

UC Davis

IDAV Publications

Title

Improved Significance Test for DNA Microarray Data: Temporal Effects of Shear Stress on Endothelial Genes

Permalink

<https://escholarship.org/uc/item/3ch2m87d>

Journal

Physiol Genomics, 12

Authors

Zhao, Yihua
Chen, Benjamin P.C.
Miao, Hui
[et al.](#)

Publication Date

2002

Peer reviewed

Improved significance test for DNA microarray data: temporal effects of shear stress on endothelial genes

YIHUA ZHAO,¹ BENJAMIN P. C. CHEN,¹ HUI MIAO,¹ SULI YUAN,¹ YI-SHUAN LI,¹ YINGLI HU,¹ DAVID M. ROCKE,² AND SHU CHIEN¹

¹Department of Bioengineering and the Whitaker Institute of Biomedical Engineering, University of California, San Diego, La Jolla 92093-0427; and ²Department of Applied Science, University of California, Davis, Davis, California 95616-8553

Submitted 6 March 2002; accepted in final form 16 October 2002

Zhao, Yihua, Benjamin P. C. Chen, Hui Miao, Suli Yuan, Yi-Shuan Li, Yingli Hu, David M. Rocke, and Shu Chien. Improved significance test for DNA microarray data: temporal effects of shear stress on endothelial genes. *Physiol Genomics* 12: 1–11, 2002. First published October 22, 2002; 10.1152/physiolgenomics.00024.2002.—Statistical methods for identifying differentially expressed genes from microarray data are evolving. We developed a test for the statistical significance of differential expression as a function of time. When applied to microarray data obtained from endothelial cells exposed to shearing for different durations, the new multi-group test (*G*-test) identified three times as many genes as the one-way ANOVA at the same significance level. Using simulated data, we showed that this increase in sensitivity was achieved without sacrificing specificity. Several genes known to respond to shear stress by Northern blotting were identified by the *G*-test at $P \leq 0.01$ (but not by ANOVA), with similar temporal patterns. The validity and utility of the *G*-test were further supported by the examination of a few more example genes in relation to the present knowledge of their regulatory mechanisms. This new significance test may have broad application for the analysis of gene-expression studies and, in fact, to other biological studies in general.

biological variability; microarray data analysis; differential gene expression; longitudinal analysis; statistical test

THE DNA MICROARRAY TECHNOLOGY developed in recent years provides a powerful and efficient tool to rapidly compare the differential expression of a large number of genes (8, 10, 12, 23, 26, 29), and the statistical methods for identifying differentially expressed genes from microarray data are still evolving (18, 21, 22, 24, 25). Using the DNA microarray approach, we investigated gene expression profiles in cultured human aortic endothelial cells (HAECs) in response to a laminar shear stress at 12 dyn/cm². We performed the experiments at 1, 4, and 24 h after shearing, as well as on static controls. When the ANOVA test was applied to analyze the results, a very small number of genes were found to have significant changes after shearing, when

compared with the static control. We realized that the ANOVA test was too stringent for the detection of significant changes in gene expression in the face of biological variations among experiments. In the absence of a better way to analyze the results of all three time periods, we chose to work on the comparison of only one pair of data (24 h shear vs. static control) by the use of the paired Student's *t*-test (6). At the suggestion of one of us (D. M. Rocke), we used the method of longitudinal analysis for statistical analysis (9) as a starting point to develop an improved significance test. This new test provides a more sensitive way to analyze DNA microarray data obtained from sets of data from different experiments with several time points. In the present paper, this new significance test has been applied to the entire data set obtained from three different experiments, each with three time points, including the 24-h results we previously reported (6). The results indicate that this new significance test can detect three times as many genes as the ANOVA test at the same cutoff point of *P* value. Many of the additional genes detected by this new method are in agreement with the known temporal changes that have been reported. Since most biological studies involve experiments with several time points and considerable individual variations, this new significant test should have significant application to the study of temporal modulation of gene expression in response to experimental manipulation. The same test can also be applied to the comparison of results obtained on different organ/tissues or to experiments with different degrees of treatment for assessing the dose-response relationship in gene expression. Therefore, this new significance test may have broad application for the analysis of gene-expression studies and, in fact, to other biological studies in general.

ANALYTICAL METHODS

Rationale. In a DNA microarray study, the observed variations in the expression levels of a given gene can be attributed to the following: *A*) the effects of the treatments, *B*) the biological variability among the subjects tested, and *C*) the experimental/measurement errors. The focus of this work is on testing the statistical significance of the effects of a treatment on gene

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: S. Chien, Whitaker Institute of Biomedical Engineering, Univ. of California, San Diego, Mail code 0427, 9500 Gilman Dr., La Jolla, CA 92093-0427 (E-mail: schien@bioeng.ucsd.edu).

expression (A), in which both B and C are regarded as noise. We postulate that each genome is a redundant network that provides flexibility in the regulation of the gene expression, thus contributing to the variability among gene expression profiles obtained from individual experiments. Consequently, the biological variability B could be greater than the experimental errors C . Therefore, the power of statistical tests for the significance of treatment effects (A) can be greatly improved if B can be isolated. The achievement of this goal requires an appropriate experimental design. Repeats of the same experiment on different subjects produce a group of data sets, in which the subject-subject variation (i.e., B) can be identified. In the experiments reported here, each data set comprises data collected from experiments on the same batch of cultured cells subject to shearing of different time durations or kept as static control, as described below. In this case, a subject means a cell batch.

The subject-subject variation (B) may have two components: the variation among subjects in the baseline of a gene's expression and that in the responsiveness of the gene to the treatment. Correspondingly, we can isolate, at least partially, the effects of B by subtracting from each data set, gene by gene, the gene-specific means and then dividing the results by the gene-specific standard deviations, as described in the section below on *Statistical tests*.

Global normalization. Let x_{ijk} denote the intensity measured (readout) for gene i , with treatment j , from subject k , where $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; and $k = 1, 2, \dots, K$. For the experiments analyzed here (see EXPERIMENTAL MATERIALS AND METHODS), $I = 1,185$ (no. of genes), $J = 4$ (static control plus three time points after shearing), and $K = 3$ (no. of experiments). Let x_{*jk} denote the data vector obtained from a single microarray which corresponds to treatment j and subject k , with its i th component being x_{ijk} . Let x_{i*} and x_{i**} denote, in a similar way, the corresponding data vectors. Similarly, x_{i**} denotes the data matrix for gene i , with rows for treatments and columns for subjects. Also used is the conventional dot notation, in which a dot replacing an index indicates an average over the whole range of the index. For example, $x_{.jk}$ denotes the mean of the I components of x_{*jk} , and $x_{i..}$ denotes the mean of the JK components of x_{i**} . Summation will be explicitly stated, and repeated indices do not automatically indicate a summation.

The measured intensity x_{ijk} has an arbitrary scale. To compare the intensity readings for a gene between different microarrays, we performed an array-wide global normalization, i.e., dividing each x_{ijk} by $x_{.jk}$. All intensities considered below are globally normalized, and hereafter the notations x_{ijk} , x_{i*} , etc., always refer to the globally normalized intensities.

Statistical model. We state explicitly the statistical model here to provide clarity for the construction of a test statistic and its null distribution. As reasoned in *Rationale* (above), in the absence of experimental errors, the variability in measured intensities can be explained by $x_{ijk} = \alpha_{ij}\gamma_{ik} + \beta_{ik}$, with a constraint $\alpha_i =$

0, where α_{ij} is due to the treatment effects A, and β_{ik} and γ_{ik} are, respectively, the baseline expression and the responsiveness of gene i in subject k . If preferred, one can set the standard deviation of α_{ij} to 1 to specify the scale of γ_{ik} . With the inclusion of experimental errors, we have

$$x_{ijk} = \alpha_{ij}\gamma_{ik} + \beta_{ik} + e_{ijk} \quad (1)$$

In Eq. 1, α_{ij} is a fixed effect because it only applies to the specified gene and treatment, whereas β_{ik} and γ_{ik} are random effects because the K subjects are a sample representing a larger population. The experimental errors are assumed to be independently distributed as

$$e_{ijk} \sim N(0, \sigma_e^2) \quad (2)$$

No assumption is needed about the distributions of β_{ik} and γ_{ik} for the statistical test described below.

