Special Issue – Imaging Cell Biology

A single molecule view of gene expression

Daniel R. Larson, Robert H. Singer and Daniel Zenklusen

Department of Anatomy and Structural Biology and The Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

Analyzing the expression of single genes in single cells appears minimalistic in comparison to gene expression studies based on more global approaches. However, stimulated by advances in imaging technologies, single-cell studies have become an essential tool in understanding the rules that govern gene expression. This quantitative view of single-cell gene expression is based on counting mRNAs in single cells, monitoring transcription in real time, and visualizing single proteins. Parallel advances in mathematical models based on stochastic, discrete descriptions of biochemical processes have provided crucial insights into the underlying cellular mechanisms that control expression. The view that has emerged is rooted in a probabilistic understanding of cellular processes that quantitatively explains both the mean and the variation observed in gene-expression patterns among single cells. Thus, the close coupling between imaging and mathematical theory has established single-cell analysis as an essential branch of systems biology.

Introduction

Gene expression refers to the sum of processes that result in a particular level of a specified mRNA and protein in the cell. In many cell biological studies, gene expression is the starting point for elucidation of mechanism at the microscopic and molecular levels, while the gene expression profile is a 'parts list' compiled at the macroscopic level. Describing the coordination of gene expression is therefore a central step towards understanding cellular systems. Classical gene expression studies use the isolation of mRNAs or proteins from cell populations to determine expression profiles. These methods, however, lack spatial resolution, are not able to detect cell-to-cell differences within a population, and can only represent a static picture. To understand biological processes fully more direct methods have to be applied, ideally giving the researcher the ability to monitor individual molecules within single cells in real time.

In recent years it has become possible to analyze gene expression at the single-cell and single-molecule levels [1,2]. Such studies reveal that the expression of individual genes is highly variable, even within a clonal population of cells, and understanding the mechanisms that cause these differences has thus become an area of active research. The

quantitative accounting of mRNA and protein expression has been made possible by a rich interaction between biologists, physicists, and mathematicians as ever more precise measurements capable of counting single molecules have advanced in concert with mathematical descriptions of gene expression. The view of gene expression that has emerged from these studies is one in which a small number of molecules of both mRNA and protein leads to randomness and variation within populations which can have direct phenotypic consequences [3,4].

In this review we focus on recent experimental developments and microscopy techniques now being used to understand the rules that govern gene expression. In particular we focus on methods that aim to count *the total number of molecules* – either mRNA or protein – in a single cell. We also summarize the theoretical approaches used to describe this experimental data and show how stochastic models are able to quantitatively describe gene expression at the single-cell level. These mathematical models provide a framework for understanding how the relative balance between the kinetic steps in expression (rates of transcription, RNA decay, translation, protein decay) contributes to the differential regulation of RNA and protein in the cell. The remarkable developments in this field – both experimental and theoretical – have led to a quantitative description of gene expression in a context that can be readily utilized by cell biologists.

Quantifying protein levels in single cells

Following the development of GFP as a tool in cell biology, the imaging of reporter protein fluorescence quickly became the most quantitative experimental method for measuring gene expression in single cells. Even though protein accumulation is the final step in gene expression, quantification of the abundance and variation in protein levels has been used to infer the mechanisms of gene expression acting at the level of individual genes and mRNA [1,2].

The first effort to obtain absolute protein numbers on a genomic scale however did not use single-cell methods but was rather an ensemble measurement. Ghaemmaghami and colleagues determined global protein abundance in *S. cerevisae* using a library of strains tagged with protein A and quantitative Western blotting, so providing the first glimpse into the scale of protein abundance in a complete organism [5]. Using a similar library where proteins were tagged with GFP, two studies then measured relative

Corresponding author: Singer, R.H. (rhsinger@aecom.yu.edu).

Box 1. Fluorescent techniques for single cell gene expression analysis

Fluorescent techniques are the most direct tools for measuring protein and mRNA concentrations in single cells. Detection of single mRNAs has been used in multiple labs as a method for quantitative gene expression analysis [3,25,26]. Single-protein detection in cells, however, is technically still very challenging, and most measurements of protein distributions are achieved by quantifying relative protein abundance [5]. For both protein detection and RNA detection, one can obtain frequency distributions by looking at a single cell over time – or by looking at many cells at a particular instant in time [46]. In live cell measurements, one can measure kinetics directly; in fixed cell measurements, dynamics are inferred from the frequency distributions.

Protein detection
The two commonly used methods to measure protein concentration in single cells are Fluorescence-activated cell sorting (FACS) and live cell microscopy. Both measure signals emitted by florescent proteins. FACS has been used extensively to determine the expression variation of a collection of >2500 yeast genes [6,7]. However, its limited sensitivity does not allow the detection of low-abundance proteins. In addition, each cell is only analyzed once. Live cell microscopy on the other hand acquires time series of individual cells,

resulting in a direct measure of expression kinetics, and fluorescence microscopy is more sensitive than FACS. However, fewer cells are analyzed compared to studies using FACS [13].

