Chapter 23

Noise in Biological Systems: Pros, Cons, and Mechanisms of Control

Yitzhak Pilpel

Abstract

Genetic regulatory circuits are often regarded as precise machines that accurately determine the level of expression of each protein. Most experimental technologies used to measure gene expression levels are incapable of testing and challenging this notion, as they often measure levels averaged over entire populations of cells. Yet, when expression levels are measured at the single cell level of even genetically identical cells, substantial cell-to-cell variation (or "noise") may be observed. Sometimes different genes in a given genome may display different levels of noise; even the same gene, expressed under different environmental conditions, may display greater cell-to-cell variability in specific conditions and more tight control in other situations. While at first glance noise may seem to be an undesired property of biological networks, it might be beneficial in some cases. For instance, noise will increase functional heterogeneity in a population of microorganisms facing variable, often unpredictable, environmental changes, increasing the probability that some cells may survive the stress. In that respect, we can speculate that the population is implementing a risk distribution strategy, long before genetic heterogeneity could be acquired. Organisms may have evolved to regulate not only the averaged gene expression levels but also the extent of allowed deviations from such an average, setting it at the desired level for every gene under each specific condition. Here we review the evolving understanding of noise, its molecular underpinnings, and its effect on phenotype and fitness – when it can be detrimental, beneficial, or neutral and which regulatory tools eukaryotic cells may use to optimally control it.

Key words: Noise in gene expression, noise control mechanisms, regulatory networks, network biology.

1. Relevance of Noise in Biological Systems

Since the earliest discoveries of the basic mechanisms that control gene expression, biologists have been intensively engaged in measuring mRNA and protein levels. Such studies, driven by a

J.I. Castrillo, S.G. Oliver (eds.), *Yeast Systems Biology*, Methods in Molecular Biology 759,
DOI 10.1007/978-1-61779-173-4_23,

broad range of technologies, have established links between the programs that regulate gene expression and the corresponding molecular phenotype – the level of expression of each individual gene and its response to external signals and, ultimately, to the phenotype of the organism. The advent of genomics revolutionized the study of gene expression: technologies such as DNA and protein microarrays, and most recently, RNA sequencing enables the measurement of gene expression for every gene, providing rich information about transcription (1), mRNA degradation (2), translation (3), RNA editing (4), and more. Examination of control regions of genes will enable researchers to decipher the regulatory programs that underlie the observed behavior.

However, common to all such technologies is the fact that measurements represent averaged RNA and protein levels over large populations of cells. For instance, a typical microarray experiment requires more than 100,000 cells. Such measurements thus provide very reliable estimates of the *average* level of expression of a given gene over an entire population of what are typically genetically identical cells. If all cells in a population expressed a given gene at the same or even very similar levels, then the average alone would, indeed, capture the reality well. However, if different cells in the population expressed a given gene at different levels, then information about cell-to-cell variation would be lost. If cells take control of the extent to which they allow or restrict such diversity, deciphering the mechanisms that exert such control would require alternative models and technologies.

From a theoretical perspective, what is the potential for cell-to-cell variation in gene expression levels, among genetically identical cells? Stochasticity and randomness govern the microscopic world inside cells, the world of molecular recognition that is driven by interactions between molecules in a crowded environment. The effects of such random events may be particularly dramatic when it comes to molecules that are represented in just a few copies per cell, as is the case with many regulators of gene expression. For instance, if a particular regulator is present, on average, in two copies per cell, we should not be surprised to find cells that have four copies and others with one or even zero copies of this molecule. It is also easy to imagine that the targets of such a regulator would also display corresponding, and perhaps even greater, fluctuations as a result. Cell-to-cell variation at the level of gene expression may constitute a real possibility.

What would be the interest of studying this possible variation? We might think that if the population averages out such fluctuations then they will have no functional consequence. Here we will argue for the converse. Consider, for instance, a population of genetically identical *Escherichia coli* cells that are attacked by an antibiotic drug. While the majority of the population may die, a portion may survive the attack. Note that we do not consider

here the case of resistant cells that are genetically different from the majority of the population; rather, only those *persisters* (5) that are genetically identical to the rest of the cells. Further investigation revealed a stochastic split of the population into two sub-populations, one that is sensitive to the drug and another that is not, and a purely random event that allows cells to switch from one fate to the other (5). Another example, for instance, sporulation in yeast, a potential defensive response to stress, manifests itself in only a portion of genetically identical cells in a population. It turns out that a single protein stochastically expressed at varying levels in different cells determines the different cellular fates (6). To take examples from multi-cellular organisms, a combined theoretical–experimental approach involving immune T-cells found that stochastic variation in the expression of key signaling proteins generates substantial cell-to-cell variability in the antigen responsiveness needed among cells in a clonal population (7). Another example relates to populations of cells in tumors. After chemotherapy treatment the majority of the cells may die, but a smaller sub-population may survive. Recently, an individual protein was found that, due to stochastic effects, may be temporarily present at varying levels in particular cancer cells (8). As a consequence, some of the cells are rendered resistant to a drug, whereas others are unable to survive the same therapy. In all these systems, a seemingly random event at the molecular level – often the choice of expression level for key genes in particular cells – gave rise to dramatically different phenotypes at the cellular and organism levels. Without measuring the levels of relevant proteins at the population level and the average over all the cells in the population, the underpinnings of the sophisticated environmental response can be altogether missed.

