

DEBATE

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Do DNA Microarrays Tell the Story of Gene Expression?

Simon Rosenfeld

National Cancer Institute, EPN 3108, 6130 Executive Blvd., Rockville, Maryland 20892.

Corresponding author email: sr212a@nih.gov

Abstract: Poor reproducibility of microarray measurements is a major obstacle to their application as an instrument for clinical diagnostics. In this paper, several aspects of poor reproducibility are analyzed. All of them belong to the category of interpretive weaknesses of DNA microarray technology. First, the attention is drawn to the fact that absence of the information regarding post-transcriptional mRNA stability makes it impossible to evaluate the level of gene activity from the relative mRNA abundances, the quantities available from microarray measurements. Second, irreducible intracellular variability with persistent patterns of stochasticity and burstiness put natural limits to reproducibility. Third, strong interactions within intracellular biomolecular networks make it highly problematic to build a bridge between transcription rates of individual genes and structural fidelity of their genetic codes. For these reasons, the microarray measurements of relative mRNA abundances are more appropriate in laboratory settings as a tool for scientific research, hypotheses generating and producing the leads for subsequent validation through more sophisticated technologies. As to clinical settings, where firm conclusive diagnoses, not the leads for further experimentation, are required, microarrays still have a long way to go until they become a reliable instrument in patient-related decision making.

Keywords: DNA microarrays, post-transcriptional mRNA stability, genetic regulatory networks, reproducibility, mRNA abundance, transcription rate, gene activity, stochasticity, burstiness, clinical applications

Gene Regulation and Systems Biology 2010:4 61–73

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Clinical promise of DNA microarray measurements

A key premise of DNA microarray ideology is that the mRNA abundances harvested from the cells are indicative of the levels of activity of the genes from which they are transcribed. In somewhat simplified terms, it seems natural to think that the more active is a gene the more mRNA molecules it transcribes in a given time interval, and therefore the higher concentration of those can be detected in the mRNA assay. This premise further propagates into the microarray differential expression probing in which higher mRNA fluorescent intensity is believed to indicate a higher mRNA concentration, thus reflecting higher activity of the corresponding gene. This notion is solidified in terminology: the microarray spots are frequently called *genes*. However, such terminology obscures the fact that actual biological material the microarrays work with are mRNAs, not the genes themselves.

Unfortunately, the simple scheme outlined above is confounded by many factors and intricacies of gene expression machinery. Some of these factors have been discussed in previous works by this author.¹⁻³ In this paper, special attention is given to one important confounder, namely to the post-transcriptional stability of mRNA molecules. It is quite obvious that if a certain mRNA molecule is chemically unstable then it gets quickly dissolved in aggressive cytoplasmic environment. As a result, the gene this mRNA has been transcribed from loses its proxy in the transcriptome, whether this gene is active or not. In the straightforward schema outlined above, the low mRNA abundance would be interpreted as reflecting low gene activity, whereas in reality it is only reflecting a low mRNA post-transcriptional stability. This simple idea can hardly be called new; it has been touched upon in many previous publications. However, an unpleasant truth is that its importance for the interpretation of DNA microarray measurements has not yet been properly explored and has not yet made its way into routine measurements. This means that an omnipresent and powerful player behind the scene of *transcriptomics* remains unknown and unchecked in microarray measurements, thus leading to contradictions, ambiguities and misinterpretations.

A common concern in microarray data analysis is poor reproducibility. In the editorial⁴ preceding the report summarizing the large-scale Microarray Quality Control Project,⁵ this aspect of microarray measurements has been characterized as follows: *“Doubts linger about the reproducibility of microarray experiments at different sites, the comparability of results on different platforms and even the variability of microarray results in the same laboratory. After 15 years of research and development, broad consensus is still lacking concerning best practice not only for experimental design and sample preparation, but also for data acquisition, statistical analysis and interpretation... Clearly, microarrays have a long way to go before they can be used to support regulatory decision-making or accurate and consistent prediction of patient outcomes in the clinic.”*

Microarrays are considered as a great hope in the bench-to-bedside translational research. In its annual, 2004, report “Critical Path”, FDA identifies microarrays technology as crucial in advancing medical product development and personalized medicine. The regulation of DNA microarray-based medical diagnostic devices and the review of toxicogenomic data submitted as part of an Investigational New Drug, New Drug Applications or Biologics License Application are expected to become an essential regulatory responsibility of the FDA in the advancement of personalized medicine.⁶ As a sign that this program hit the road of implementation, the U.S. Food and Drug Administration cleared for marketing a test that determines the likelihood of breast cancer returning within 5 to 10 years after a woman’s initial cancer. It is the first cleared molecular test that profiles genetic activity. This test relies on microarray analysis. (www.fda.gov/NewsEvents, 2007).

These high stakes and hopes attached to microarrays are in stark contrast to a chorus of skeptical opinions regarding microarray diagnostic capabilities and technical performance. Thus J. Ioannidis, in his paper with a telling title “Microarrays and molecular research: noise discovery?”,⁷ writes: *“The promise of microarrays has been of apocalyptic dimensions... All diseases are to be redefined, all human suffering reduced to gene-expression profiles. Cancer has been the most common early target of this revolution and publications in the*



most prestigious journals have heralded the discovery of molecular signatures conferring different outcomes and requiring different treatments. Yet, on close scrutiny, in five of the seven largest studies on cancer prognosis, this technology performs no better than flipping a coin. The other two studies barely beat horoscopes.” This austere opinion is echoed in:⁸ “The potential advantages of improving tumour classification by expression profiling has been central for several large-scale breast cancer studies that have reported identification of signature gene lists with potential for prediction of clinical outcome over the past few years. The microarrays used in different studies generally have several thousand genes in common and the underlying principles of the measurement technologies are the same. The most striking finding when comparing the signature lists is the virtually complete lack of agreement in the included genes.”