RNA detection
Fluorescent *in situ* hybridization (FISH) allows the detection of single mRNAs in intact cells (Figure 1A) [3,25,56]. FISH is the most direct way to acquire quantitative mRNA expression data because no genetic manipulations are required. However, cells have to be fixed and FISH only provides a snapshot of mRNA abundance and gene activity. Similar to FACS measurements of protein distribution, FISH provides information on the kinetics of expression by considering many cells at a snapshot in time. Using probes labeled with different dyes, FISH can also be used to measure the expression levels of multiple genes within the same cell [55]. In contrast to FISH, the MS2 system allows mRNA detection in living cells (Figure 1B). Single cytoplasmic mRNAs as well as nascent transcripts at the site of transcription are detected in real time using this fluorescent-protein-based approach [27,28]. Yet, single mRNA counting in living cells is challenging: simultaneous single mRNA abundance and nascent mRNA quantification has not yet been described in living eukaryotes. However, MS2 is the only system that can directly visualize mRNA expression in real time (see also Table 1).

---

protein concentrations by FACS and determined the variability of protein expression for proteins in *S. cerevisiae* (Box 1) [6,7]. Measuring the abundance and variation of individual proteins provided a quantitative basis for assessing different models of expression regulation. By combining absolute numbers and expression variability, and complemented by earlier studies measuring expression variation of small sets of proteins, these studies concluded that variation in protein expression is dominated by the stochastic production/destruction of messenger RNAs on a scale that reflects natural protein abundance [8–12]. This variation in expression arising from stochastic fluctuations has been referred to as 'noise' in gene expression. These studies suggested further that protein-specific differences in noise correlate with a protein's 'mode' of transcription, in other words the kinetic details of how a gene is transcribed into mRNA [8–12].

These models still contained unknown parameters. One is how many proteins are translated from a single mRNA. Knowing the number of proteins per cell and the number of proteins per mRNA is crucial to understanding the stochastic variation of gene expression that has been observed.

Visualization of single proteins in single cells defines the ultimate sensitivity of quantifying gene expression. However, in addition to simply observing single molecules, one must be able to record each and every protein molecule in the cell at a given time, or produced from a given mRNA. This mandate exceeds the already stringent experimental conditions required for single-molecule microscopy, and demands a new set of experimental approaches. One approach was realized in a landmark paper in 2006 by Yu and coworkers in *E. coli* [13]. The authors attached the fluorescent protein Venus to a membrane protein, Tsr, constituting a reporter for monitoring *lac* operon activity. The membrane localization of Venus slowed the diffusion of the reporter protein so that it could be visualized under the microscope. After the protein was imaged, it was immediately bleached in preparation for observing the

next membrane-localized Venus. Protein production was based on the dissociation event of the repressor from the operator region of DNA. Using this system the authors were the first to observe that protein molecules are generated in bursts from a single mRNA. These bursts of protein production are due to the relative rates of protein translation and mRNA degradation, and have fundamental consequences for gene expression (see below). Protein-bursting amplifies the variation that already occurs from the stochastic production of RNA, because each mRNA can produce multiple proteins before it is degraded (4.2 in the case measured by Yu *et al.*). The same group also showed, using a different method of cell lysis followed by enzymatic amplification, that the yeast protein β-galactosidase is synthesized in geometric bursts of 1.7 proteins/mRNA [14].

To date direct microscopic visualization and counting of proteins for single-molecule gene expression measurements has only been demonstrated in prokaryotes. Such an approach may be quite challenging to implement in eukaryotes where translational burst sizes are likely to be larger because of the longer lifetimes of eukaryotic mRNA. One approach to overcome this problem has been proposed by Rosenfeld and coworkers [15]. Their method to obtain absolute protein numbers is based on long-term observation of dividing cells. If partitioning of proteins at cell division obeys a binomial distribution – where each protein-partitioning event is independent – it is possible to empirically determine the number of proteins present before cell division. However, any approach seeking to count fluorescent proteins will be confounded by protein folding and maturation of the fluorescent protein chromophore [16].

Measuring mRNA levels in single cells
Measuring protein distributions and determining the amplitude of translation bursts has revealed different parameters necessary for modeling gene expression [1,2]. However, accumulation of proteins is the last step of gene expression and is influenced by multiple upstream

processes. Changes in protein levels can be caused by changes in transcription rate, RNA or protein half-life, translation efficiency or any combination thereof. To understand the entire expression pathway, the individual steps have to be analyzed independently, necessitating direct measurements of transcriptional output by determining mRNA levels.

As with protein level measurements, *in vitro* ensemble measurements were first used to quantify RNA and were crucial to the understanding of gene expression at the single-cell level, especially in yeast. Quantitative microarrays showed that expression levels for more than 80% of genes are very low, with fewer than two mRNA copies per cell [17]. Combining mRNA copy number and half-life allows the calculation of average transcription frequencies for each gene [17]. These numbers, often in combination with measurement of protein concentration and/or protein noise, were used in many studies to model gene expression kinetics. However, knowing only these numbers, compared to the whole distribution at the protein level, have limited the descriptive power of the models; in consequence it became essential to determine the variability in mRNA expression levels [6].