The paradigm that emerges is that certain critical biological phenomena, such as drug persistence, stress response, immune response, and cancer cell proliferation, are rooted in stochastic molecular events, which ultimately lead to phenotypic variation among genetically identical cells. Yet these cases might represent the exception rather than the rule. The perception that cells tightly control the expression of their genes, to such an extent that cell-to-cell variation would be limited, may indeed be correct in the case of many genes that must be expressed at precisely fixed levels. The dozens of proteins that make up large macromolecular complexes such as the ribosome, for example, should conceivably be kept under tight control, not least to eliminate wasteful production. It is thus conceivable that cells may have evolved mechanisms to determine the extent to which they can safely permit variations in the expression of some genes in their genomes and restrict variation in others.

Stochastic variation in gene expression levels among genetically identical cells grown under the same conditions is often

dubbed "noise" (9–12). Quantification of noise in gene expression often requires measurements of mRNA or protein levels in single cells. In recent years, experimental techniques to measure gene expression in single cells at high throughput and with great accuracy have matured (9–13, 18), along with theoretical models to rationalize the results (14, 15). The aims of those studying noise in gene expression are thus to measure noise in various biological setups, to decipher the regulatory means by which cells control noise, and to study the biological consequences of such non-genetic variation. Another substantial bonus from the study of noise comes from the fact that the statistical properties of noise can reveal basic principles of the molecular processes that govern gene expression. For instance, by analysis of noise spectra it was possible to deduce that gene expression occurs in bursts, whose two key parameters are the burst size and frequency (15). This information is essential to understanding noise behavior but its implications extend to the basics of gene expression mechanisms.

2. Noise in Gene Expression

Consider the expression level of protein X in a unicellular organism such as yeast in the following thought experiment: Let us measure its expression in two genetically identical cells. Since the two cells are genetically identical, we might expect that the expression level of our protein would be identical in both. Suppose that we have measured the actual copy number of protein X in each cell and found it to be twice as high in one of the cells, compared to the other. Our first suspicion might be that a measurement error occurred or, to put it more quantitatively, that the measurement error is larger than the true variation between our two cells. Let us then assume that the difference is reproducible, even after many repetitions; moreover, it is not seen with respect to another protein, Y, which on average shares the same expression level as X.

So, why the level of expression of protein X appears to differ so significantly between the two cells? First, even if our two cells are identical at the DNA level, errors do occur at the level of transcription and translation: on average, 1 in every 10,000 transcribed nucleotides, and 1 in every 1,000 translated amino acids, is expected to be wrong (16, 17). Such errors may affect the expression level of a protein, e.g., by affecting the stability of the mRNA and the protein or by affecting protein X's regulators. A dramatic phenotypic consequence of such errors was recently demonstrated in a study in which transcription error rate was increased by a positive feedback (17).

Let us assume that in the two cells, all copies of X were transcribed and translated without a single error. What other reason may be responsible for the different levels of expression of protein X? The cells might have been subjected to different "micro-environments" (for instance, one might have been closer to the edge of a colony or they might have been at different stages of the cell cycle). Let us further assume that the two cells had been exposed to the same micro-environment, they were synchronized relative to the cell cycle, and they were also equal in size and mass. Remarkably, the two cells might still express protein X in varying amounts, because they could differ at the microscopic level (for instance, the existence of two copies of a transcription factor controlling the expression of the X gene in one cell, while three in the other). Let us assume that the regulators are found at the same level in all cells too. Moreover, the two cells are identical not only with respect to DNA, mRNA, and protein sequence but also with respect to the concentration, location, and molecular dynamics of every molecule within them, including every transcription factor and ligands. From this perfectly identical starting point, let each of the cells live out its natural lifespan. After a short while, would our cells show identical or different levels of protein X? Even under these "identical" conditions, basic physical chemistry shows that differences might still be possible. All processes involving the propagation of genetic information, including the unwinding of the DNA for transcription, transcription itself, processing of RNA, degradation of transcripts, translation, protein modification, and protein degradation, are based on interactions between molecules and inevitably include a stochastic component (i.e., while in one cell the transcription factor may initiate, e.g., two transcription events at a given time interval, in the other it may occur only once. The binding constant between that factor and the promoter of the gene encoding X may be the same in the two cells, but this is merely a macroscopic constant that relates to a ratio of probabilities). Of particular interest is the fact that small stochastic differences may be amplified or canceled out due to further random events (for instance, in one cell each transcript is translated 10 times before it is degraded, resulting in 20 copies of protein per cell and only 10 in the other). An example in which the initial difference between the two cells may be diminished may be that RNA degradation, often mediated by the binding of an RNA binding protein, may occur faster in the first cell with higher transcript levels.

In summary, there are many mechanisms by which two genetically identical cells may express a specific protein (protein X) at different levels. Can we measure the effect of stochasticity if all other factors such as genotype and cell size are kept constant or due to environmental changes (for instance, by applying environmental stress)? If so, what patterns would be expected for the

levels of protein X and other proteins (for example, protein Y, expressed on average at the same level as protein X across all cells but with different functions and involved in a different regulatory network, or protein Z, expressed at higher levels)? More importantly, can noise play a part in central biological processes and be responsible for cell fates and specific phenotypes that could be investigated? For example, do cells set different noise levels for each gene, according to both its function and environmental conditions? Can cells and organisms control the extent of variation and stochasticity so as to minimize it when harmful or amplify it and benefit from it when possible?

