Published in final edited form as:

Trends Cell Biol. 2009 November ; 19(11): 630–637. doi:10.1016/j.tcb.2009.08.008.

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

Abstract

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 critical insight into the underlying cellular mechanisms that control expression. The view that has emerged is rooted in a probabilistic understanding of cellular processes and 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 results in a particular level of a specified mRNA and protein in the cell. For many cell biological studies, gene expression is the starting point for elucidation of mechanism at the microscopic, molecular level, while the gene expression profile is the parts list compiled at a macroscopic level. Describing the coordination of gene expression is therefore a central step towards understanding cellular systems. Classical gene expression studies use isolation of mRNAs or proteins from cell populations to determine expression profiles. Those 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 fully understand biological processes, 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, single molecule level ^{1,2}. Such studies reveal that expression of individual genes, even within clonal population of cells is highly variable; 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 small numbers

Corresponding author: Robert H Singer, rhsinger@aeom.yu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of molecules of both mRNA and protein leads to randomness and variation within populations which have direct phenotypic consequences^{3,4}.

In this review, we focus on the recent experimental developments and microscopy techniques which are 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 of 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 which can be readily utilized by cell biologists.

Quantifying Protein Levels in Single Cells

After the development of GFP as a tool in cell biology, observation of fluorescent proteins 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 abundance and variation in protein levels was used to infer the mechanisms of gene expression which acted 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. Ghaemmighami and colleagues determined global protein abundance in *S. cerevisiae* using a library of ProteinA tagged strains and quantitative western blotting, providing the first glimpse into the scale of protein abundance for a complete organism⁵. Using a similar library where proteins were tagged with GFP, two studies then measured relative 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 the quantitative basis for assessing different models of expression regulation. Combining absolute numbers and expression variability, 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 and scales with 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, meaning the kinetic details of how a gene was 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 critical for understanding the stochastic variation that has been observed in gene expression.

The visualization of single proteins in single cells provides the ultimate sensitivity in 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 in 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 variation that occurs from 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 laboratory 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/mRNA¹⁴.

Thus far, 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, due to the longer lifetimes of mRNA. One approach for overcoming this problem has been proposed by Rosenfeld and coworkers¹⁵. Their method for obtaining absolute protein numbers is based on long-time observation of dividing cells. If partitioning of proteins at cell division obeys a binomial distribution – every protein partitioning event is independent – it is possible to empirically determine the number of proteins which were present before cell division. However, any approach which seeks to count fluorescent proteins will be confounded by protein folding and maturation of the fluorescent protein chromophore¹⁶.

Measuring mRNA levels in single cells

Measuring protein distributions and determining the amplitude of translation bursts revealed different parameters necessary to model 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 altering transcription, 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.

Similar to measuring the protein level, *in vitro* ensemble measurements were first used to quantify RNA and were crucial to understand 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, fewer than two copies per cell¹⁷. Combining mRNA copy number and half-life allows the calculation of average transcription frequencies for each gene¹⁷. These numbers, often in combination with measurement of protein concentration and/or protein noise measurements were used in many studies to model gene expression kinetics. However, knowing only these numbers, compared to the whole distribution at the protein level, limited the descriptive power of models, and determining the variability in mRNA expression became essential⁶.

Different approaches were introduced to determine mRNA concentration 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 showed to be a very fruitful approach (Figure 1/Box 1). Pioneered in a study by Femino et al., single mRNA sensitivity FISH allowed 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. In one extreme, genes are transcribed in bursts, where periods of activity are interspaced 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 eukaryotes²⁶. Using a tetracycline induced reporter gene, the authors demonstrated that mRNA levels vary considerably when the gene was 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 expression pattern. These results suggested that transcription bursting might be the prominent expression mode in higher eukaryotes. On the other hand, experiments in yeast showed a very narrow distribution of the expression levels for three housekeeping genes, suggesting these genes do not burst but are constitutively transcribed³. 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 allowing high temporal resolution expression data. The MS2 system has only been applied to mRNA counting in bacteria but will likely 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 considering only 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. However, transcriptional responses are often fast and, depending on the length of a gene, require only a few minutes to produce mRNA²⁹.

Higher temporal resolution observation of transcription kinetics can be obtained only 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 constant³⁰. 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 time of activation³¹. The constant transcription was rather surprising when compared to the binding behavior of the transcriptional activator Ace1p that regulated *CUP1* transcription. Using fluorescence recovery after photobleaching (FRAP, described in this issue by Lidke & Wilson) the authors showed that Ace1p bound only transiently to the *CUP1* promoter, with a residence time of less than two min suggesting that constant rebinding of Ace1p was required to ensure efficient transcription.