Meticulous analysis and sober evaluation of the state-of-the-art in microarray technology presented in⁶ is summarized as follows: “Profound problems in data quality have been observed from analyzing published data sets, and many laboratories have been struggling with technical troubleshooting rather than generating reliable data of scientific significance... These fundamental issues must be adequately addressed before microarray technology can be transformed from a research tool to clinical practices.”

In the above cited report,⁶ all the *profound problems* of microarrays are categorized in four big classes: technical (microarray manufacturing, sample collection, RNA extraction, cDNA and cRNA synthesis, fluorescent labeling and hybridization); instrumental (laser intensity, scanner calibration, image acquisition and spot quantification); computational (data preprocessing, normalization, statistical analysis of differential expression); and interpretive (biologic reasoning, pathway analysis, bioinformatics tools). The authors point out that “A single hidden, uncontrolled factor may completely negate an experiment.” Obviously, in clinical settings the cost of such a *single, uncontrolled factor* may be much higher and lead to wrong diagnosis with potentially harmful consequences for patient-related decision making.

In this paper, several aspects of the DNA microarray methodological weakness are analyzed. First,

the attention is drawn to the fact that absence of the information regarding the post-transcriptional mRNA stability makes it highly problematic to evaluate the level of gene activity from the relative mRNA abundances. Second, irreducible intracellular variability with persistent patterns of stochasticity and burstiness put natural limits to reproducibility. Third, strong interactions within intracellular biomolecular networks make it difficult, if possible at all, to build a bridge from the transcription rates of individual genes to structural fidelity of their genetic codes. Among these three topics, the post-transcriptional mRNA stability is the central one. It is the author’s view that this problem is a sort of elephant in the room; it is utterly important in many contexts, it is well known to experimental and theoretical biologists, and yet it is largely unaddressed in the context of routine microarray measurements and data interpretation.

Post-transcriptional mRNA stability belongs to the fourth category among the “*profound problems*” listed above,⁶ that is, to the category of biological interpretations. Notably, despite tremendous difficulties of purely technical nature, the authors⁶ believe that this group of problems “*is the most difficult task in microarray studies.*” Let us imagine for a moment that in some, hopefully not so distant, future all the technical problems associated with microarrays are solved, and the measurements became perfectly intra-laboratory repeatable, between-laboratories reproducible, cross-platform compatible and FDA-approved for using in clinical settings. At this point, a cluster of bigger questions will come into focus: What exactly will microarray measurements tell us about the state of the cell? Is it really true that mRNA assays provide us with valid and comprehensive information regarding the status of genome? Isn’t it misleading to equate *transcription* profiling with *expression* profiling? What are the actual relations between the genome and transcriptome that are elucidated by the DNA microarrays? Are they just weak spurious correlations or something more tangible? What is the diagnostic value of the DNA microarray measurements?

In pre-technological era, medical doctors often relied on such biomarkers as odor of the body, wetness of hands, color of the skin, texture of the nails, etc. Undoubtedly, in the absence of more definitive markers, even these diagnostic tools could serve as a basis



for patient-related decision making. These tools continue to be of value in many domains of modern, especially holistic, medicine. However, it is also obvious that these holistic markers are wide open to a variety of interpretations and produce just vague hints on possible cause of disease. In modern terminology, one would call these tools *hypotheses-generating*, rather than *diagnostic*. Since DNA microarrays are expected to be a workhorse of future personalized medicine, a fundamental question arises: whether or not the DNA microarrays may in principle be anything more than a hypotheses-generating instrument appropriate only for scientific laboratory experimentation?

Basic Concepts: Transcription Levels and Transcription Rates

Prior to reviewing known facts pertaining to post-transcriptional mRNA stability, we need to introduce some terminology. According to the *Central Dogma* in molecular biology,⁹ there is a unidirectional flow of biological information from genes to proteins, with mRNA being an intermediary. Per copying from genetic code by RNA Polymerase (transcription) and prior to protein synthesis by ribosomes (translation), the mRNA molecules undergo numerous transformations, some sequence-specific, some not. The key steps in these transformations are export of nascent mRNA to cytoplasm assisted by ribonucleoprotein complexes (mRNPs),¹⁰ removal of non-coding regions of mRNA (splicing), maturation, editing, and other processes preceding binding of mRNA molecules to the ribosomal sites.⁹ It is within this time span between transcription and translation, the mRNA is being isolated for preparation of microarray assays. Obviously, there is not much that can be standardized here; rich gamut of biochemical processes and their natural variability will inevitably be imprinted in the mRNA assay.

For quantitative description of transcription, two concepts should be introduced, that is, transcription level (TL) and transcription rate (TR). TL is a number of the mRNA copies of a certain type present in the cell at a certain moment of time, t . TR is a number of transcripts produced within the cell per unit of time. TR and TL are two independent characterizations of transcription having different dimensions (a detailed discussion of these concepts may be found in the author's.)¹¹ Obviously, it is the TR that is a direct characterization of the gene's activity. The higher TR , the more transcripts are

produced per unit of time, the more active is the gene. In the opposite case, one may say that the gene is shut down if its TR is zero. Mathematically, the roles of TR and TL are expressed by the equation

$$\frac{d}{dt}TL(t) = TR(t) - \frac{1}{\tau}TL(t) \quad (1)$$

where τ is the characteristic time of degradation. All the quantities above are assumed to be mRNA-specific. In the simplest case, when τ is independent of other TLs (linear system), solution to the equation (1) is

$$TL(t) = TL(0)\exp\left(-\frac{t}{\tau}\right) + \int_0^t TR(t')\exp\left(\frac{t'-t}{\tau}\right)dt' \quad (2)$$

In the case when transcription rate is time-independent, we have

$$TL(t) = TL(0)\exp\left(-\frac{t}{\tau}\right) + \tau TR\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] \quad (3)$$

which means that with time increasing, $t \rightarrow \infty$, TL changes from its initial value, $TL(0)$, to its asymptotic limit, τTR . Even from this grossly simplified example one may see that the stable transcription level

$$TL(\infty) = \tau TR, \quad (4)$$

depends not only on the transcription rate, TR , but also on the characteristic time of degradation, τ . For a stable transcript, that is for one with large τ , the cell may be abundantly populated by the corresponding mRNAs, regardless of the transcription rate of gene. For two genes with the same transcription rates, their respective mRNA abundances are proportional to the times of degradation. If in a certain experiment, the quantity being measured is TL , then, generally speaking, this measurement tells nothing about the TR without an *a priori* knowledge of the time of degradation, τ .