3. Measuring Noise in Living Cells

A landmark work that transformed the study of cellular stochasticity into a quantitative biophysical science was that of Elowitz et al. (18). While finding two cells with the same concentration, location, and dynamics of all molecules is still impossible, they found a clever practical solution. By using two fluorescent proteins (cyan and yellow) under the control of two identical promoters, placed in two similar locations in the *E. coli* genome, they generated a cellular environment that was practically identical for the two genes – thus, not two identical cells with the same gene as in our thought experiment (see previous section), but rather the same cell that serves the expression of two distinguishable genes. Since the two proteins fluoresce in separate colors, comparing the intensity from the two channels in the same cell was sufficient to identify differences that must be attributed, for the most part, to stochastic events that happened inside that cell. Elowitz and coworkers then used fluorescent-activated cell sorting (FACS) to measure fluorescence in the two channels that correspond to the two genes in individual cells. The difference in expression between a pair of two such proteins expressed in the same cells was termed the "intrinsic noise" of the system – only stochastic processes within each cell could give rise to differences in the expression level of these two probes in each cell. With the same experimental method, focusing on one wavelength at a time, the researchers could then compare the different cells and examine the variations among them at the level of protein expression. This variation was called the "extrinsic noise," as it captured external sources of variation, such as varying concentrations of a relevant transcription factor and number of ribosomes in each cell. It was then possible to estimate not only each noise source but its relative contribution to the final level of variation. The authors concluded that extrinsic noise was a major factor, but that intrinsic

noise had a significant contribution too to the overall cell-to-cell variation. Applying a similar system to *Saccharomyces cerevisiae*, Raser and O'Shea were able to measure the relative contribution of intrinsic and extrinsic sources of noise in this model eukaryote and found that the intrinsic noise is gene-specific and may constitute an evolvable trait that can be optimized to balance fidelity and diversity in eukaryotic gene expression (19).

4. Expanding the Scope: Measuring Noise for Many Genes and Conditions

The studies described thus far have focused on exogenous genes and measured noise levels under a limited set of environmental conditions. Yet to penetrate deeper into the biological significance of noise, three additional steps were necessary: first, to measure noise of endogenous genes, insofar as was possible, with minimal interruption of their native controls; second, to scale up measurements so as to probe as many genes as possible; and third, it was essential that the noise of a given native gene would be measured under varying growth conditions. Bar-Even et al. (20) and Newman et al. (21) accomplished these aims in a complementary fashion. For this purpose, they used a library of *S. cerevisiae* strains, each expressing one of the endogenous genes of this species (22) fused to a green florescent protein (GFP). Examining the fluorescence of cells from each such strain by FACS, it was possible to measure the cell-to-cell variations in expression of each gene in the genome. Here, since a single type of fluorescent protein (GFP) was used, intrinsic and extrinsic noise were no longer separated and the integrated contribution from the two sources of variation was measured as a single number [yet when a sample of the genes was also measured in two colors, a predominant contribution from the intrinsic noise was actually found (20)]. In all, Bar-Even and colleagues studied 43 genes; yet they examined noise under a diversity of conditions. These authors selected their sample genes so as to represent several genetic modules that were originally defined, based on classical microarray experiments: these included stress-related genes, genes encoding structural constituents of the proteasome, genes involved in ergosterol metabolism, and genes responsible for the processing of ribosomal RNA (rRNA). They then exposed the cells to 11 different conditions, the majority of which were stressful, yet a few actually involved recovery from stress conditions. Finally, they measured, for each gene, the distribution of expression values under each condition at the single cell level. Newman and colleagues did not explore as many conditions; instead, they measured the expression of a most significant portion of the

entire yeast genome (21). The two papers focused on two parameters that characterized each gene: the mean of the distribution and the "noise coefficient" – namely, the variance divided by the square of the mean. Plotting the noise coefficient vs the mean expression, the two groups found the same intriguing, "scaling" relationship: the noise declined as the reciprocal of the mean of the distribution. That is, not only were highly expressed genes less noisy (as might be expected, given the mean of the distribution, with the value obtained from a standard microarray experiment) it was possible to rather accurately predict the amount of noise for most genes under most conditions. Theoretical analysis in both works suggested that such scaling might result from variations in mRNA copy numbers, caused by the stochasticity of "birth and death" of mRNA molecules or from fluctuations in promoter activity (20, 21).

If noise can be inferred from the mean, could we avoid measuring noise in future experiments and settle for the more conventional measurements? The fact is that, although most genes under most conditions obeyed the scaling law, interesting deviations were found. The "noise residual" was thus defined as the difference between the amount of noise that a gene actually displayed and the amount of noise that could be predicted for the gene, given its mean expression and the general scaling between noise and mean that holds true for most genes in most conditions (**Fig. 23.1**). For instance, stress-related genes were consistently above the scale (i.e., for these genes, noise was typically higher than the expected value, given their own mean and the general scaling).

What could be the rationale behind this enhanced noise in stress genes? One intriguing possibility is that cells implement a "risk distribution strategy" with these genes – that different cells in the isogenic population provide stochastically different "responses" (i.e., expression levels of these genes). According to this hypothesis, the cells that happened to express these genes at the optimal level would be more likely to survive. In fluctuating environments such approach might, under some circumstances, constitute the most feasible strategy (23). Note, however, that since the cells are genetically identical, such changes would not be inherited (see below on the "memory" of such fluctuations and on the combination of genetic and non-genetic diversity). The fact that in the experiments where stress was alleviated (20), the stress genes typically showed reduced noise (**Fig. 23.1**) supports that control of the noise in these genes may constitute a cellular response to changes in environmental conditions from non-stressful to stressful conditions and vice versa.

