Research Focus

Gene expression and the myth of the average cell

Jeffrey M. Levsky and Robert H. Singer

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

We all know that gene expression occurs within cells, yet we do not think of expression in terms of its fundamental unit – a single cell. Instead, we understand the expression of genes in terms of a cell population as all of our information comes from samples containing millions of cells. From a complex mixture of cells, we attempt to infer the probable state of an average cell in the population. In truth, what we obtain is an *averaged* cell, a contrivance for representing biological knowledge beyond the limits of detection. We never know the variation among the members of the population that our methods average into a mean. Recent technological advances allow the precise measurement of single-cell transcriptional states to study this variability more rigorously. How genes are expressed in the population is strikingly different to what we have assumed from extrapolating to an average cell. Does the average cell actually exist? As we discuss, it is becoming increasingly clear that it doesn't.

The beginning of most contemporary studies of gene expression is the derivation of RNA from a sample (reviewed in Ref. [1]). The very basis of how we understand the submicroscopic realm of cellular processes such as transcription is made possible by adding together enough material to reach a measurable threshold. *Ex vivo* samples are invariably composed of different cell types that are closely juxtaposed, if not entirely mixed together. Some of these cell types are in vanishingly small amounts by mass, but at the same time they can be extremely important in terms of information content. The disparity between the ability to measure and the desire to make biological conclusions calls out for a new approach to expression.

In this article, we summarize previous methods of measuring expression and contrast these with a novel single-cell assay of gene activity. We go on to present advances in our understanding of single-cell behavior and proffer questions regarding physiology and gene expression that result from the newly observed single-cell states.

Cell population assays

Modern microarray technologies employ phenomenal miniaturization to detect differences in population levels of thousands of different transcripts; changes as small as 1.3 to twofold are detectable [2,3]. This sensitivity is more than sufficient to measure ramped up expression of many genes with a measurable basal level, as long as a considerable portion of the population actively expresses. Shortcomings of the technique are seen when the target exists at too high levels in the reference sample, such that incremental changes cannot be resolved, or when the target is too sparse in either sample, such that the measurable fluorescent signal is lost in background noise.

Another contemporary and widely used approach, serial analysis of gene expression (SAGE), allows quantitative measurement of entire transcriptomes [4]. It is based on sequencing tags that are also derived from a population of cells, which involves averaging, or obscuring, the individual cell responses. While this is acceptable for study of homogeneous materials, it is known that medical diagnostics would benefit from increased resolution. For example, the few HIV-carrying blood cells in early infection or the tiny malignant focus within a lymph node metastasis are rare, but few would ignore them by choice.

Closer analysis

There are a few important attempts at clarity in the face of the obstacles inherent to tissue analysis using adjuncts to conventional means of detecting expression. The development and commercialization of laser-dissection microscopy allows one to sort out tissue types manually, enabling isolation of even a single cell [5]. However, there are several reasons why extracting information on gene expression from a single cell is not yet possible. First, producing the amount of nucleic acid necessary for microarray studies would clearly require pooling of many cells; the resultant cocktail cannot be representative of any individual cell, *per se*. Second, deriving usable, reproducible RNA from these microdissections has proven to be a non-trivial process [6]. Refinements, robotic automation and the use of additional cell-sorting schemes should make this approach increasingly sensitive and specific, providing for a higher-resolution understanding of events at the level of the cell.

The smaller the sample isolated, the fewer the transcripts available for assay and the more difficult the detection. Regardless of the organism or model system, the majority of mRNAs are thought to be present at less than ten copies per cell [7]. A solution to this problem is reverse transcriptase–polymerase chain reaction (RT–PCR) amplification of transcripts. This procedure has its own peril, which is that amplification varies across RNAs. The mixture produced can contain ratios of transcripts that are far from the original starting material. One approach to correct this problem is real-time or quantitative PCR, a means of reliability assurance by monitoring synthesis

