INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY

Phys. Biol. **3** (2006) 172–182 doi:10.1088/1478-3975/3/3/002

Dynamics of protein distributions in cell populations

Naama Brenner[1], **Keren Farkash**[2] **and Erez Braun**[2]

[1] Department of Chemical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel
[2] Department of Physics, Technion-Israel Institute of Technology, Haifa 32000, Israel

E-mail: nbrenner@tx.technion.ac.il

Received 17 April 2006
Accepted for publication 21 August 2006
Published 8 September 2006
Online at stacks.iop.org/PhysBio/3/172

Abstract

A population of cells exhibits wide phenotypic variation even if it is genetically homogeneous. In particular, individual cells differ from one another in the amount of protein they express under a given regulatory system under fixed conditions. Here we study how protein distributions in a population of the yeast *S. cerevisiae* are shaped by a balance of processes: protein production—an intracellular process—and protein dilution due to cell division—a population process. We measure protein distributions by employing reporter green fluorescence protein (*gfp*) under the regulation of the yeast GAL system under conditions where it is metabolically essential. Cell populations are grown in chemostats, thus allowing control of the environment and stable measurements of distribution dynamics over many generations. Despite the essential functional role of the GAL system in a pure galactose medium, steady-state distributions are found to be universally broad, with exponential tails and a large standard-deviation-to-mean ratio. Under several different perturbations the dynamics of the distribution is observed to be asymmetric, with a much longer time to build a wide expression distribution from below compared with a fast relaxation of the distribution toward steady state from above. These results show that the main features of the protein distributions are largely determined by population effects and are less sensitive to the intracellular biochemical noise.

Introduction

In a population of organisms, each individual is unique. Genetic variation, maintained by various forces, is an important factor for the survival of the population over long time scales in the context of ecology and evolution. However, even a genetically homogeneous (clonal) population exhibits a large degree of phenotypic variation among its individuals [1–5]. While the fundamental questions concerning phenotypic variation were raised many years ago [1, 2, 6], there has recently been a renewed interest in quantifying variation by single-cell measurements [7–13]. Microorganisms provide a convenient model system to study phenotypic variation in clonal populations. This variation can be characterized in many ways, such as by the gene expression level, or by morphological or metabolic characteristics; here we focus on variation in the expression of a particular protein among individual yeast cells in a genetically homogeneous population.

Proteins are produced inside cells in a process that is regulated at several levels. In particular in eukaryotic cells, transcription regulation has evolved to exquisitely sensitive and precise mechanisms that can respond to multiple inputs [14–18]. However, even precisely regulated genes are variably expressed in cells grown under homogeneous conditions. This fact raises fundamental questions about how gene regulation can remain functional in a variety of conditions; indeed, protein expression variation has been extensively characterized in recent years (for a review, see [5]). However, we note that the functional significance of expression variation can be assessed only under conditions where the expressed gene plays an essential role in the cellular function. Thus in experiments where proteins are expressed from promoters that are induced by synthetic agents (such as IPTG), or by external signals irrelevant to metabolism (such as induction

of the GAL system by galactose in cells grown in raffinose), only local biochemical aspects of noise in gene expression are in fact probed. In this work, since we are interested in variation in a functionally regulated protein, a central feature of our experiments is the measurement of expression from GAL promoters under conditions of galactose as a *sole carbon source*. This experimental design allows us to gain insight into the variation inherent to a regulatory system that carries an essential metabolic role.

At the level of the single cell, there are multiple mechanisms rendering functionality immune to variation in gene expression [11, 19, 20]; cell metabolism and growth are the integrated outcomes of many interacting processes and under normal conditions cellular homeostasis buffers against variations in protein content [21–23]. In some cases, however, variation in the content of particular proteins across a population can have significant effects on functionality, for example by allowing a population to adapt to transient stress [24, 25], or by flipping a genetic switch and initiating a split of the population into distinct phenotypic subgroups [26–28]. Thus, the functional implications of protein variation are most crucial at the level of a cell population: it can enable population survival by providing a substrate for phenotypic selection under changing environments [24, 29–31]. These effects are generally important over physiological time scales, namely times longer than a generation but short relative to the time typical of mutations and their fixation in the population. A quantitative characterization of regulated protein variation from the population point of view is required in order to understand its role over these time scales.

At the level of the population, one is interested in the characteristics of protein distributions, their steady states and stability properties, and their dynamics in response to perturbations. Recent studies have focused on characterizing the sources of variation internal to the cell and related to the small number of molecules involved [4, 5, 7–11, 32–34]. These include many interacting sources with different stochastic characteristics that generally depend on the details of underlying biochemical systems. A cell population, however, is not a statistical ensemble of independent individuals but a dynamical system. Cells are continually dividing and proteins are inherited to the newborn cells; their gene expression state is dependent, among other things, on these initial conditions determined at cell division. Thus from the population point of view both the growth history and the metabolic conditions play important roles in shaping the population protein distribution. This dynamic problem spans a wide range of time scales that need to be addressed both experimentally and theoretically.

Characterizing protein distributions in proliferating cell populations with sensitivity to the population's history and details of cell division necessitates special measurement systems. It requires following the population at single-cell resolution over several generations under controlled conditions. It is also necessary to distinguish between transients—which may last longer than a cell generation time—and steady states of the population [12]. Environmental signals regulating gene expression, such as inducing or repressing substances, need to be applied in a controlled way and their concentration maintained stably over long time scales. Standard growth techniques do not allow such control of the environment. In this work we use a chemostat, a continuous culture technique, to grow yeast populations under controlled conditions—a mixed, homogeneous environment—for many generations, allowing for quantitative investigation over the relevant time window spanning several generations ($\sim$10) or longer. We study variations in a protein that is expressed from a promoter regulating a metabolically essential system—the galactose utilization system under conditions where galactose is the sole carbon source. In this context the chemostat plays an additional important role: yeast cells that are not metabolically well fitted, namely that do not exhibit an adequate growth rate, cannot compete in the chemostat and will be represented poorly in the population. We employ a reporter green fluorescence protein (*gfp*) expressed from a GAL promoter and high-resolution fluorescence microscopy to quantify the protein distribution in the population, and to characterize both their steady states and their dynamic responses to various perturbations. Previous studies have shown that this method accurately reflects the pattern of the native GAL system proteins [12, 35]. Direct analysis of microscope images allows us to overcome artifacts that may occur while using automated analysis methods [36] and to analyze in detail the physiological properties of the population (see below).

We find that the population steady state is universally characterized by a broad distribution of expressed protein with a peak near zero and an exponential tail. Increasing the mean expression by multiple promoters increases the standard-deviation-to-mean ratio of the distribution. In response to several perturbations, the relaxation back to steady state is found to be asymmetric. Repression is much faster than induction, and is established within two to three generations as opposed to approximately ten generations. A high-expressing subpopulation relaxes back to the steady state also within approximately three generations, while a low-expressing subpopulation takes much longer, more than ten generations, to build up the steady-state distribution. Our results, in particular the time scales that emerge in the dynamics of distribution, show that population dynamic effects dominate over the intracellular noise in shaping the protein distribution.