In a standard approach of longitudinal analysis, it would be assumed that subjects differ only in their baseline levels, equivalent to setting $\gamma_{ik} = 1$ in Eq. 1. In that case, Eq. 1 is reduced to a model of two-way ANOVA. We introduce the multiplicative factor γ_{ik} to account for the fact that subjects may also differ in their responsiveness to a treatment, as reasoned in *Rationale* (above). Indeed, experimental data show that the responsiveness of many genes vary significantly from subject to subject, as exemplified by RGS3 ("regulator of G protein signaling 3") in Fig. 2B.

Statistical tests. The null hypothesis to be tested is $\alpha_{i1} = \alpha_{i2} = \dots = \alpha_{iJ} = 0$. Statistical tests are performed for each individual gene. Given data x_{ijk} , compute

$$y_{ijk} = (x_{ijk} - x_{i..}) / \sqrt{\sum_{j=1}^J (x_{ijk} - x_{i..})^2 / (J - 1)} \quad (3)$$

Note that in the absence of experimental error

$$y_{ijk} = y_{ij.} = \alpha_{ij} / \sqrt{\sum_{j=1}^J \alpha_{ij}^2 / (J - 1)}$$

Now, for each gene i , construct an ANOVA table for the derived data matrix y_{i**}

Source of Variation	Sum of Squares	Degree of Freedom
Between treatments	$BTSS = K \sum_{j=1}^J (y_{ij.} - y_{i..})^2$	$J - 2$
Residual	$RSS = TSS - BTSS$	$(J - 1)(K - 1)$
Total	$TSS = \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{i..})^2$	$(J - 1)K - 1$

where the degrees of freedom have been adjusted because of the data transformation via Eq. 3. The y_{ijk} is constructed in a way so that BTSS is a measure for the treatment effects and RSS for the experimental errors. We define a G statistic analogous to Fisher's F statistic

$$G = \frac{BTSS/(J-2)}{RSS[(J-1)(K-1)]} \quad (4)$$

This likelihood ratio is called G instead of F to emphasize that the null distribution of this G statistic is not an F -distribution.

Under the null hypothesis $\alpha_{ij} = 0$, Eq. 1 reduces to

$$x_{ijk} = \beta_{ik} + e_{ijk} \quad (5)$$

For the error distribution given in Eq. 2, it can be shown that y_{ijk} computed from the x_{ijk} given by Eq. 5 is equivalent in distribution to y_{ijk} computed from $x_{ijk} \sim N(0,1)$, independent of the distributions of β_{ik} and γ_{ik} . This allows us to readily compute the null distribution of the G statistic by computer simulation (see *Simulated data sets*, below). The upper tail of the G 's null distribution gives the P value, which is the probability of having a G score randomly drawn from the null distribution greater than the experimental value of G . We will call this G -statistic-based test as G -test, and denote the P value it yields as P_G .

For comparison, we also computed P value for gene i by directly applying the conventional one-way ANOVA to the data matrix x_{i**} , which will be denoted as P_{anova} .

Simulated data sets. A random number generator is used to assign $x_{ijk} \sim N(0, 1)$. Then, the y_{ijk} values defined in Eq. 3 are calculated from the simulated x_{ijk} , and a G statistic is computed for each i as described above. We computed 100,000 G statistics in this way to generate a null distribution for the G statistic.

To compare the sensitivities and specificities of different statistical tests, simulated data sets with pre-scribed expression patterns were generated based on Eq. 1, with errors and random effects being assigned by a random number generator.

To test the effects of subject-subject variation, simulated data sets were generated from the experimental data by permuting the identities of subjects. Let x_{ijk} and x'_{ijk} be the experimental and simulated data, respectively. A simulated group of data sets was constructed by letting each vector x'_{ij*} be a random permutation of vector x_{ij*} , with the random permutation for different (i, j) performed independently. One hundred groups of subject-identity-permuted data sets generated from the same experimental group of data sets were used.

EXPERIMENTAL MATERIALS AND METHODS

Experimental design. Of particular importance to the present study is the experimental design. Three independent sets of experiments were performed. In each set, HAECs were subjected to a shear stress at 12 dyn/cm² for 1, 4, and 24 h together with a paired control experiment with the HAECs kept under static condition. All cells were first cultured to passage 3 (*Cell culture*, below) and kept frozen in liquid nitrogen. Then, for each set of experiments, the cells were thawed and further cultured for two more passages to obtain sufficient cells so that all the experiments within the same set used the same batch of HAECs at passage 5. However, three different batches of cells were used in the three sets of experiments.

Atlas Human Cancer 1.2 arrays (Clontech) were used in these experiments. Each Atlas 1.2 array consists of four separate filters. The four filters from the same set were used to perform a set of experiments, one for each of the four experiments. A total of three sets of filters were used, thus providing three sets of experiments.

Cell culture and shear stress experiments. The experimental materials and methods were previously reported in detail (6). Briefly, HAECs were obtained from Clonetics (San Diego, CA) and cultured in endothelial cell growth medium-2 (EGM-2) supplemented with 2% fetal bovine serum (FBS), hydrocortisone, hFGF- β , VEGF, R3-IGF-I, ascorbic acid, hEGF, GA-1000, and heparin (Clonetics). Cell cultures were maintained in a humidified 95% air-5% CO₂ incubator at 37°C. To impose shear stress on cultured cells, a glass slide seeded with a confluent monolayer of HAECs was mounted to a parallel-plate flow chamber and connected to a flow system as described by Frangos et al. (11). A laminar shear stress of 12 dyn/cm² was generated by the flow resulting from a hydrostatic pressure difference between two reservoirs. This level of shear stress is encountered under physiological conditions in the straight part of the aorta and frequently used to study the effects of shear stress on endothelial cells. To maintain the same cell environment, the flow experiments were performed by using the same culture medium as in the static culture, thus controlling the possible effects of growth factors on gene expression. The whole flow system was kept at 37°C and ventilated with 95% humidified air and 5% CO₂.

Microarray experiments and imaging. For microarray experiments, the total RNA was isolated by using STAT60 total RNA purification reagent (Tel-Test "B"). The cells were lysed in phenol-containing STAT60 solution and centrifuged. The RNA-containing aqueous phase was isolated, and 0.6 vol of isopropanol was added to precipitate RNA. The 20–30 μ g of total RNA was subjected to reverse transcription (RT) reaction and [³³P]dCTP labeling using an Atlas Pure Total RNA Labeling System (Clontech) and [³³P]dCTP. The ³³P-labeled cDNA was hybridized to Atlas Human Cancer 1.2 arrays at 68°C for 16 h, followed by washing once in 2 \times SSC/1% SDS at 68°C for 30 min and three times in 0.1 \times SSC 0.5% SDS at 68°C for 30 min. The hybridized array was then exposed to Phospho Storage Screen (Molecular Dynamics), and the images were analyzed and quantified by using an Atlas Image 1.01 software (Clontech) to quantify the intensity of each of the 1,185 spots.

RESULTS

Validation of the global normalization. The validity of the global normalization is confirmed by the results for GAPDH. The mRNA level of GAPDH in endothelial cells is known to be not changed by shear (14) and has been widely used as the negative control in studying the regulation of gene expression in endothelial cells by shear stress (4, 14, 15, 23). In this work, the normalized intensity for GAPDH ($P_G = 0.78$, $P_{anova} = 0.98$) is essentially constant across all the time points tested, as shown in Fig. 1.