Different approaches were then introduced to determine mRNA concentrations in individual cells (Table 1): single-cell quantitative PCR, single-cell microarrays, *in situ* fluorescent PCR, the MS2 system (described below), and single-molecule resolution fluorescent *in situ* hybridization (FISH) [3,18–26]. FISH proved to be a very fruitful approach (Figure 1 and Box 1). Pioneered in a study by Femino *et al.*, single mRNA sensitivity FISH allowed the detection of individual mRNAs in fixed cells and was able to determine the exact number of mRNAs per cell for any gene of interest [3,25,26].

Determining mRNA distribution in single cells showed that the variability in expression levels for different genes was much larger in higher eukaryotes than in yeast [3,26]. Integrating expression variability into kinetic models revealed the existence of a range of kinetic modes by which mRNAs are expressed. At one extreme, genes are transcribed in bursts, where periods of activity are interspersed by long periods of inactivity. In another mode, transcription events are uncorrelated and uniformly distributed in time. Raj and colleagues showed bursting expression for two genes in higher eukaryotes [26]. Using a tetracycline-induced reporter gene, the authors demonstrated that mRNA levels vary considerably when the gene is activated and showed that those distributions can only be explained by bursting transcription. The second gene, the endogenous gene coding for RNA polymerase II, showed a similar bursting pattern of expression. These results suggested that bursting of transcription might be the predominant mode of expression in higher eukaryotes. On the other hand, experiments in yeast showed a very narrow distribution of the expression levels for three housekeeping genes, suggesting that these genes do not burst but are constitutively transcribed [3]. Their variability was low enough to be explained by pure Poisson noise. Interestingly, the same study also found a gene in yeast showing much higher variability, suggesting that constitutive as well as bursting transcription exists in yeast.

Bursting transcription was also described in *E. coli* [27]. Here, RNA was not detected by FISH but rather by using the MS2 system. This approach uses a unique genetically encoded tag that, when inserted into RNA and bound by specific fluorescent proteins, allows mRNA detection in living cells (Figure 1B) [28]. The advantage of this system is that expression levels are monitored in real time, and thus can provide expression data with high temporal resolution. The MS2 system has only been applied to mRNA counting in bacteria but will probably also be a powerful tool in other organisms.

Observing transcription kinetics in single cells
With the inclusion of mRNA distributions, one can achieve a more complete description of gene expression than is possible by only considering protein distributions. However, mRNA levels are not a direct measure of transcription *per se*. The inference of transcriptional dynamics that comes from counting mRNA in fixed cells is limited by the half-life of mRNA. For an mRNA with a 30 min half-life, the steady-state cytoplasmic mRNA level reflects almost an hour of mRNA expression. Moreover, transcriptional responses are often fast and, depending on the length of a gene, require only a few minutes to produce mRNA [29].

The observation of transcription kinetics at higher temporal resolution can only be obtained by measuring transcription directly. Single-cell methods for studying transcription rely on the ability to detect nascent mRNAs. Using the MS2 system, Chubb *et al.* studied the expression of the developmentally regulated *dscA* gene in *Dictyostelium* (Figure 1B). The study found that transcription occurred in irregularly-spaced bursts, with the length and amplitude of these bursts staying fairly constant [30]. Transcription of the yeast *CUP1* gene on the other hand was shown to be achieved in a different manner. Upon induction, mRNA production was constant over the duration of activation [31]. The constant level of transcription was rather surprising when compared to the binding behavior of the transcriptional activator Ace1p that regulates *CUP1* transcription. Using fluorescence recovery after photobleaching (FRAP, described in this issue by Lidke and Wilson) it was shown that Ace1p bound only transiently to the *CUP1* promoter, with a residence time of less than two minutes, suggesting that constant rebinding of Ace1p was required to ensure efficient transcription.

The low stability of promoter complexes in living cells (determined by FRAP, as reviewed in Ref. [32]) appears to be a common phenomenon, and might be one important factor that defines transcription kinetics. Many activators have very short dwell times at the transcription site, some for only a few seconds, suggesting that activators do not need to be bound stably to the promoter to stimulate transcription [33–36]. Their affinity however might regulate transcription frequency. Binding of the HSP activator, that regulates *Drosophila* heat shock genes, becomes very tight upon heat shock [37]. Heat-shock genes are very efficiently transcribed, and when transcription is fully activated new transcripts are initiated approximately every four seconds [29]. It is possible that tight binding of activators allows efficient transcription but simultaneously reduces the flexibility to fine-tune the

