremely delicate balance has never been measured even for a single gene. That all genes would be characterized by exactly the same exotic instabilities is inconceivable and is also ruled out by the fact that a vanishingly small change in transcription or translation rates (say 0.1%) then would have substantial effect on protein levels (approximately 100%).

The second term of equation 3 describes fluctuations in x that indirectly force the protein to fluctuate as well. Below and in the main text, we analyze the potential of such noise to explain the observed scaling behavior, considering different molecular sources of noise. Note that this noise will dominate at high protein abundances, where spontaneous fluctuations in protein production become insignificant.

Contribution of averaging times and susceptibility to noise generated by global factors. Even if the upstream noise η_x^2 were constant, could differential susceptibility to this noise explain the observations? As described above, the steady-state susceptibility factor measures how strongly the upstream factor affects the steady-state protein levels and is related to Hill coefficients of activation or inhibition. We see no reason why they should systematically correlate with mean expression levels at all, let alone follow the observed scaling pattern.

The last factor to consider is time-averaging. The τ parameters indicate whether the level of p is defined by the instantaneous x fluctuation, or whether it depends on a long history of x fluctuations. If the upstream fluctuations are slower than those of the proteins ($H_{xx}/\tau_x \ll H_{pp}/\tau_p$), the time-averaging factor will be close to unity and will not significantly affect the protein variance η_p^2. If upstream fluctuations instead are fast, the time-averaging factor will be inversely proportional to the average protein lifetime: $H_{pp}/H_{xx} \times \tau_x/\tau_p$. In this case, the observed scaling behavior, $\eta_p^2 = C/\langle p \rangle$, will follow if protein abundance is proportional to protein lifetime. To provide a consistent explanation of

our measurements (**Fig. 2a**), the observed differences in protein abundance in our data set, both between the different proteins at steady state as well as during the response to perturbations, would need to be explained by changes in the lifetimes of the various proteins, rather than by changing gene expression or translation rates. Moreover, once dilution of proteins owing to cell growth dominates over actual degradation, differential degradation rates become impossible, and scaling will break down. We conclude that although time-averaging could explain the results kinetically, we find it biologically unlikely.

Contribution of spontaneous mRNA fluctuations to protein noise. Assume that transcription occurs with probability λ_{mRNA} per time unit and that translation occurs with rate λ_p per transcript. Further assume exponential decay of both transcripts and proteins with average lifetimes τ_{mRNA} and τ_p. This corresponds to the general model of equation (3), with all elasticities set to unity, but can also be calculated exactly from the full Markov process:

$$\eta_p^2 = \eta_{\mathrm{mRNA}}^2 \times \frac{1/\tau_p}{1/\tau_p + 1/\tau_{\mathrm{mRNA}}} \approx \eta_{\mathrm{mRNA}}^2 \times \frac{\tau_{\mathrm{mRNA}}}{\tau_p} \qquad (4)$$

where the approximation assumes that mRNAs are short-lived compared to proteins ($\tau_{\mathrm{mRNA}} \ll \tau_p$). Assuming that mRNA noise is due to random birth and death of individual transcripts, we obtain $\eta_{\mathrm{mRNA}}^2 = 1/\langle \mathrm{mRNA} \rangle$, with $\langle \mathrm{mRNA} \rangle = \lambda_{\mathrm{mRNA}} \tau_{\mathrm{mRNA}}$ as the mean mRNA level. Note that the mean protein levels are given by $\langle p \rangle = \langle \mathrm{mRNA} \rangle \lambda_p \tau_p$, gives equation (1) in the main text.

Fluorescence per molecule. To translate the fluorescence intensity into protein abundance, we used the protein abundance as measured in ref. 26. For each gene, we took, as representative fluorescence, the median of its fluorescence in the steady-state time points. The correlation between the known abundances and the representative fluorescence gives $R^2 = 0.44$. This R^2 indicates that there is a significant correlation; the hypothesis that the actual $R^2 \sim 0$ has a P value of 0.005 (**Supplementary Fig. 3**). For each gene, we calculated the ratio between its representative fluorescence and its measured protein abundance. The average fluorescence per GFP molecule was taken as the average of those ratios. According to this analysis, each GFP has a fluorescence of 0.16 arbitrary units.

Estimating the proportionality factor for low-copy mRNA noise. In the mRNA model (equation (1) in the main text, and equation (4)), parameter C is given by

$$C = \frac{\langle p \rangle}{\langle \mathrm{mRNA} \rangle} \times \frac{1/\tau_p}{1/\tau_p + 1/\tau_{\mathrm{mRNA}}} \qquad (5)$$

For each protein in our experiments, we extracted these parameters from large-scale data sets. Protein abundances $\langle p \rangle$ were taken from ref. 26, whereas the mRNA abundances $\langle \mathrm{mRNA} \rangle$ and the average mRNA lifetime τ_{mRNA} were taken from ref. 36. The average protein lifetime τ_p was estimated to be 100 min (roughly one cell cycle period). The value reported is the average over all the chosen genes: $C = 1{,}278 \approx 1{,}300$. Using other data sets of mRNA abundance[37] and mRNA half-life[38] gave similar expected C values. We did not use the mRNA abundances and half-life values from ref. 39 because those distributions are substantially different from those of the data sets in refs. 36–38 (see **Supplementary Methods** and **Supplementary Fig. 4** online).

General trend in transient response. To obtain the best-fitted curve for the transient response, we used an iterative procedure, similar to the one described above, implemented in **Figure 2a**. In this case, an outlier was defined as a point whose distance from the best-fitted line was >0.2. In this process, 15% of the points were excluded. Both slope and intersection points were determined by the fitting process and were not fixed beforehand.

