[Frontiers in Bioscience 7, a126-134, June 1, 2002]

SYSTEMS ANALYSIS OF MATRIX METALLOPROTEINASE mRNA EXPRESSION IN SKELETAL TISSUES

Lei Qian [1], Yunlong Liu [1], Hui Bin Sun [1,2], and Hiroki Yokota [1,2,3]

[1] Biomedical Engineering Program, [2] Departments of Anatomy and Cell Biology, and [3] Mechanical Engineering, Indiana University - Purdue University Indianapolis, Indianapolis, IN

TABLE OF CONTENTS

1. ABSTRACT

The availability of human genome sequences provides life scientists and biomedical engineers with a challenging opportunity to develop computational and experimental tools for quantitatively analyzing biological processes. In response to a growing need to integrate experimental mRNA expression data with human genome sequences, we present here a unique analysis named "Promoter-Based Estimation (PROBE)" analysis. The PROBE analysis is "systems analysis" of transcriptional processes using control and estimation theories. A linear model was built in order to estimate the mRNA levels of a group of genes from their regulatory DNA sequences. The model was also used to interpret two independent datasets in skeletal tissues. The results demonstrated that the mRNA levels of a family of matrix metalloproteinases can be modeled from a distribution of cis-acting elements on regulatory DNA sequences. The model accurately predicted a stimulatory role of cis-acting elements such as AP1, NFY, PEA3, and Sp1 as well as an inhibitory role of AP2. These predictions are consistent with biological observations, and a specific assay for testing such predictions is proposed. Although eukaryotic transcription is a complex mechanism, the two examples presented here support the potential use of the described analysis for elucidating the functional significance of DNA regulatory elements.

2. INTRODUCTION

A complete set of eukaryotic genome sequences provides a unique opportunity to develop analytical and computational tools for interpreting and evaluating complex biological responses (1, 2). Two of the immediate engineering challenges in a post-genomic era are building user-friendly databases and developing computational algorithms to effectively utilize those databases.

Transcriptional regulation is tightly linked to regulatory DNA sequences and is a critical step in the cascade of gene expression (3). Computer programs such as PROSCAN and SIGSCAN have been developed to identify putative promoter regions and to determine binding sites of transcription factors (4, 5). To our knowledge, however, few mathematical models are formulated to relate the mRNA expression levels to the role of DNA regulatory elements, and to model cellular transcriptional states.

In response to a growing need to integrate experimental mRNA data with human genome sequences, we developed a unique computational approach named "PROmoter-Based Estimation (PROBE)" analysis and conducted a "systems analysis" of mRNA expression in skeletal tissues. Previously a non-model-driven approach such as cluster analysis had been developed, where quantitative expression profiles among genes are classified into hierarchical clusters based on expressional similarity (6, 7). A computational method was then developed for discovering cis-regulatory elements responsible for each cluster (8-10). However, few works have attempted to build a holistic model suitable for performing "systems analysis" of transcriptional activities. A mathematical model would be useful to life scientists and biomedical engineers if the model could evaluate the functional role of DNA regulatory elements in growth and differentiation of various tissues.

With the understanding that individual DNA regulatory elements can be regulated with diversity and precision, we built and evaluated a linear least-square model. Focusing on the 5'-flanking regulatory region, we first counted the number of cis-acting elements for individual genes. We then modeled the experimentally observed mRNA levels using a weighed sum of the frequency of the selected cis-acting elements. In this promoter-based model, optimal weights were determined using a standard linear estimation technique. The model was used to predict a stimulatory or inhibitory role for each cis-acting element and the specific combination of cis-acting elements that would most effectively regulate mRNA expression. Mathematical formulation is described in the Appendix.

To examine the PROBE algorithm, two mRNA expression datasets in skeletal tissues were used. One dataset consisted of 14 matrix metalloproteinase (MMP) genes, an influential proteolytic enzyme that degrades collagen, and collagen-associated molecules in an extracellular matrix (11-13). Controlling MMP expression is critical in preserving or remodeling skeletal tissues (14-16). In the second example, we selected a heterogeneous set of genes such as MMPs, tissue inhibitors of metalloproteinases, and growth factors involved in inflammation and degradation of skeletal tissues (17). We evaluated the effects of 5-7 cis-acting elements including AP1, AP2, NFY, PEA3, Sp1, TFIID, and TIE using the PROBE algorithm and demonstrated in the two examples that the promoter-based linear model can represent at least in part complex regulatory mechanisms. The merits and limitations of the PROBE analysis for skeletal tissue engineering are discussed and a biological assay for testing the predictions is proposed.

3. MATERIALS AND METHODS

The PROBE algorithm receives two inputs such as "mRNA expression data" and "information on cis-acting DNA regulatory elements." The linear model was built to minimize mismatches between the observed mRNA levels and the modeled mRNA levels, where three mathematical entities such as a promoter matrix (H), a promoter-associated matrix (H_A), and a weighting matrix (R) were defined (Figure 1). Mathematical formulation is described in the Appendix. The multidimensional scaling analysis in 2D Euclidian space was performed using SPSS (version 11.0, LEAD Technologies, Inc.).

In order to evaluate the PROBE model, we conducted a leave-one-out cross-validation test and a Monte Carlo simulation. In leave-one-out cross-validation, the expression level of one gene in the dataset was predicted from the expression levels of the other genes. In the Monte Carlo simulation, the observed expression levels were randomly re-assigned among genes and samples, and the error estimated from 10,000 random trials were compared to the true model error for the correctly assigned expression levels.