Table 1. Summary of the most common methods for mRNA quantification and transcription analysis

|	Northern	Microarray	Real-time PCR	RNA seq	ChIP-ChIP	FISH	MS2
	Ensemble	Ensemble	Ensemble	Ensemble	Ensemble	Single cell	Single cell
Determining mRNA expression							
Measuring	Total mRNA	Total mRNA	Total mRNA	Total mRNA		Total mRNA	Total mRNA
Detection method	Blotting to membrane/ hybridization with synthetic probe	RNA labeling/ hybridization to array	Reverse transcription/ PCR	mRNA fragmentation/ adaptor ligation/ amplification/ sequencing		Hybridization using fluorescent probes	Insertion of repeats/binding of fluorescent protein
mRNA quantification	Relative intensity	Relative intensity	Absolute numbers requires standard	Single-molecule counting		Single-molecule counting	Single-molecule counting
Number of genes	Multiple	Genome wide	Many	Genome wide		1–3 per cell	1 per cell
Measuring transcription							
Measuring					Polymerase association	Nascent mRNAs	Nascent mRNAs
Detection method					Fragmentation/ IP/amplification/ hybridization to array	Hybridization using fluorescent probes	Insertion of repeats/Binding of fluorescent protein
Quantification					Relative changes in polymerase loading	Counting of nascent mRNAs	Relative intensity of mRNA signal
Number of genes					Genome wide	1–3 per cell	1 per cell
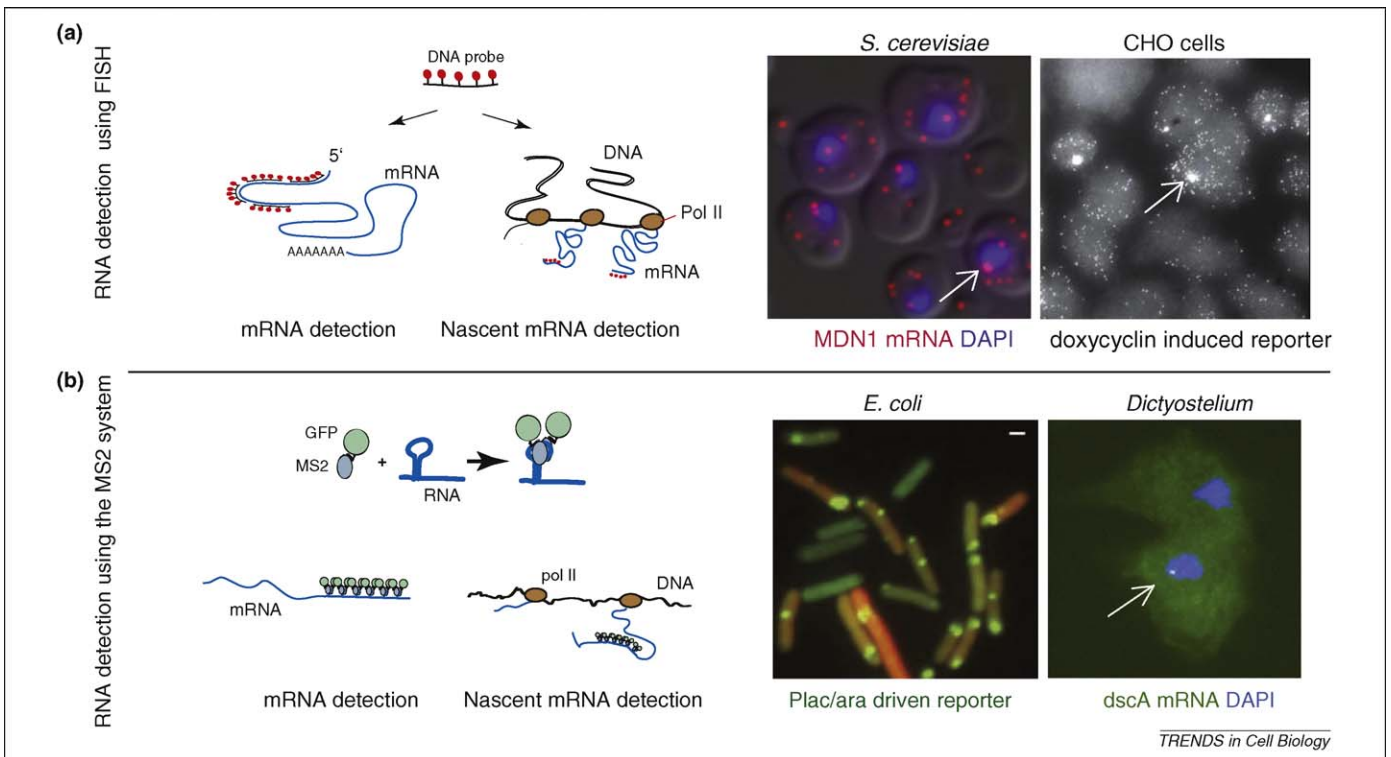
Figure 1. mRNA detection in single cells. Two common methods for single cell gene expression analysis using imaging. **(a)** Single-molecule-resolution fluorescence *in situ* hybridization (FISH) uses synthetic oligonucleotides labeled at multiple positions with fluorescent dyes to detect single mRNAs. Multiple fluorescent probes are hybridized to paraformaldehyde-fixed cells. FISH allows the detection of single mRNAs in the cytoplasm as well as nascent mRNAs at the site of transcription. On the right, yeast cells expressing *MDN1* mRNA and mammalian CHO cells (hamster cell line) expressing a doxycycline-induced reporter are shown [3,26]. **(b)** The MS2 system uses the specific interaction between the MS2 RNA hairpin and a fusion of a fluorescent protein and the MS2 phage coat protein to create a fluorescent labeled mRNA. Inserting multiple binding sites into an mRNA allows the detection of single mRNAs in living cells. The MS2 system has been used to count single mRNAs in different organisms, for example in *E. coli* as shown here (with permission from Ref. [27]), or to determine transcription kinetics in *Dictyostelium* in real time (with permission from Ref. [30]). Sites of transcription are marked by arrows.

transcriptional response. In addition, the position of the activator binding sites with respect to histones was shown to affect both transcription initiation and transcription rate [38,39]. To further underscore the dynamic, probing nature of molecular interactions at the gene, Darzacq *et al*. showed that only about 1% of polymerase-gene interactions lead to a completion of an mRNA [40]. There seem to be many different dynamic ways to modulate the transcriptional outcome, and a combination of methods will probably be required to dissect this process, probably on a gene by gene basis.

