IDENTIFICATION OF ACTIVATED TRANSCRIPTION FACTORS FROM MICROARRAY GENE EXPRESSION DATA OF KAMPO MEDICINE-TREATED MICE

RUI YAMAGUCHI[1]
ruiy@ims.u-tokyo.ac.jp

MASAHIRO YAMAMOTO[2,3]
yamamoto_masahiro@mail.tsumura.co.jp

SEIYA IMOTO[1]
imoto@ims.u-tokyo.ac.jp

MASAO NAGASAKI[1]
masao@ims.u-tokyo.ac.jp

RYO YOSHIDA[4]
yoshidar@ism.ac.jp

KENJI TSUIJI[2]
di055036@sc.itc.keio.ac.jp

ATSUSHI ISHIGE[2]
ishige@sc.itc.keio.ac.jp

HIROAKI ASOU[3]
asou@tmig.or.jp

KENJI WATANABE[2]
toyokeio@sc.itc.keio.ac.jp

SATORU MIYANO[1]
miyano@ims.u-tokyo.ac.jp

[1] *Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[2] *Department of Kampo medicine, Keio University School of Medicine 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan*
[3] *Department of Neuro-glia cell biology, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakaecho, Itabashi-ku, Tokyo 173-0015, Japan*
[4] *The Institute of Statistical Mathematics 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan*

We propose an approach to identify activated transcription factors from gene expression data using a statistical test. Applying the method, we can obtain a synoptic map of transcription factor activities which helps us to easily grasp the system's behavior. As a real data analysis, we use a case-control experiment data of mice treated by a drug of Kampo medicine remedying degraded myelin sheath of nerves in central nervous system. Kampo medicine is Japanese traditional herbal medicine. Since the drug is not a single chemical compound but extracts of multiple medicinal herb, the effector sites are possibly multiple. Thus it is hard to understand the action mechanism and the system's behavior by investigating only few highly expressed individual genes. Our method gives summary for the system's behavior with various functional annotations, e.g. TFAs and gene ontology, and thus offer clues to understand it in more holistic manner.

Keywords: transcription factor activity; gene expression; Kampo medicine; multiple sclerosis; statistical absolute evaluation; MetaGP; myelination; remyelination.

1. Introduction

Transcription factors (TFs) have a central role in gene regulatory systems. The transcription factor activities (TFAs) affect the transcription of genes which are regulated by the TFs. To investigate when and under what condition they become effective is important for understanding the system's behavior. However, to observe TFAs is difficult, since the gene expression levels of TFs are often low. Moreover,

TFAs are activities of proteins, and thus, it would be hard to interpret the TFs' gene expressions directly into the TFAs. Several methods have been proposed to estimate the TFAs from the gene expression data, e.g. kinetic models [14], network component analysis [4, 7, 8, 12], state-space models [11].

As typical biological studies, experiments are often conducted in case-control manner. The main interest is to find differences between the cases and the control. To this purpose, a simple way is to find genes that are differentially expressed between the cases and the controls. For more precise understanding of such genes, the activities of TFs that regulate them is essential. Therefore, we can extend the problem from the finding differentially expressed genes to the discovery of TFs that show different activities between the cases and the controls. Although we may not observe difference in TFAs directly, we can observe that of gene expressions. Thus it is beneficial to develop a method to identify TFs having significantly different activities from the gene expression data. Here we introduce an approach for that purpose based on a statistical test.

To test differentially activated TFs, we use a statistical test for a set of genes which was proposed by Gupta *et al.* [6]. We name the method the Meta Gene Profiler (MetaGP) method. The MetaGP method accumulates statistical evidences from a set of genes in order to build a test with higher power than those for individual genes. There exist many other testing methods for a set of genes, which is often annotated by a term in Gene Ontology (GO) [3], have been proposed for the similar purpose. Unlike the most of existing methods, the main characteristic of MetaGP is that it evaluates the statistical evidences of a set of genes independently from information of the other set of genes. In this regard, we refer it as statistical absolute evaluation. Due to the characteristic, it makes sense to compare the results of tests for the same set of genes but observed in different conditions, e.g. case-control experiments with multi cases, time-course experiments, because the statistical evidences are evaluated with the same standard. The MetaGP method will be explained in detail in the following section.