3.1. Example 1

The mRNA level of 14 MMPs including MMP-1, -2, -3, -7, -8, -9, -10, -11, -12, -13, -14, -16, -19, and -20 was obtained from the study of rheumatoid arthritis and traumatic disease conducted by Konttinen *et al*. (11). The extraction of mRNA was performed two or more times and the results were shown to be reproducible. The expression level was represented by a value in the range of 0 to 1 and a continuous gray level was used to illustrate the expression profile. Ten tissue samples (k = 1 to 10) were derived from rheumatoid arthritis patients, and nine tissue samples (k = 11 to 19) were isolated from traumatic disease patients. The MMP expression of rheumatic tissue was on average higher than that of traumatic patients.

The 5'-end upstream regulatory region, 500 bp in length, was used for the PROBE analysis. MMP-15 and MMP-17 were excluded, since the size of a dominant PCR fragment differed from the control and the promoter was not retrievable from the currently available human genome. The accession numbers were AJ002550 (MMP-1), AJ298926 (MMP-2), U51914 (MMP-3), NT009151 (MMPs-7, -10, -12, and -20), AF059679 (MMP-8), NT011375 (MMP-9), NT011520 (MMP-11), U52692 (MMP-13), NT024615 (MMP-14), NT008256 (MMP-16), and NT009458 (MMP-19). Cis-acting regulatory elements such as AP1, AP2, NFY, PEA3, Sp1, TFIID, and TIE were considered.

3.2. Example 2

The mRNA expression data were obtained from the study conducted by Bunker *et al*. (17). The data included three sample groups such as chronic fibrosing patients, control individuals, and Dupuytren's disease

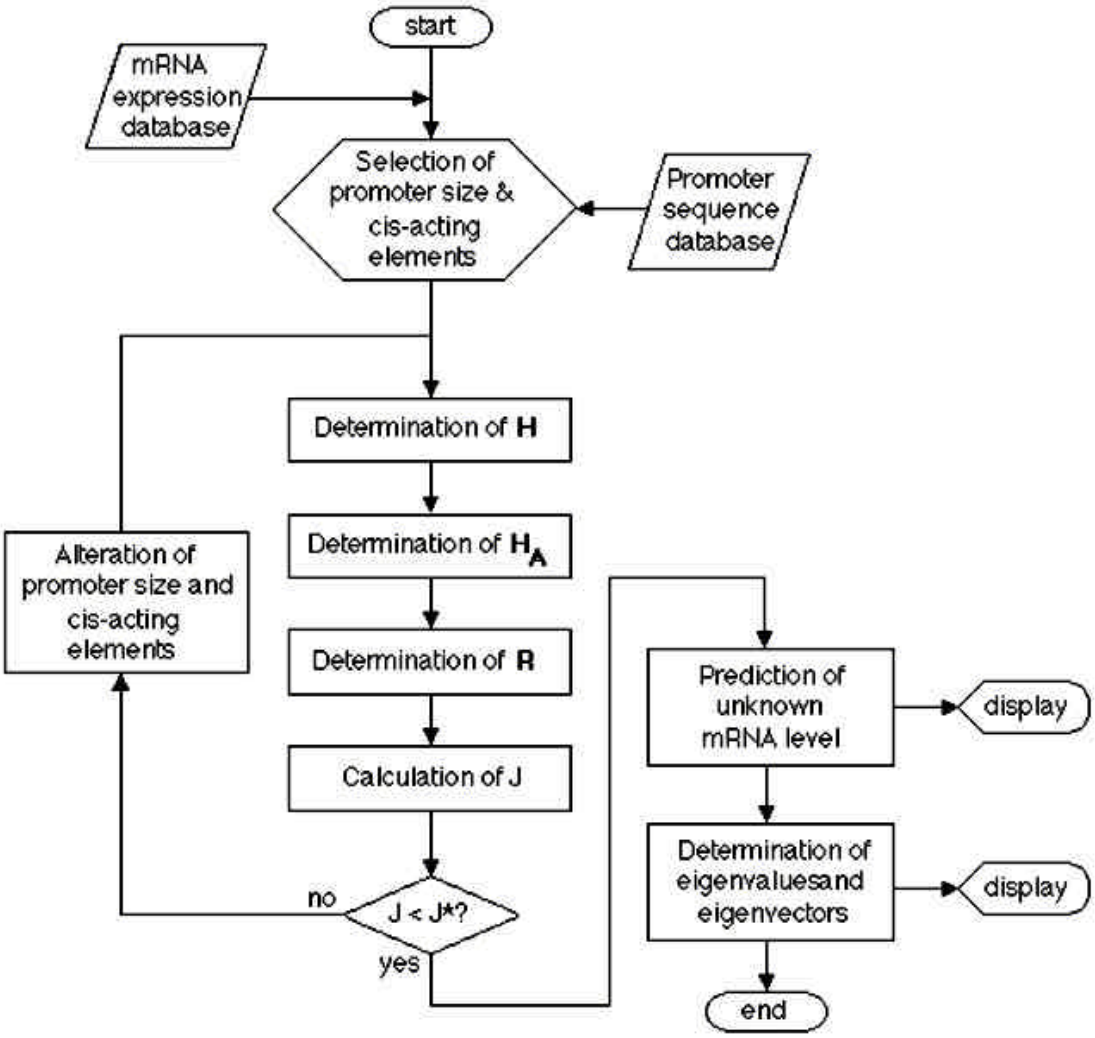

Figure 1. Flowchart of the described promoter-based estimation algorithm.

patients. We focused on 13 genes such as MMP-1, MMP-2, MMP-3, MMP-9, MMP-14, TIMP-1, β-2 microglobulin, acidic fibroblast growth factor (a-FGF), basic fibroblast growth factor (b-FGF), interleukin-6 (IL-6), platelet driven growth factor-α (PDGF- α), transforming growth factor-β (TGF- β), and tumor necrosis factor-α (TNF-α).