Materials and methods

Plasmid and strain constructions

All experiments were carried out with the haploid yeast strain YPH499 [*Mata, ura3-52, lys2-801, ade2-101, trp1-*Δ*63, his3*Δ*200, leu2*Δ*1*]. Multiple promoters were inserted into the plasmid vector pESC-LEU (Stratagene) containing pGAL1-pGAL10 divergent promoters, on which the *gfp* was inserted under the GAL10 side of the promoter [12]. Alternatively, a single *gfp* copy was integrated into the genome at the LEU2 locus by using the plasmid pRS405 cloned with the *gfp* downstream *GAL10* promoter [12].

Cell growth and microscopy

Cells were grown in homemade chemostats as described in [12]. Dilution rates were varied between experiments in the range of 0.06–0.2 l h^{-1}. This corresponds to a generation time between 11 and 3.5 h. We used synthetic dropout medium lacking *leucine* and 2% of pure galactose (or pure glucose) as a sole carbon source, 1.7 g l^{-1} yeast nitrogen base, 5 g l^{-1} ammonium sulfate, 1.4 g l^{-1} amino acids dropout powder (Sigma), 0.02 g l^{-1} L-histidine, 0.04 g l^{-1} L-tryptophan, 0.02 g l^{-1} uracil, 20 g l^{-1} galactose or 20 g l^{-1} glucose. The chemostat was amino acid limited. Typical populations in the chemostat contained 10^9–10^{10} cells. The turbidity and the fluorescence level of the culture were monitored during the experiment. Fluorescence measurements were performed using an inverted microscope (Zeiss Axiovert 135) with 100× oil-immersed objective and 100 W mercury lamp.

Time-lapse measurements of single cells were done by a homemade temperature-stabilized incubator on the stage of the fluorescence microscope. Bright-field and fluorescence images were captured at regular time intervals with minimal exposure of the cells to light.

Image analysis

Basic image analysis was done as described in [12]. The procedure employs a combination of analysis by ImagePro with our homemade-developed software. It produces the histogram of cell sizes and cell fluorescence in samples of the population consisting of several thousand cells at each time point. In addition, we identified and tagged pairs of mother and bud cells for separating subpopulations and for following the inheritance of proteins across generations. From each experiment, several hundred such pairs could be identified.

Plasmid copy number measurements

The mean plasmid copy number was measured by real-time PCR (AB 7700). Measured amounts of *ACT1* prepared by PCR served as a ruler. Total DNA was purified from cells harvested from the chemostat (Invitek Tissue kit). The number of *gfp* DNA sequences was compared to that of *GAL4* and *ACT1*. The *GAL4* to *ACT1* ratio was used to estimate the measurement errors. Maximal errors were less than 3% in duplicates at the same PCR measurement and less than 15% between separate PCR measurements.

Cell separation

The separation of cells was done using a *FACSVantageSE* cell sorter (Becton Dickinson). Approximately the same number of cells was sorted from the two ends of the fluorescence distribution. The cells were immediately grown in the same galactose-rich medium described above at 30 °C. The cell growth rate was estimated by plating cells from the two subpopulations from the onset of separation at regular time intervals on YPD plates and then counting colonies. Samples of cells were scanned under the fluorescence microscope at regular time intervals for fluorescence distribution estimates.

Monte Carlo simulations

A large population of cells (∼100 000) was followed across generations. In each generation, each cell produced an amount of protein drawn from a given distribution representing the internal cellular protein production stochastic properties. At the end of each generation the cells divided and their protein content was distributed according to the binomial distribution with a fixed mean fraction f. This enabled us to describe the asymmetric division in the budding yeast. The population was kept at a constant size by random washout. The convergence was monitored by computing the first two moments as a function of generations.

Results and discussion

Steady-state distributions

The chemostat is the only growth technique that enables us to distinguish clearly between transient responses and steady states of a regulatory system in proliferating populations [12, 23, 25]. We have used the chemostat to grow populations of yeast cells to steady state with high-concentration galactose (2%) as the sole carbon source; under these conditions the GAL regulation system is expected to be strongly induced. Using fluorescence microscopy, we measured the single-cell fluorescence level of a *gfp* reporter expressed from the GAL10 promoter in a sample of cells and constructed a histogram representing the population distribution (figure 1). At steady state the population is characterized by a broad expression distribution with a peak at a relatively low fluorescence level and an exponential tail. Figure 1(*a*) compares the distributions of expression from a single promoter integrated into the genome with that from promoters on multiple-copy plasmids. Both distributions exhibit exponential tails, with standard-deviation-to-mean ratios of $\sigma/\mu = 0.6$ and 1 respectively. These values are surprisingly large considering the metabolic role of the GAL system in pure galactose. It is noted that the ratio σ/μ increases with increasing mean, in contrast to the expected behavior of microscopic random fluctuations. This is the result of the long exponential tail that builds as the source of protein production is increased. In the experiments that follow, it will be shown that this empirical relation between σ/μ and μ is maintained under various conditions.

Figure 1(*b*) shows the fluorescence histograms from many separate experiments with different chemostat dilution rates and limiting nutrients. The scaled fluorescence distributions (normalized to unit mean) exhibit two universality classes, corresponding to those shown in figure 1(*a*), independent of the growth medium or dilution rate. This universality enables us to pool together data from several experiments with a single promoter, thus revealing in more detail the exponential tail of this distribution (figure 1(*b*), inset). We note that previously published results [12] show that these main features of the expression distribution are not a result of the growth technique using a chemostat—growing cells in a serially diluted batch culture results in similar distributions.

What are the microscopic mechanisms underlying the broad expression distribution at steady state? The yeast