Although the method was used with GO terms in [6], as they mentioned in the paper, it is applicable for any predefined set of genes like gene set enrichment analysis (GSEA) [13, 17]. To test TFAs, we define a set of genes annotated by a TF as the genes which have the binding site of the TF in the promoter regions. We compile this information from a TFs' data base, TRANSFAC [10]. In the analysis, 306 unique TFs were found for the genes mapped on the microarrays that we used. Applying MetaGP with TFs exhaustively, we can obtain a synoptic map of the activities and expect to achieve deeper knowledge of the gene regulations characterizing the systems of the subject.

As a real data analysis, we apply the method to a microarray data obtained from a set of case-control experiments for mice [16]. In one case, mice were dosed with cuprizone, a well-known neuro toxin, inducing demyelination for nerves in the central nervous system, thus, mimicking demyelinating diseases, e.g. multiple sclerosis (MS). In another case, a drug, Ninjin'yoeito (NYT), was additionally dosed to mice

with the cuprizone during the same period. NYT is a drug of Kampo medicine which is Japanese traditional herbal medicine originating in ancient Chinese traditional medicine. It is the only drug known to remedy demyelination induced by aging or dosage of cuprizone up to the present date [15, 16]. However, little is known regarding the action mechanism since the drug is not a single molecule but a mixture of extracts from medicinal plants. Hence, obtaining the synoptic map of TFAs by MetaGP and comparing the activities across different cases would help to understanding the action mechanism of the drag and the system's behavior. We also apply the method with GO terms for the same purpose but to obtain information regarding the system from another aspect.

The organization of this paper is as follows. In Section 2, we explain the method to test the group of genes, e.g. those which correspond to TFAs or GO terms. In Section 3, we apply the method to the expression data obtained from the case-control experiment, in which demyelination-induced mice were dosed with a herbal medicine leading to demyelination and remyelination, and compared the significance of TFAs and GO terms in the several different experimental conditions. Section 5 is the concluding remarks.

2. Method

2.1. *MetaGP*

In this paper, to evaluate significance of differentially activated TFs from gene-expression data, we employ a testing procedure for a set of genes proposed by [6]. We call the method MetaGP. To this end, we define the gene sets corresponding to TFAs and apply MetaGP with them to data set. As a result, we can obtain p-values for TFAs.

The MetaGP method accumulates statistical evidences from a set of genes in order to build a test with higher power than those for individual genes and to help the biological interpretation with functional annotation terms. The tests for individual genes, e.g. t-test for differentially expressed genes, tend to have low statistical powers, due to the small sample size. For the similar purpose, many other testing methods for a set of genes have been proposed, e.g. GSEA [13, 17], FatiGO [1, 2], and so on. See for a review [9].

Most of those methods use a test based on a 2×2 contingency table for a list of genes e.g. [1]. Each of the four cells in the table is characterized by two properties of genes, that is, (i) annotated by the term or not and (ii) differentially expressed or not. For the table, each of the cells is filled by the number of genes categorized by the two properties. Then the p-value for over-representation of the set of genes annotated by the term is calculated by using an independence test of the 2×2 contingency table, e.g. Fisher's exact test. The categorization of the differential expression of each genes is determined by setting a threshold value for a ranked list of genes, e.g. a list ordered by t-statistic. Therefore the test for the gene set with the annotation depends on the rank information relative to the gene set without the annotation.

On the other hand, the main characteristic of MetaGP is that it evaluates the statistical evidences of a set of genes independently from information of the other set of genes. Due to the characteristic, it makes sense to compare the results of tests for the same set of genes but observed in different conditions, e.g. case-control experiments with multi cases, time-course experiments, since the statistical evidences are evaluated in an absolute manner. On the contrary, it is not appropriate to compare the results of 2×2 tests for the same set of genes observed in different conditions, since the result of each test depends on the rank statistics relative to the complementary set and the rank information can not be compared in an absolute manner across different conditions.
The formulation of MetaGP is explained according to [6] in the following section.