The expression level in each sample group was defined by N_{ik}^{+}/N_k, where N_{ik}^{+} = the number of the tissue samples expressing the i-th gene in the k-th group, and N_k = the total number of tissue samples in the k-th group. The accession numbers were NT019712 (TIMP-1), NT010302 (β-2 microglobulin), NT016788 (a-FGF), NT016354 (b-FGF), AF869204 (IL-6), M59423 (PDGF-α), NT011139 (TGF-β), and NT023426 (TNF-α). Five cis-acting regulatory elements, AP1, AP2, NFY, PEA3, and Sp1, were considered in this example. The PROBE analysis is not designed to model post-transcriptional processes. Therefore, genes such as IL-1 on which the mRNA level is regulated after transcription were excluded.

4. RESULTS

4. 1. Example 1

4.1.1. Modeling of MMP transcript levels

Using the seven cis-acting elements (AP1, AP2, NFY, PEA3, Sp1, TFIID, and TIE) on the 500-bp upstream DNA sequences, the least-square estimator was applied to the dataset consisting of the transcript levels of 14 MMPs for 19 samples (Figure 2). Two simulated expression patterns were derived from the observed expression pattern in a continuous gray code. The expression pattern illustrated in Figure 2C was generated by the "leave-one-out" cross-validation procedure, and the expression pattern in Figure 2D was modeled using all available data. The multidimensional scaling analysis was performed to locate the 19 tissue samples in 2D Euclidian space, where the black and white circles represented the samples from the patients with rheumatoid arthritis and non-rheumatoid arthritis, respectively, for the observed (Figure 3A) and predicted expression patterns (Figure 3B). The model error, defined by "J" in the Appendix, was calculated as