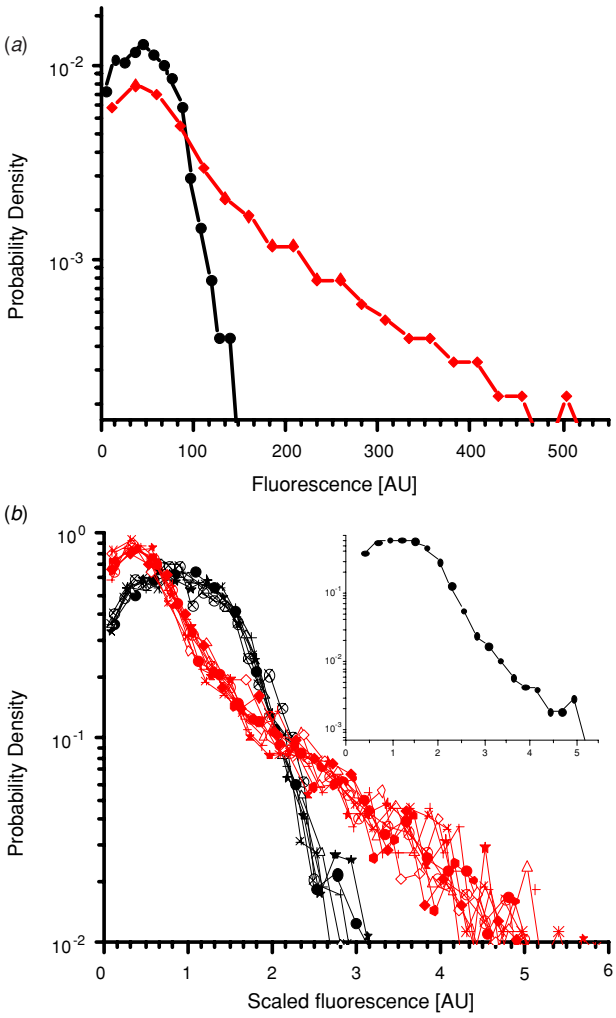

Figure 1. Steady-state fluorescence distributions. *gfp* was expressed from GAL10 promoters in populations of yeast cells grown to steady state in chemostats with high concentrations of galactose (2%) as the sole carbon source. (*a*) Comparison of expression distributions from a single promoter (integrated into the chromosome; black circles) and multiple promoters (on 2μ plasmids; red diamonds). The two distributions have a near-zero maximum, with a common crossover point to an exponential tail. (*b*) Different populations grown under different conditions, such as chemostat dilution rate and nutrient composition, cluster into two shapes of the distribution according to the promoter number and independent of other conditions. The standard-deviation-to-mean ratios in the two families of distributions are 0.62 ± 0.02 and 1.03 ± 0.04. Inset: pooling together statistical data from several experiments with the single promoter strain reveals an exponential tail in these distributions over several decades in probability density.

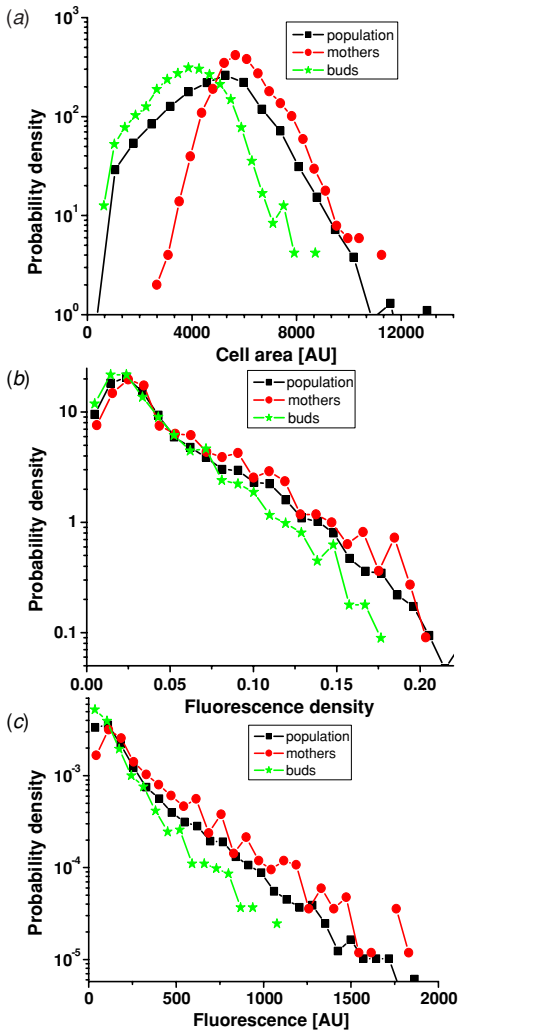

Figure 2. Yeast population structure—mother cells and buds. By identifying pairs of mother cells and buds in microscope images (see Methods), we constructed histograms of the different yeast subpopulations. (*a*) Cell projected area clearly reflects the population structure: buds are small in area and make up a relatively broad distribution (green stars; $\sigma/\mu = 0.34$) since they are sampled at different stages of growth. Mother cells are larger and make up a narrower (red circles; $\sigma/\mu = 0.19$), approximately Gaussian distribution of sizes. (*b*) Fluorescence density is distributed independently of position along the cell cycle; distributions of mother cells (red circles; $\sigma/\mu = 0.8$), buds (green stars; $\sigma/\mu = 0.86$) and total population (black squares; $\sigma/\mu = 0.86$) are very similar. (*c*) The distribution of total fluorescence carries a signature of the differences in cell area distributions, but all parts of the population have an exponential tail: buds (green stars; $\sigma/\mu = 1.19$), mother cells (red circles; $\sigma/\mu = 0.94$) and total population (black squares; $\sigma/\mu = 1.1$).

S. cerevisiae proliferates by budding, and thus at a given time the population is a mixture of adult cells and buds at different growth stages. Recent work on the yeast GAL system utilizing FACS measurements has indicated that this population structure is an important source of variation in gene expression in the regime of high mean expression [13]. In this work we analyze microscope images of the yeast population rather than measuring the fluorescence distribution by automatic FACS. This enables us to assess the contribution of the population structure directly, by examining the cell-size distribution and by analyzing separately subpopulations of mother cells and buds (see Methods). The cell-size distribution (figure 2(*a*)) clearly reflects the structure of the population: it is a superposition of the size distribution of the mother cells, a narrow distribution centered on a high value (red circles), and the size distribution of buds, a broad distribution of small sizes (green stars). In contrast, the

gfp density distribution (fluorescence per unit projected area; figure 2(*b*)) is practically indistinguishable between the two subpopulations. The total fluorescence (figure 2(*c*)) shows some difference due to the differences in cells size, but both subpopulations have similar exponential tails and standard-deviation-to-mean ratios. We therefore conclude that the main features of the steady-state expression distribution, the peak near zero and the exponential tail, are not a result of the cell-size distribution or the presence of a mixture of mothers and buds in the population. We note that in our experiments, in contrast to most previous work (in particular [13]), galactose is the sole carbon source and therefore the expression of the GAL genes is strongly connected to other parts of cell metabolism; it is this expression, which is of functional significance, that is of interest here. If galactose is used only as an inducing agent and a more readily metabolizable carbon source is supplied (e.g. raffinose), the expression of the GAL genes is mainly determined by local biochemical properties of the expression circuit and by the population size structure. Indeed, the variation in that case ($\sigma/\mu \cong 0.25$) [13] is much smaller than that observed here ($\sigma/\mu = 0.6$ or 1).