The high sensitivity of the G-test. When applied to the experimental data, the G -test was found to be three times more sensitive than ANOVA, as shown in Table 1. For example, at $P \leq 0.01$, 61 genes were called significant by G -test, whereas ANOVA called only 18 genes significant. The 61 genes that showed $P_G \leq 0.01$ are listed in Table 2. The results for all the 1,185 genes are available as an online data supplement at <http://>

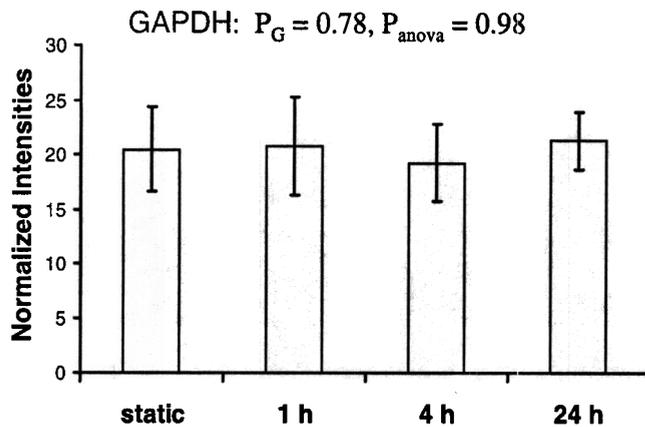


Fig. 1. The gene expression level of GAPDH, a negative control, was not changed by shear stress. Bars represent the globally normalized intensities for GAPDH from three independent experiments (means \pm SE). P_G is the P value yielded by the G -statistic-based test. P_{anova} is the P value computed directly applying the conventional one-way ANOVA to the data matrix.

wibe.ucsd.edu/resources/microarray.shtml, which are also published at the *Physiological Genomics* web site.¹

To examine how the G -test improves the sensitivity of the statistical testing, we used simulated data with prescribed patterns of gene expression. Data sets for “positive” and “negative” genes were generated in accordance of Eq. 1. For positive genes, effects $\alpha_{i*} = (3/2, -1/2, -1/2, -1/2)$, $\beta_{ik} \sim N(1, \sigma_\beta^2)$, $\gamma_{ik} = 1$, and error distribution $\epsilon_{ijk} \sim N(0, 1)$; γ_{ik} was set to 1 for simplicity. Negative genes were generated in the same way, except that effects $\alpha_{i*} = (0, 0, 0, 0)$. For each group of simulated data set, the index ranges were set to $I = 100,000$ to have a large sample size and $J = 4, K = 3$ to mimic the design of the experiments. Note that the standard deviation of the experimental error was set to unity and served as the measurement unit for the effects α_{ij} and β_{ik} . The prescribed α_{i*} for the positive genes is such that the hypothesized gene responds only to one of the treatments (which one is immaterial) by increasing its expression level by an amount equal to two standard deviations of the experimental error. The subject-subject variation was modulated in the simulation by using different values of σ_β . The G -test and ANOVA were applied to the simulated data, and the sensitivity and specificity of a test at each fixed cutoff point of P were computed as the fractions of, respectively, positive and negative genes correctly classified by the test. The receiver operating characteristic (ROC) curves of both tests were plotted in Fig. 2A for various values of σ_β . ROC curves are commonly used for comparing competing tests (1). In Fig. 2A, $\sigma_\beta = 0$ through 4 were used to examine the effects of subject-subject variation on the performance of the tests. It is interesting to note that, in the absence of the subject-

subject variation ($\sigma_\beta = 0$), ANOVA is actually more (though only slightly) sensitive than the G -test. However, as the subject-subject variation increases, the sensitivity of ANOVA (see the dashed lines) drops rapidly; at $\sigma_\beta = 3$ [which is greater than the magnitude of the treatment effect on positive genes, namely, $3/2 - (-1/2) = 2$], the sensitivity of ANOVA becomes too low to distinguish the populations of positive and negative genes, as indicated by the closeness of the dashed line for $\sigma_\beta = 3$ to the diagonal dot line. In contrast, as expected, the sensitivity of G -test is essentially not affected by the subject-subject variation, as the ROC curves produced by G -test with different values of σ_β are practically indistinguishable (the solid line in Fig. 2A). In summary, for data with sizable subject-subject variations, the G -test is more sensitive than ANOVA at the same level of specificity.

An example will help to illustrate the above observations. Figure 2B shows the actual experimental data for gene RGS3. It is evident that both the baseline expression and the responsiveness to shear stress of the gene varied markedly among the three repeated sets of experiments, each using a different batch of cells. These observed variations are likely to be biological in origin rather than experimental errors, for the time course of regulation by shear was very similar in the three sets of experiments, as shown in Fig. 2C, which plots the same data transformed by Eq. 3. However, it should be noted that, if care is not taken to minimize systematic experimental errors, then a nonrandomly distributed analytical variation could produce a similar pattern. For example, applying labeled probes with different specific activities or quantities to an array could produce changes in the observed baseline (due to background) or responsiveness (due to saturation), respectively. Without isolating the large biological and nonrandomly distributed analytical variations between subjects (cell batches), ANOVA failed to detect the consistent pattern which shows the effect of shear and yielded a $P = 0.18$. In contrast, the G -test yielded a $P < 0.01$, indicating that the effect of shear is statistically significant. The upregulation of RGS3 by 1-h shear was confirmed by a Northern analysis (data not shown).

The other side of the coin of what is shown in Fig. 2A is that if many more genes are found to be significant in the same sets of data by the G -test than ANOVA, then the biological variations between subjects are not small compared with the effect of the treatment, i.e., shear-

Table 1. Number of genes showing statistical significance at the indicated levels

P	G -test	ANOVA	G -test + Permutation*
≤ 0.01			
≤ 0.02			
≤ 0.05			
≤ 0.1			

¹The Data Supplement (results for all the 1,185 genes) for this article is available online at <http://physiolgenomics.physiology.org/cgi/content/full/12/1/1/DC1>.

*Obtained by applying G -test to subject-identity-permuted experimental data (see ANALYTICAL METHODS). Numbers shown are the average of 100 computations.

Table 2. Genes whose expression was significantly ($P_G \leq 0.01$) regulated by shear