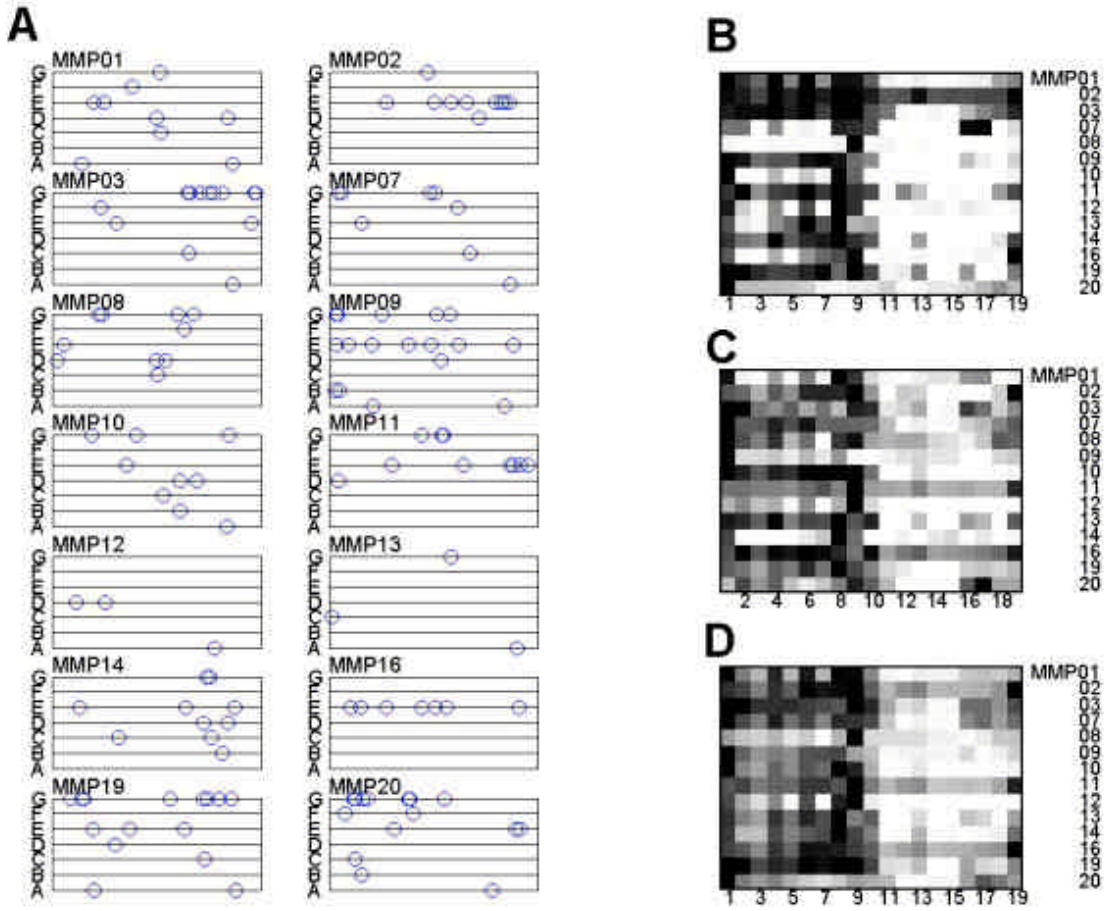

Figure 2. Map of cis-acting elements and mRNA expression patterns for 14 MMPs in example 1. (A) Distribution of 7 cis-acting motifs on the 500-bp upstream sequences where the indexes A through G represent AP1, AP2, NFY, PEA3, Sp1, TFIID, and TIE, respectively. The right end of the horizontal axis corresponds to a transcription initiation site. (B) Observed mRNA expression pattern. Using 266 squares corresponding to 14 MMPs in 19 tissue samples, the mRNA levels are illustrated in a gray-code where darker color indicates higher expression. (C) Predicted mRNA expression pattern using a "leave-one-out" cross-validation procedure. (D) Modeled mRNA expression pattern using all MMP data.

11.3 for the modeled pattern depicted in Figure 2D. When the expression levels were randomly re-assigned among 266 data points, the mean error and the standard deviation were 22.2 and 2.4 for 10,000 cases in the Monte Carlo simulation (Figure 3C).

4.1.2. Sensitivity analysis

We next conducted a sensitivity analysis using eigenvalues and eigenvectors as indicators where the effectiveness of each cis-acting element on MMP expression was examined. There are seven sets of eigenvalues and unit eigenvectors corresponding to the seven selected cis-acting elements. An eigenvector represents a specific combination of seven cis-acting elements and an associated eigenvalue indicates effectiveness of the combination in regulating mRNA levels. The calculated eigenvalues were 1.44, 0.35, 0.11, 0.07, 0.02, 0.004, and 0.0001, and the eigenvector corresponding to the largest eigenvalue was $(0.209, -0.209, 0.240, 0.004, 0.921, -0.055, 0.068)^T$ in the 7-dim space of AP1, AP2, NFY, PEA3, Sp1, TFIID, and TIE. The positive values in the elements of the eigenvector indicated a stimulatory role of the corresponding cis-acting elements (AP1, NFY, PEA3, Sp1, and TIE) and the negative values suggested an inhibitory role (AP2, and TFIID).

Since each skeletal tissue exhibited a unique MMP mRNA pattern, the estimate of active cis-acting elements must differ among the 19 tissue samples. In order to examine the role of the selected cis-regulatory elements, we determined a component of active cis-acting elements projected onto the eigenvector with the largest eigenvalue. In a linear estimation analysis, this projected component serves as an indicator of the MMP expression levels. Indeed, the component was positively correlated to each tissue's mean MMP expression level with a correlation coefficient of 0.98 (Figure 3D). The average value was 0.567 and 0.169 for the tissues derived from rheumatic arthritis and traumatic diseases respectively, consistent with the observation that the rheumatic arthritis tissues present a higher level of MMP mRNAs than the traumatic disease tissues.