Cartoon illustration of noise propagation. The expected dependencies between the noise and the mean protein abundance, considering different cellular sources, are plotted in **Figure 3**. In all cases, we assume that dots represent different genes, under a (singular) given condition. All graphs should be considered cartoons rather than full analyses and were simulated using MATLAB. In **Figure 3b–d**, we show 200 genes with various protein abundances, selected randomly on the logarithmic x axis. In **Figure 3e**, we show four modules, each containing 50 genes, whose mean abundances are distributed normally (on the logarithmic x axis) around a certain center. The noise level was calculated using the formulas given below, where the parameter C was randomly chosen, for each gene, from the distributions given below. **Figure 3b**: protein intrinsic noise of genes with various protein abundances, obeying $\eta_p^2 = 1/\langle p \rangle$. **Figure 3c**: mRNA intrinsic noise of genes with various protein abundances, obeying $\eta_p^2 = C/\langle p \rangle$, where $C \sim N(190,95)$. **Figure 3d**: global extrinsic noise of genes with various protein abundances, obeying $\eta_p^2 = C/\langle G \rangle$, where $C \sim N(1,1)$ and $\langle G \rangle = 30$. **Figure 3e**: module extrinsic noise of genes belonging to different modules, obeying $\eta_p^2 = C/\langle R \rangle$, where $C \sim \mathrm{N}(1,1/3)$ and $\langle R \rangle$ is equal to 5 (×), 50 (triangles), 300 (circles) and 5,000 (+).

Fluorescence microscopy. Diploid two-color yeast strains grown overnight were diluted to OD ~0.05 and grown 5 h in YPD medium. Cells were harvested, washed with synthetic complete (SC) medium and immediately imaged on glass slides. Microscopy was conducted using a Zeiss Axiovert 200M with a 63× oil-immersion objective (Zeiss) and motorized XY stage (Ludl) driven by Metamorph 6.3r5 imaging software (Molecular Devices). Separate fluorophores were excited using a xenon arc lamp (Sutter Instruments) and an appropriate filter set (Chroma). Data acquisition was automated, using a custom script written in Metamorph that repeatedly took background-corrected bright-field, CFP, YFP and RFP images; segmented cells; and output morphology and fluorescence intensity information. Filtering on the cell size, morphology, and RFP fluorescence eliminated small buds, debris and dead cells. For each strain, over 1,000 cells were imaged on at least two separate days. Independence and equivalence of reporters was verified by comparing fluorescence distributions of haploid strains with each color and the corresponding diploid strain.

Comparison of GFP fusion protein and GFP driven by promoter. See **Supplementary Note** and **Supplementary Figure 5** online.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the members of the Barkai and Pilpel labs for discussions and help in the experiments. This work was supported by the Tauber fund through the Foundations of Cognition Initiative. N.B. acknowledges the hospitality of the Bauer Center at Harvard, where part of this research was performed. Y.P. is an incumbent of the Aser Rothstein Career Development chair in Genetic Diseases. Y.P. acknowledges financial support from EMBRACE (a European Model for Bioinformatics Research and Community Education), funded by the European Commission within its FP6 Program.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. McAdams, H.H. & Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999).
2. McAdams, H.H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **94**, 814–819 (1997).
3. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
4. Barkai, N. & Leibler, S. Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
5. Berg, O.G., Paulsson, J. & Ehrenberg, M. Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys. J.* **79**, 1228–1236 (2000).
6. Rao, C.V., Wolf, D.M. & Arkin, A.P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).
7. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
8. Spudich, J.L. & Koshland, D.E. Jr. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
9. Maloney, P.C. & Rotman, B. Distribution of suboptimally induces -D-galactosidase in *Escherichia coli*. The enzyme content of individual cells. *J. Mol. Biol.* **73**, 77–91 (1973).
10. Lobner-Olesen, A. Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control. *EMBO J.* **18**, 1712–1721 (1999).

11. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
12. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
13. Raser, J.M. & O'Shea, E.K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
14. Kaern, M., Elston, T.C., Blake, W.J. & Collins, J.J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
15. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98**, 8614–8619 (2001).
16. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
17. Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
18. Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).
19. Becskei, A., Kaufmann, B.B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* **37**, 937–944 (2005).
20. Blake, W.J., Kaern, M., Cantor, C.R. & Collins, J.J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
21. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
22. Raser, J.M. & O'Shea, E.K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
23. Pedraza, J.M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
24. Swain, P.S., Elowitz, M.B. & Siggia, E.D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* **99**, 12795–12800 (2002).
25. Huh, W.K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
26. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
27. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* (2005).
28. Rigney, D.R. & Schieve, W.C. Stochastic model of linear, continuous protein synthesis in bacterial populations. *J. Theor. Biol.* **69**, 761–766 (1977).
29. Berg, O.G. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* **71**, 587–603 (1978).
30. Peccoud, J. & Ycart, B. Markovian modelling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).
31. Argollo de Menezes, M. & Barabasi, A.L. Separating internal and external dynamics of complex systems. *Phys. Rev. Lett.* **93**, 068701 (2004).
32. de Menezes, M.A. & Barabasi, A.L. Fluctuations in network dynamics. *Phys. Rev. Lett.* **92**, 028701 (2004).
33. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J. & Eisen, M.B. Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2**, e137 (2004).
34. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
35. Paulsson, J. Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175 (2005).
36. Holstege, F.C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
37. Velculescu, V.E. *et al.* Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
38. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860–5865 (2002).
39. Garcia-Martinez, J., Aranda, A. & Perez-Ortin, J.E. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol. Cell* **15**, 303–313 (2004).
40. Newman, J.R.S. *et al.* Single-cell proteomics of the budding yeast *Saccharomyces cerevisiae*. *Nature* (in the press).