Gene	Shear/Static			P_G	P_{Anova}
	1 h	4 h	24 h		
Matrix metalloproteinase 1 (MMP1)	1.20 ± 0.07		2.77 ± 0.35	<0.01	
Regulator of G protein signaling 3 (RGS3)	1.73 ± 0.29		0.71 ± 0.03	<0.01	
Tie-2, tyrosine-protein kinase receptor	0.95 ± 0.11		1.90 ± 0.01	<0.01	
Zinc finger 5 protein	0.75 ± 0.07		0.78 ± 0.05	<0.01	
Large fibroblast proteoglycan	0.60 ± 0.15		1.19 ± 0.08	<0.01	
CD59 glycoprotein precursor	1.01 ± 0.08		1.46 ± 0.15	<0.01	
Microtubule-associated protein 1B	1.05 ± 0.06		0.53 ± 0.11	<0.01	
Importin α 3 subunit	1.02 ± 0.02		0.61 ± 0.05	<0.01	
G ₁ /S-specific cyclin D1 (CCND1)	1.09 ± 0.07		1.53 ± 0.18	<0.01	
Caveolin-1	0.85 ± 0.08		0.42 ± 0.04	<0.01	
ATP synthase coupling factor 6	1.03 ± 0.13		1.74 ± 0.15	<0.01	
Netrin-2	1.29 ± 0.23		0.78 ± 0.14	<0.01	
Macrophage inhibitory cytokine 1 (MIC1)	1.47 ± 0.18		0.49 ± 0.14	<0.01	
Jun-B	1.81 ± 0.15		1.37 ± 0.02	<0.01	
Tissue-type plasminogen activator precursor	1.22 ± 0.01		0.61 ± 0.11	<0.01	
Mitogen-responsive phosphoprotein DOC2	0.80 ± 0.04		0.59 ± 0.18	<0.01	
Lysophospholipase homolog	1.19 ± 0.12		1.04 ± 0.13	<0.01	
Extracellular signal-regulated kinase 1 (ERK1)	0.91 ± 0.03		1.15 ± 0.05	<0.01	
IMP dehydrogenase 2 (IMPD2)	1.33 ± 0.13		0.56 ± 0.12	<0.01	
IFN- γ transducer 1 (IFNGT1)	0.89 ± 0.15		0.83 ± 0.08	<0.01	
Eph-family protein	1.70 ± 0.25		1.70 ± 0.20	<0.01	
Hepatoma-derived growth factor (HDGF)	0.79 ± 0.10		0.48 ± 0.08	<0.01	
Integrin β 3	0.78 ± 0.06		0.54 ± 0.06	<0.01	
VEGF receptor 1 (VEGFR1)	0.90 ± 0.08		1.91 ± 0.10	<0.01	
Shb proto-oncogene	1.31 ± 0.09		1.76 ± 0.26	<0.01	
Delta-like protein precursor (DLK)	1.02 ± 0.08		0.76 ± 0.10	0.01	
Bone morphogenetic protein 4 (BMP4)	0.43 ± 0.05		0.36 ± 0.14	0.01	
Endothelial PAI-1 precursor	1.04 ± 0.04		1.30 ± 0.11	0.01	
Butyrophilin precursor (BT)	0.56 ± 0.26		1.14 ± 0.01	0.01	
Protein P84	1.10 ± 0.15		1.81 ± 0.37	0.01	
Nuclear hormone receptor	1.31 ± 0.08		0.81 ± 0.17	0.01	
Desmoplakin I and II	0.41 ± 0.11		0.92 ± 0.12	0.01	
Glutathione-S-transferase homolog	1.21 ± 0.06		1.59 ± 0.12	0.01	
Frizzled 5	0.88 ± 0.47		1.32 ± 0.24	0.01	
Early growth response alpha (Egr- α)	2.13 ± 0.11		1.16 ± 0.22	0.01	
Fos-related antigen (FRA1)	1.92 ± 0.33		1.15 ± 0.09	0.01	
Thymus-expressed chemokine precursor	0.61 ± 0.11		0.54 ± 0.07	0.01	
Ninjurin-1	1.02 ± 0.06		0.73 ± 0.06	0.01	
C-fgr proto-oncogene (p55-FGR)	1.33 ± 0.14		2.18 ± 0.34	0.01	
CD40 ligand (CD40-L)	0.63 ± 0.07		0.67 ± 0.10	0.01	
Adenylosuccinate synthetase	0.80 ± 0.08		0.60 ± 0.07	0.01	
VHL-binding protein 1 (VBP1)	0.95 ± 0.10		1.98 ± 0.31	0.01	
Adenovirus E1A enhancer-binding protein	0.82 ± 0.08		0.76 ± 0.03	0.01	
Active BCR-related protein	0.98 ± 0.03		1.32 ± 0.12	0.01	
Apoptosis antigen ligand (APTL)	0.82 ± 0.09		0.54 ± 0.12	0.01	
G ₁ /S-specific cyclin E (CCNE)	0.88 ± 0.17		1.78 ± 0.18	0.01	
Inducible protein	0.74 ± 0.08		0.40 ± 0.02	0.01	
L-Lactate dehydrogenase H subunit (LDHB)	1.34 ± 0.25		0.80 ± 0.06	0.01	
Interferon- α 2 precursor (IFN- α)	1.12 ± 0.10		0.51 ± 0.20	0.01	
BRCA1-associated ring domain protein	0.77 ± 0.03		0.82 ± 0.11	0.01	
Platelet-derived growth factor A (PDGF-A)	1.46 ± 0.16		0.73 ± 0.18	0.01	
G protein-coupled receptor GPR-NGA	0.85 ± 0.19		0.58 ± 0.13	0.01	
EB1 protein	1.42 ± 0.02		1.54 ± 0.08	0.01	
Transforming protein rhoA H12 (RHO12)	1.10 ± 0.09		1.48 ± 0.10	0.01	
Myxovirus (influenza) resistance 2	1.32 ± 0.21		2.40 ± 0.66	0.01	
UMP synthase	0.85 ± 0.04		0.67 ± 0.02	0.01	
CD5 precursor	1.17 ± 0.35		0.71 ± 0.12	0.01	
High mobility group protein (HMG-I)	1.14 ± 0.07		0.64 ± 0.06	0.01	
CD27 ligand (CD27LG)	0.98 ± 0.15		0.79 ± 0.03	0.01	
IL-6R- α precursor	1.45 ± 0.24		2.14 ± 0.41	0.01	
Early growth response protein 1 (Egr-1)	2.10 ± 0.23		0.77 ± 0.26	0.01	

ing. Such biological variations have important implications in the modeling of genetic network. For the microarray data present here, we tested whether this converse reasoning holds by applying the G -test to the

experimental data with the subject identities (indices for cell batch) being randomly permuted (see subject-identity-permuted data sets in *Simulated data sets*, above). If the better performance of G -test over

ANOVA is indeed due to the existence of a large subject-subject variation, then the permutation should eliminate the advantage of G -test over ANOVA, since the subject-identity-permutation has no effect on

ANOVA. As can be seen by comparing the last two columns of Table 1, this is exactly what happened. This result provides experimental evidence that supports the existence of relatively large individual variations in the genome-wide regulation of gene expression in response to a given stimulus.

Egr-1 and related genes. The purpose of this paper is to present a method for analyzing microarray data rather than a detailed analysis of the effect of shear stress on gene expression. We would like to use several genes that have known time-dependent variations to illustrate the utility of our new significance test. Two of the genes most extensively studied for shear-modulated gene expression in endothelial cells are early growth response-1 (Egr-1) and platelet-derived growth factor-A (PDGF-A). Both were picked up by the G -test but missed by ANOVA (for Egr-1, $P_G = 0.01$ and $P_{\text{anova}} = 0.07$; for PDGF-A, $P_G = 0.01$ and $P_{\text{anova}} = 0.15$). The shear-stimulated expression of Egr-1 was transient, peaked at 1 h, and returned to basal level at 4 h and 24 h (Fig. 3A). This is in excellent agreement with the published time course (0 to 4–6 h) for shear-stimulated Egr-1 expression in various types of vascular endothelial cells obtained by Northern blot analyses; all those studies found Egr-1 mRNA expression to be transient, peaking at 0.5–1 h and returning to basal level after 3-h shearing (7, 19, 27). The shear-stimulated expression of PDGF-A increased progressively up to 4 h and returned to basal level at 24 h (Fig. 3B). Again, this agrees very well with the published time course (0 to 4 h) obtained by Northern blots (15, 19). It is worth mentioning here that Egr-1 is a transcription factor that binds to the proximal PDGF-A promoter as a shear-stress-responsive element (19), which explains the findings that the transient expression of Egr-1 mRNA precedes the PDGF-A gene expression.

It has been shown that Egr-1 also binds to the cyclin D1 promoter and that this contributes to the enhancement of cyclin D1 transcription by transforming growth factor- α (30) and/or by angiotensin II (13) in an extracellular signal-regulated kinase (ERK)-dependent manner. It is well established that shear stress