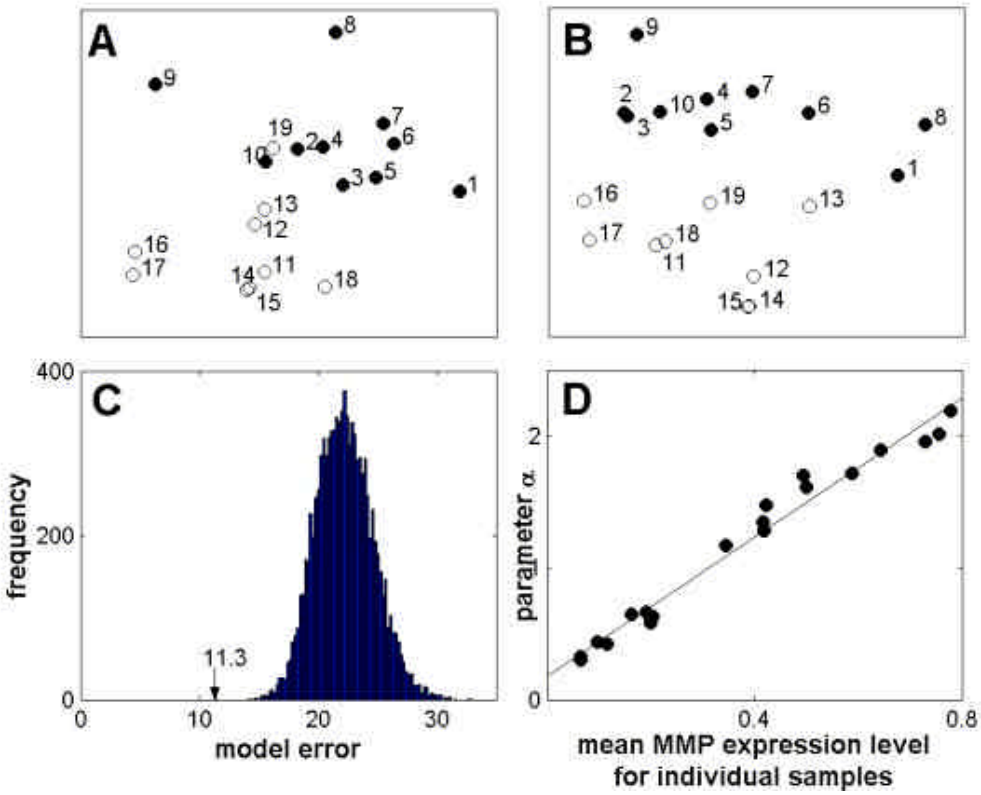

Figure 3. 2D scaling analysis and error analysis in example 1. (A) 2D Euclidian representation of 19 tissue samples based on the observed MMP expression pattern. The black circles represent rheumatoid arthritis patients, and the white circles represent non-rheumatoid arthritis patients. (B) 2D Euclidian representation based on the predicted MMP expression pattern. (C) Monte Carlo simulation for model error with the randomly assigned expression levels. The mean error ± standard deviation for 10,000 cases was 22.2 ± 2.4. The arrow indicates the true model error of 11.3 for the expression pattern illustrated in Figure 2D. (D) Positive correlation between the parameter α and mean MMP expression level for individual samples. The best-fit-line is $y = 2.63x + 0.18$ with $r^2 = 0.98$.

4. 2. Example 2

4.2.1. Modeling of a heterogeneous group of transcripts

In the second example, a heterogeneous set of genes including MMPs, tissue inhibitor of metalloproteinases, and various growth factors was modeled. Three groups of tissues were derived from chronic fibrosing patients, normal control, and Dupuytren's disease patients. We first predicted the mRNA level of one gene from the mRNA level of the other genes. When the mRNA level was assigned to 3 levels, the prediction by the least-square estimator gave the correct level in 26 (67%) out of 39 total cases (Figs. 4A and 4B). Twelve cases were incorrect by a single expression level, and one case was off by two expression levels. Without a weight imposed for compensating variations among genes, the correct cases were reduced to 22 (56%). When the expression was assigned to 2 levels, the maximum rate of successful prediction increased to 34 cases (87%) (Figs. 4C and 4D). Without any weighting factor, the correct prediction was limited to 28 cases (72%).

4.2.2. Variations among genes

Weighting factors were introduced to evaluate variations among genes. Since an element in the weighting matrix was assigned inversely proportional to the mean-square error (see Appendix), each element should serve as a fitness indicator of each gene. We have shown that performance of the least-square modeler was enhanced by the weighting matrix. A large element assigned to the genes such as TIMP-1, IL-6, MMP-9, and MMP-1 indicated that their measured mRNA levels fit well with the linear estimation model (Figure 5A). A poor fitting of MMP-2, MMP-14, aFGF and TGF-β, on the other hand, was suggested by a low value. We excluded four genes such as IL-1α, IL-1β, TNF-β, and PDGF-β because of poor fitting.

4.2.3. Estimation of active cis-acting elements

The last step was to estimate a level of active cis-acting elements for three tissue groups derived from chronic fibrosing patients, control, and Dupuytren's disease patients (Figure 5B). In chronic fibrosing patients, the active level of AP1 was significantly higher than the other two groups, while the estimated level of NF-Y was highest in Dupuytren's disease patients. Five sensitivity values (eigenvalues) corresponding to the selected cis-acting elements were 9.7, 2.1, 0.8, 0.005, and 0.003. The eigenvector (a combination of cis-acting elements) corresponding to the largest eigenvalue was $(0.03, -0.30, -0.03, 0.65, 0.70)^T$ in a 5-dim space of AP1, AP2, NFY,

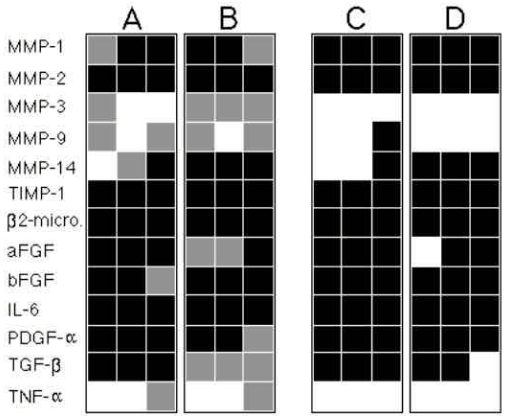

Figure 4. Comparison of the measured mRNA level and the predicted mRNA level in Example 2. (A) Measured mRNA expression in three levels: white - the level lower than 1/3, gray - the level between 1/3 and 2/3, and black - the level higher than 2/3. (B) Predicted mRNA expression in three levels. (C) Measured mRNA expression in two levels. (D) Predicted mRNA expression in two levels.