This issue of meaning and interrelation between the TR and TL has been elucidated in a number of previous works by this author (e.g.).¹¹ This difference is similar to that between someone's income (transcription rate) and his/her account balance



(transcription level). Even complete knowledge of someone's account balance generates no information regarding his/her income. Likewise, complete knowledge of the transcription levels creates no information regarding transcription rates. In other words, one may say that measurements of mRNA abundances are not fully representative of gene activity. In particular, abnormal mRNA abundance cannot serve as an evidence of the gene's abnormality. Mathematically, TR and TL are the quantities of different dimensions; they are neither comparable nor directly linked to each other.¹¹

The simplified way of reasoning described by equations (1)–(4) is only applicable to linear systems. The things change drastically if the system is nonlinear. In biochemistry, it is often the case that production and/or degradation of certain constituents is supported by, or even happens only because of, interactions with other constituents. Let us imagine a coupled pair of chemical reactions in which the production terms are independent while degradation happens only because of their interaction with each other. Instead of (1), we will have now a system of two coupled nonlinear differential equations

$$\begin{aligned}\frac{d}{dt}TL_1(t) &= TR_1(t) - r_1TL_1(t)TL_2(t), \\ \frac{d}{dt}TL_2(t) &= TR_2(t) - r_2TL_1(t)TL_2(t)\end{aligned}\quad (5)$$

where r_1 and r_2 are kinetic rates. Detailed analysis of this system goes beyond the scope of this paper; general theory may be found in.¹² Here, we only mention several fundamental differences with the linear case. Firstly, it is no longer possible to unconditionally ascribe to a constituent a certain degradation rate; generally speaking, this degradation rate becomes dependent on the level of other constituent. Secondly, asymptotic, $t \rightarrow \infty$, solution to the system (5) may not exist. This means that it may happen that a simple asymptotic relation similar to (4) may be unavailable. Thirdly, in contrast to the linear case, the degradation rate may be dependent on initial conditions; e.g. it may happen that high initial concentrations disappear quickly, whereas the low ones may stay unchanged for a long time. This would mean that the degradation rate would no longer be a descriptor of a certain type of molecules; it would also be a function of their concentrations.

In biochemistry of gene expression, there is a vast universe of events that may happen only if a large number of constituents simultaneously enter the chemical interaction. In mathematical terms, this means that chemical constituents form a highly nonlinear biochemical network. For such systems, simple concepts borrowed from the linear theory, strictly speaking, are not applicable. If they are nevertheless used for quantitative analysis of experimental data, with heavy reliance on intuition and common sense, then inevitably they will be prone to contradictions and ambiguities. Hence, they should be used with extreme caution. Interpretive difficulties of this kind are well known even with respect to much simpler systems, such as systems of enzymatic reactions described by the Michaelis-Menten kinetics.^{13,14} In the dynamics of large systems, only a carefully formulated mathematical model may provide a solid basis for self-consistent interpretation.

Kinetics of mRNA Degradation

Traditionally, the speed of mRNA degradation (as well as of any other chemical constituent) is measured by the quantity called *half-life*, $t_{1/2}$, that is, the time required for a concentration (or a copy number) to drop to the half of its initial level. In linear systems, when the characterization in terms of a single degradation time is valid, there is a simple relation between the half-lives and degradation times: $t_{1/2} = \tau \ln 2$. In more complex systems, when the concepts of linear theory are not applicable, such a simple relation may not exist, and other chemical constituents may influence the half-lives. In the context of microarray measurements, the crucial questions are whether or not degradation rates of different mRNAs are substantially different, and whether or not they are sequence-specific. If all the half-lives were approximately of the same order of magnitude for all mRNAs and thus could be considered as a non-specific constant for the entire transcriptome, then, due to equation (4), a direct relation would exist which links the transcription level and transcription rate. It would be a favorable situation for microarray measurements; in this case it would be correct to claim that microarrays provide information regarding gene activity. Available experimental evidence, however, shows that it is very far from being the case.



Thus, it has been shown in¹⁵ that in the mouse embryonic stem cells, the mRNA half-lives vary within the range from 2 to 24 hours and beyond. In yeast, according to,¹⁶ mRNA half lives range from 3 min to 1.5 hours. The study of time-course kinetics and half-lives of ~1500 mRNAs in microbial cells has been undertaken in.¹⁷ The half-lives have been found to vary from less than 30 seconds to more than 20 minutes. An important observation made in this work is that half-lives may cluster into comparatively tight groups. This means that what is traditionally interpreted as *gene co-regulation* may in fact be nothing more than a reflection of this clustering of half-life times. It is not yet known what biochemical factors influence degradation rates. Preliminary considerations, mixed with intuition, suggest that the mRNA decay rates are not sequence-specific; rather they are dependent on the mRNA's length and some structural properties such as the number of exon junctions per open reading frame.¹⁵ Wide variations of mRNA half-lives ranging from several minutes to several hours in human T lymphocytes have been also reported in.¹⁸

Another aspect of the mRNA degradation kinetics is that it is epigenetically regulated. For example, it was found in¹⁹ that there is no direct connection between gene activity and mRNA degradation rate. In more formal language, this means that τ is independent of TR . If such a dependence, $\tau = \tau(TR)$, did exist then equation (4) would reduce to $TL(\infty) = \tau(TR)TR$, and therefore there would be some hope that, per solving this equation with respect to TR , it would be possible to directly link TR to $TL(\infty)$. Since τ is regulated epigenetically, such a hope is baseless. To the same end, mRNAs may have different stabilities at various cell stages, at various levels of its protein product, or at various developmental stages.²⁰ As a drastic example, these authors indicate that immunoglobulin p mRNA is at least sixfold(!) more stable in the late versus early stage of B cells. We specifically emphasize that all these variations occur post-transcriptionally, that is, without direct connection to the parent gene's activity. In the microarray context, this means that regardless the gene's transcription rate, the mRNA's half-lives have been modulated by factors other than gene activity. Broad study of factors influencing mRNA stability has been undertaken in.¹⁶ The authors report that there is no simple correlation between the mRNA half-lives and their abundances. They also conclude that the

control of mRNA decay rates is a fundamental feature of gene expression machinery.