Examination of the response of other genes revealed an opposite trend: negative noise residuals (**Fig. 23.1**). Take, for instance, the genes encoding constituents of the proteasome, a multi-subunit cellular complex. Under stress conditions, these

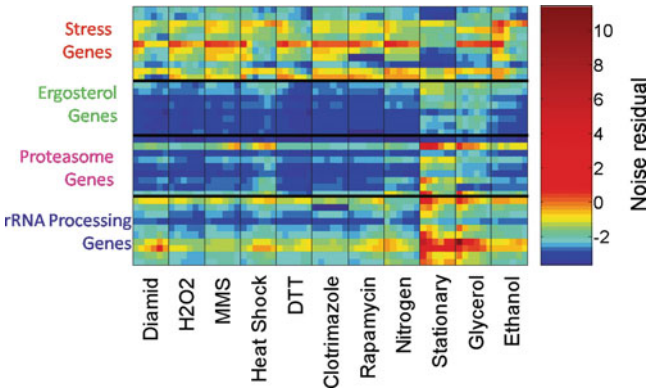

Fig. 23.1. Variations in levels of noise of specific yeast genes under different conditions. Levels of noise residuals of 43 *S. cerevisiae* genes in 11 different environmental conditions. Here, each *horizontal line* corresponds to a gene. Each condition is represented as a set of consecutive columns with each column corresponding to a time point. The noise was measured at six time points following the environmental change (20), and noise propagation can be traced for each gene at each condition. The color code depicts the "noise residual" of each gene in each condition, namely the difference between the actual noise level of a gene and the noise level expected for that gene given its mean expression level and the general scaling between noise and mean expression (20). The 43 genes were sampled from the entire yeast genome and they represent four modules: stress genes, ergosterol metabolism, constituents of the proteasome, and genes involved in the processing of ribosomal RNA (rRNA). The environmental conditions represent different perturbations and stress relaxing conditions (1st through the 7th, and the 11th sets of columns, and 8th–10th column, respectively) (20). The stress genes show higher noise levels throughout the conditions, especially under stress, while genes involved in ergosterol metabolism and proteasome and rRNA biosynthesis show noise being kept at controlled, low levels, particularly under stress conditions.

genes featured very tight distribution, with negative noise residuals. The example of genes encoding structural constituents of the proteasome shows that in some cases high levels of noise may be actually undesirable or should be kept under tight, controlled levels, for instance, when a fine coordination and stoichiometry of synthesis of specific subunits of a multi-subunit complex is necessary. In other genes in which the extent of noise implied by the mean is neither helpful nor detrimental, cells may not attempt to control the levels, and noise may be set by simple probabilistic rules (21).

5. How Cells Control Noise and Set a Desired Level for Each Gene

The aforementioned studies (18–21) show that cells appear to be able to set the noise levels and their enhanced or reduced extent compared to mean expression values. Genes that belong to the

same functional category or structural complex often show similar changes in noise under a specific condition. Moreover, a given gene may show different levels of noise amplification or reduction under different conditions. This requires to be finely regulated. In recent years, new studies are revealing a rich array of strategies to regulate noise levels in cells. The means to control cellular noise can be related to the kinetic parameters of the processes governing gene expression and the topology of the regulatory networks and can sometimes be inferred from the genomic features inside and in the proximity to genes.

5.1. A Model for the Propagation of Noise

As an alternative to the works by Elowitz and colleagues (18) and Raser and O'Shea (19) which measured intrinsic and extrinsic noise, Paulsson took a different approach and developed a mathematical model that described the propagation of noise in a cellular pathway (14). Consider two molecules, A and B, such that A is either a regulator of B or that A and B are, respectively, the mRNA and protein encoded by a particular gene. In either of these cases, noisy fluctuations in A might be further propagated into B. Focusing on the downstream component B, Paulsson suggested an alternative to the dichotomy between intrinsic and extrinsic noise, realizing that the noise in B would arise from a combination of two components: the noise generated by B itself and the noise that B "inherits" from A. While Paulsson's model was predominantly theoretical, a similar conclusion was reached by Pedraza and van Oudenaarden (24) who experimentally measured expression correlations between genes in single cells. These authors also found that noise in the expression of a gene was determined by its intrinsic fluctuations, noise transmitted from upstream genes, and global noise affecting all genes.

Understanding the contribution of noise in A to noise in B is of particular interest, since such knowledge enables the description of noise propagation along genetic chains. According to Paulsson's model, one relevant parameter that governs such propagation is the response dynamics of A and B. Intuitively, if A is a very rapidly changing molecule, then B will "inherit" the fluctuation only if B, too, is rapidly fluctuating. If, on the other hand, B has a very slow rate of turnover, it will not trace the fluctuations in A over time and will thus not inherit the noise (i.e., B will be said to have "time-averaged" fluctuations in A). What governs the response times of specific molecules? Some response times may be largely governed by the degradation kinetics. Consider an mRNA and a protein encoded by a given gene as the "A" and "B" molecule in the above formalism. If the protein had was rapidly fluctuating (e.g., due to a relatively high degradation rate) then noise at the mRNA level would be effectively propagated to the protein. In recent years, techniques to measure the stability of both mRNA (2, 25) and protein (26) at the genome level

have begun to mature, and comprehensive studies to test the possibility that noise propagation can be deduced from elementary parameters and other additional factors affecting the accuracy at which fluctuations might be propagated (14) are becoming more feasible.