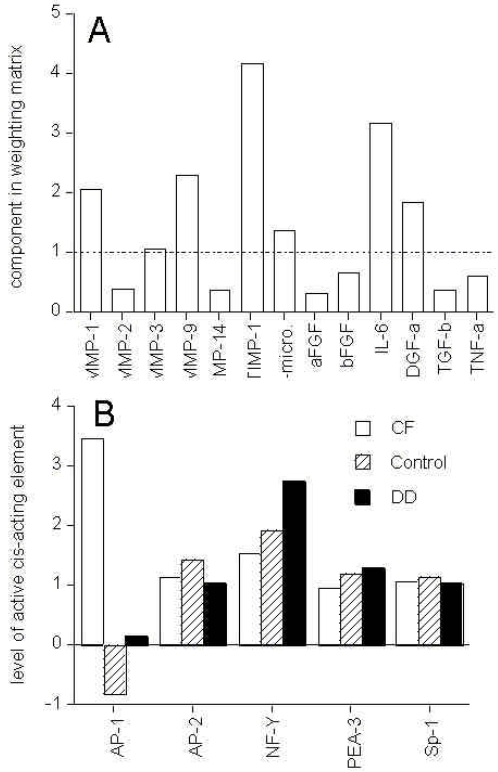

Figure 5. Weighting matrix and estimate of active cis-acting elements in Example 2. (A) Diagonal components of the weighting matrix for each gene. Among 13 genes, 7 genes, MMP-1, MMP-3, MMP-9, TIMP-1, β2-microglobulin, IL-6, and PDGF-α, had a weighting factor greater than 1. This suggests that their mRNA expression pattern fits to the linear model better than the other genes such as MMP-2, MMP-14, aFGF, bFGF, TGF-β, and TNF-α. (B) Estimate of active cis-acting elements such as AP1, AP2, NFY, PEA3, and Sp1 for three tissue groups. Three tissue groups are: CF - chronic fibrosing patients, Control – control individuals, and DD - Dupuytrenís disease patients.

PEA3, and Sp1. A large positive value for PEA3 and Sp1 suggested a stimulatory role and a large negative value for AP2 indicated an inhibitory role.

5. DUSCUSSION

In an attempt to establish a systematic model for eukaryotic transcription activities, a promoter-based estimation algorithm was developed and a sensitivity analysis for the selected cis-acting regulatory elements was conducted. The described mathematical formulation allowed us to highlight the merits and limitations of linear approximation in analyzing complex eukaryotic transcriptional regulation.

Two merits of the described promoter-based estimation analysis are the capability of modeling and predicting mRNA levels and the unique sensitivity analysis for active cis-acting elements. One major difference between the current work and other linear regression models is a system's formulation (10). In our formulation a direct mRNA level rather than a logarithm of an expression ratio was used as a measurement variable, and an activation level of cis-acting elements was defined as a state variable. This formulation allowed us to estimate the state variables (cellular states) and to model and predict the measurement variables (mRNA levels) from the promoter matrix and the associated matrices. Our least-square modeler can accommodate, if necessary, supplementary data in the form of *a priori* information or weighting factors, and it can be extended into a dynamical model without altering the definition of state and measurement variables. In this study, 5-7 cis-acting elements were chosen from 500-bp promoters (Example 1) and 800-bp promoters (Example 2). A careful determination of promoter length and cis-acting elements seems to further improve performance of the described least-square linear estimator. Although a model with 7 cis-acting elements was presented in Example 1, the combination of 5 elements such as AP1, AP2, NFY, PEA3, and Sp1 gave the minimum model error for the 500-bp upstream regulatory sequences and different combinations of cis-acting elements were better for other regulatory regions (data not shown).

The sensitivity analysis provided a good measure for the combinatorial effects of cis-acting elements. A set of combinations of cis-acting elements, eigenvectors, represent independent (orthogonal) combinations in a space of cis-acting elements, and associated sensitivity values (eigenvalues) indicate the effectiveness of particular combinations of cis-acting elements in altering mRNA expression. The primary eigenvector corresponding to the largest eigenvalue indicates the most effective combination of cis-acting elements to regulate a mean-square sum of mRNA levels. For instance, the positive value in the primary eigenvector for AP1, PEA3, and Sp1 in two examples suggested a stimulatory effect of transcription factors such as c-fos, c-jun, and ets-1. On the other hand, AP2 had a negative value in both examples, suggesting an inhibitory effect. Although the above interpretation of AP1, PEA3, Sp1, and AP2 appears consistent with several lines of biochemical observations, a role of these elements

depends on tissue samples, individual genes, and growth conditions (16-18). The linear PROBE model should be able to predict the role of cis-acting elements specific to individual tissues or genes.

The model in the PROBE analysis is a coarse approximation of eukaryotic regulatory networks. Genes that are regulated on a post-transcriptional level such as IL-1α and IL-1β did not fit the model (19). When an expression level was randomly assigned in the Monte Carlo simulation, the predicted mRNA level also became nearly random. Therefore, the linear model represents, at least in part, complex transcriptional machinery related inflammation and degeneration of skeletal tissues. The following three reasons provide justification for our approach. First, cis-acting elements are indispensable in transcription activities and the 5'-end regulatory promoter focused in this analysis represents a core region besides other regulatory regions located in 3'-ends or introns (20). Second, eukaryotic transcriptional activities are controlled by a combination of multiple cis-acting elements and a weighed sum of the number of cis-acting elements appears as a simplified representation of their contribution. Third, transcriptional assays such as a reporter gene assay and an electrophoretic mobility shift assay are able to simulate the functional significance of cis-acting elements using shorter DNA fragments in 20 – 500 bp than complete gonomic DNA sequences (21).

6. FUTURE WORK

The promoter-based estimation analysis described here is at an infancy stage, and some future works are therefore suggested. First, a set of cis-acting elements has to be carefully chosen. Although representative 5-7 elements were chosen in our examples, other elements such as NF-κB are shown to affect some of the MMP expression (22-24). A cluster analysis may help identify active cis-regulatory elements from a cluster of co-regulated genes. Second, the size of the promoter and the degree of sequence degeneracy need to be evaluated. Third, implementing any nonlinear effect needs to be investigated. Modeling a nonlinear system using a linear perturbation from a reference state may facilitate the incorporation of cooperative or competitive binding of multiple transcription factors.