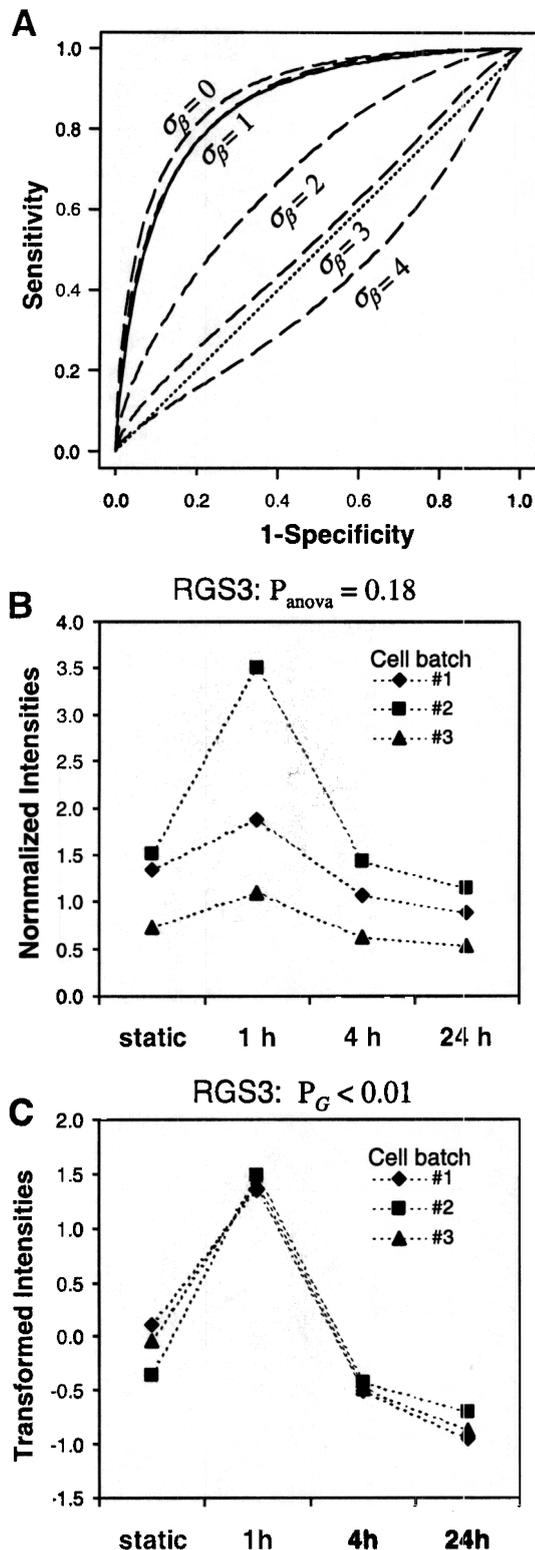


Fig. 2. G -test has a higher sensitivity than ANOVA in detecting treatment effects from microarray data in which the biological variability between subjects is large. **A**: G -test and ANOVA were applied to simulated data that had prescribed the effects of treatments and the subject-subject variation, with the latter being quantified by σ_{β} (see text). Dashed lines are for ANOVA, with the values of σ_{β} being labeled. The solid line is for the G -test, which does not change with σ_{β} . The plot indicates that the G -test has a higher sensitivity at each level of specificity than ANOVA when the subject-subject variation is not small compared with the experimental error, i.e., when $\sigma_{\beta} \geq 1$, and that this improvement in power increases as the subject-subject variation increases. **B**: the data of microarray experiments indicate that the subject-subject (i.e., between-cell batch) variation in the expression levels of individual genes could be very large, as exemplified by the gene expression of RGS3. Because of the large subject-subject variation, ANOVA failed to detect an apparent consistency in the expression pattern of RGS3, namely, the transient upregulation by 1-h shear. **C**: the expression pattern of RGS3 across all time points was consistent, as indicated by closeness of the time courses after the data were transformed via Eq. 3. This is reflected by a small P value returned by the G -test. RGS3, regulator of G protein signaling 3.

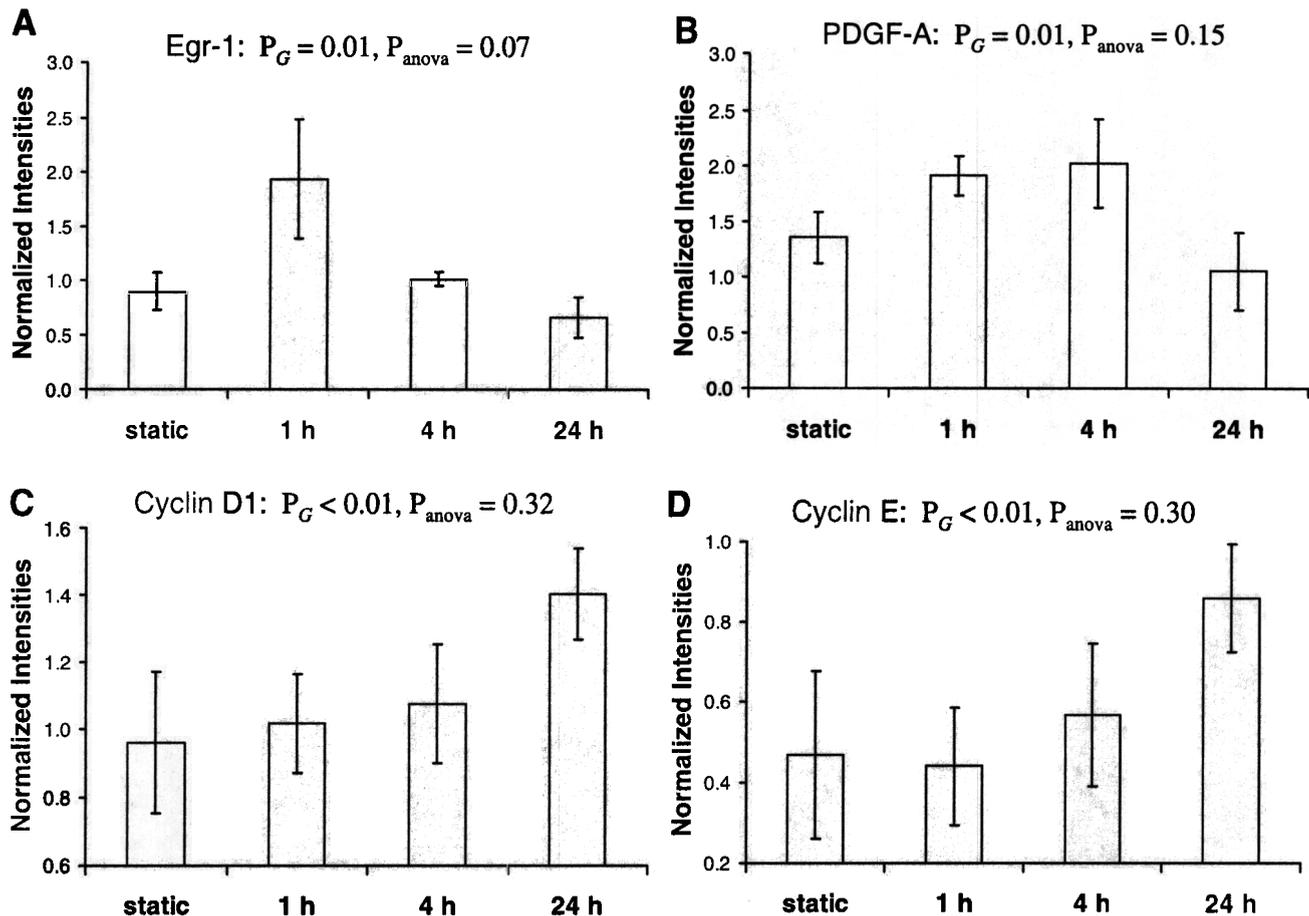


Fig. 3. Egr-1 and related genes. Egr-1 is a transcription factor that binds to the promoter regions of many genes, including PDGF-A and cyclin D1. The gene expressions of Egr-1 (A) and PDGF-A (B) were upregulated by shear stress, with the peak expression of Egr-1 preceding that of PDGF-A. The time courses of Egr-1 and PDGF-A expression shown in the microarray data agree very well with the time courses of their mRNA expression in endothelial cells established by Northern blotting analysis (14, 19, 27). The gene expressions of cyclins D1 (C) and E (D) were upregulated by shear stress with a time course suggesting a temporal lag following the transient upregulation of Egr-1, as in the established case of PDGF-A (see text). For each of these four genes, the G -test gave a $P \leq 0.01$, while ANOVA produced a $P > 0.05$.

activates ERK (5, 16, 17, 20, 28). Therefore, it is reasonable to postulate that the shear stress-induced transient expression of Egr-1 may lead to an enhancement in the gene expression of cyclin D1 at later times. There are 11 cyclins presented in the Atlas 1.2 Cancer array, of which two were significantly modulated by shear according to the G -test: one is cyclin D1 ($P_G < 0.01, P_{\text{anova}} = 0.32$), and the other is cyclin E ($P_G < 0.01, P_{\text{anova}} = 0.30$). Both were upregulated by shear at the later time points (Fig. 3, C and D), as expected for cyclin D1. For a more liberal cutoff point $P = 0.1$, two more cyclins were found to be significant by G -test: cyclins C ($P_G = 0.06, P_{\text{anova}} = 0.42$) and D3 ($P_G = 0.07, P_{\text{anova}} = 0.13$); both displayed a time course similar to those of cyclins D1 and E. It is interesting to note that 5 of the 11 cyclins in the microarray we used are G_1/S specific, that all four cyclins picked up by G -test are G_1/S specific, and that these four G_1/S -specific cyclins showed similar temporal changes after shearing. In

comparison, ANOVA did not pick up any of these cyclins ($P > 0.1$ for all 11 cyclins).