An attempt to tackle the problem of possible connection (or lack thereof) between the TL and TR experimentally has been undertaken in.²¹ In this time-course experiment, the TL (mRNA copy numbers) and TR (rate of elongation) have been measured simultaneously in budding yeast. It was found that from a total of 5,500 TR and TL time-series pairs, about half turned out to be uncorrelated with each other. Partial explanation to this phenomenon has been given in^{22,23} by this author.

We conclude this brief review by the proposition that post-transcriptional life of the mRNA molecules is largely independent of the status of parent genes. The mRNA half-lives are modulated by a large number of epigenetic factors and may vary within order of magnitude. It is evident from here that gene activity measured by its transcription rate is just one of many factors determining the mRNA abundances. There are no, and cannot be, any one-to-one correspondences between the mRNA abundances and corresponding gene's levels of activity.

Crowding, Burstiness, Stochasticity and Noise

The intracellular environment is a fairly chaotic place in which no molecular event ever happens in accordance with strict, well defined rules. In,^{24,25} Goodsell represents a vivid picture of the interior of a living cell compiled from electron microscopy and X-ray crystallography. He likens this picture to a crowded airport terminal where each person slowly moves to his destination pushing and shoving other people, and where each individual trajectory is hundreds of times longer than it would be in the space free of other passengers. Technically speaking, about 70% of intracellular space is filled with water. On this basis, a macromolecule's motion is often portrayed as if it is the motion of a colloidal particle in an aqueous solution disturbed only by the *collisions* with other particles.^{26,27} However, it should be taken into consideration that the water molecules in the cytoplasmic region are surrounded by much bigger molecules with typical molecular mass well above 40 KDa.²⁴ Under these conditions, the tiny water molecules (18 Da) behave entirely differently from those in bulk water. Their strong polarity as well as numerous phenomena of protein-water interactions^{26,28} produce a much more intricate behavior



of the macromolecule's motions than simple friction described by the Stokes' law. Effects of crowding has been discussed in depth in.²⁹ This author concludes that crowding exerts profound quantitative effects on macromolecular interactions in living systems.

The premise that the trajectories of macromolecules within the cytoplasm represent the patterns of a *random walk* is well appreciated in the literature.³⁰ However, many models utilizing this concept also adhere to the idea of *collisions* for characterizing molecular interactions. The concept of molecular collisions originates in the kinetic theory of gases and generally assumes that the molecular interactions are comparatively fast and localized events, that is, they take much less time than that between the collisions and unfold within a space much smaller than that between the molecules. None of these assumptions can be true with respect to intracellular interactions. As is known from the theory of chemical reactions, when two (or more) molecules approach each other, they first create a meta-stable transition state (a.k.a. activated complex), which may be considered as an intermediate, usually short-lived, compound with its own chemical properties. Given the high density and big size of macromolecules within the cell, it can be easily imagined that, in fact, the molecules spend a substantial amount of their lifetimes being clustered in a variety of temporary intermediate associations and exploring their very complex free energy landscapes. The complexity of this picture is a far cry from a naive collision scheme adopted in many theoretical models.

A key premise of molecular biology known as *Central Dogma* assumes that there is a unidirectional flow of genetic information from the DNA to proteins with mRNAs being the intermediaries. Although it is generally accepted that the DNA contains instructions for assembling the proteins and not vice versa, there is a fundamental, though implicit, reciprocal role of proteins in producing mRNAs: they serve as transcription factors in the very process of decoding the genes. In order for this circulation of matter and information to be possible, the proteins produced by ribosomes in the cytoplasmic area should be able to reach their pre-specified places in the gene's regulatory regions in a timely manner. Since typically from 30 to 100 regulatory proteins per gene are used as transcription factors,³¹ a corresponding number of genes should go through

their individual cycles of expression in a perfectly synchronized manner; otherwise, a mere shortage of a few transcription factors may lead to the gene's drop-out from the regulatory process and a halting of big sections of transcription machinery. Obviously, a simultaneous random walk of thousands of molecular species through all kinds of impediments in a densely packed intracellular environment is not very conducive to an orderly assembly-line-like behavior. Spontaneous failures like traffic jams, bottlenecks, backlogs, delays, loss of synchronization, etc., are inevitable circumstances of intracellular biochemical dynamics.

A large number of molecular events require cooperative participation of a team of molecular species, with transcription itself being the best example of the kind. Since each of these team members makes its way towards the designated regulatory sites through a densely tangled intracellular environment, each such event can be only thought of as a sporadic coincidence. In²³ these events have been termed as instances of *stochastic cooperativity*. A number of observations support the view of transcription as a sequence of sporadic events (see more detailed discussions in).^{2,3,11,32} Recent experiments³³ demonstrated that even in an individual cell, the production of a protein and supporting enzymes is a stochastic process following a complex pattern of *bursting* with random distribution of intensities and durations. Similarly, it was found in³⁴ that quantitative relations between transcription factor concentrations and the rate of protein production "fluctuate dramatically in individual living cells, thereby limiting the accuracy with which genetic transcription circuits can transfer signals." The phenomenon of burstiness is wide spread in genetic regulation. Thus, the authors³⁵ report that "transcription occurs in pulses in muscle fibers." In Ref.,³⁶ it was found that "transcription of individual genes in eukaryotic cells occurs randomly and infrequently." Similar observations have been made in.³⁷⁻⁴⁰

Oftentimes, all these innumerable peculiarities and intricacies of intracellular machinery are covered by the blanket term *stochasticity*⁴¹⁻⁴⁴ The researchers with engineering background, in their explorations of analogies between intracellular dynamics and electronic circuitry, go even further and prefer the word *noise* for the same purposes. Sometimes, even biologists express frustration that in the cell's life



things are not so easily formalizable and standardizable as in electronic engineering.⁴⁵ Although the motivation for using engineering terminology is quite obvious, yet it should be recognized that these analogies should not be stretched too far: there are fundamental and irreconcilable differences between the man-made engineering devices and living organisms developed evolutionarily (see detailed discussion in the author's).²

Natural stochasticity and noise are great enemies of reproducibility in any domain of science, including biology and DNA microarrays. It is quite uncertain what exactly is supposed to be reproducible if the system is stochastic and noisy by its very nature. Later in this paper, this topic will be explored in more detail.