2.2. *Formulation of MetaGP*

Suppose that we test the null hypothesis $H_0^{(i)}$ for the ith gene where the total number of genes is denoted by d $(i = 1, \cdots, d)$. For example, to identify differentially expressed genes between case and control samples, one may apply t-test under the null hypothesis $H_0^{(i)} : \mu_0^{(i)} = \mu_1^{(i)}$ for $i = 1, \cdots, d$. Here $\mu_0^{(i)}$ and $\mu_1^{(i)}$ denote group means of the control and the case samples for the ith gene.
The objective of MetaGP is to retrieve the annotations corresponding to the set of genes which are relevant to the underlying gene regulation mechanism on the basis of the statistical evidences. As remarked in [6], the method can be applicable not only for GOs but also for retrieving generic biological knowledge by defining the set of genes, e.g. TFAs, biological pathways.
Let us denote a set of genes, which are classified in a term of a generic biological knowledge by $\mathcal{F}$. For instance, $\mathcal{F}$ represents a set of genes annotated by a term in GO. In this study, $\mathcal{F}$ denotes a set of genes annotated by a TF's name. The annotation procedure is described in Section 3.3. The problem of evaluating the significance of $\mathcal{F}$ is stated by the statistical test with the null hypothesis H_0 and the alternative H_1 as follows:

$$H_0 : H_0^{(i)} \text{ is true for all } i \in \mathcal{F},$$
$$H_1 : H_0^{(i)} \text{ is false for one or more } i \in \mathcal{F}.$$

For instance, to understand functional gene regulations, one aims to retrieve GO terms in which more genes indicating the false $H_0^{(i)}$ are involved.
In order to evaluate this test, they present a testing procedure which exploits a technique of the statistical meta-analysis, known as the normal inversion method. Let p_i denote the p-values of the ith gene. The testing procedure for a $\mathcal{F}$ is then described as follows:

(1) Transform p_i to the random deviate z_i according to

$$z_i = \Phi^{-1}(1 - p_i), \quad i \in \mathcal{F},$$

where Φ^{-1} stands for the inverse function of the standard normal cumulative distribution function.

(2) Compute z-score by

$$z = \sum_{i \in \mathcal{F}} z_i \Big/ \sqrt{|\mathcal{F}|},$$

where $|\mathcal{F}|$ denotes the number of genes labeled by the term $\mathcal{F}$.

(3) Compute an integrated p-value of the annotation $\mathcal{F}$, denoted by $p_{\mathcal{F}}$, by the reverse transformation of z-score as

$$p_{\mathcal{F}} = 1 - \Phi(z).$$

The basic theory of statistics indicates that each z_i is independently and identically distributed according to the standard normal distribution if and only if the individual null hypothesis $H_0^{(i)}$ is true. Moreover, under the assumption that the null hypotheses for all of the genes in $\mathcal{F}$ are true (H_0 is true), and independent to each other, the integrated z-score also follows the standard normal distribution. Based on these statistical properties, the integrated p-value captures the enrichment of the term in the following way: For a part of genes in $\mathcal{F}$, of which the null hypotheses are true, the p-values are uniformly distributed over $[0, 1]$, while p-values for the other genes, of which the null hypotheses are not true, in $\mathcal{F}$ are clustered around the region close to zero. This property can be formally stated if a set of tests is statistically unbiased. Hence, as the proportion of the genes in $\mathcal{F}$, for which the null hypothesis are not true, becomes larger, the computed z-score is shifted towards a higher value. Correspondingly, the integrated p-value becomes smaller.

3. Data and Analysis

To investigate the efficiency of MetaGP with TFAs, we applied the method to a real data set obtained by case-control experiments with mice [16]. As mentioned in Section 1, the experiment were conducted to see the dose effect of NYT on the gene expressions, which is a herbal drug and known to be the only drug to remedy demyelination of myelin sheath of the central nervous system up to the present date. However, the action mechanism has not been clarified. One of the difficulty comes from that the target of NYT might be multiple, since it consists of extract from a mixture of 12 raw medicinal plants. Thus obtaining a synoptic map of multiple TFs' activities for different dosage and time conditions by MetaGP, we can overview the system's behavior and expect to achieve clues to reveal the action mechanism.