The other member of the Egr family present in the microarray is Egr- α , which is highly expressed in the early G_1 phase of the cell cycle (3). In the present study, the expression of Egr- α in endothelial cells was found to be modulated significantly ($P_G < 0.01, P_{\text{anova}} = 0.02$) by shear with a time course (Fig. 4A) very similar to that of Egr-1. Egr-1 and Egr- α are immediate early genes. Another immediate early gene found to be upregulated with a peak at the early time point is Jun-B (Fig. 4B, $P_G < 0.01, P_{\text{anova}} < 0.01$), with a time course similar to the reported time course of c-Jun mRNA expression in response to a steady flow (15). c-Jun was upregulated at the early time point in the present study, but not statistically significant ($P_G = 0.2, P_{\text{anova}} = 0.5$). To our knowledge, there is no report in the literature about the effects of shear on the time course of gene expression of Egr- α and Jun-B, as well as the cyclins.

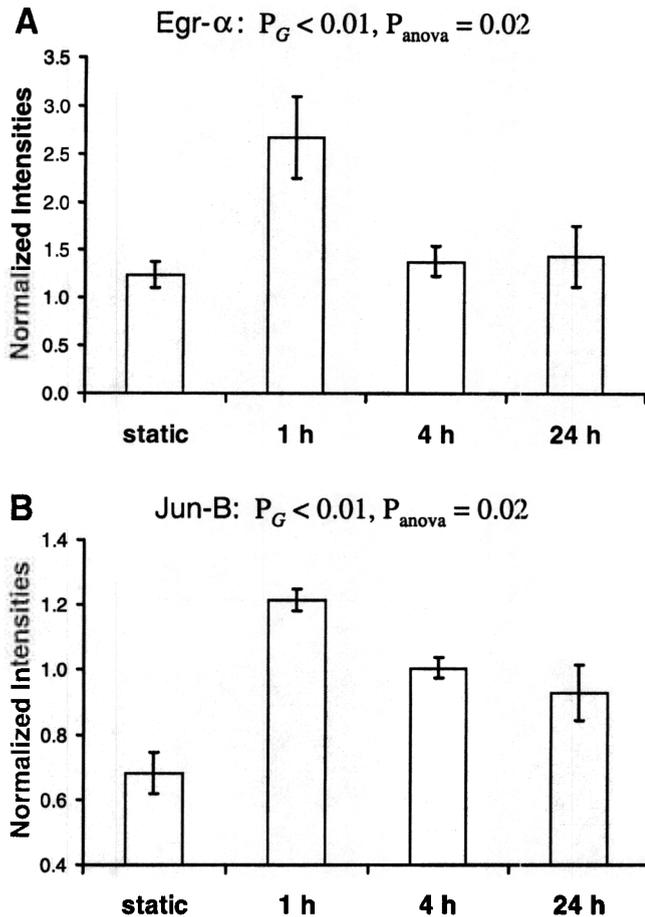


Fig. 4. Microarray experiments showed that, in addition to the Egr-1 shown in Fig. 3A, shear stress also transiently upregulated Egr- α (A) and Jun-B (B) ($P_G < 0.01$, $P_{\text{anova}} = 0.02$), two other immediate early genes.

Other example genes. Shown in Fig. 5 are three more examples that illustrate the superior performance of the G -test regardless of the temporal pattern of gene expression. The time courses of the expression of genes shown in Fig. 5 differ from those shown in Figs. 3 and 4. The gene expression level of caveolin-1 decreased progressively ($P_G < 0.01$, $P_{\text{anova}} = 0.08$) under shear. Tyrosine-protein kinase receptor (Tie-2, $P_G < 0.01$, $P_{\text{anova}} = 0.42$) and matrix metalloproteinase 1 (MMP1, $P_G < 0.01$, $P_{\text{anova}} < 0.01$) were not responsive to short-term shear but upregulated by 24-h shear. The differential expressions of these genes between static and 24-h shear have been confirmed by RT-PCR and/or Northern blot analysis (6).

DISCUSSION

We have proposed a new model for differential gene expression that explicitly specifies the effects of experimental treatments and biological variability and developed a corresponding multi-group test (G -test) for the statistical significance of the treatment effects. Using experimental and simulated data, we showed that this new test is much more sensitive than the

conventional one-way ANOVA in detecting treatment effects from data produced by microarray experiments. This improvement in sensitivity could be of critical importance, as exemplified by the present study on the temporal modulation of gene expression in endothelial cells in response to shear stress. As shown in Table 1, the number of genes showing significant differential expression as determined by ANOVA is very close to

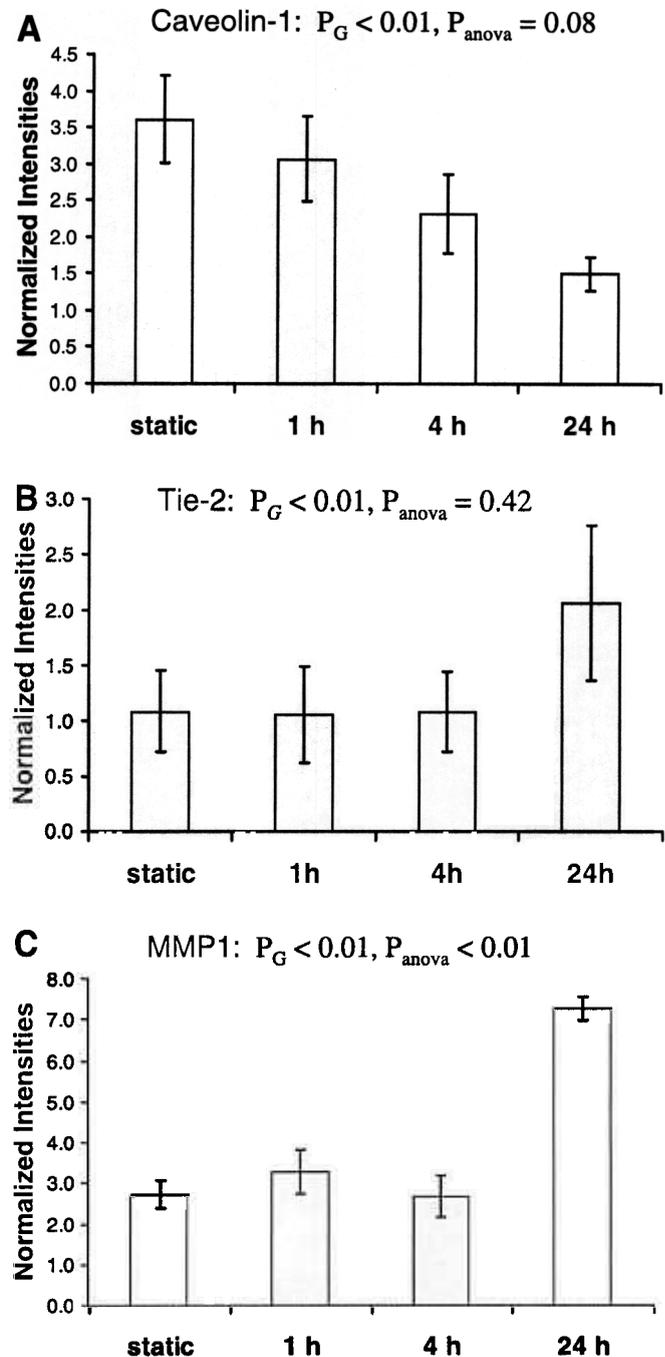


Fig. 5. The time courses of the expression of genes encoding caveolin-1 (A), Tie-2 (B), and matrix metalloproteinase 1 (MMP1) (C), whose modulation by 24-h shear had been confirmed by RT-PCR and/or Northern blotting (6).