The low stability of promoter complexes in living cells (determined by FRAP, and reviewed in³²) 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 have to be stably bound to their promoter to allow transcription^{33–36}. Their affinity however might regulate transcription

frequency. Binding of the HSP activator which regulates *Drosophila* heat shock genes becomes very tight upon heat shock³⁷. Heat shock genes are very efficiently transcribed, with new transcripts initiated about every four seconds at full activation²⁹. It is possible that tight binding of activators allows efficient transcription but simultaneously reduces the flexibility to fine tune the transcriptional response. In addition, the position of 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 mRNA⁴⁰. Thus, there seem to be many different dynamic ways to modulate the transcriptional outcome, and a combination of methods will likely be required to dissect this process, probably gene by gene.

A relatively simple expression control seems to occur at constitutively expressed housekeeping genes in yeast. Zenklusen and colleagues used single molecule resolution FISH to determine the exact number of nascent mRNAs on constitutively expressed genes³. This analysis showed that on short genes expressed at a low level, only a single nascent mRNA is detected at the gene. At a transcription elongation velocity of less than 1kb per minute, this suggested that initiation of individual mRNAs was separated by minutes. Taken together with the stability of promoter complexes described above, it seems likely that assembled transcription factor complexes often fall apart after initiation of a single mRNA. Combining polymerase occupancy data (determined from nascent mRNA at a transcription site) with the counting of mRNAs within the same cell further allowed modeling of the expression kinetics of these genes and showed that individual initiation events were uncorrelated with each other for most genes³. This simple regulation might suggest the existence of a stochastic limiting step that controls the 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* of molecules corresponding to each stage of the central dogma for a single gene. The probability of observing a certain number of proteins or RNA molecules 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 probability 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 (for example, first-order kinetic decay of mRNA and protein) is physically the same in both cases⁴¹. Indeed, there has been a tremendous amount of parallel development both in the theoretical models which predict single molecule distributions and the experimental techniques which can measure these single molecule distributions. In many cases, there is excellent agreement between the model and the experiment, enabling a distillation of a large body of work on gene expression, for example in *S. cerevisiae*, into a few numbers.

The Random Telegraph Model

The gene expression description which 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 can be produced with a certain initiation rate (v_0 , following the notation of⁴²). These transcripts are

degraded with a rate d_0 and translated into protein with a rate v_I , which likewise is degraded at a rate of d_I . This model of gene induction, sometimes called a Random Telegraph Model⁴¹, was first proposed by Ko⁴³ and later expanded by Peccoud and Ycart⁴⁴. This model results in a set of stochastic differential equations known as the master equation which explicitly takes into account the random nature of events associated with single molecules⁴⁴. 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 coworkers²⁶ 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 Swain⁴² and by Iyer-Biswas and coworkers⁴⁵.

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 preinitiation 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 observed³. A translational burst is due to the fact that translation frequency (v_I) 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 once⁴⁸. The result is a long-tailed decaying distribution for number of proteins/mRNA which is very different from the peaked distribution of protein/cell (Figure 2). In the former case, the *mean* number of proteins to come from a single mRNA is the ratio of translation/mRNA decay (v_I/d_0), but the *most likely* number of proteins to come 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 pressure¹⁰.

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 probability distributions, can be illustrated by considering the *S. cerevisiae* gene *MDN1*, which codes for a protein involved in ribosome biogenesis. The gene is a housekeeping gene which is necessary for survival and 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 events³. 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.³), the RNA decay rate (d_0 , obtained from Holstege et al.)¹⁷, the translation rate (v_I , obtained from Arava et al.

⁵¹), and the protein decay rate (d_I , obtained from Belle et al.⁵²). The final output is the protein abundance and variation, from Ghaemmamaghami 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 state or the active and inactive state? In some cases, it has been possible to connect an on/off state with a direct biological correlate, for example nucleosome remodeling around the promoter¹². However, other scenarios may apply for different genes, and may be as simple as the kinetic dwell time of a specific factor or as global as a stage of the cell cycle³. 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 desirable⁵³, 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 thereof⁵⁴), provide an abstract intermediate for classification which 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 more 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 which make the RNA increasingly likely to be degraded⁵⁷. At present, there is an order of magnitude disagreement between estimated protein burst sizes in *S. cerevisiae*. Bar-Even and coworkers report an average calculated protein burst size for > 40 genes to be ~ 1200⁷, Cai et al. measure a burst for a single gene of 1.7¹⁴, and the *MDN1* gene has a calculated burst size of 30^{3,17,51}. To better understand how protein production is controlled from single mRNAs, it will be necessary to achieve both single RNA imaging 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.