Nonlinear Dynamics Perspective

With the advent of high throughput measurements in biology (microarrays, protein mass spectrometry, etc.) it became increasingly clear that it is no longer possible (as it used to be in the past) to associate the origins of disease and prevention strategies with the individual genes, or proteins, or other molecular *markers*. It is now widely recognized that only a systemic view can provide an appropriate vehicle for understanding complex biology and navigating the oceans of available data. Lack of such a systemic view is often seen as a major bottleneck for making progress in understanding cancer.⁴⁶

From the systemic viewpoint, intracellular biochemical machinery is a high-dimensional highly nonlinear system in which interactions are governed by the laws of chemical kinetics and thermodynamics. A brief term for denoting such systems is *biochemical networks*. During the last decades, an impressive arsenal of mathematical, data-analytical and computational tools have been amassed for studying the dynamics of biochemical networks.⁴⁷ Nonlinear dynamics of genetic regulatory systems is a central topic of a number of publications by the author of this manuscript.^{2,11,32,48}

Biochemical networks have much in common with other networks of natural, technological and societal origins. Among them are the internet, predator-prey food chains, social networks, traffic systems, epidemics, rumor networks, electric power grids and many others. A fundamental feature of all the networks is that no event within them can occur

in isolation, independently of other events. Figuratively speaking, each individual event creates a domino-effect of events propagating throughout the system. Genetic regulatory networks represent a perfect example. Each transcription event is supported by a large number of transcription factors with many of them being the proteins translated from other mRNAs; these, in turn, are transcribed from other genes. This means that indirectly each transcript is a product of the teamwork of many genes, not just the one from which the mRNA has been directly encoded. For example, if a certain transcription factor originated from gene-A failed to report to regulatory site of gene-B, then the latter will be halted, or shut down, thus mimicking its low or zero activity. In fact, it is gene-A to blame for the failure of gene-B. In somewhat loose terms, it may be said that a network acts as a whole and reacts as a whole. This tight interconnectedness makes it difficult, both technically and conceptually, to analyze the observational and experimental data pertaining to complex systems. In particular, mRNA abundances, to a large extent, are the products of the system's behavior, rather than of activity of individual parent genes.

Another important issue pertaining to large biochemical networks is dynamic stability. A fundamental question posed by R. May in his seminal paper⁴⁹: "Will a large complex system be stable?" gets a negative response for a vast majority of such systems. Only in highly specialized conditions, such as those stipulated by the famous *Deficiency Zero Theorem* (e.g.),⁵⁰ a large biochemical network may be dynamically stable. Behavior of a dynamically unstable system cannot be anything else but a stochastic process.¹¹ It should be emphasized that this kind of stochasticity cannot be eliminated through precise measurements. Like waves in the ocean, or turbulence in the atmosphere, they are intrinsic properties of the system itself.

Reproducibility Conundrum: What Exactly Should Be Reproducible?

In natural sciences, the terms *reproducibility* and *repeatability* are referred to the capability of producing identical results in a series of repeated measurements. In experimental biology, it is sometimes drawn a distinction between these two terms. Usually,

the former refers to the agreement of measurements between different sites (say, laboratories) made with the same or identical assays, and the latter refers to the agreement between repeated measurements at the same site with the same assay. The concept of *noise* has an implicit connotation that a system by itself, i.e. without noise, generally possesses some properties of self-identity, and it is only a set of uncontrollable nuisance parameters that are to blame for destroying an idyllic picture of precise reproducibility. In intracellular biomolecular dynamics, various types of variability are often called *intrinsic noise*.⁵¹ Schematically, intrinsic noise includes both stochasticity due to internal nonlinear instabilities and random variations caused by uncontrollable random forces. Another component of the overall variability is called *extrinsic noise*. This type of variability includes reaction of the system to uncontrollable random forces external of the system. It may also include sample heterogeneity and spatial variability. There is a blanket term *natural variability* covering all these types of variability. Accurate measurements can only produce a precise picture of this variability, but not eliminate it. By standardizing the protocols of sample preparation, applying temporal, spatial or across-sample averaging, it is possible to reduce some components of natural variability, but residual variability will be always present in experimental results. This means that, even with precise measurements, a series of identical experiments would never produce identical results due to the omnipresence of uncontrollable natural variability. In other words, the measurements cannot be completely reproducible even in principle if the object under exploration possesses the property of natural variability.

On top of natural variability, there is another source of errors, the one associated with inaccuracies of measurements. These are usually called *instrumental noise*. Detailed classification and analysis of instrumental errors in microarray experiments is given in the above cited work.⁶ Microarray measurements involve a number of delicate physical and chemical processes, such as scanning of fluorescent labels, laser calibration, spot quantification, etc. These processes are sometimes very difficult to standardize and maintain their precision and stability. It should be noted also, that although the errors of this type are called instrumental they can also include

some irreducible natural variability associated with the physical processes within these instruments. For example, stability of laser illumination can be reduced only down to the level of thermal noise and quantum fluctuations. Accuracy of spot quantification always would be limited by granular structure of underlying substrate glasses, etc. This means that instrumental accuracy always will be limited by natural variability within the measurement device itself. Due to this reason, measurements can never be entirely reproducible even if the subject of measurements is perfectly stable and invariable.