3.1. *Data: Demyelination-/Remyelination-Induced Mice Treated with Kampo Medicine*

The summary of the experiments is as follows. There are one control-experiment and three case-experiments. The experiments will be denoted by E1, E2, E3, and E4, respectively (see Table 1). The experimental period was 7 weeks. Male C57BL/6J

Table 1. The types and conditions of experiments.

Experiment ID	E1	E2	E3	E4
Case/Control	Control	Case	Case	Case
NYT		√		√
Cuprizone			√	√

mice were used. At the beginning of the period, they are 8-week-old.
In the control experiment E1, mice were fed with a normal diet during the 7 weeks. In the case experiment E2, mice were fed with a diet containing 1% NYT during the 7 weeks, which is a spray-dried extract of a mixture of 12 raw medicinal plants. In the case experiment E3, to model the demyelination and remyelination process, mice were fed with a diet containing 0.2% cuprizone, a well-known neuro toxin, for the first 5 week. The demyelination started at the 3rd week and the degree of demyelination had a peak at the 5th week. During the following two weeks, the normal diet were fed. In the following two weeks, the remyelination process occurred. The myelin sheath looks almost recovered at the 7th week.
In the case experiment E4, NYT was additionally dosed with cuprizone to mice during the first 5 weeks. As a result the demyelination was suppressed. During the following two weeks, the dosage of cuprizone was quitted and that of NYT was continued. At the 7th week, the shape of the regenerated myelin sheath were better than those for the above cuprizone-dosed experiment (E3).
To obtain gene expression data, for each experiment (E1, E2, E3, and E4), three mice were sacrificed at the 3rd and the 7th week. White matter of each mouse's brain was homogenized in Trizol reagent with a Phycontron, and total RNA was isolated, labeled, and prepared for hybridization with Mouse Expression 430A 2.0 Array according to the recommended protocol. As the result, three DNA microarray samples were obtained for each experiment and time point. See [16] for more detail about the data and experiments.

3.2. *Annotation to Gene Groups*

3.2.1. *TFs*

We defined a gene group to be annotated with a TFA as a set of genes in which each gene has the binding site of the TF in the promoter region. To select such genes placed on the microarray, we searched a database of TFs, TRANSFAC.
Each searching region spanned 2500 bps from the transcription starting site (TSS) toward the upstream. The confidence level of the sequences to be accepted as binding sites was 0.98. As a result, the number of the gene sets annotated by unique TFAs became 306. We note that it is highly possible not all genes are really regulated by the TFs, even if they have the particular sequences with the confidence level. The accuracy to select genes regulated by the TFs would be improved by incorporating other information from e.g. ChIP-chip experiments.

3.2.2. *GO terms*

As a set of genes to be annotated with a GO term, we used the predefined group of genes for each GO term [3]. In the analysis, we used all three gene ontology, i.e. Biological Process (BP), Cellular Component (CC), Molecular Function (MF). As a result, the number of the gene sets annotated by GO terms became over 20000.

3.3. *MetaGP for TFs and GO terms*

We applied the MetaGP method to the gene sets annotated by TFAs or GO terms with the gene expression data as follows:

(i) To obtain the p-values for individual genes, Welch's t-test was done for the relative difference between μ^{ijk} and μ^{i1k}, where μ^{ijk} is the mean expression value of the ith gene of the jth case experiment i.e. Ej ($j \in \{2, 3, 4\}$), at the kth week ($k \in \{3, 7\}$) and μ^{i1k} is that for the control experiment, i.e. E1. Under the null hypothesis $H_0^{ijk} : \mu^{ijk} = \mu^{i1k}$, the p-value, p^{ijk} was obtained for each gene, case sample, and time.