the number of false-positive genes one would expect for a hypothetical situation in which none of the genes are modulated by shear. For example, the number of genes called significant by ANOVA is 18 at $P \leq 0.01$ and 49 at $P \leq 0.05$. Assume that the 1,185 genes were not regulated by shear, and that the measurements for the 1,185 genes were statistically independent (both are not true). Then, on average, a statistical test would classify $1,185 \times P$ genes as regulated by shear because P can be interpreted as the rate of false positives (type I error); $1,185 \times 0.01$ yields 12, and $1,185 \times 0.05$ yields 59; these numbers are almost the same as what are called significant by ANOVA. Obviously, the performance of ANOVA was not satisfactory. In contrast, when the G -test was used, the number of genes with positive test results is 61 at $P \leq 0.01$ and 161 at $P \leq 0.05$, which are more than three times the numbers of false positives expected.

Naturally, one would ask whether the G -test may have included too many genes that are not really significant. To answer this question, we must first define the meanings of positive and negative genes. When the effects of experimental treatments are of interest, we should regard a gene as a positive if some of the treatments change the expression level of the gene, i.e., at least one of the α_{ik} in Eq. 1 is not zero. It is in this sense that Fig. 2A compares the sensitivities of G -test and ANOVA at each fixed level of specificity, and the results show that the G -test can markedly improve sensitivity without sacrificing specificity when the biological variability is not small. We can address the same question from another perspective by comparing the positive predictive values (PPVs, i.e., the proportion of genes tested positive that are true positives) of the different tests. Unlike the sensitivity and specificity, the PPV of a test is a function of the prevalence of positives (i.e., the proportion of the total number of genes tested that are positive). Using Bayes' theorem, it can be shown that (1)

$$\text{PPV} = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})} \quad (6)$$

Figure 6 plots PPV as a function of the sensitivity and prevalence, with the specificity fixed at 0.95. It is evident that PPV increases with increasing sensitivity. Therefore, the set of genes selected by G -test actually had a smaller expected percentage of false positives than that by ANOVA.

Equation 6 points to a challenge one would face in identifying differentially expressed genes from microarray data. The power of the microarray approach lies in the fact that the number of genes assayed in a single experiment can be very large, often in the order of thousands or tens of thousands. When used to select specific genes for further in-depth analysis, one would like to have a large percentage of the selected genes to be truly positive, i.e., having a high PPV. This could be a challenge, because in many cases the prevalence of positive genes on the microarrays used with respect to

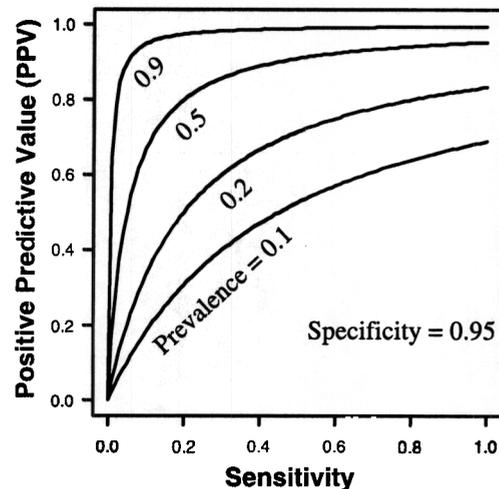


Fig. 6. The proportion of genes tested positive that are truly positive (PPV) increases as the sensitivity of the test increases, when the specificity of the test is fixed.

the particular experimental treatments is small, say, less than 0.1. As shown in Fig. 6, the low prevalence will lead to a small PPV, especially when the sensitivity is also low. To compare with the traditional gene-by-gene approach, note that for an individual gene the prevalence can be interpreted as the prior probability that the gene is positive, and PPV as the posterior probability that the gene is positive after it is tested to be positive. In the gene-by-gene approach, e.g., Northern analysis, the small number of genes studied are often selected exactly because they are expected to be positives (differentially expressed) by the researcher based on his/her prior knowledge; in other words, the prevalence of positives is high, say, 0.5 or higher. Therefore, as illustrated in Fig. 6, with the same sensitivity and specificity, the PPV of a gene tested positive in a microarray study is likely to be lower than the PPV of a gene tested positive in a gene-by-gene approach. For this reason, having a sensitive test for the statistical significance of observed differential expression is a more critical issue for data from microarray experiments than for data acquired by traditional methods.

PPV is a measure of confidence one can place in a positive call. Manduchi et al. (22) developed a novel method to assign confidence to a differential gene expression assessed from microarray data. In a similar way, one can show that

$$\text{PPV} \geq 1 - \frac{\text{Prob}(\text{called positive}|\text{negative})}{\text{Prob}(\text{called positive})} \quad (7)$$

The probability "Prob(called positive|negative)" is the preset cutoff point of P value, and "Prob(called positive)" can be estimated empirically from data as the fraction of genes called positive by the test. When applied to the results in Table 1, Eq. 7 yields a $\text{PPV} \geq 81\%$ for the 61 genes passed G -test at $P = 0.01$ and a $\text{PPV} \geq 34\%$ for the 18 genes passed ANOVA at the same cutoff point of P . This difference in PPV between

the two tests increases dramatically as the cutoff point of P is raised. At $P = 0.02$, PPV becomes $\geq 71\%$ for G -test and 1% for ANOVA; this very low confidence level for the ANOVA test further illustrates the point discussed at the beginning of this section. At the larger cutoff points of P , Eq. 7 is no longer applicable to the ANOVA results because it yields negative values. For G -test, however, one obtains a PPV $\geq 53\%$ even at the cutoff point of $P = 0.1$.

To compare with other previously described statistical tests designed for analysis of microarray data sets, we have used PaGE (PaGE_4.0.pl) obtained from Computational Biology and Informatics Laboratory at the Center for Bioinformatics at the University of Pennsylvania (<http://www.pcbi.upenn.edu>) to analyze the present data set. The PaGE_4.0 software by Manduchi et al., which has been updated from their original publication (22), can be used to identify differentially expressed genes with a preset minimum level of confidence. The application of PaGE_4.0 to our data yielded only two genes as differentially expressed at a minimum confidence level of 81%, and seven genes at a minimum confidence level of 53%. These genes do not include Egr-1, PDGF-A, caveolin-1, and other genes mentioned in RESULTS, whose expressions have been known to be regulated by shear. Note that at the same confidence levels (81% and 53%) G -test yielded, respectively, 61 and 252 differentially expressed genes. The better performance of the G -test over the method of Manduchi et al. (22) on the present data sets can be explained by the same reasons given above for comparing G -test and ANOVA, namely, the ability of G -test to separate the subject-subject variation from other random variations of experimental error, thus improving the power of a statistical test. The comparison of G -test with the method of Manduchi et al. again supports the concept that the subject-subject variation in gene expression is not small and hence efforts should be made to identify such variation in designing a statistical tests for gene expression data. Of course, such a test will require certain features in experimental design, as described in ANALYTICAL METHODS.

The approach used in our microarray experiments can be regarded as a "one-color" system. Another platform of microarray experiments is the "two-color" system, in which the reverse-transcribed cDNA from two samples labeled with two different fluorescent dyes are mixed and then hybridized to a single microarray. This allows the gene expressions in the two samples to be measured simultaneously, and the ratio of the two measurements from the same spot on the same microarray is used as a data point. In experiments using the two-color system, one of the two channels is often used for a reference sample, and the other for an experiment sample. The G -test can be applied to analyze the data (i.e., the experiment/reference ratios) obtained from such experiments in a similar manner as in the one-color system. For more complex experimental designs for the two-color system, modifications of the statistical model would be needed.