Schematically, the total error budget can be expressed as follows

$$\sigma^2 = \underbrace{\sigma_0^2 + \sigma_1^2}_{\text{sample}} + \underbrace{\theta_0^2 + \theta_1^2}_{\text{instrument}} \quad (6)$$

where the first pair of terms depicts natural sample variability and the second pair represents instrumental noise, that is,

σ_0^2 is variance associated with natural stochasticity due to molecular interactions;

σ_1^2 is variance associated with sample heterogeneity and extrinsic noise;

θ_0^2 is variance associated with natural variability underlying physical principles of measurements;

θ_1^2 is variance associated with errors in calibration and environmental factors.

This schema is, of course, fairly crude and may be broken down into many subtler details. Some of the components in (6) may be reduced through perfection of protocols, standardization, careful experimental design, technological advances. Others are irreducible and will have their footprint in the measurements under any circumstances. Thus, one may imagine that with time, technological progress may help to reduce the terms σ_1^2 and θ_1^2 to negligibly small values. However, the terms σ_0^2 and θ_0^2 reflect the very nature of the measurements and of the object under exploration. They can never be eliminated or even reduced (without killing the subject or turning off the instrument) whatever time, labor, money and intellectual efforts are poured into this improvement. After the technology attains the status when $\sigma_1^2 \ll \sigma_0^2$ and $\theta_1^2 \ll \theta_0^2$, further efforts directed towards improving reproducibility will become worthless. The irreducible part of combined



natural variability, $\sigma^2 = \sigma_0^2 + \theta_0^2$, puts an ultimate limit to attainable reproducibility.

Differential Expression Testing and Elements of Statistical Mythology

With the advent of microarrays, a vast body of statistical literature has emerged in which the problem of differential expression testing has been tackled with the most sophisticated statistical methodologies available. From a purely statistical point of view, analysis of microarray data is a new type of problem largely unfamiliar to traditional mathematical statistics. The core problem has been termed a *curse of dimensionality*. In statistics, this term denotes the situation when the number of samples available for the analysis is smaller, or even much smaller, than the number of parameters to be estimated.⁵² In the microarray context, the number of mRNA abundances to be analyzed may be as large as 40,000, whereas the number of microarrays available for the analysis is typically in the dozens, and only very rarely may approach hundred. In order to defeat the curse of dimensionality and increase statistical power, a common idea of *borrowing strength* from the totality of all the data available is being employed. Such a paradigm requires strong *a priori* assumptions regarding probabilistic properties of parameters; it also requires development of special significance scores for detection of differentially expressed genes. This big problem has opened numerous opportunities for statistical creativity, and inspired development of a large variety of statistical methods. This multitude of methods, however, is not of much help in the routine work of an experimental biologist, unless a professional statistician is a member of his team. Not only does a biologist face the problem of deciding which statistical method is more appropriate for his/her experimental situation, but, regrettably, statisticians themselves still lack a consensus regarding comparative merits of various approaches (e.g. see discussion in).⁵³ It is a sort of irony that the very statistical method of assigning a significance score to a differential expression turns out to be difficult to standardize. However, this part of overall standardization is crucial because generally different significance scores would produce different lists of differentially expressed genes.

A simple example of the kind is the selection between two alternative methods, both very popular,

the one based on the fold change and the one based on the p -values of t-test. This question is discussed in much detail in the work⁵⁴ by this author. Understandably, biologists are more inclined to trust what they see and rely on fold change as a significance measure. Although it is well appreciated that big fold changes may be spurious and originate from pure noise, yet the fold-change-based estimates may serve as valuable leads for subsequent experimental verification using more accurate (and usually more expensive) methods such as Quantitative RT PCR or Northern Blot. An alternative to the fold-change approach is to compute the gene-specific t-tests and rank significance of differential expression in the reverse order of p -values. This would be a preferable choice for a statistician. In contrast, a biologist would be leery about such a criterion because there is always a suspicion that small p -values may originate not from the differences in transcription levels in the assay but from the ubiquitous small uncontrollable biases. As discussed in the above mentioned report,⁶ there are numerous sources from which such biases may originate. An attempt to use both criteria simultaneously usually results in a very meager list of differentially expressed genes or none at all. (In order to reconcile these two extremes, in⁵⁴ a combination score, the *bio-weight test statistic*, has been proposed by this author.) Importantly, the p -value-based score and the fold-change-based score take into focus different, largely alternative, properties of the assay, and inevitably produce different lists of differentially expressed genes.

It is not always clearly understood that assigning different significance score to differential expression would actually mean assigning different meanings to the very notion of differential expression. (As a crude analogy, it is the same as comparing two groups of subjects by their BMIs, or alternatively, by the sizes of their shoes.) Some authors even go so far as to propose *validation* of one statistical method by assessing its agreement with another. Thus, the authors⁵⁵ claim that pessimistic view of microarrays as a diagnostic tool may originate from the fact that a *single statistical practice is used without alternative validation*. In the above discussed example, that would mean that the fold-change-based score should be *validated* by the p -values-based score. (Or, in the above crude analogy, the criterion based on shoe size should be *validated* by that of BMI) Obviously, as



long as the choice of statistical methodology is not restricted by some sort of consensus or tradition, it will always be possible to continue the search for such an approach that would allow declaring significant any group of genes *a priori* recognized as significant. Ironically, it may easily happen that the variability associated with the selection from the pool of available statistical methodologies will make its own contribution towards poor reproducibility of microarray measurements. Essentially, this is a reflection of the well known in mathematical statistics effect of inflation of variance due to model selection, only applied in a different context. In anticipation of using the microarrays in clinical practice, it is easy to imagine a nightmarish situation when two statisticians, at the patient bedside, dispute whose statistical method is more reliable and whose list of genes should be selected as targets, with a physician in the corner of the room waiting for the verdict that he needs for administering a life-saving treatment.