With high-copy plasmids in yeast, the precise number of promoters varies from cell to cell. One may expect this variation to play an important role in determining the expressed protein distribution. However, in eukaryotic regulated modules such as the GAL system, this is not necessarily true because expression from a large number of promoters is controlled by an extremely small number of transcription factors. In the GAL system, for example, there are typically a few copies of the GAL4 transcription factor [37], which bind in dimer form, and more than ten binding sites on the different GAL promoters [38]. In addition, there are many other GAL4 transcription factor binding sites, not related to the GAL genes, throughout the genome [39]. In pure galactose, all GAL promoters need to be active; therefore the natural system is strongly limited by the availability of transcription factors rather than by promoter copy number. In this regime, the sensitivity of expression to the promoter copy number is expected to be very low. We have verified this hypothesis by the direct measurement of mRNA levels of the native GAL genes by real-time PCR, showing that these levels are the same in both our yeast strains, with a single additional GAL promoter and with the addition of ∼40 GAL promoters on plasmids ([12], table 1). A comparison of the *gfp* mRNA levels between the two strains shows only a sixfold increase in *gfp* mRNA expression (see table 1 in [12]). Thus, there is a nonlinear relation between the number of promoters and the level of expression, presumably due to limitation of transcription factors. Note that the situation described here is typical of many eukaryotic regulation modules [40].

To study directly the sensitivity of the expression distribution to the mean number of promoters, we used the plasmid copy number as a control parameter. We performed an experiment in which the yeast cells were switched to an environment that is non-selective for the plasmids. Figure 3 shows the *gfp* fluorescence distributions at different time points along this experiment (3(*a*)), together with a direct measurement of the mean *gfp*-expressing plasmid copy

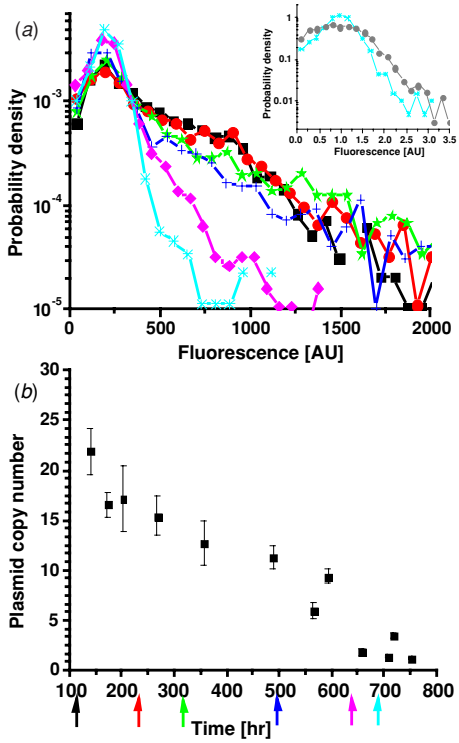

Figure 3. Relation between plasmid copy number and fluorescence distribution. In our yeast strains, the 2μ plasmids are selected for by lack of leucine in the medium. Here the selection has been removed by adding leucine to the medium, and fluorescence distributions were followed over time together with changes in the plasmid copy number (see Methods). (*a*) Fluorescence distributions at different times following removal of selection. The distributions are stable for approximately 100 generations before it changes significantly. The exponential tail then decreases in weight relative to the head of the distribution, until it converges to a shape similar to the single-promoter strain (inset). (*b*) Plasmid copy number, measured directly by real-time PCR. Colored arrows indicate time points where the distributions, plotted with the corresponding colors, were measured in (*a*). It is seen that the plasmid copy number changes by almost a factor of 2 before the distribution changes significantly.

number in the population (3(*b*)). It is clearly seen that the distribution is completely insensitive to the plasmid copy number within a wide range: loss of half of the initial plasmids, over a time scale of approximately 80 generations, does not affect the distribution at all. Note that with the release of the selection pressure not only the mean but also the variation in plasmid copy number is expected to change (rise) significantly. This implies that, in the regime of high mean copy number, the direct contribution of plasmid copy number variation to protein variation is negligible.

When the copy number dropped below a critical value, (∼7), the protein distribution changed its shape and converged, at a low copy number, to a distribution similar to that exhibited by the integrated single-copy promoter (figure 3(*a*), inset). At intermediate levels of mean plasmid copy numbers, the mean fluorescence in the population decreased but the exponential tail was preserved with a similar slope, as seen in figure 4. The ratio between standard deviation and mean in this dynamic process remained fixed. This implies that there is no direct

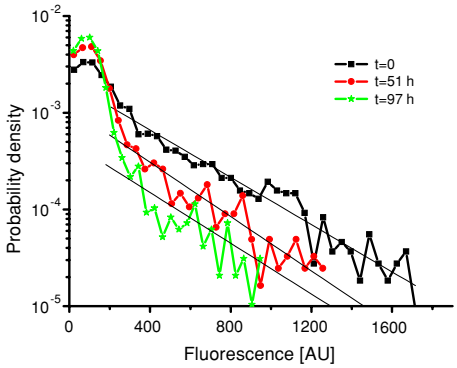

Figure 4. Distributions during plasmid loss by non-selective medium. The chemostat was switched to a medium containing leucine at $t = 0$, causing plasmid loss by the non-selective medium (similar to the experiment of figure 3). Distributions were constructed from particularly large samples of the population, thus highlighting the changes in the exponential tail. The loss of plasmids results in a change in proportion between the head and tail of the distribution while maintaining the slope of the tail. Lines correspond to best fit of the tail to $y = A\,\mathrm{e}^{-x/\xi}$. Parameter values: $A = 0.002, 0.0011, 0.0005$; $1/\xi = 0.0028, 0.0032, 0.003$ for the black, red and green curves respectively. $\sigma/\mu = 1.08$, 1.16 and 1.11 respectively. The dynamics of the expression distribution during plasmid loss are very different from the dynamics of repression by glucose (figure 9), which results in a strong change of the slope of the exponential tail accompanied by a significant change in σ/μ.

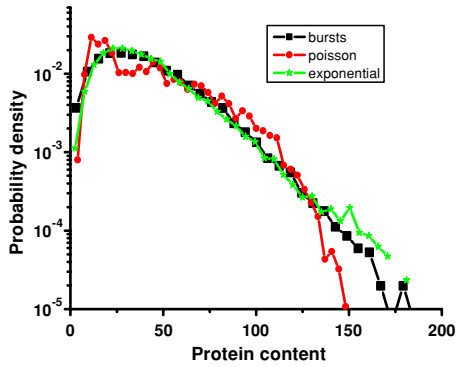

Figure 5. Steady-state protein distributions from Monte Carlo simulations of a dividing population. The figure shows that an exponential tailed distribution, similar in nature to the one observed in the experiments, can be obtained by different characteristics of the internal cellular production process and division properties: stochastic production and degradation of mRNA, which produces protein by a Poisson process, with symmetric division ($f = 0.5$, black. Here we took an average of two mRNA molecules produced per generation and a production rate to match the total protein production rate per generation with the other models.); Poissonian production at a constant rate with asymmetric division ($f = 0.75$, red); and protein production drawn from an exponential distribution with symmetric division (green). These results show that measuring the steady-state distribution does not provide sufficient information about the internal cellular and population processes.

relation between the exponential tail and the plasmid copy number, because if this were the case the loss of plasmids could not be uniform throughout the tail but would depend on the position in the tail [41]. Moreover, the protein variation, as reflected in the ratio between standard deviation and mean, is independent of the mean plasmid copy number, even in this regime where the distribution changes its shape. In conclusion, the dependence of the expression distribution on plasmid copy number is strongly saturated. In particular, the hallmark of the expression distributions—an exponential tail and a large standard deviation over mean—is insensitive to the promoter copy number in the regime of our experiment.