5.2. The Relative Efficiency of Transcription vs Translation

The extent (efficiency) of transcription and translation of a given gene and the relative contribution (ratio) between them deserve special attention. As a case example, assume that a given protein is needed at an average of 200 copies per cell. Imagine two extreme (probably not very "realistic") strategies to obtain that desired level of protein expression: The first one is when the transcription of the corresponding mRNA takes place at a very low extent: so, for instance, only two mRNA molecules are present, on average, per cell, requiring each one to be translated, on average, 100 times. The other extreme is that transcription is extensive, resulting in 200 copies of the corresponding mRNA. To get the 200 proteins, in principle, it will be enough that each mRNA will be directed to the protein biosynthesis machinery only once (less if polyribosomes are acting).

The first case, with a very low transcription rate and the need for extensive translation, is economical in terms of RNA synthesis, but what will happen to the noise at the protein level? While the mRNA is present, on average, at two copies per cell, fluctuations with one, three, or four copies are likely. Translation has the potential to further amplify such fluctuation, yielding high predicted noise. Also, the need to reuse the same mRNAs for translation may lead to a slow global response, with low rates of protein biosynthesis (in some cases higher than the mRNA and protein turnover, in which case the degraded molecules will need to be re-synthesized). Target protein levels may be difficult to achieve in some cases, with high predicted noise.

On the other hand, a higher, efficient transcription (which will provide a "pool" of ready-to-use mRNAs) coupled with limited translation (which may be controlled by different mechanisms: polyadenylation, subcellular localization, and polyribosome levels) will result in few fluctuations and a relatively noise-free protein population. This strategy may also ensure fast responses in shorter times and, once target levels are obtained, the possibility of activation of mechanisms of control (e.g., feedback regulation or others, see below). Efficient, extensive transcription coupled with balanced translation can provide quick and fine control of protein levels.

From here, a simple prediction might be that cases of extensive transcription coupled with limited translation would exhibit low noise and high level of control of expression. The possibility that low levels (less efficient) of transcription coupled with extensive translation may constitute in some cases a means to

enhance noise levels should not be discarded. There are several lines of support for the argument that the extent of translation (translation levels) is directly linked to noise levels. Ozbudak and coworkers independently modified the transcription and translation of a reporter gene and found that an increase in translation mainly increased noise level (11). Black et al. also provided a convincing demonstration of the idea (12), where they changed the nucleotide coding sequence of a gene without affecting amino acid sequence and shifted it toward higher or lower translation efficiency codons. They found that noise indeed increases with translation extent (efficiency), provided that the gene was not fully induced at the transcription level. Interestingly, these results could also be predicted from a theoretical model that these authors provided that deals with the propagation of noise from the mRNA to the protein level (12). In accordance with this, Fraser et al. (27) have also found a related trend in yeast: essential genes and genes involved in cellular complexes tend to minimize their predicted noise level by employing a strategy that maximized the ratio of transcription to translation. Furthermore, an inspection of the noise residuals of the 43 *S. cerevisiae* genes measured by Bar-Even et al., against their sequence-based calculation of translation efficiency (the tRNA adaptation index) (28), shows a correlation between them (**Fig. 23.2**). Genes with high

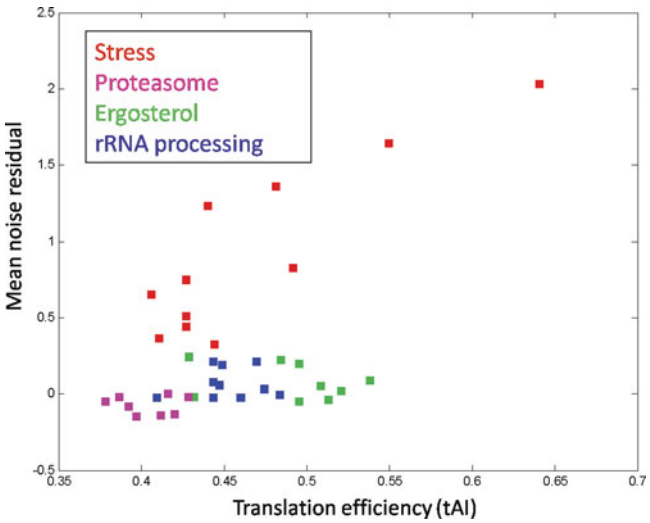

Fig. 23.2. Correlations of noise residuals and tRNA adaptation index. Noise residuals correlate with the tRNA adaptation index of yeast genes, particularly with genes involved in response to stress. Here, the same set of genes as in **Fig. 23.1** is analyzed, with colors depicting association to the four modules in **Fig. 23.1**. The noise residual of each gene is calculated as the mean of its values across all time points in all six conditions. The tRNA adaptation index (tAI) of each gene was calculated as in (28). The tAI captures the extent to which the codons in a gene are biased toward the more abundant tRNAs in the genome (high translated genes).

predicted translation efficiency (highly translated) indeed show some tendency toward higher noise residuals, whereas genes with low predicted efficiency of translation corresponded to the least noisy genes. Of particular interest are the proteasomal genes that under stress conditions display little noise, in fact even lower than expected from their means. These genes, under stressful conditions, are typically induced at the mRNA level (29). The results suggest that a means to obtain tight, noise-filtered distribution for proteasomal genes under stress is to employ the strategy of high transcription rate with limited, coupled translation. Similar conclusions have been reached by a mathematical treatment of a related system (30).