There are several persistent patterns in statistical thinking that are usually taken for granted, but in fact are nothing else than elements of some sort of statistical mythology. Thus, in cluster analysis of microarray data, the genes belonging to the same cluster are thought to be *co-regulated*. It is not out of order to remind again, that in the DNA microarray technology, one is not dealing with the genes themselves but only with the fluorescent intensities presumably proportional to the mRNA abundances. As discussed above in Sections 2 and 3, the latter may cluster due to many reasons, with the mRNA half-lives being a dominant factor. It is, therefore, an unwarranted logical leap over many intermediate steps from the clustering of fluorescent intensities to the co-regulation of parent genes.

The genes that are up- or down-regulated, as compared to some standard or normal behavior, are often thought to be abnormal, faulty, perhaps mutated. As discussed in Section 5, each transcription event in genetic regulatory system is a result of the team work of a large number of transcription factors. Deficiency in any of these factors may slow or even shut down transcription of each particular gene. Figuratively speaking, in addition to the core reason that the parent gene may be faulty, there may be from 30 to 100 other reasons for the abnormal behavior of this gene. Therefore, there is no and cannot be any unambiguous relations between anomalies in fluorescent

intensities observed in microarray experiments and fidelities of the corresponding genetic codes. At best, such anomalies indicate a target for further exploration by alternative, more advanced, techniques.

In statistics, there may be many different estimators for the same quantity of interest. Their comparative merits are measured by their *asymptotic relative efficiency*. In statistical analysis of microarray data, however, the consideration of asymptotic efficiency is not directly applicable. This is because the number of microarrays usually available is miserably small, so small that any extrapolation of asymptotic efficiency to the experiment with just several subjects at hand would be preposterous. What should be actually done for demonstrating superiority of certain estimates is to perform a simulation in which sample size is the one actually available. To the best of this author's knowledge, such an approach is very rare in the statistical literature on microarray data analysis. Such a stance has been adopted in the above cited work by this author.⁵⁴ It has been shown by simulation that for the sample size smaller than ten, the bio-weight test statistic has a higher power than that of the t-test. For a larger sample size, the advantages of the bio-weight test statistic disappear. In formal terms, this means that the bio-weight test statistic and the t-test have the same asymptotic efficiency; however, the former is superior for the small sample size. All this means that considerations of asymptotic relative efficiency cannot be used as an argument in favor of one or another statistical method, as far as microarrays are concerned.

Some authors report high specificity in classification of cancers using the DNA microarrays (see⁵⁶ as a recent example.) However, it often happens that efficiency is measured by the specificity in clustering the groups known *a priori*. It is not usually the case in clinical settings. Cancer is a highly heterogeneous disease; it cannot be always known *a priori* whether or not all the conceivable clinical outcomes have been adequately presented in training of the classifier. Therefore, in principle, successful classification of *a priori* known outcomes may demonstrate some clinical potential but does not represent a tool for clinical diagnostics by itself.

In summary, the lack of consensus in statistical methodologies leaves wide latitude for different interpretations of precisely the same DNA microarray data. This means that not only ambiguities in biological



interpretation but also the very statistical procedures that are supposed to articulate the outcome may contribute to uncertainties of the DNA microarray measurements thus posing difficulties in clinical setting.

Conclusion. Diagnostic Value of DNA Microarrays

All the above is not to say that the DNA microarrays are useless in clinical settings. It is only cautionary notes that DNA microarrays should be regarded for what they are, that is, the tool for studying the transcriptome, not the genome. After all, it is the mRNA transcripts what are directly involved in the protein synthesis. Therefore, DNA microarrays may provide valuable predictions regarding the state of proteome, its imbalances and abnormalities. The upstream relations between transcriptome and genome are much more vague and ambiguous.

This situation may be improved. As shown above, a key element in the interplay between the genome and transcriptome is the post-transcriptional mRNA stability. If the mRNA half-lives become available as routine measurements, then at least one important obstacle on the way towards evaluating transcription rates will be removed. It does not mean, however, that other obstacles automatically disappear as well. It would only mean that given the mRNA abundances and mRNA half-lives, the evaluation of transcription rate in equilibrium becomes possible. However, a number of important questions still will remain unresolved. Is asymptotic stable equilibrium achievable within the lifecycle of the cell? Does it exist at all? How to disentangle the functionality of the gene of interest from the influences of other genes acting through the transcription factors translated from them? How to infer reliable information regarding abnormality of genetic codes from observed abnormalities of transcription rates?

DNA microarray proved to be an indispensable tool in scientific and laboratory settings. Following the leads provided by DNA microarrays, innumerable discoveries have been made in experimental biology, despite all the complexities and unresolved issues. DNA microarrays provide important leads for follow-up studies using more advanced technologies such as PCR, ChIP, protein mass spectrometry, reporter plasmid analysis, and Northern blot. However, clinical practice requires firm cost-effective diagnoses, not just the leads for further experimentation. It is the author's opinion that

there is still long way to go until this becomes possible. Following,^{57,58} one may say that our knowledge is still *too limited*; available techniques are *too expensive*; and it is *too soon* to introduce microarray-based molecular profiling into clinical practice.

Acknowledgements

The author would like to express his gratitude to Drs. I. Kapetanovic and P. Prorok (National Cancer Institute, Division of Cancer Prevention) for useful comments helping the author to better articulate his message.

Disclosure

This manuscript has been read and approved by the author. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The author and peer reviewers of this paper report no conflicts of interest. The author confirms that they have permission to reproduce any copyrighted material.