Processes shaping the steady-state distribution

The previous section has ruled out several hypotheses concerning the molecular mechanism underlying the broad expression distribution at steady state, including variations in cell size and in plasmid copy number. Together with the independence of the distribution on growth conditions, our results suggest that the distribution is not a direct result of a single molecular mechanism, and that it is insensitive to the details of the intracellular noise. To further test this hypothesis, we performed simulations of a dividing cell population with different intracellular protein production characteristics and different division characteristics, and computed the steady-state distributions (see Methods). Figure 5 shows that the main features of the protein distributions—a large standard-deviation-to-mean ratio, a peak near zero and an exponential tail—can be reproduced with different models of intracellular noise source and cell division. These simulations show that the steady-state properties are not determined by the details of the intracellular noise.

Therefore, rather than trying to trace the sources at the molecular level, we adopted a phenomenological approach to the problem: we perturbed the population and measured how the total distribution relaxed back to the steady state. Using fluorescence-activated cell sorting (FACS), we separated out two subpopulations from the two extreme ends of the steady-state distribution, as illustrated in figure 6(*a*). These subpopulations were then separately grown in the same galactose-rich medium, and the fluorescence distribution and growth rates were monitored over several generations from the time of separation. The growth rates of the two subpopulations were found to be indistinguishable; in yeast cells, the growth rate is a direct consequence of the total metabolism, showing that the distribution does not correspond simply to variations in metabolic state across the population. Figure 6(*b*) shows that the dynamics of the fluorescence distributions of the two subpopulations were significantly different. The subpopulation starting at the high-fluorescence end converged back to the steady-state distribution within three to five generations. By contrast, for the low-fluorescence end subpopulation, the distribution steadily increased toward the steady state but had not reached it even after as much as nine to ten generations. The long time scales and asymmetry of this relaxation clearly show that a dividing cell population is very different from an ensemble of independent realizations of biochemical noise. The measured distribution reflects not only the internal cellular mechanisms, but necessarily also dynamic processes with typical time scales longer than

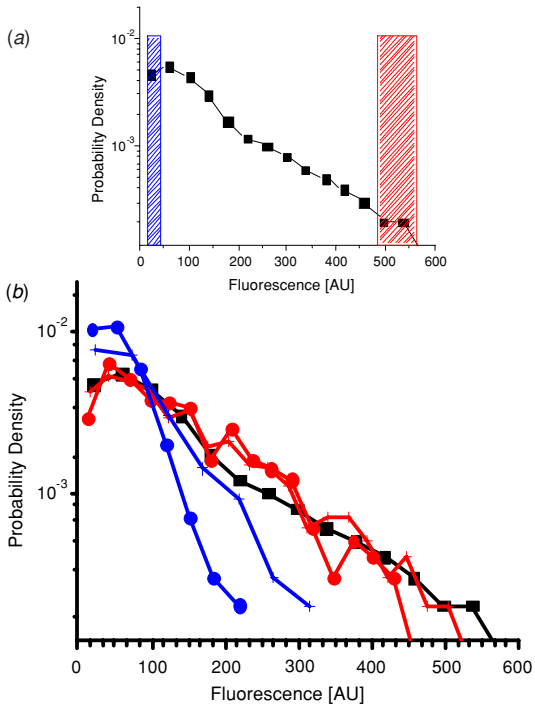

Figure 6. Asymmetric relaxation to steady state following a sorting perturbation. (*a*) Starting from a steady-state distribution (black histogram), fluorescence activated cell sorting (FACS) was used to separate the high-end 1% from the low-end fluorescence 1% of the population, illustrated here schematically by red and blue shaded regions respectively. These subpopulations were then grown separately by serial dilution in the same galactose-rich medium, and their growth rate and fluorescence distributions were measured over time. The growth rate was indistinguishable in the two subpopulations. (*b*) The fluorescence distributions, however, relaxed in an extremely asymmetric manner: the high-fluorescence end (red curves) has already returned to the steady-state distribution after approximately three generations, and the low-fluorescence end (blue curves) was still relatively low after as much as nine generations (circles: $t = 23$ h, plus: $t = 38$ h).

a cell generation and involving cell division and protein inheritance.

The budding yeast provides a unique opportunity to directly measure the protein inheritance across consecutive generations. Figure 7(*a*) shows the *gfp* density in mother cells and corresponding daughter cells, measured from identified pairs in steady-state populations (see Methods). There is a strong correlation between them (correlation coefficient ∼0.9), with a slope close to 1 and with relatively small dispersion. Figure 7(*b*) shows a time-lapse microscopy measurement of a single mother–daughter pair during bud growth. In agreement with figure 7(*a*), it shows that the protein density tends to equilibrate between the two cells and reaches approximately the same value at the end of bud growth, corresponding to protein diffusion that is fast relative to bud growth. These results show that the budding process ensures similar *gfp* densities in mature buds and their mother cells, rendering the new generation essentially a duplicate of the old one in terms of density. This reliable protein inheritance process induces strong dissipative loss of protein proportional to the existing resources in the mother cells and enables a fast convergence of the distribution to steady state after perturbation.

The protein distribution in the population is thus formed by a balance between two opposing processes: a dilution process (due to cell division) and a production process (protein production). The fractional nature of loss at division—the mother cell loses a fixed fraction of its existing protein content—leads to adjustment of the dilution process to the production process. The time scale of dilution is expected to be relatively short, and thus the population can undergo transitions from a high distribution to a lower one within a small number of generations. This is consistent with the rapid decay of the high-end fluorescence subpopulation, described above in the FACS experiment (figure 6). It is also consistent with previously observed rapid transitions that occur spontaneously (over three to five generations) from a broad to a narrower distribution in similar chemostat experiments [12].

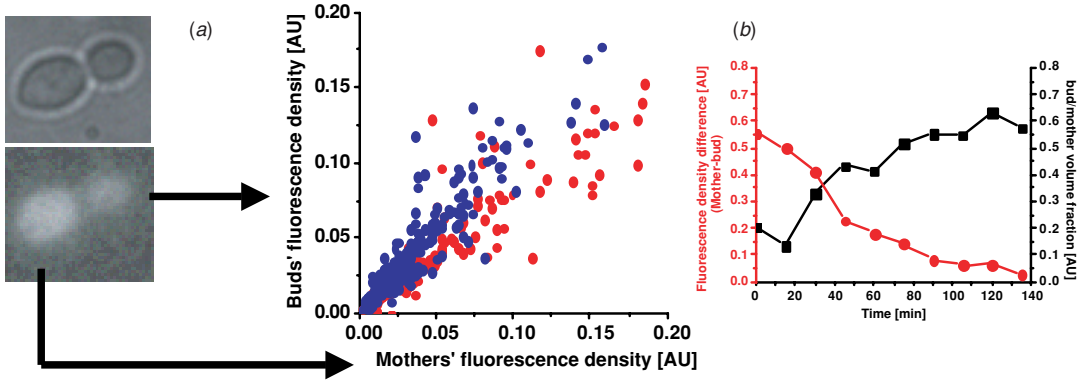

Figure 7. Protein correlations among generations in a steady-state yeast population. (*a*) Protein density, defined as the fluorescence signal per unit projected area in microscope images, is plotted for identified mother–daughter pairs in two steady-state populations, red and blue dots. (*b*) Protein content in a mother and bud during bud growth, measured by time-lapse microscopy. The ratio between mother and bud cell volume is shown as a function of time during bud growth (black line), together with the difference in fluorescence density (red line), which decreases to zero at the end of the process. These results show that protein transfer is dominated by fast diffusion, resulting in a high correlation among generations, so that dilution in the population is essentially a fractional process: at cell division, an approximately fixed fraction of the total protein is lost to the next generation.