The described PROBE analysis for modeling and analyzing transcription activities offers a computational tool for life scientists and biomedical engineers to integrate experimental expression data with available genome information. It was a unique application of a linear estimation theory popularly used in navigating spacecraft or processing electric signals (25). In this study we started with the smallest number of essential components, since an elegant model can often have greater intrinsic value than an accurate one overloaded with detail (26). We did not follow a commonly accepted scheme of modeling that requires a number of parameters related to binding affinity and stability of trans-acting regulatory elements (27). Although the linear model described here is a coarse approximation of complex eukaryotic regulation, a simple, but general mathematical framework provided logical insights in a combinatorial role of cis-acting elements.

In order to examine the prediction of the promoter-based estimation analysis regarding the role of cis-acting regulatory elements, a biological assay using DNA oligonucleotides can be considered (28). In this assay DNA fragments consisting of specific cis-acting elements are transferred into cultured cells. Since exogenous DNA fragments act as a competitor of genomic cis-acting elements, reduction in specific gene transcripts in the assay suggests that the transferred cis-acting elements mimic the binding capacity of endogenous cis-acting elements. The integrated approach of the promoter-based estimation analysis with the biological competition assay will likely provide a power tool for life scientists and biomedical engineers to elucidate molecular mechanisms underlying tissue growth and differentiation.

7. APPENDIX

Formulation of Promoter-Based Linear Model: A transcript level of "n" genes and a level of "m" functional cis-acting elements are represented by a vector $\underline{z}_k$ and a vector $\underline{x}_k$, respectively, and they are linearly linked:

$$\underline{z}_k = HH_A\underline{x}_k + \underline{v}_k$$

where H is an (n x m) promoter matrix, H_A is an (m x m) promoter-associated matrix, $\underline{v}_k$ is a vector for measurement error, and subscript k designates tissue samples. The (i, j) component of H corresponds to the number of the j-th cis-acting element for the i-th gene. The software SIGSCAN (Version 4.05, Advance Biosciences Computing Center, University of Minnesota) was used to identify H from 5'-end regulatory regions (Figure 2A and Table 1). H_A is a diagonal matrix whose j-th diagonal component weighs a contribution of the j-th cis-acting element to transcript levels. We determined the vector $\underline{h}_A = (H^TH)^{-1}H^T\underline{z}^{av}$ and set the j-th component of $\underline{h}_A$ to the j-th diagonal component of H_A. The vector $\underline{z}^{av}$ represents the mean mRNA level among tissue samples.

In order to estimate $\underline{x}_k$ from the observed $\underline{z}_k$, the function J is defined:

$$J = (\underline{z}_k - HH_A\underline{x}_k)^TR^{-1}(\underline{z}_k - HH_A\underline{x}_k)$$

where R^{-1} is a diagonal weighting matrix. The i-th diagonal component of R^{-1} is set to $1/\sigma_i^2$, where σ_i^2 is the approximate mean-square variation of the mRNA level for the i-th gene. The least-square estimate of $\underline{x}_k$ is obtained by $?J/?\underline{x}_k = 0$:

$$\underline{x}_k^e = (H_A^TH^TR^{-1}HH_A)^{-1}H_A^TH^TR^{-1}\underline{z}_k$$

Eigenvalue Analysis: In order to evaluate the effectiveness of the selected cis-acting elements on the mRNA expression, the eigenvalue and the eigenvector of the matrix, $A = H_A^TH^TR^{-1}HH_A$, was analyzed. In a linear equation such as $A\underline{y} = \lambda\underline{y}$, a scalar λ and a unit vector $\underline{y}$ are called an eigenvalue and an eigenvector. There are in

Table 1. Promoter matrix H for Example 2

Gene	cis-acting element				
	AP1	**AP2**	**NFY**	**PEA3**	**Sp1**
MMP-1	2	0	1	2	3
MMP-2	0	0	0	1	18
MMP-3	1	0	2	1	3
MMP-9	3	3	0	2	9
MMP-14	0	1	3	3	4
TIMP-1	2	1	1	3	7
β2-microglobulin	0	0	1	2	7
a-FGF	0	1	1	2	6
b-FGF	1	1	0	0	20
IL-1	1	0	2	2	8
PDGF-α	0	1	0	0	18
TGF-β	0	1	1	0	18
TNF-α	1	4	0	3	5

The promoter 800 bp in length was used.

general m sets of eigenvalues and eigenvectors corresponding to "m" cis-acting elements. In evaluating effectiveness of a particular combination of cis-acting elements for each tissue sample, we determined α_k:

$$\alpha_k = \underline{v}_1^T \underline{x}_k^e$$

where α_k represents the component of $\underline{x}_k^e$ parallel to the primary eigenvector. Matrix operation such as multiplication, transposition, and inversion as well as solving eigenvalues and eigenvectors were conducted using MATLAB (version 6, The Math Works, Inc.).

8. ACKNOWLEDGEMENTS

The authors would like to thank Y.T. Konttinen and T.D. Bunker for use of the expression data, J. Chen, Z. Ben Miled, and O. Bukhrus for their valuable advice, and Heather Ramsey for proofreading. This study was in part supported by Twenty-First Century Research and Technology Funds.