Constitutively expressed housekeeping genes in yeast appear to use a relatively simple mechanism of expression control. Zenklusen and colleagues used single-molecule resolution FISH to determine the exact number of nascent mRNAs located on constitutively expressed genes [3]. For short genes expressed at a low level, only a single nascent mRNA was detected on the gene. Given a transcription elongation velocity of less than 1 kb per minute, this suggested that the initiation of individual mRNAs was separated by minutes. Taken together with the stability of promoter complexes as described above, it is very likely that assembled transcription factor complexes might fall apart after initiating the production of a single mRNA. By combining polymerase occupancy data (determined from nascent mRNA at a transcription site) with the counting of mRNAs within the same cell has further allowed modeling of the expression kinetics of these genes; this revealed that,

for most genes, individual initiation events were uncorrelated with each other [3]. This simple regulation might suggest the existence of a stochastic limiting-step that controls expression behavior. Such a step may constitute the binding of an activator, opening of chromatin, assembly or stability of a pre-initiation complex, or the efficiency of a polymerase to enter elongation. Measuring transcriptional responses in real time with single-mRNA resolution will be necessary to dissect these different possibilities.

Theoretical models of single-molecule gene expression
The advantage of counting single molecules is that one obtains the *probability distribution* (or, more properly, the *frequency* distribution) of molecules corresponding to each stage of the central dogma for a single gene. The frequency at which a certain number of proteins or RNA molecules are observed in a single cell carries more information than the mean alone: one is able to infer general rules and mechanisms for expression based on comparisons between mathematical models and the observed frequency distributions. These mathematical models differ from those that cell biologists are accustomed to encountering. Instead of continuous, deterministic models of kinetic behavior, the mathematics of gene expression is described by discrete, stochastic models. This latter class of models takes into account the small numbers of molecules involved – at both the mRNA and protein level – even though the basic kinetic mechanisms (e.g. first-order kinetic decay of mRNA and
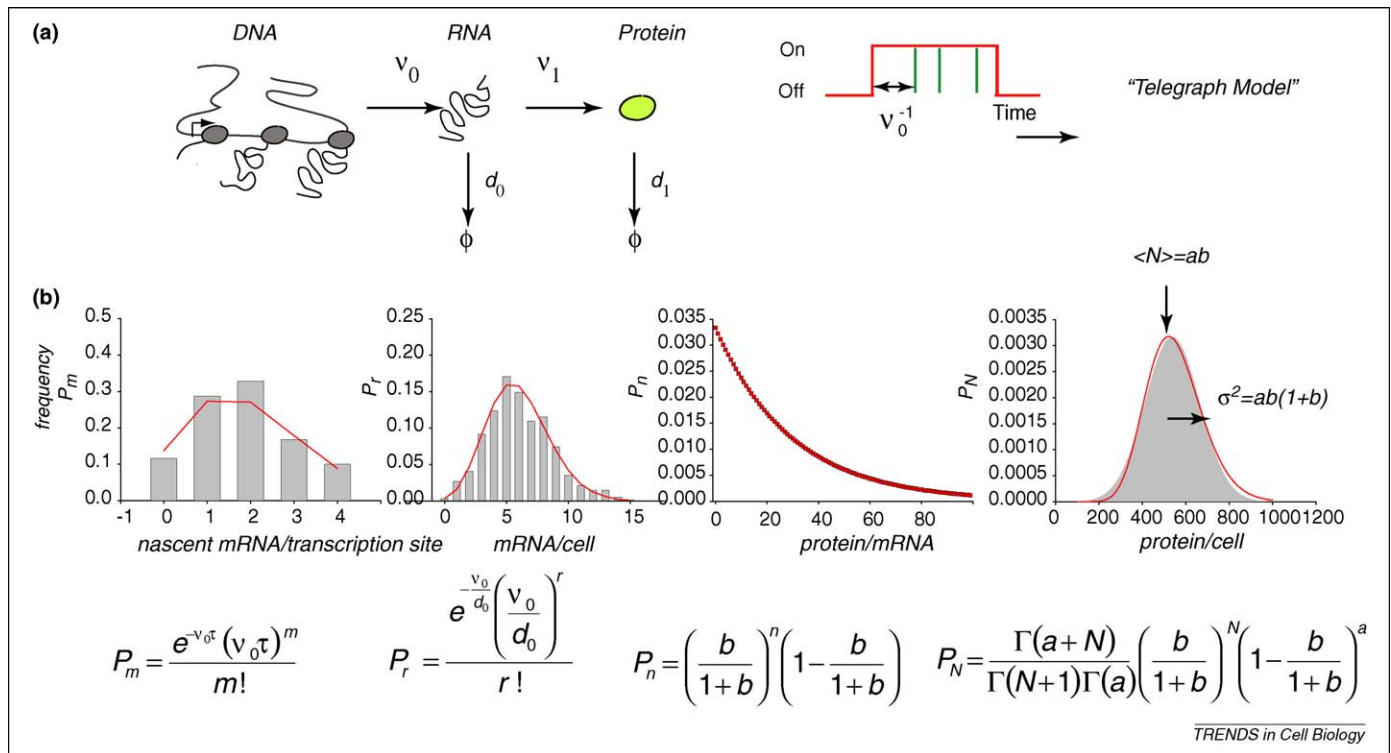
Figure 2. A stochastic model of gene expression. **(a)** *Left*, the central dogma of molecular biology – DNA to RNA to protein – is shown with rate constants of production and degradation: the rate of transcription v_0, the rate of RNA degradation d_0, the rate of translation v_1, and the rate of protein decay d_1. *Right*, a model of gene induction known as the Random Telegraph model. The gene transitions between an inactive 'off' state and an active 'on' state (red line). From the active state, transcript initiation events (vertical green lines) are separated by an average time interval v_0^{-1}. **(b)** The probability distributions (P_x) for each step of *MDN1* expression, from left to right: nascent RNA at a transcription site (P_m, m = number of nascent chains), total cellular mRNA (P_r, r = number of mRNA), proteins/mRNA (P_n, n = number of proteins/mRNA), and total protein/cell (P_N, N = number of proteins/cell). The gray symbols are published data; the red lines are theoretical fits using the equations shown below each panel. The data for nascent transcripts and total mRNA are from Ref. [3]. There are no experimental data for proteins/mRNA for *MDN1*. The data for total protein/cell were reported as a mean and variance (σ^2) (Refs. [5,6]), shown here as a Gaussian distribution. The parameters in the probability distributions are shown in panel A. τ is the time to synthesize an *MDN1* transcript. Γ denotes the gamma function. Upon degradation, neither the RNA nor the protein is considered in the probability distribution, so the state of degraded RNA and degraded protein is indicated by the symbol ϕ. Two additional symbols are used for simplicity: a is the ratio of transcription rate/protein degradation rate (v_0/d_1); b is the translation rate/RNA decay (v_1/d_0), also known as the protein burst size. For *MDN1*, a = 19, b = 29.