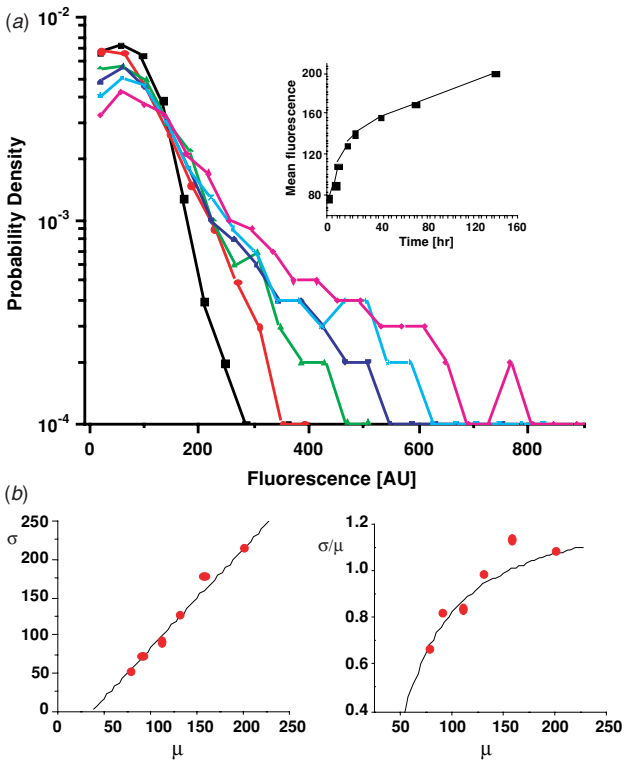

Figure 8. Induction of the GAL system—a population view. (*a*) Fluorescence distributions during induction. Starting from the non-inducing, non-repressing raffinose (black histogram), the population was switched to pure galactose which induces the GAL system. It is seen that along the process the exponential tail of the distributions increases, whereas the crossover between the head and exponential tail remains fixed. Inset: mean fluorescence, calculated from the distributions, as a function of time. The generation time is approximately 5 h. (*b*) Relations between standard deviation and mean of distributions along the induction process. The standard deviation is a shifted linear function of the mean (left), corresponding to an increasing standard-deviation-to-mean ratio (right).

Dynamics of protein distributions

The GAL system, being a metabolic switch that responds to carbon source, enables us to characterize induction and repression dynamics as reflected in protein distributions. To induce the GAL system we inoculated the chemostat with cells grown first in pure raffinose, which is non-inducing and non-repressing to the GAL system; we operated the chemostat in pure galactose medium and monitored the time evolution of the *gfp* distribution (expressed, as before, from the GAL10 promoter). Figure 8(*a*) shows that the exponential tail of the distribution is extremely sensitive to the protein production process, determined here by the induction of the promoters. During this time the culture density in the chemostat remained stable. This implies that the GAL promoters have been induced within a short time and that cell metabolism had reacted quickly (within a single dilution time) to the medium switch from raffinose to galactose. Still, the protein distribution in the population takes an extremely long time (order of ten generations) to build up toward the *gfp* fluorescence steady-state distribution. The inset to figure 8(*a*) shows the mean fluorescence in the population as a function of time along the buildup process of the distribution. It shows that there exists a significant barrier to be passed when increasing the protein content in the population against the dilution forces, consistent with the long time scale for increase starting from a low-expression subpopulation in the FACS experiment described above (figure 6). Examining the shape of the distribution throughout the induction process, it is seen that the crossover point between the head and the exponential tail remained approximately fixed; the slope of the exponential, as well as the relative weight of the two components, changed as a function of time. The standard deviation and mean of the distributions obey an empirical linear relation along the induction process and their ratio is an increasing function of the mean (figure 8(*b*)). This is in contrast to a global scaling relationship found recently over many different genes and conditions [36, 42]. The particular relation observed here is consistent with the distributions being made up of two independent components, one relatively narrow and independent of the induction, and the other with an exponential tail that increases along the induction process (see the appendix).

The opposite dynamics of repression was observed by switching the chemostat medium from pure galactose to pure glucose. Figure 9(*a*) shows the time evolution of the distributions, starting from the broad galactose steady state and decreasing toward a narrower distribution in glucose. Once again, the crossover point between the head of the distribution and the exponential tail remained approximately fixed, while the tail responded sensitively to the repression conditions. We have shown in a previous work that the lower distribution shown here is not the true steady state of the system in glucose, due to long-term adaptation dynamics [12]. Here, we concentrate on the transient repression in response to the medium switch to glucose. The inset of figure 9(*a*) shows that the mean fluorescence decays rapidly and reaches the repressed population state in approximately one generation; in this process, the population kinetics parallels the intracellular promoter repression kinetics. This is in marked contrast to the long time scale revealed in the population induction process. The strong dissipation induced by protein dilution in cell division acts against the induction process while it acts in the same direction as repression. This fast repression is consistent with the short relaxation time scale of the distribution starting from a high-expression subpopulation in the FACS experiment described above (figure 6).

The main features of the distributions throughout the repression process are similar to those during induction, as shown in figure 9(*b*). In both cases, the exponential tail is sensitive to the activity of the GAL system, which is the source of protein production. The standard-deviation-to-mean ratio, once again, is an increasing function of the mean during the process—in this case, decreasing as a function of time.