9. REFERENCES

1. Collins, F.S., Patriano, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., the members of the DOE and NIH planning groups: New goals for the U.S. human genome project: 1998-2003. *Science* 282, 682-689 (1998)

2. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001)

3. Ptashne, M. & Gann, A.: Transcriptional activation by recruitment. *Nature* 386, 569-577 (1997)

4. Prestridge, D.S.: SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. *CABIOS* 7, 203-206 (1991)

5. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T.: MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nuc. Acids. Res*. 23, 4878-4884 (1995)

6. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D.: Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14863-14868 (1998)

7. Alter, O., Brown, P.O. & Botstein, D.: Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10101-10106 (2000)

8. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M.: Systematic determination of genetic network architecture. *Nature Genet*. 22, 281-285 (1999)

9. Hughes, J.D., Estep, P.W., Tavazoie, S. & Church, G.M.: Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol*. 296, 1205-1214 (2000)

10. Bussemaker, H.J., Li, H. & Siggia, E.D.: Regulatory element detection using correlation with expression. *Nature Genet*. 27, 167-171 (2001)

11. Konttinen, Y.T., Ainola, M., Valleala, H., Ma, J., Ida, H., Mandelin, J., Kinne, R.W., Santavirta, S., Sorsa, T., Lopez-Otin, C. & Takagi, M.: Analysis of 16 different matrix metalloproteinases (MMP-1 to MMP-20) in the synovial membrane: different profiles in trauma and rheumatoid arthritis. *Ann. Rheum. Dis*. 58, 691-697 (1999)

12. Shingleton, W.D., Hodes, D.J., Bzrick, P. & Cawston, T.E.: Collagenase: a key enzyme in collagen turnover. *Biochem. Cell Biol*. 74, 759-775 (1996)

13. Borden, P. & Heller, R.A.: Transcriptional control of matrix metalloproteinases and the tissue inhibitors of matrix metalloproteinases. *Critical Rev. Eukaryotic Gene Expression* 7, 159-178 (1997)

14. Arnett, F.C.: Rheumatoid arthritis, pp. 1492-1499. In Textbook of Medicine, Goldman, L., and Bennett, J.C. (eds.), W.B. Saunders Co., Philadelphia (2000)

15. Yoshihara, Y., Nakamura, H., Obata, K., Yamada, H., Hayakawa, T., Fujikawa, K. & Okada, Y.: Matrix metalloproteinases and tissue inhibitors of metalloproteinases in synovial fluids from patients with rheumatoid arthritis or osteoarthritis. *Ann. Rheum. Dis*. 59, 455-461 (2000)

16. Feldmann, M. & Maini, R.N.: The role of cytokines in the pathogenesis of rheumatoid arthritis. *Rhematol*. 38 (suppl. 2), 3-7 (1999)

17. Bunker, T.D., Reilly J., Baird, K.S. & Hamblen, D.L.: Expression of growth factors, cytokines and matrix metalloproteinases in frozen shoulder. *J. Bone Joint Surgery*, 82B, 768-773 (2000)

18. Sun, H.B. & Yokota, H.: Messenger-RNA expression of matrix metalloproteinases, tissue inhibitors of metalloproteinases, and transcription factors in rheumatic synovial cells under mechanical stimuli, *Bone* 28, 303-309 (2000)

19. Dinarello, C.A.: Interleukin-1. *Cytokine Growth Factor Rev*. 8, 253-265 (1997)

20. Stuuhl, K.: Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98, 1-4 (1999)

21. Mueller, J.M. & Pahl, H.L.: Assaying NF-κB and AP-1 DNA-binding and transcriptional activity. *Methods Mol. Biol*. 99, 205-216 (1999)

22. Benbow, U. & Brinckerhoff, C.E.: The AP-1 site and MMP gene regulation: What is all the fuss about? *Matrix Biol*. 15, 519-526 (1997)

23. Westermarck, J. & Kahari, V.M.: Regulation of matrix metalloproteinases expression in tumor invation. *FASEB J.* 13, 781-792 (1999)

24. Pendas, A.M., Santamaria, I., Alvarez, M.V., Pritchard, M. & Lopez-Otin, C.: Fine physical mapping of the human matrix metalloproteinase genes clustered on chromosome 11q22.3, *Genomics* 37, 264-265 (1996)

25. Gelb, A.: Applied Optimal Estimation, pp. 102-105. The MIT Press, Cambridge (1984)

26. Nowak, M.A. & Bangham, C.R.M.: Population dynamics of immune responses to persistent viruses. *Science* 272, 74-79 (1996)

27. Hammond, B.J.: Quantitative study of the control of HIV-1 gene expression. *J. Theor. Biol*. 163, 199-221 (1993)

28. Sun, H.B., Malacinski, G.M. & Yokota, H.: Promoter competition assay for analyzing gene regulation in joint tissue engineering (Unpublished data)

Key Words: Gene Regulation, Promoter, Cis-Acting Element, Least-Square Estimation, Eigenvalue, MMP

Send correspondence to: Hiroki Yokota, Ph.D., Department of Anatomy and Cell Biology, Indiana University School of Medicine, 635 Barnhill Drive, MS-504, Indianapolis, IN 46202, Tel : 317-274-2448, Fax: 317-278-2040, E-mail: hyokota@iupui.edu