protein) is physically the same in both cases [41]. Indeed, there has been a tremendous amount of parallel development both in the theoretical models that predict single-molecule distributions and in the experimental techniques able to measure these distributions. In many cases there has been excellent agreement between the model and the experiment, enabling the distillation of a large body of work on gene expression, for example in *S. cerevisae*, into a few numbers.

The Random Telegraph model
A description of gene expression that has gained wide popularity, both for its simplicity and generality, is one in which a gene can be considered 'off' (incapable of producing transcripts), or 'on' (capable of producing transcripts) (Figure 2A). When the gene is 'on', transcripts are produced at a certain initiation rate (v_0, following the notation of Ref. [42]). These transcripts are degraded at a rate d_0, and are translated into protein at a rate v_1, that likewise is degraded at a rate of d_1. This model of gene induction, sometimes called a Random Telegraph model [41], was first proposed by Ko [43] and later expanded by Peccoud and Ycart [44]. This model results in a set of stochastic differential equations known as the master equation that explicitly takes into account the random nature of events associated with single molecules [44]. The solution to this master equation describes gene expression – from gene to mRNA to protein – at the single-molecule level and takes the form of a probability distribution. Obtaining this solution under various limiting cases is the basis for a quantitative understanding of gene expression.

The steady-state solution was first obtained by Raj and coworkers [26] after Paulsson and coworkers [58] who used it to explain the distribution of PolII mRNA in fixed cells. This elegant work, both experimental and theoretical, demonstrated how variation in expression begins with the process of transcription. Recently, a time-dependent solution to the master equation was reported by Shahrezaei and Swain [42] and by Iyer-Biswas and coworkers [45].

The primary implication of the telegraph model is that variation in gene expression is greatly increased through the process of transcriptional or translational bursting. Mathematically, transcriptional bursting means simply that many transcripts are produced from a single transcription 'on' state [26,27,43,46,47]; translational bursting means that many proteins are produced from a single mRNA [11,13,14,48–50]. Biologically, a transcriptional burst may be due, for example, to the stability of a transcription pre-initiation complex, leading to many transcripts produced from a stable complex [26,27,43]. Transcriptional bursting does not occur for all genes but is rather one limiting kinetic case that can be observed [3].