(ii) The individual p-values corresponding to an annotation $\mathcal{F}$ were integrated by MetaGP. The integrated p-value for the annotation $\mathcal{F}$ of the jth case experiment and the kth week denoted by $p_{\mathcal{F}}^{jk}$ were calculated by integrating p^{ijk}'s, $i \in \mathcal{F}$.

(iii) Finally, we obtained the integrated p-values for all TFs and GO terms in our lists.

4. Discussion

After obtaining the integrated p-values for 306 TFs, in order to see the characteristics of the significance of all TFs, we did a cluster analysis of the p-values as shown with the heatmap in Fig. 1. That is a synoptic map of the TFAs.

As shown in the upper half, a number of TFs are differentially activated for both the cuprizone-dosed (W3E3, W7E3) and the cuprizone-NYT-dosed (W3E4, W7E4) cases, while NYT-dosed (W3E2, W7E2) cases are not. Thus these TFAs might be induced by the dose effect of cuprizone.

In the lower half, we can find more variety of patterns. Investigating these TFs activated for particular combinations and the corresponding gene expressions systematically, we can expect to obtain newer insight about the action mechanism of the medicines and processes of demyelination and remyelination.

Since the experiments intended to find a way to cure demyelination diseases, e.g. MS, it is interesting to compare our result with TFAs reported in existing literature regarding the disease. Gobin *et al.* [5] reported upregulated TFs in MS lesion of patient's brain, i.e. RFX, NF-κB, IRF1, STAT1, USF, CREB, and CIITA, which control major histocompatibility complex (MHC). In the paper, the existence of the TFs were confirmed by immunohistochemical staining.

Fig. 2 shows the integrated p-values of the TFs in our data; except for CIITA which was not included in our list. Note that, in each panel, a hight of bar represents

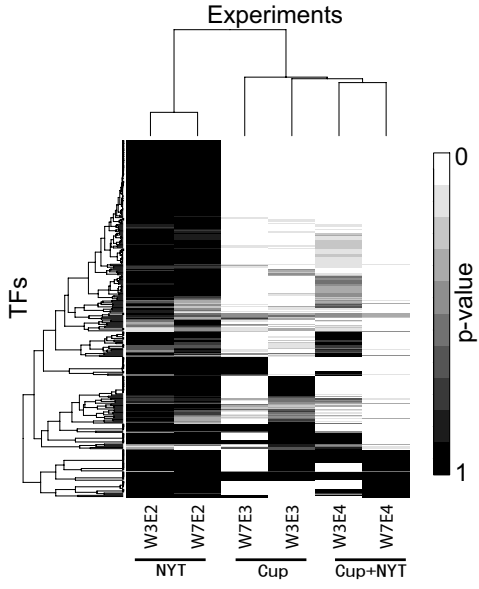

Fig. 1. The clustered profiles of the integrated p-values of the differentially activated TFs, $p^{jk}_{TF_i}$, for each of TFs (rows) and experiments (columns). The number of TFs is 306, i.e. $i = 1, \cdots, 306$, and the names of TFs are not shown. At the bottom, the label "WkEj" suggests the experiment at the kth week by the experiment Ej.

$1 - p^{jk}_{\mathcal{F}}$, thus the higher one corresponds to the smaller $p^{jk}_{\mathcal{F}}$. The mean of the label "WkEj" is the same as in Figure 1. Interestingly, we can see variety of patters in the panels. These observation have helpful information for the system's response to the dosage and the action mechanism of the drugs as discussed below. In the panel (a) RFX, the TF is activated in the experiment E4 (cuprizone + NYT dosage), but not in the experiments E2 (NYT dosage) and E3 (cuprizone dosage). It suggests the existence of interaction effect from both medicines to the pathway of RFX. In the panels (b) NF-κB and (c) IRF-1, we can see the coherent pattern in the experiment E3 and the E4. It may suggest that mainly cuprizone affected on the pathways of these TFs, and thus, NYT could not control the cuprizone's effect. In the panel (d) STAT1, at the former period (the 3rd week), there are no significant activities. At the latter period (the 7th week), the TF is activated in the experiment E3 but not in E4. It suggests the existence of a time effect. In the latter period, NYT seems to repress the effect of cuprizone and to stabilize the system. In the panel (e) USF and (f) CREB, there are also time effect for the effect of NYT. At the former period. NYT was not effective for both TFs. At the latter period, the interaction effect existed for USF. On the other hand, the repressing effect by NYT can be seen for CREB.