TEXT 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 established 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 abundance⁵. For both protein detection and RNA detection, one can obtain probability distributions by looking at a single cell over time *or* by looking at many cells at a particular instant in time⁴⁶. In live cell measurements, one can measure kinetics directly; in fixed cell measurements, dynamics are inferred from the probability 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 fluorescent 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 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 FACS¹³.

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 as 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 cell⁵⁵. 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, the MS2 system is the only to directly visualize real time mRNA expression (see also Table 1).

References

1. Kaufmann BB, van Oudenaarden A. Stochastic gene expression: from single molecules to the proteome. *Current Opinion in Genetics & Development* 2007;17 (2):107–112. [PubMed: 17317149]
2. Locke JCW, Elowitz MB. Using movies to analyse gene circuit dynamics in single cells. 2009;7(5): 383–392.
3. Zenklusen D, et al. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 2008;15 (12):1263–1271. [PubMed: 19011635]
4. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. 2008;135(2):216–226.
5. Ghaemmaghami S. Global analysis of protein expression in yeast. *Nature* 2003;425:737–744. [PubMed: 14562106]

6. Newman JRS, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. 2006;441(7095):840–846.
7. Bar-Even A, et al. Noise in protein expression scales with natural protein abundance. 2006;38(6):636–643.
8. Elowitz MB, et al. Stochastic Gene Expression in a Single Cell. *Science* 2002;297 (5584):1183–1186. [PubMed: 12183631]
9. Blake WJ, et al. Noise in eukaryotic gene expression. 2003;422(6932):633–637.
10. Fraser HB, et al. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biol* 2004;2 (6):e137. [PubMed: 15124029]
11. Ozbudak EM, et al. Regulation of noise in the expression of a single gene. 2002;31(1):69–73.
12. Raser JM, O'Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. *Science* 2004;304 (5678):1811–1814. [PubMed: 15166317]
13. Yu J, et al. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* 2006;311 (5767):1600–1603. [PubMed: 16543458]
14. Cai L, et al. Stochastic protein expression in individual cells at the single molecule level. 2006;440 (7082):358–362.
15. Rosenfeld N, et al. A Fluctuation Method to Quantify In Vivo Fluorescence Data. 2006;91(2):759–766.
16. Dong GQ, McMillen DR. Effects of protein maturation on the noise in gene expression. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* 2008;77 (2):021908–021911.
17. Holstege FCP, et al. Dissecting the Regulatory Circuitry of a Eukaryotic Genome. 1998;95(5):717–728.
18. Tang F, et al. mRNA-Seq whole-transcriptome analysis of a single cell. 2009;6(5):377–382.
19. Wang Z, et al. RNA-Seq: a revolutionary tool for transcriptomics. 2009;10(1):57–63.
20. Cloonan N, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. 2008;5(7): 613–619.
21. Eberwine J, et al. Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences of the United States of America* 1992;89(7):3010–3014. [PubMed: 1557406]
22. Bengtsson M, et al. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biology* 2008;9 (1):63. [PubMed: 18631407]
23. Warren L, et al. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proceedings of the National Academy of Sciences* 2006;103 (47):17807–17812.
24. Kurimoto K, et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucl Acids Res* 2006;34(5):e42. [PubMed: 16547197]
25. Femino AM, et al. Visualization of Single RNA Transcripts in Situ. *Science* 1998;280 (5363):585–590. [PubMed: 9554849]
26. Raj A, et al. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol* 2006;4 (10):e309. [PubMed: 17048983]
27. Golding I, et al. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* 2005;123 (6):1025–1036. [PubMed: 16360033]
28. Bertrand E, et al. Localization of ASH1 mRNA Particles in Living Yeast. 1998;2(4):437–445.
29. O'Brien T, Lis JT. Rapid changes in *Drosophila* transcription after an instantaneous heat shock. *Mol Cell Biol* 1993;13(6):3456–3463. [PubMed: 8497261]
30. Chubb JR, et al. Transcriptional Pulsing of a Developmental Gene. 2006;16(10):1018–1025.
31. Karpova TS, et al. Concurrent Fast and Slow Cycling of a Transcriptional Activator at an Endogenous Promoter. *Science* 2008;319 (5862):466–469. [PubMed: 18218898]
32. Darzacq X, et al. Imaging Transcription in Living Cells. *Annual Review of Biophysics* 2009;38 (1): 173–196.
33. McNally JG, et al. The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science* 2000;287 (5456):1262–1265. [PubMed: 10678832]