Conclusion and outlook

Much recent work has been devoted to dissecting the different sources of variation in gene expression. In some

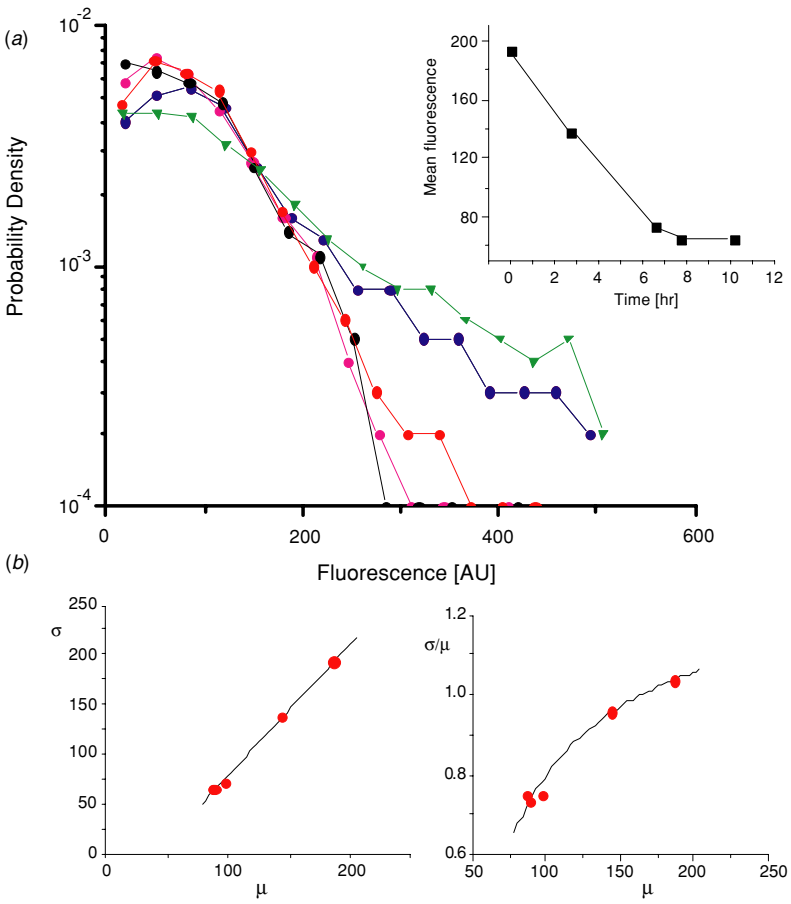

Figure 9. Glucose repression of the GAL system—a population view. (*a*) Starting from the pure galactose steady state (green triangles), the chemostat was switched to pure glucose as the sole carbon source, which rapidly repressed the GAL system, reflecting a decrease in the fluorescence distribution. As in the induction process, the exponential tail changes its slope during repression, while the crossover point between the head and tail of the distribution remains fixed. Inset: mean fluorescence as a function of time, showing that in contrast to induction, repression is extremely rapid in the population. (*b*) Relations between standard deviation and mean during repression.

studies artificial expression systems were used, or induction was applied artificially (where the expressed gene was not functionally essential). In these experiments it is often possible to relate statistical properties of gene expression to the detailed biochemistry of the specific regulation circuit since it is detached from many other cellular processes. Other studies have considered the large-scale behavior of natively regulated proteins by using automated FACS measurements. It was shown that particular gating procedures can separate out the intrinsic component of variation, which is dominated by biochemical noise [36]. In these studies generally growth and preparation were not characterized in detail, for example the difference between transient and steady state was not emphasized. Moreover, in recent work measurements from different time points relative to different applied conditions were pooled in one statistical analysis, making it difficult to characterize particular cellular processes [42].

In the current work we studied the behavior of a protein under the regulation of a system that is essential for metabolism—the GAL system in pure galactose medium. Special attention was paid to culture preparation and growth. Under these conditions, it was shown that several simple mechanisms that could account for the distribution are in fact not the main sources of variation. In particular, the biochemical noise in gene expression, heterogeneity in cell size and variations in promoter copy number are not the main mechanisms underlying the observed variation. From the population point of view, it is the total variation that is functionally important; we have therefore directed the focus of our research from tracing and separating molecular mechanisms to a phenomenological approach which emphasizes dynamics of distributions along time.

We have characterized the total variation in the population in terms of the protein distributions, their steady states and dynamics in response to perturbations over intermediate (∼ten generations) time scales. We found that steady-state distributions of a regulated essential system are universally broad, in contrast to a naïve single-cell picture in which each and every cell must strongly 'turn on' expression under inducing conditions. The shape of the steady-state distribution was independent of medium composition and growth characteristics over a wide range of conditions. Interestingly, the broad exponential-tail distribution is not unique to the budding yeast but was also observed in a

different regulatory system in bacteria dividing by fission [43]. Such common behavior of the distribution tails in different regulatory systems supports the view that it is not the internal cellular processes that determine the steady-state distribution but rather a more general population balance of processes.

We have found in several different experiments a marked asymmetry between the relaxation dynamics of subpopulations depending on their initial conditions: transitions are much slower for subpopulations that need to ignite protein sources and build the steady-state expression distribution from below than for those that need to dilute high stocks of proteins and build the steady-state distribution from above. This asymmetry clearly shows that, under conditions where the underlying regulatory system is functionally essential, the distribution is not a realization of an ensemble of independent particles, each with its own internal noise source. Rather, it demonstrates the important role of population dynamics in constructing and maintaining the steady state. The distribution is a result of a balance between a source of protein production, which is intracellular and noisy, and a strong dilution process that dissipates a constant fraction of the protein content in each generation by cell division.

The population point of view presented here calls for a quantitative understanding of gene regulation at the population level, complementing the single-cell level. At the population level, the history of the population and the mechanisms of protein inheritance among generations play crucial roles in determining the main features of the protein expression distributions. These features are most crucial for the population fate, since the expression distribution provides the substrate for selection and adaptation.

Acknowledgments

We would like to thank T Dror for help in measurements, E Stolovicki, S Behar, S Stern, T Friedlander and Y Shokef for useful discussions. This research was supported by the FIRST (Bikura) program of the Israel Science Foundation and by the Center for Complexity Science.

Appendix. Histogram analysis—relation between standard deviation and mean

During induction and repression, the standard deviation and mean of the protein distributions obey the linear relation $\sigma = a\mu - b$, with positive constants a and b (fitting lines in figures 5 and 6; $a \cong 1.3$ and $b \cong 50$ in both cases). Accordingly, the ratio σ/μ is an increasing function of μ:$\sigma = a - b/\mu$ (fitting lines in figures 5 and 6). This relation is consistent with the population distribution being composed of two independent parts: an exponential tail that is sensitive to induction and repression, and a narrower 'head' of the distribution that is less sensitive to protein production: $x = x_0 + x_1$. Assuming that $\langle x_0 x_1 \rangle = \langle x_0 \rangle \langle x_1 \rangle$ (statistical independence), $\sigma(x_1) \approx \langle x_1 \rangle$ (exponential tail) and $\sigma^2(x_0) \ll \sigma^2(x_1)$ (much narrower head), one can estimate the standard deviation of the total random variable: $\sigma(x) \cong \sigma(x_1) = \langle x_1 \rangle = \langle x \rangle - \langle x_0 \rangle$. This is exactly the relation observed in the experiments with $a = 1$.

Glossary

Phenotypic variation. Differences among cells or organisms other than in their DNA sequences. Phenotypic variation can exist among cells in clonal (genetically homogeneous) populations grown under homogeneous conditions and thus do not stem from differences in the genomes of these cells.

Gene regulation. The processes controlling the expression of genes from their coding in DNA sequence into proteins.