At the molecular level, the TATA box in gene promoters was found to be another key factor with which cells may tune noise levels. The TATA sequence is not present in every gene's promoter, but genes that do contain it typically show high levels of noise (19, 20, 31, 32). The prevailing explanation is that the TATA sequence amplifies fluctuations through facilitation of transcription re-initiation (33, 34).

5.3. Local Network Connectivity, Redundancy, and the Effect on Noise

An additional regulatory attribute that affects the noise level and pattern displayed by genes is their connectivity within the regulatory network. Consider two genes that are identical with respect to many of the aforementioned properties, such as the efficiency of their translation and transcription, but are nonetheless embedded in two different regulatory network motifs. In one case, the gene exerts negative feedback regulation on itself (either directly or through a mediator) and in the other, the gene exerts a positive feedback on its own level of expression. In which case would the gene manifest a higher amount of noise? Intuitively, the negative regulatory scheme would tend to counteract, or "correct," noisy fluctuations – when the amount of the autoregulated gene stochastically increases the negative regulation will counteract the fluctuation by increasing the extent of inhibition, while a fluctuation in the other direction would result in lower inhibition (10). In the positive feedback case, fluctuations in either direction are expected to intensify themselves, resulting in higher noise levels. Theoretical work, however, has recently elegantly refined these notions, predicting that while negative feedback eliminates noise, it comes at a price: reduced sensitivity to changes in environmental signals (35). When comparing circuits with the same level of sensitivity to environmental changes, it was found that a positive feedback design actually buffers noise better than negative feedback. It was further suggested that the improved capacity of a positive feedback circuit to buffer noise at a given level of environmental sensitivity comes from its time-averaging capacity (35).

Most recently, researchers faced an intriguing question pertaining to the architecture of regulatory networks: often two

or more alternative circuit designs can produce the same outcome – in particular a negative regulatory effect on a gene can be obtained either by repressing the inducer of the gene or by inducing its repressor. Why is it that in reality one of the designs, and not the other, appears to operate in a particular system (36)? The authors focused on the genetic network that governs competence – the ability of microbes to uptake DNA from the environment, typically upon stress – in *Bacillus subtilis*. Like any regulated system it needs to control not only induction but also its shutoff. It turns out that the shutoff of the competence network in these bacteria is obtained by repressing the system's inducer, not by inducing a potential repressor. The authors synthesized a seemingly equivalent variation on the circuit in which shutoff happens through induction of a repressor. While the averaged properties of the native and synthetic designs were designed to be similar, the native circuit showed enhanced diversity between single cells, while the synthetic design was precise and relatively noise free. Why did evolution prefer the nosier version? The fact is that the accuracy of the synthetic design came at a price – bacteria that artificially expressed it were fit (i.e., could take up DNA from the environment) only at a narrow range of environmental parameters, compared to the cells that expressed the native system. The higher noise obtained in the native system thus appears to be adaptive as it allows higher population diversity in variable situations (36). Presumably, the negative regulation of the inducer in the native circuit is responsible for a lower expression level, possible high relative noise level, in this regulator. It was concluded that noise actually facilitates the response of the network to a variable environment. The more precise synthetic design, in which the repressor is induced, is more similar to circuits such as at the heart of the circadian clock, where accuracy and control are the main issue.

Apart from the connectivity in regulatory networks, the challenge of noise control may have constituted a driving force explaining the unexpected conservation of redundancy in biological systems. It was recently suggested that partially redundant duplicate genes may have been selected for preservation in genomes, so as to filter noise in regulatory networks (37, 38). Why are redundant genes often preserved, especially in pivotal nodes of regulatory networks, if only one member of the gene pair would suffice? Although redundant genes can often back each other up if mutations occur in one of them, it is entirely possible that the partially redundant genes will cover for each other when, due to stochasticity, one of them showed a temporary fluctuation that either increased or decreased its level, a far more likely event.

In many cases of redundancy in regulatory proteins, it was found that the regulators also negatively regulate one another (37, 38). Such a design could serve to reduce the effect of

fluctuations: when the expression level of one of the two regulators goes up, it would further inhibit its partner, and when it is decreased, it would exert less of an inhibitory effect. Modeling results (37) show that the sum, or the product, of the concentrations of the two regulators may be kept relatively constant in such a regime; furthermore, if a joint target of the two regulators is only affected by the sum, or product, of the concentrations of its two upstream regulators, it may experience a low amount of noise. In other words, the negative regulatory cross-talk that is often observed between semi-redundant gene duplicates (*see* (37) for a review) may serve to transform noise fluctuations in each of them into a relatively constant sum (or product) of the two, thus minimizing further propagation of the noise. This design may also explain why, in the extreme case of a constant reduction of a gene's expression level, e.g., due to a deletion mutation in the gene, the semi-redundant duplicate may respond by increased expression, in compensation (38). Such a capacity might, in some cases, constitute a byproduct of a selectable tunable capacity to respond to random, temporal fluctuations in its counterpart's expression level (38).