We also assessed GO terms by MetaGP. The number of GO terms considered is

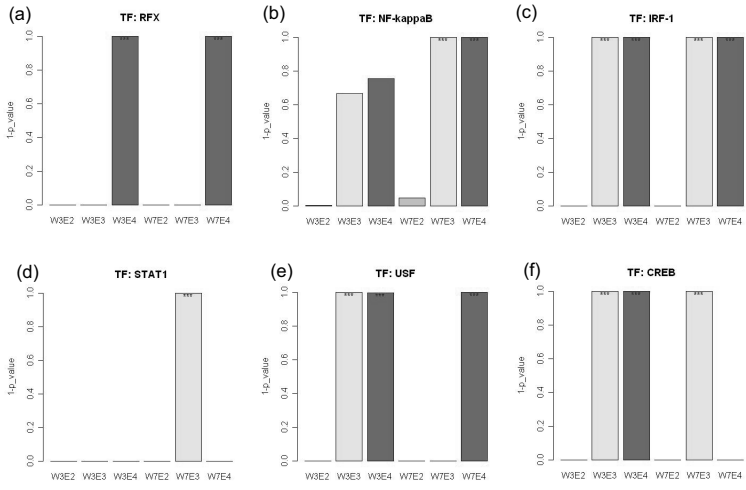

Fig. 2. The integrated p-values of the differentially activated TFs for each experiment relative to the control. The TFs, (a) RFX, (b) NF-κB, (c) IRF1, (d) STAT1, (e) USF, and (f) CREB, are reported as activated TFs in the brain of multiple sclerosis patient [5]. Note: The vertical axis is for $1-p$. The marks "$*$", "$**$", and "$***$" in the bars mean "$p \leq 0.1$", "$p \leq 0.05$", and "$p \leq 0.01$", respectively.

more than 20000. Here we only show two of them, i.e. "myelination" and "structural constituent of myelin sheath" (Fig. 3), because of limitation of space. Since these terms must relate to myelination/remyelination, it becomes a validation of the method. Investigating the results of other terms in more detail, we can acquire information regarding the system's behavior from another point of view.

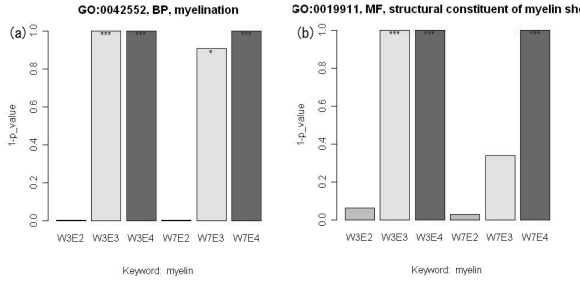

Fig. 3. The integrated p-values of the GO terms including the word "myelin" in its definition. The format is the same as in Fig. 2.

5. Concluding Remarks

In this study, we proposed a method to identify activated TFs, that possibly decide or control cell status, from gene expression data by using the statistical test-based method named MetaGP. MetaGP tests a set of genes with an annotated notion and is an absolute evaluation method [6]. Unlike the rank-based methods, the MetaGP method makes sense to compare the results of the same set of genes obtained in different conditions, e.g. data of case-control experiments with multiple cases, and time course experiments, etc. Applying the method with TFs, we can obtain a synoptic map of the activity of multiple TFs in different conditions.

The main contribution of this study is three folds. i) We applied MetaGP with TFs to a real data. We defined the annotated groups of genes by TFs and applied MetaGP. We showed the results in a synoptic map of TFAs by which we can easily grasp the tendency of those activities in the system. ii) We applied the MetaGP method to the data obtained from case-control experiments including multiple cases and time courses. Because the results of the tests are fairly comparable between different conditions, then we compared the activities and discussed action mechanism of the drugs. The information would be useful for drug discovery and improvement. iii) In our knowledge, this is possibly the first study of Kampo medicine data analyzed with a systems-biological method and view.