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

 0962-8924/02/$ - see front matter PII: S0962-8924(02)00002-8

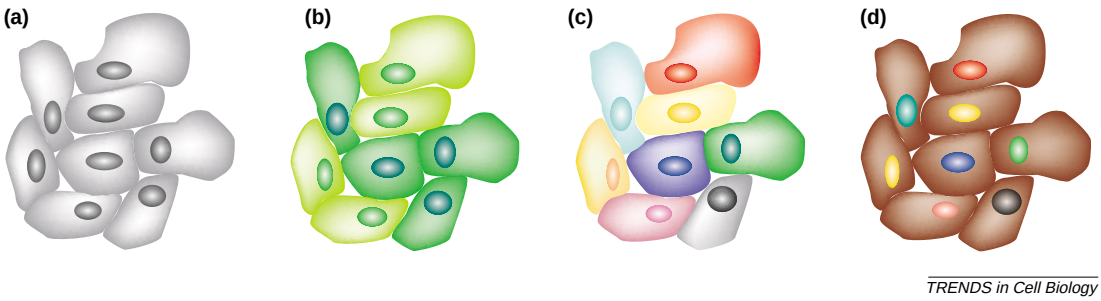

Fig. 1. Conceptualizing single-cell biology. (a) Before assays of individual cells were available, one would imagine that each cell acts in basically the same way. When expression is detected, for example, by northern blot, the averaged cells would be assumed to be actively producing transcripts to some level such that a detectable threshold is reached. (b) Once transcription sites were visualized using *in situ* techniques [13], heterogeneity in single-cell expression was apparent. This allowed for cells to be categorized in one of two fundamental states – 'gene on' and 'gene off'. (c) Once multigene cellular transcriptional profiles were detected [11], it became apparent that perhaps no two cells' precisely measured expression repertoires would be the same. If this is so, biological variability is less difficult to explain than commonality. The complement of mRNAs and proteins might vary considerably from cell to cell. In the context of relatively homogeneous cell physiology, this means that there is high tolerance for fluctuations in the pool of biomolecules. (d) It is possible that variable expression activity is offset by redundancy between genes, integration of expression over time or relatively stable protein levels. Expression changes can be severe, but their physiological effects are dampened by functional overlap or posttranscriptional controls.

products at each step of amplification [8]. Applications of similar corrections in one-cell expression assays [9] could potentially allow for more precise *in vitro* study of the biology of single cells, albeit indirectly through the magnifying lens of PCR.

A novel view

Obvious exceptions to the requirement of deriving an averaged or artificially amplified population of mRNA are *in situ* techniques, in which the species of interest is detected in its location in individual cells. While these methods allow precise localization and quantitation of nucleic acids [10], simultaneous assay of many transcripts has not been possible and the image data produced are usually difficult to interpret in an objective, high-throughput manner. Recent developments in fluorescence *in situ* hybridization (FISH) surpass these limitations to yield single-cell information largely unavailable to other techniques. Results using a multiplexed expression FISH approach show considerable cell-to-cell variability. Even so-called synchronized culture conditions result in a diverse population of expression profiles that defy simple statistical modeling. When a transcriptional program is initiated, there is no common, ordered pathway of gene activations within every step for each cell. Instead, each nucleus is probabilistic in its expression repertoire. Correlated transcription in individual cells might depend on seemingly random events, like cellular levels of a certain transcription factor. For the most part, each gene can be conceptualized as an independent variable with a likelihood of expression associated with it [11].

Single-cell biology turns out to be drastically different from our expectations of highly ordered, consistent expression. If we consider a snapshot in time, even measuring only a small cohort of genes, most cells are not alike. Given, for example, 30 different genes, there can be every variation possible, from no genes expressed, to all at once, and every combination in between. Assuming, for simplicity, that the activity of each gene is a binary event, there would be 2^{30} possible permutations, or over a billion possible outcomes. Given the ability to measure the entire transcriptome, one might venture that no two cells in our body respond to a stimulus in exactly the same way.

Concluding remarks

It appears that our vision of gene expression must be substantially revised (see the scheme presented in Fig. 1). How can we reconcile this view with our concept of the average cell? One possibility recently supported by evidence is that there is a 'normal' response to any cellular condition, but how each cell responds is stochastic and subject to various types of noise [12]. Another explanation for variance in expression is that many genes are functionally redundant and the possible biological outcomes from these variations in gene expression are less significant. A third possibility is that while variability exists at each time point, ultimately all cells will end up expressing the necessary genes over time, thereby resulting in similar protein complements. According to any of these approaches, gene expression at the cellular level does not require tight specifications, and the tolerance for variation is high. This model predicts that monitoring expression of a gene at a given time point cannot precisely predict a particular phenotype. This poses two fundamental questions: how is it that cells can individually make a suitable response when single-cell behavior appears so random? Even more so, how can many stochastically determined cell states be coordinated into a functional organ or organism that seemingly requires uniformity? The introduction of a means to interrogate expression in single cells has revealed how complex behavior is at the fundamental unit of life. We can only continue to speculate as to how, amidst such disorder, the genome is faithfully translated into life.

Acknowledgements

We thank Mike Lorenz for help with the figure. Research in the authors' laboratory is supported by the Innovative Molecular Analysis Technologies program at the National Cancer Institute, and research and training grants from the NIH.

References

1 Bowtell, D.D. (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nat. Genet.* 21 (Suppl. 1), 25–32
2 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
3 Lockhart, D.J. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675–1680
4 Velculescu, V.E. *et al.* (1995) Serial analysis of gene expression. *Science* 270, 484–487
5 Emmert-Buck, M.R. *et al.* (1996) Laser capture microdissection. *Science* 274, 998–1001
6 Mikulowska-Mennis, A. *et al.* (2002) High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques* 33, 176–179
7 Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836
8 Higuchi, R. *et al.* (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* 11, 1026–1030
9 Klein, C.A. *et al.* (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* 20, 387–392
10 Femino, A.M. *et al.* (1998) Visualization of single RNA transcripts *in situ*. *Science* 280, 585–590
11 Levsky, J.M. *et al.* (2002) Single-cell gene expression profiling. *Science* 297, 836–840
12 Elowitz, M.B. *et al.* (2002) Stochastic gene expression in a single cell. *Science* 297, 1183–1186
13 Lawrence, J.B. *et al.* (1989) Highly localized tracks of specific transcripts within interphase nuclei visualized by *in situ* hybridization. *Cell* 57, 493–502