A translational burst is due to the fact that translation frequency (v_1) is greater than mRNA decay frequency (d_0) for most genes [9,42,47,48]. The translational burst from a single mRNA follows a geometric distribution [48,50] (see also Figure 2B), and has been observed directly [13,14]. Intuitively, this geometric distribution can be understood as the relative frequency of encounter of a single mRNA with either the translation machinery or the RNA decay machinery. When translation frequency is greater than RNA decay frequency, the mRNA is more likely to be translated than degraded. So the probability of a burst of n proteins is the probability of encountering the translation machinery n times in a row before encountering the decay machinery once [48]. The result is a long-tailed decaying distribution for the number of proteins per mRNA that is very different from the peaked distribution of proteins/cell (Figure 2). In the former case, the *mean* number of proteins translated from a single mRNA is the ratio of translation/mRNA decay (v_1/d_0), but the *most likely* number of proteins translated from a single mRNA is zero. Thus, the balance of production and decay not only determines the mean, but also the relative variation, providing the cell with a means of limiting or enhancing variability according to selective pressure [10].

Variation in the central dogma: comparing theory and experiment

The consequences of stochastic gene expression, and the success of the stochastic model in explaining measured frequency distributions, can be illustrated by considering the *S. cerevisae* gene *MDN1* that encodes a protein involved in ribosome biogenesis. The gene is a housekeeping gene which is necessary for survival and is present at low copy number in every cell (Figure 2B). For this gene, the model can be simplified even further because the gene is always 'on', producing transcripts in single uncorrelated events [3]. The steady-state solution to the master equation for mRNA distribution then becomes a Poisson distribution. The measured distribution of nascent chains, mRNA/cell, and protein/cell are shown in Figure 2B as gray bars. The theoretical probability distributions are shown as red lines, with the corresponding equation shown underneath. There are no free-fitting parameters in these curves – the kinetic rate constants are the initiation frequency (v_0, obtained from Zenklusen *et al.*, Ref. [3]), the RNA decay rate (d_0, obtained from Holstege *et al.*, Ref. [17]), the translation rate (v_1, obtained from Arava *et al.*, Ref. [51]), and the protein decay rate (d_1, obtained from Belle *et al.*, Ref. [52]). The final output is the protein abundance and variation, from Ghaemmaghami *et al.* [5] and Newman *et al.* [6], respectively. Although *MDN1* is a simple example of gene expression, the complete agreement between theoretical, biochemical, and microscopic data from multiple laboratories is a milestone in our description of gene expression.

The immediate question that arises from the telegraph model of gene expression is: what is the biological interpretation of the 'on' and 'off' states or the 'active' and 'inactive' states? In some cases it has been possible to connect an on/off state with a direct biological correlate, for example nucleosome remodeling around the promoter [12]. However, other scenarios might apply for different genes, and could be as simple as the kinetic dwell-time of a specific factor or as global as a stage of the cell cycle [3]. The strength of this mathematical description lies in the ability to classify a wide range of behaviors in a few generic rate-constants. Although a complete thermodynamic description of a particular regulatory unit based on kinetic rate-constants is always desirable [53], for a great many genes, especially in eukaryotes, this description requires a level of detailed understanding of the constituent elements that is not yet available. Therefore, models such as the telegraph model (and multi-state extensions thereof, Ref. [54]), provide an abstract intermediate for classification that seems particularly suited to the complexity of cell biological studies.

Outlook

The ability to count molecules within cells is an important step towards a more quantitative analysis of gene expression. Just as high-throughput sequencing markedly advanced our knowledge of gene expression by counting sequence tags, single-molecule counting in cells has introduced a new era in quantitative gene expression analysis (Table 1). Integrating these numbers into mathematical models will reveal important insights into the mechanisms of gene expression. One limitation is that single mRNA and protein counting is still limited to a few genes per cell, compared to entire genome capability in techniques such as RNA sequencing. The ability to analyze several genes simultaneously within single cells will provide a systems-level understanding at the single-molecule level [55,56].

One of the experimental challenges in a complete quantitative description of gene expression is to obtain measurements of the distribution of proteins translated from a single mRNA. Implicit in the theoretical model above is the assumption that translation events and mRNA decay events are independent. This assumption results in the geometric distribution of protein/mRNA. However, in many cases this assumption may not hold, and there is a competing model where translation leads to modifications of the mRNA that make the RNA increasingly likely to be degraded [57]. At present, there is an order of magnitude disagreement between different estimates of protein burst sizes in *S. cerevisae*. Bar-Even and coworkers report an average calculated protein burst size for > 40 genes to be $\sim$ 1200 (Ref. [7]), Cai *et al.* measure a burst for a single gene of 1.7 (Ref. [14]), and the *MDN1* gene has a calculated burst size of 30 (Refs. [3,17,51]). To better understand how protein production is controlled from single mRNAs it will be necessary to achieve both single-RNA and single-protein imaging in the same cell.

This combination of systems biology, computational biology, and single-molecule microscopy lays the groundwork for a quantitative understanding of gene expression that will expand rapidly in the coming years.