Although the focus of the present work is on testing the significance of the treatment effects, the model described by Eq. 1 also provides a basis for quantifying and testing the biological variability in the regulation of expressions of individual genes. Quantitative knowledge about the biological variability at the level of individual genes can be useful and might be necessary in appropriately modeling the interaction of genes as a network. We hypothesize that as a result of evolution, rather than by design, a genetic network has a high degree of redundancy which ensures its robustness and that this redundancy provides flexibility in regulating gene expression. As a result, there are great variabilities in gene expression profiles between experiments on different individuals or even between repeated experiments on the same individual at different times.

Indeed, the finding that the G -test yielded many more positive genes than ANOVA when applied to the experimental data suggests that the biological variability in gene expressions is large compared with the experimental random errors. This is consistent with the previous findings that the sample-sample variation of biological origin is greater than the spot-spot and slide-slide variations of technical origins (2). The microarray technology is still evolving rapidly. Its future advances will certainly reduce the extent of experimental errors. On the other hand, the biological variability is simply a property of the system under study and will remain large for any study in which conclusions are meant to be applicable to a general population represented by the sampled subjects. Therefore, the ratio of the variances due to the biological variability and experimental errors will be increased as the technology improves and so will be the difference between the sensitivities of the G -test and ANOVA, as indicated in Fig. 2A.

This work was supported by National Heart, Lung, and Blood Institute Grants HL-43026, HL-62747, and HL-64382.

REFERENCES

1. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.
2. Bartosiewicz M, Trounstein M, Barker D, Johnston R, and Buckpitt A. Development of a toxicological gene array and quantitative assessment of this technology. *Arch Biochem Biophys* 376: 66–73, 2000.
3. Blok LJ, Grossmann ME, Perry JE, and Tindall DJ. Characterization of an early growth response gene, which encodes a zinc finger transcription factor, potentially involved in cell cycle regulation. *Mol Endocrinol* 9: 1610–1620, 1995.
4. Bongrazio M, Baumann C, Zakrzewicz A, Pries AR, and Gaegtgens P. Evidence for modulation of genes involved in vascular adaptation by prolonged exposure of endothelial cells to shear stress. *Cardiovasc Res* 47: 384–393, 2000.
5. Butler PJ, Tsou TC, Li JY, Usami S, and Chien S. Rate sensitivity of shear-induced changes in the lateral diffusion of endothelial cell membrane lipids: a role for membrane perturbation in shear-induced MAPK activation. *FASEB J* 16: 216–218, 2002.
6. Chen BPC, Li YS, Zhao Y, Chen KD, Li S, Lao J, Yuan S, Shyy JYJ, and Chien S. DNA microarray analysis of gene expression in endothelial cells in response to 24-h shear stress. *Physiol Genomics* 7: 55–63, 2001.
7. Chiu JJ, Wung BS, Hsieh HJ, Lo LW, and Wang DL. Nitric oxide regulates shear stress-induced early growth response-1.

- Expression via the extracellular signal-regulated kinase pathway in endothelial cells. *Circ Res* 85: 238–246, 1999.
8. **DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, and Trent JM.** Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14: 457–460, 1996.
 9. **Diggle PJ, Liang KY, and Zeger SL.** *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1994.
 10. **Eisen MB, Spellman PT, Brown PO, and Botstein D.** Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863–14868, 1998.
 11. **Frangos JA, Eskin SG, McIntire LV, and Ives CL.** Flow effects on prostacyclin production by cultured human endothelial cells. *Science* 227: 1477–1479, 1985.
 12. **Garcia-Cardena G, Comander J, Anderson KR, Blackman BR, and Gimbrone MA Jr.** Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. *Proc Natl Acad Sci USA* 98: 4478–4485, 2001.
 13. **Guillemot L, Levy A, Raymondjean M, and Rothhut B.** Angiotensin II-induced transcriptional activation of the cyclin D1 gene is mediated by Egr-1 in CHO-AT(1A) cells. *J Biol Chem* 276: 39394–29403, 2001.
 14. **Hsieh HJ, Li NQ, and Frangos JA.** Shear-induced platelet-derived growth factor gene expression in human endothelial cells is mediated by protein kinase C. *J Cell Physiol* 150: 552–558, 1992.
 15. **Hsieh HJ, Li NQ, and Frangos JA.** Pulsatile and steady flow induces c-fos expression in human endothelial cells. *J Cell Physiol* 154: 143–151, 1993.
 16. **Ishida T, Takahashi M, Corson MA, and Berk BC.** Fluid shear stress-mediated signal transduction: how do endothelial cells transduce mechanical force into biological responses? *Ann NY Acad Sci* 811: 12–23 (discussion 23–24), 1997.
 17. **Jalali S, Li YS, Sotoudeh M, Yuan S, Li S, Chien S, and Shyy JY.** Shear stress activates p60src-Ras-MAPK signaling pathways in vascular endothelial cells. *Arterioscler Thromb Vasc Biol* 18: 227–234, 1998.
 18. **Kerr MK and Churchill GA.** Statistical design and the analysis of gene expression microarray data. *Genet Res* 77: 123–128, 2001.
 19. **Khachigian LM, Anderson KR, Halnon NJ, Gimbrone MA Jr, Resnick N, and Collins T.** Egr-1 is activated in endothelial cells exposed to fluid shear stress and interacts with a novel shear-stress-response element in the PDGF A-chain promoter. *Arterioscler Thromb Vasc Biol* 17: 2280–2286, 1997.
 20. **Li S, Kim M, Hu YL, Jalali S, Schlaepfer DD, Hunter T, Chien S, and Shyy JY.** Fluid shear stress activation of focal adhesion kinase. Linking to mitogen-activated protein kinases. *J Biol Chem* 272: 30455–30462, 1997.
 21. **Long AD, Mangalam HJ, Chan BY, Toller L, Hatfield GW, and Baldi P.** Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem* 276: 19937–19944, 2001.
 22. **Manduchi E, Grant GR, McKenzie SE, Overton GC, Surrey S, and Stoeckert CJ Jr.** Generation of patterns from gene expression data by assigning confidence to differentially expressed genes. *Bioinformatics* 16: 685–698, 2000.
 23. **McCormick SM, Eskin SG, McIntire LV, Teng CL, Lu CM, Russell CG, and Chittur KK.** DNA microarray reveals changes in gene expression of shear stressed human umbilical vein endothelial cells. *Proc Natl Acad Sci USA* 98: 8955–8960, 2001.
 24. **Newton MA, Kendzierski CM, Richmond CS, Blattner FR, and Tsui KW.** On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J Comput Biol* 8: 37–52, 2001.
 25. **Rocke DM and Durbin B.** A model for measurement error for gene expression arrays. *J Comput Biol* 8: 557–569, 2001.
 26. **Schena M, Shalon D, Davis RW, and Brown PO.** Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470, 1995.
 27. **Schwachtgen JL, Houston P, Campbell C, Sukhatme V, and Braddock M.** Fluid shear stress activation of egr-1 transcription in cultured human endothelial and epithelial cells is mediated via the extracellular signal-related kinase 1/2 mitogen-activated protein kinase pathway. *J Clin Invest* 101: 2540–2549, 1998.
 28. **Traub O, Monia BP, Dean NM, and Berk BC.** PKC-epsilon is required for mechano-sensitive activation of ERK1/2 in endothelial cells. *J Biol Chem* 272: 31251–31257, 1997.
 29. **Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, and Schoolnik GK.** Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* 96: 12833–12838, 1999.
 30. **Yan YX, Nakagawa H, Lee MH, and Rustgi AK.** Transforming growth factor-alpha enhances cyclin D1 transcription through the binding of early growth response protein to a cis-regulatory element in the cyclin D1 promoter. *J Biol Chem* 272: 33181–33190, 1997.