References

1. Rosenfeld S. Stochastic Oscillations in Genetic Regulatory Networks. Applications to Microarray Experiment, in: Rizzi A, Vichi M Edtiors. COMPSTAT-2006. *Physica-Verlag*. 2006; p. 1609–18.
2. Rosenfeld S. Why do high-dimensional networks seem to be stable?-A reflection on stochasticity of dynamically unstable nonlinear systems, in: Gauges R, Kummer U, Pahle J, Willy P, Edtiors. Fifth Workshop on Computation of Biochemical Pathways and Genetic Networks, University of Heidelberg, Heidelberg, 2008, p. 101–12.
3. Rosenfeld S. Origins of Stochasticity and Burstiness in High-Dimensional Biochemical Networks, *EURASIP Journal of Bioinformatics and Systems Biology*, (2009).
4. Editorial, Making the most of microarrays. *Nature Biotechnology*. 2006; 24:1039.
5. Shi L, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology*. 2006;24:1151.
6. Shi L, et al. QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Rev Mol Diagn*. 2004;4:761.
7. Ioannidis JP. Microarrays and molecular research: noise discovery? *Lancet*. 2005;365:454.
8. Jenssen TK, Hovig E. Gene-expression profiling in breast cancer. *Lancet*. 2005;365:634.
9. Lewin B. *Genes VIII*, Upper Saddle River NJ. 2004.
10. Stutz F, Izaurralde E. The interplay of nuclear mRNA assembly, mRNA surveillance and export. *Trends Cell Biol*. 2003;13:319.
11. Rosenfeld S. Characteristics of transcriptional activity in nonlinear dynamics of genetic regulatory networks. *Gene Regul Syst Bio*. 2009; 3:159.
12. van Kampen. *Stochastic Processes in Physics and Chemistry*, North Holland, 2006.
13. Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng*. 2005;91:534.
14. Steuer R, Junker B. Computational Models of Metabolism: Stability and Regulation in Metabolic Networks, *Advances of Chemical Physics* (in press) (2009).



15. Sharova LV, et al. Database for mRNA half-life of 19,977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* 2009;16:45.
16. Wang Y, et al. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A.* 2002;99:5860.
17. Hambræus G, von WC, Hederstedt L. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol Genet Genomics.* 2003;269:706.
18. Raghavan A, et al. Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Res.* 2002;30:5529.
19. Raghavan A, Bohjanen PR. Microarray-based analyses of mRNA decay in the regulation of mammalian gene expression. *Brief Funct Genomic Proteomic.* 2004;3:112.
20. Harrold S, Genovese C, Kobrin B, Morrison SL, Milcarek C. A comparison of apparent mRNA half-life using kinetic labeling techniques vs. decay following administration of transcriptional inhibitors. *Anal Biochem.* 1991;198:19.
21. Garcia-Martinez J, Aranda A, Perez-Ortin JE. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell.* 2004;15:303.
22. Rosenfeld S. Stochastic Oscillations in Genetic Regulatory Networks. *EURASIP Journal of Bioinformatics and Systems Biology.* 2006:1.
23. Rosenfeld S. Stochastic cooperativity in non-linear dynamics of genetic regulatory networks. *Math Biosci.* 2007;210:121.
24. Goodsell DS. Inside a living cell. *Trends Biochem Sci.* 1991;16:203.
25. Goodsell DS, Olson AJ. Soluble proteins: size, shape and function. *Trends Biochem Sci.* 1993;18:65.
26. Pederson T. Diffusional protein transport within the nucleus: a message in the medium. *Nat Cell Biol.* 2000;2:E73.
27. Wang YM, Austin RH, Cox EC. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett.* 2006;97:048302.
28. Parsegian VA. Protein-water interactions. *Int Rev Cytol.* 2002;215:1.
29. Ellis RJ. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci.* 2001;26:597.
30. Gillespie DT. Stochastic Simulation of Chemical Kinetics. *Annu Rev Phys Chem.* 2006.
31. Kadonaga JT. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell.* 2004;116:247.
32. Rosenfeld S. Patterns of stochastic behavior in dynamically unstable high-dimensional biochemical networks. *Gene Regul Syst Bio.* 2009;3:1.
33. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature.* 2006;440:358.
34. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. *Science.* 2005;307:1962.
35. Newlands S, et al. Transcription occurs in pulses in muscle fibers. *Genes Dev.* 1998;12:2748.
36. Ross IL, Browne CM, Hume DA. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* 1994;72:177.
37. Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature.* 2003;422:633.
38. Golding I, Cox EC. RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci U S A.* 2004;101:11310.
39. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell.* 2005;123:1025.
40. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science.* 2006;311:1600.
41. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A.* 1997;94:814.
42. McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* 1999;15:65.
43. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet.* 2005;6:451.
44. Kepler T, Eiston T. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophysical Journal.* 2001;81:3116.
45. Lazebnik Y. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell.* 2002;2:179.
46. Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J. Cancer: a Systems Biology disease. *Biosystems.* 2006;83:81.
47. Junker B, Schreiber F. Analysis of Biological Networks. John Wiley and Sons, 2008.
48. Rosenfeld S. Mathematical Description of Biochemical Networks: instability, stochasticity, evolution, in: Thiangliam S Editor. Systems Biology of Cancer, Oxford University Press, 2009.
49. May RM. Will a large complex system be stable? *Nature.* 1972;238:413.
50. Kirkilionis M. Reaction systems, graph theory and dynamical networks, in: Gauges R, Kummer U, Pahle J, Willy P Editors. Fifth Workshop on Computation of Biochemical Pathways and Genetic Networks, University of Heidelberg, Heidelberg, 2008.
51. Scott M, Ingalls B, Kaern M. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos.* 2006;16:026107.
52. Donoho DL. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality, Aide-Memoire, 2000.
53. Ge Y, Dudoit S, Speed T. Resampling-based Multiple Testing for Microarray Data Analysis. *Test.* 2003;12:1.
54. Rosenfeld S. Detection of Differentially Expressed Genes In Small Sets of cDNA Microarrays. *Journal of Data Science.* 2007;5:451.
55. Fan X, et al. DNA microarrays are predictive of cancer prognosis: a re-evaluation. *Clin Cancer Res.* 2010;16:629.
56. Glas AM, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics.* 2006;7:278.
57. Ioannidis JP. Is molecular profiling ready for use in clinical decision making? *Oncologist.* 2007;12:301.
58. Ioannidis JP. Personalized genetic prediction: too limited, too expensive, or too soon? *Ann Intern Med.* 2009;150:139.

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>