References
1 Kaufmann, B.B. and van Oudenaarden, A. (2007) Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.* 17, 107–112

2 Locke, J.C.W. and Elowitz, M.B. (2009) Using movies to analyse gene circuit dynamics in single cells. *Nat. Rev. Micro.* 7, 383–392

3 Zenklusen, D. *et al.* (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15, 1263–1271

4 Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or bhance: stochastic gene expression and its consequences. *Cell* 135, 216–226

5 Ghaemmaghami, S. (2003) Global analysis of protein expression in yeast. *Nature* 425, 737–744

6 Newman, J.R.S. *et al.* (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846

7 Bar-Even, A. *et al.* (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38, 636–643

8 Elowitz, M.B. *et al.* (2002) Stochastic gene expression in a single cell. *Science* 297, 1183–1186

9 Blake, W.J. *et al.* (2003) Noise in eukaryotic gene expression. *Nature* 422, 633–637

10 Fraser, H.B. *et al.* (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol.* 2, e137

11 Ozbudak, E.M. *et al.* (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73

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

13 Yu, J. *et al.* (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311, 1600–1603

14 Cai, L. *et al.* (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358–362

15 Rosenfeld, N. *et al.* (2006) A Fluctuation Method to Quantify In Vivo Fluorescence Data. *Biophys. J.* 91, 759–766

16 Dong, G.Q. and McMillen, D.R. (2008) Effects of protein maturation on the noise in gene expression. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phy.* 77, 021908–021911

17 Holstege, F.C.P. *et al.* (1998) Dissecting the Regulatory Circuitry of a Eukaryotic Genome. *Cell* 95, 717–728

18 Tang, F. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Meth.* 6, 377–382

19 Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63

20 Cloonan, N. *et al.* (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Meth.* 5, 613–619

21 Eberwine, J. *et al.* (1992) Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. U. S. A.* 89, 3010–3014

22 Bengtsson, M. *et al.* (2008) Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Mol. Biol.* 9, 63

23 Warren, L. *et al.* (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17807–17812

24 Kurimoto, K. *et al.* (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucl. Acids Res.* 34, e42

25 Femino, A.M. *et al.* (1998) Visualization of single RNA transcripts in situ. *Science* 280, 585–590

26 Raj, A. *et al.* (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, e309

27 Golding, I. *et al.* (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025–1036

28 Bertrand, E. *et al.* (1998) Localization of ASH1 mRNA Particles in Living Yeast. *Mol. Cell* 2, 437–445

29 O'Brien, T. and Lis, J.T. (1993) Rapid changes in *Drosophila* transcription after an instantaneous heat shock. *Mol. Cell. Biol.* 13, 3456–3463

30 Chubb, J.R. *et al.* (2006) Transcriptional Pulsing of a Developmental Gene. *Current Biol.* 16, 1018–1025

31 Karpova, T.S. *et al.* (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* 319, 466–469

32 Darzacq, X. *et al.* (2009) Imaging transcription in living cells. *Annu. Rev. Biophys.* 38, 173–196

33 McNally, J.G. *et al.* (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* 287, 1262–1265

34 Rayasam, G.V. *et al.* (2005) Ligand-specific dynamics of the progesterone receptor in living cells and during chromatin remodeling in vitro. *Mol. Cell. Biol.* 25, 2406–2418

35 Bosisio, D. *et al.* (2006) A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J.* 25, 798–810

36 Sharp, Z.D. *et al.* (2006) Estrogen-receptor-alpha exchange and chromatin dynamics are ligand- and domain-dependent. *J. Cell. Sci.* 119, 4101–4116

37 Yao, J. *et al.* (2006) Dynamics of heat shock factor association with native gene loci in living cells. *Nature* 442, 1050–1053

38 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

39 Lam, F.H. *et al.* (2008) Chromatin decouples promoter threshold from dynamic range. *Nature* 453, 246–250

40 Darzacq, X. *et al.* (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806

41 van Kampen, N.G. (1981) *Stochastic Processes in Physics and Chemistry*, Elsevier

42 Shahrezaei, V. and Swain, P.S. (2008) Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17256–17261

43 Ko, M.S.H. (1991) A stochastic model for gene induction. *J. Theor. Biol.* 153, 181–194

44 Peccoud, J. and Ycart, B. (1995) Markovian modeling of gene product synthesis. *Theor. Popul. Biol.* 48, 222–234

45 Iyer-Biswas, S. *et al.* (2009) Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 79, 031911-1–031911-9

46 Pedraza, J.M. and Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319, 339–343

47 Paulsson, J. (2004) Summing up the noise in gene networks. *Nature* 427, 415–418

48 McAdams, H.H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 94, 814–819

49 Kepler, T.B. and Elston, T.C. (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81, 3116–3136

50 Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8614–8619

51 Arava, Y. *et al.* (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3889–3894

52 Belle, A. *et al.* (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13004–13009

53 Bintu, L. *et al.* (2005) Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15, 125–135

54 Sánchez, Á. and Kondev, J. (2008) Transcriptional control of noise in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5081–5086

55 Levsky, J.M. *et al.* (2002) Single-cell gene expression profiling. *Science* 297, 836–840

56 Raj, A. *et al.* (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Meth.* 5, 877–879

57 Parker, R. and Sheth, U. (2007) P Bodies and the Control of mRNA Translation and Degradation. *Mol. Cell* 25, 635–646

58 Paulsson, J. *et al.* (2000) Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7148–7153