34. Rayasam GV, et al. Ligand-specific dynamics of the progesterone receptor in living cells and during chromatin remodeling in vitro. *Molecular and Cellular Biology* 2005;25 (6):2406–2418. [PubMed: 15743833]
35. Bosisio D, et al. A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *Embo J* 2006;25:798–810. [PubMed: 16467852]
36. Sharp ZD, et al. Estrogen-receptor- α exchange and chromatin dynamics are ligand- and domain-dependent. *J Cell Sci* 2006;119 (19):4101–4116. [PubMed: 16968748]
37. Yao J, et al. Dynamics of heat shock factor association with native gene loci in living cells. 2006;442 (7106):1050–1053.
38. Kim HD, O'Shea EK. A quantitative model of transcription factor-activated gene expression. 2008;15 (11):1192–1198.
39. Lam FH, et al. Chromatin decouples promoter threshold from dynamic range. 2008;453(7192):246–250.
40. Darzacq X, et al. In vivo dynamics of RNA polymerase II transcription. 2007;14(9):796–806.
41. van Kampen, NG. *Stochastic Processes in Physics and Chemistry*. Elsevier; 1981.
42. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences* 2008;105 (45):17256–17261.
43. Ko MSH. A stochastic model for gene induction. *Journal of Theoretical Biology* 1991;153 (2):181–194. [PubMed: 1787735]
44. Peccoud J, Ycart B. Markovian Modeling of Gene Product Synthesis. *Theor Popul Biol* 1995;48:222–234.
45. Iyer-Biswas S, et al. Stochasticity of gene products from transcriptional pulsing. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* 2009;79(3)
46. Pedraza JM, Paulsson J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* 2008;319 (5861):339–343. [PubMed: 18202292]
47. Paulsson J. Summing up the noise in gene networks. 2004;427(6973):415–418.
48. McAdams HH, Arkin A. Stochastic mechanisms in gene expression *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(3):814–819.
49. Kepler TB, Elston TC. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophysical Journal* 2001;81 (6):3116–3136. [PubMed: 11720979]
50. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* 2001;98 (15):8614–8619. [PubMed: 11438714]
51. Arava Y, et al. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100 (7): 3889–3894. [PubMed: 12660367]
52. Belle A, et al. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences* 2006;103 (35):13004–13009.
53. Bintu L, et al. Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development* 2005;15 (2):125–135. [PubMed: 15797195]
54. Sánchez Á, Kondev J. Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences* 2008;105 (13):5081–5086.
55. Levsky JM, et al. Single-Cell Gene Expression Profiling. *Science* 2002;297 (5582):836–840. [PubMed: 12161654]
56. Raj A, et al. Imaging individual mRNA molecules using multiple singly labeled probes. 2008;5(10): 877–879.
57. Parker R, Sheth U. P Bodies and the Control of mRNA Translation and Degradation. 2007;25(5): 635–646.

Figure 1. mRNA detection in single cells

Two common methods for single cell gene expression analysis using imaging: A) Single molecular 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 doxycyclin 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 fluorescently labeled mRNA. Inserting multiple binding sites into an mRNA allows detection of single mRNAs in living cells. The MS2 system has been used in different organisms, shown here, for example, in *E. coli* to count single mRNAs (with permission from Ref 27) and in *Dictyostelium* to determine transcription kinetics in real time (with permission from Ref 30). Sites of transcription are marked by arrows.

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 . Once degradation, the RNA and protein are Right, a model of gene induction known as the telegraph model. The gene transitions between an inactive off state and an active on state (red line). From the active state, transcripts 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 is from ³. There is no experimental data for proteins/mRNA for *MDN1*. The data for total protein/cell was reported as a mean and variance (σ^2) ^{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*, $a=19$, $b=29$.

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

	Northern ensemble	Microarray ensemble	Real time PCR ensemble	RNA seq ensemble	ChIP-CHIP ensemble	FISH Single cell	MS2 Single cell
Determining mRNA expression							
measuring Detection method	Total mRNA Blotting to membrane/ hybridization with synthetic probe	Total mRNA RNA labeling/ Hybridization to array	Total mRNA Reverse transcription/ PCR	Total mRNA mRNA fragmentation/ adaptor ligation/ amplification/ sequencing		Total mRNA Probe hybridization in situ	Total mRNA 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	1 per cell	Genome wide	many	Genome wide		1–3 per cell	1 per cell
Measuring transcription					Polymerase association	nascent mRNAs	nascent mRNAs
measuring 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