5.4. A Potential Role of Non-coding RNAs

Non-coding RNAs are transcripts that are not translated into proteins. During the last decade, it has become apparent that in essentially every organism, a considerable portion of the transcribed RNAs are not translated. Many of these newly discovered non-coding RNAs function as regulators, that is, they may regulate gene expression at multiple levels. Two relevant examples of regulatory RNAs are microRNAs and antisense RNAs. Such transcripts are known to interact with their targets by means of base pair complementarity, mainly affecting the levels of stability of the RNA target (e.g., an mRNA), and the efficiency of its translation.

It is conceivable that a non-coding RNA would affect not only the average expression level of its target but also the noise that the target would display. A non-coding regulatory RNA present in excess, relative to its target, may actually serve to buffer noisy fluctuations in the target. Assume, for instance, that the regulatory RNA is present in a high average copy number, say 20 copies, in each cell in a population of isogenic cells. Now consider an mRNA target of this regulator that is gradually induced from very close to 0 copies to 50 copies per cell. The noise coefficient of that target would be high at the beginning of the induction process, due to its presence in low copy number. However, if the regulatory RNA efficiently sequesters the target when the latter is at a lower copy number, random fluctuations in the target would be dumped. Yet, as the target's concentration continues to rise, at some point it may exceed the level of the regulatory RNA, and from that point on, the buffer would be unable to effectively sequester the excess mRNAs. This would lead to a

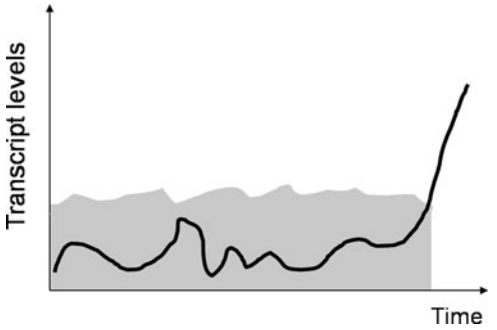

Fig. 23.3. A conceptual model of fluctuations and control of noise in RNA expression. The expression of a regulatory RNA (e.g., a microRNA or an antisense RNA) is depicted by the *gray area*, while the expression of the regulated transcript (e.g., an mRNA) is depicted with a *black line*. A regulatory RNA may filter noisy fluctuations in the levels of another regulated transcript (e.g., an mRNA). Levels of a regulatory RNA higher than the levels of the target RNA may effectively buffer noisy fluctuations (due to specific binding, sequestering, or targeting to degradation). Yet, when the levels of the regulatory RNA are lower than the target levels the difference cannot be buffered by the regulatory RNA, which will lead to fluctuations in mRNA expression.

step-like function in the concentration of the free mRNA target in the cell: it would remain close to zero, so long as its total level is below that of the regulator, but would abruptly increase, once it exceeded that level (**Fig. 23.3**). Notably, microRNAs are known to affect their targets both at the level of mRNA stability, where they destine targets to degradation, and by inhibiting their translation (c.f. (39)). It is still not clear how the choice between these two separate fates is determined for a pair of regulators and a target or why, in some cases, one fate is desired over the other. Yet the realization that the ratio of translation to transcription affects noise suggests an effect on the noise level of the target, due to a choice between the two regulatory mechanisms. A potential application arising out of such considerations lies in the emerging field of synthetic biology, in which one of the challenges is to design and build small circuits with desired properties. The synthetic use of non-coding regulatory RNAs (40) may enable researchers to "tune" the desired level of noise (either high or low) of the various components in the system.

6. Genetic vs Non-genetic Variation

It has long been known that stressful conditions increase the rate of mutations among microorganisms (41). This is rationalized as an adaptive trait of such species – leading to an increase in genetic diversity, the origin of new biological innovations, and, ultimately,

survival of selected cells. Yet, depending on the selective advantage, the rate of mutations, and the effective population size, it may take tens of generations, typically many more, for specific mutations to reach a substantial portion of the population. This is a long time, compared for instance to a killing effect of the stressful condition. Thus, a faster means to increase diversity in the population may be needed.

An increase in noise levels, predominantly observed in stress-related genes under stressful conditions (20), may be envisaged as a complementary mechanism for the rapid generation of diversity. Unlike genetic diversity, which develops slowly, non-genetic diversity, or noise, is obtained instantaneously in a population (20). Furthermore, unlike mutations that are often, though not always (42), assumed to be generated at random sites, noise enhancement appears to be specifically directed toward genes that are expressed in response to changes in specific environmental parameters (43). The possibility exists that the labor involved in generating population diversity might split between non-genetic and genetic mechanisms, with the former occurring before the latter can take over. One interesting example, obtained in yeast, showed the feasibility of fitness advantage of enhanced noise under stressful conditions. In this experiment a mutation leading to increase in cell-to-cell variability in gene expression was found to be beneficial after an acute change in environmental conditions (31). In future, it would be relevant to see if spontaneous increased noise can evolve and be selected for in the lab when microorganisms adapt to stressful conditions.

If noise introduces diversity in a population, why is the need for genetic-based diversity, in the form of enhanced mutation? Noise has one obvious limitation: cells have a very short memory for noisy fluctuations; hence a stochastic increase or decrease in expression level of a gene may not be faithfully inherited to daughter cells. Since a cell that expresses a given gene at a high level is genetically identical to the rest of the cells, its descendents are expected to return to an averaged expression level in the coming generations. How long does it take the progeny to "forget" this legacy? A study in a mammalian system (44) suggested that for many genes it is one generation time.