Transcription. The process in which regulatory proteins (transcription factors) interact with each other and bind to specialized DNA sequences to control the process of gene expression. This is the first and usually most significant step in gene regulation.

Reporter gfp. A green fluorescent protein that is inserted into a cell under the control of a particular regulatory system that is of interest. This way, the gene regulation system controls the expression of a protein that is quantitatively measurable by optical methods in live cells. These measurements provide useful information about the regulation system: its level of activity, dynamics, sensitivity etc.

GAL system. A group of genes responsible for utilizing galactose as a source of energy and carbon in the yeast *S. cerevisiae*. These genes have common components to their regulation system, and are known to respond directly to the type of sugar available to the cells.

FACS. Fluorescence activated cell sorter. A technique for measuring certain cellular properties (such as the expression of a fluorescent protein) at single-cell resolution in large cell populations. The distribution of these properties in the population can then be estimated, and the cells can be physically sorted to different groups according to these properties.

Chemostat. A technique for growing cell populations (e.g., of microorganisms) in continuous culture. The proliferating population is fed at a constant rate, and cells and medium are diluted out at the same rate preserving the volume fixed. The population can be grown under fixed environmental conditions and can achieve a steady state in which the number of cells in the growth chamber is also fixed.

References

[1] Spudlich J L and Koshland D E J 1976 Non genetic individuality: chance and the single cell *Nature* **262** 467–71
[2] Wheals A E and Lord P G 1992 Clonal heterogeneity in specific growth rate of *Saccharomyces cerevisiae* cells *Cell Prolif.* **25** 217–23
[3] Korobkova E *et al* 2004 From molecular noise to behavioural variability in a single bacterium *Nature* **428** 574–8
[4] Raser J M and O'Shea E K 2005 Noise in gene expression: origins, consequences, and control *Science* **309** 2010–3

[5] Kaern M *et al* 2005 Stochasticity in gene expression: from theories to phenotypes *Nat. Genet.* **6** 451–64
[6] Berg O G 1978 A model for the statistical fluctuations of protein numbers in a microbial population *J. Theor. Biol.* **71** 587
[7] Elowitz M B *et al* 2002 Stochastic gene expression in a single cell *Science* **297** 1183
[8] Ozbudak E M *et al* 2002 Regulation of noise in the expression of a single gene *Nat. Genet.* **31** 69
[9] Blake W J *et al* 2003 Noise in eukaryotic gene expression *Nature* **422** 633
[10] Banerjee B *et al* 2004 Tracking operator state fluctuations in gene expression in single cells *Biophys. J.* **86** 3052–9
[11] Raser J M and O'Shea E K 2004 Control of stochasticity in eukaryotic gene expression *Science* **304** 1811
[12] Braun E and Brenner N 2004 Transient responses and adaptation to steady state in eukaryotic gene regulation system *Phys. Biol.* **1** 67
[13] Volfson D *et al* 2006 Origins of extrinsic variability in eukaryotic gene expression *Nature* **439** 861–4
[14] Yanofsky C 1992 Transcriptional regulation: elegance in design and discovery *Transcriptional Regulation* (New York: Cold Spring Harbor Laboratory Press) p 3
[15] Holstege F C and Young R A 1999 Transcriptional regulation: contending with complexity *Proc. Natl. Acad. Sci. USA* **96** 2
[16] Struhl K 1999 Fundamentally different logic of gene regulation in eukaryotes and prokaryotes *Cell* **98** 1
[17] Kornberg R D 1999 Eukaryotic transcriptional control *Millenium issue TCB TIBS TIG* **M46**
[18] Biggin M D 2001 To bind or not to bind *Nat. Genet.* **28** 303
[19] Atauri P de *et al* 2005 Is the regulation of galactose 1-phosphate tuned against gene expression noise? *Biochem. J.* **387** 77–84
[20] Pedraza J M and Oudenaarden A van 2005 Noise propagation in gene networks *Science* **307** 1965–9
[21] Kacser H and Burns J 1981 The molecular basis of dominance *Genetics* **97** 639–66
[22] Nijhout H F *et al* 2004 A mathematical model of the folate cycle—new insights into folate homeostasis *J. Biol. Chem.* **279** 55008–16
[23] Brauer M J *et al* 2005 Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures *Mol. Biol. Cell* **16** 2503–17
[24] Booth I R 2002 Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress *Food Microbiol.* **78** 19–30
[25] Stolovicki E *et al* 2006 Synthetic gene-recruitment reveals adaptive reprogramming of gene regulation in yeast *Genetics* **173** 75–85
[26] Ptashne M 1992 *A Genetic Switch: Phage l and Higher Organisms* (Cambridge, MA: Cell Press)
[27] Krishna S *et al* 2005 Stochastic simulations of the origins and implications of long-tailed distributions in gene expression *Proc. Natl. Acad. Sci.* **102** 4771–6
[28] Colman-Lerner A *et al* 2005 Regulated cell-to-cell variation in a cell-fate decision system *Nature* **437** 699–706
[29] Lachmann M and Jablonka E 1996 The inheritance of phenotypes: an adaptation to fluctuating environment *J. Theor. Biol.* **181** 1–9
[30] Thattai M and van Oudenaarden A 2004 Stochastic gene expression in fluctuating environments *Genetics* **167** 523
[31] Kussell E and Leibler S 2005 Phenotypic diversity, population growth and information in fluctuating environments *Science* **309** 2075–8
[32] McAdams H H and Arkin A 1997 Stochastic mechanisms in gene expression *Proc. Natl. Acad. Sci.* **94** 814–9
[33] Swain P S, Elowitz M B and Siggia E D 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression *Proc. Natl. Acad. Sci. USA* **99** 12795
[34] Paulsson J 2004 Summing up the noise in gene networks *Nature* **427** 415
[35] Li J *et al* 2000 Green fluorescent protein in *Saccharomyces cerevisiae:* real-time studies of the GAL1 promoter *Biotechnol. Bioeng.* **70** 187
[36] Newman J R S *et al* 2006 Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise *Nature* **441** 840–6
[37] Griggs D W and Johnston M 1991 Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression *Proc. Natl. Acad. Sci. USA* **88** 8597
[38] Lohr D, Venkov P and Zlatanova J 1995 Transcriptional regulation in the yeast GAL gene family: a complex genetic network *FASEB* **9** 777
[39] Harbison C T *et al* 2004 Transcriptional regulatory code of a eukaryotic genome *Nature* **431** 99–104
[40] Jones E W, Pringle J R and Broach J R 1992 *The molecular and cellular biology of the yeast saccharomyces* vol 2 (New York: Cold Spring Harbor Laboratories Press)
[41] Zhang Z, Young M and Chisti Y 1996 Plasmid stability in recombinant Saccharomyces cerevisiae *Biotechnol. Adv.* **14** 401–35
[42] Bar-Even A *et al* 2006 Noise in protein expression scales with natural protein abundance *Nat. Genet.* **38** 636–43
[43] Salman H and Libchaber A 2005 Private communication