Since a drug of Kampo medicine is not a single chemical compound but extracts of multiple medicinal herb, the effector sites are possibly multiple. Thus it is hard to understand the action mechanism and the system's behavior by investigating only few highly expressed individual genes. MetaGP with TFs and other annotations, e.g. GO terms, can give summaries for the systems' behavior from various points of view, and thus offers clues to obtain insights regarding the system in more holistic manner.

References

[1] Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J., A web tool for finding significant associations of gene ontology terms with groups of genesy. *Bioinformatics*, 20(4):578–580, 2004.

[2] Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J., Discovering molecular functions significantly related to phenotypes by combining gene expresesionon data and biological information. *Bioinformatics*, 21(13):2988–2993, 2005.

[3] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Traver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G., Gene ontology: tool for the unification of biology. *Nat. Genet.*, 25(1):25–29, 2000.

[4] Galbraith, S. J., Tran, L. M., and Liao, J. C., Trancriptome nework analysis with limited microarray data. *Bioinformatics*, 22(15):1886–1894, 2006.

[5] Gobin, J. P., Montagne, L., van Zutphen, M., van den Elsen, P. J., van der Valk, P., and de Groot, C. J. A., Upregulation of transcription factors controlling MHC expression in multiple sclerosis lesions. *Glia*, 36(1):68–77, 2001.

[6] Gupta, P. K., Yoshida, R., Imoto, S., Yamaguchi, R., and Miyano, S., Statistical ab-

solute evaluation of gene ontology terms with gene expression data. *LNBI*, 4463:146–157, Springer-Verlag, 2007.

[7] Kao, K. C., Yang, Y. L., Boscolo, R., Sabatti, C., Roychowdhury, V., Liao, J. C., Transcriptome-based determination of multiple transcription regulator activities in *escherichia coli* by using network component analysis. *Proc. Natl. Acad. Sci. USA*, 101:641–646, 2004.

[8] Kao, K. C., Tran, L. M., Liao, J. C., A global regulatory role of gluconeogenic genes in *escherichia coli* revealed by transcriptome network analysis. *J. Biol. Chem.*, 280:36079–36087, 2005.

[9] Khatri, P., Drăghici, S., Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, 21(18):3587–3595, 2005.

[10] Knäppel, R., Dietze, P., Lehnberg, W., Frech, K., and Wingender, E. Transfac retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J. Comput. Biol.*, 1:191–198, 1994.

[11] Li, Z., Shaw, S. M., Yedwabnick, M. J., and Chan, C., Using a state-space model with hidden variables to infer transcription factor activities. *Bioinformatics*, 22(6):747–754, 2006.

[12] Liao, J. C., Boscolo, R., Yang, Y.-L., Tran, L. M., Sabatti, C., and Roychowdhury, V. P., Network component analysis: reconstruction of regulatrory signals in biological systems. *Proc. Natl. Acad. Sci. USA*, 100(26):15522–15527, 2003.

[13] Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.*, 34(3):267–273, 2003.

[14] Nachman, I., Regev, A., and Friedman, N., Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, 20(Suppl. 1):i248–i256, 2004.

[15] Nakahara, J., Tan-Takeuchi, K., Seiwa, C., Gotoh, M., Kaifu, T., Ujike, A., Inui, M., Yagi, T., Ogawa, M., Aiso, S., Takai, T., and Asou, H., Signaling via immunoglobulin Fc receptors induces oligodendrocyte precursor cell differentiation. *Dev. Cell*, 4(6):841–852, 2003.

[16] Seiwa, C., Yamamoto, M., Tanaka, K., Fukutake, M., Ueki, T., Takeda, S., Sakai, R., Ishige, A., Watanabe, K., Akita, M., Yagi, T., Tanaka, K., and Asou, A., Restoration of FcRγ/Fyn signaling repairs central nervous system demyelination. *J. Neurosci. Res.*, 85(5):954–966, 2007.

[17] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102(43):15278–15279, 2005.