In summary, one potential model would suggest that following an environmental change, stochasticity in gene expression may begin to diversify the population with respect to particular genes. Such diversity may provide the substrate for the selection of cells during the initial phase of coping with the stress. In parallel, as mutations begin to appear at a slower pace, they may allow sustained diversity, from which the fittest cells will become fixated in the population. Together, the two mechanisms may provide diversity and a substrate for selection at both short and long timescales.

Acknowledgments

The author thanks the European Research Council (REC) for grant support. The author also thanks Barbara Morgenstern for editorial help with the chapter.

References

1. Nagalakshmi, U., Wang, Z., Waern, K., et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.
2. Shalem, O., Dahan, O., Levo, M., et al. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* **4**, 223.
3. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., and Weissman, J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223.
4. Li, J. B., Levanon, E. Y., Yoon, J. K., et al. (2009) Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213.
5. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004) Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625.
6. Nachman, I., Regev, A., and Ramanathan, S. (2007) Dissecting timing variability in yeast meiosis. *Cell* **131**, 544–556.
7. Feinerman, O., Veiga, J., Dorfman, J. R., Germain, R. N., and Altan-Bonnet, G. (2008) Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science* **321**, 1081–1084.
8. Cohen, A. A., Geva-Zatorsky, N., Eden, E., et al. (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511–1516.
9. Raser, J. M., and O'Shea, E. K. (2005) Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013.
10. Thattai, M., and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98**, 8614–8619.
11. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden A. (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73.
12. Blake, W. J., Kaern, M., Cantor, C. R., and Collins, J. J. (2003) Noise in eukaryotic gene expression. *Nature* **422**, 633–637.
13. Sigal, A., Milo, R., Cohen, A., et al. (2006) Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Methods* **3**, 525–531.
14. Paulsson, J. (2004) Summing up the noise in gene networks. *Nature* **427**, 415–418.
15. Cai, L., Friedman, N., and Xie, X. S. (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362.
16. Rosenberger, R. F., and Hilton, J. (1983) The frequency of transcriptional and translational errors at nonsense codons in the lacZ gene of *Escherichia coli*. *Mol. Gen. Genet.* **191**, 207–212.
17. Gordon, A. J, Halliday, J. A., Blankschien, M. D., Burns, P. A, Yatagai, F., and Herman, C. (2009) Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. *PLoS Biol.* **24**, e44.
18. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science* **297**, 1183–1186.
19. Raser, J. M., and O′Shea, E. K. (2004) Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814.
20. Bar-Even, A., Paulsson, J., Maheshri, N., et al. (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643.
21. Newman, J. R., Ghaemmaghami, S., Ihmels, J., et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846.
22. Huh, W. K., Falvo, J. V., Gerke, L. C., et al. (2003) Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691.
23. Kussell, E., and Leibler, S. (2005) Phenotypic diversity, population growth, and

information in fluctuating environments. *Science* **309**, 2075–2078.
24. Pedraza, J. M., and van Oudenaarden, A. (2005) Noise propagation in gene networks. *Science* **307**, 1965–1969.
25. Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., and Brown, P. O. (2002) Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860–5865.
26. Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E. K. (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. USA* **103**, 13004–13009.
27. Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J., and Eisen, M. B. (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2**, e137.
28. dos Reis, M., Savva, R., and Wernisch, L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044.
29. Gasch, A. P., Spellman, P. T., Kao, C. M., et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell.* **11**, 4241–4257.
30. Rodríguez Martínez, M., Soriano, J., Tlusty, T., Pilpel, Y., and Furman, I. (2010) Messenger RNA fluctuations and regulatory RNAs shape the dynamics of a negative feedback loop. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **81**, 031924.
31. Murphy, K. F., Balázsi, G., and Collins, J. J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. USA* **104**, 12726–12731.
32. Blake, W. J., Balázsi, G., Kohanski, M. A., et al. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865.
33. Segal, E., and Widom, J. (2009) What controls nucleosome positions? *Trends Genet.* **25**, 335–343.
34. Kim, H. D., and O'Shea, E. K. (2008) A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* **15**, 1192–1198.
35. Hornung, G., and Barkai, N. (2008) Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput. Biol.* **4**, e8
36. Cağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., and Süel GM. (2009) Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139**, 512–522.
37. Kafri, R., Levy, M., and Pilpel, Y. (2006) The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl. Acad. Sci. USA* **103**, 11653–11658.
38. Kafri, R., Springer, M., and Pilpel, Y. (2009) Genetic redundancy: new tricks for old genes. *Cell* **136**, 389–392.
39. Hendrickson, D. G., Hogan, D. J., McCullough, H. L., et al. (2009) Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.* **7**, e1000238.
40. Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R., and Benenson, Y. (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795–801.
41. Sniegowski, P. D., Gerrish, P. J., and Lenski, R. E. (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705.
42. Koonin, E. V., and Wolf, Y. I. (2009) Is evolution Darwinian or/and Lamarckian? *Biol. Direct.* **4**, 42.
43. Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471–475.
44. Sigal, A., Milo, R., Cohen, A., et al. (2006) Variability and memory of protein levels in human cells. *Nature* **444**, 643–646.