MATERIALS AND METHODS

MRS database

In order to perform genome-wide microRNA target prediction, we used the human UniGene (build 187) database (1), which has 54,576 unique sequences for putative 3’ UTR database construction. The unique sequence was selected from each UniGene cluster with the longest region of high-quality sequence data. There were 16,029 unique sequences with coding sequences. We selected 10,696 unique sequences with known 3’ UTR whose length was > 500 bases. The last 500 bases in the 3’ end of the unique sequence were used for the putative 3’ UTR, if the length of the unique sequence was > 500 and it did not have a known 3’ UTR whose length was > 500 bases. If the full unique sequence was less than 500 bases, the whole sequence was used for putative 3’ UTR.

Mature human microRNA sequences were obtained from the miRBase database, release 7.1 (2,3), which includes 319 human mature microRNA sequences. We used the miRanda algorithm (4) to scan all of the putative 3’ UTR sequences for all available mature microRNA sequences. The system parameters of this program referred to the study of John et al. (5). This algorithm used dynamic programming to search for maximal local alignments. A score of +5 was assigned for G:C and A:T pairs, +2 for G:U wobble pairs, and -3 for mismatch pairs, and the gap-open and gap-elongation parameters were set to -8 and -2, respectively. In addition, the scores at the first eleven positions, counting from the microRNA 5’ end, were multiplied by a scaling factor of 2.0; for example, G:C and A:T base pairs contributed +10 to the match score in these positions. The thresholds for candidate target sites had alignment scores > 90 and free energy < -17 kcal/mol, where the alignment score was the sum of the single-residue-pair match score over the alignment trace, and the free energy of duplex formation from a completely dissociated state, calculated by using the miRanda algorithm (4). Then, we applied the miRanda program to scan for all putative 3’ UTR sequences to obtain 3,096,159 potential microRNA regulatory interactions. In order to decrease the false discovery rate, the model of rank filtration in the study of Rhodes et al. (6) was used. We filtered the microRNA targets to produce only the top 2,000 targets per mature microRNA by the alignment score.
**TRS database**

The TRS database was derived from the integration of the human putative 2100-bases promoter database (PromoSer) (7) and the TF database, TRANSFAC 9.3 (8). There are 33,962 promoter sequences that map to 23,095 genes of UniGene. The 557 distinct vertebrate group matrices were used in TRANSFAC. These matrices were nucleotide distribution matrices of aligned binding sequences. Then, we applied the Match program (9) to scan for all promoter sequences with a matrix similarity threshold of 0.85 and a core similarity threshold of 0.9 to obtain 81,286,307 potential regulatory interactions. In order to decrease the false discovery rate, the model of rank filtration in the study of Rhodes et al. (6) was used. We filtered the binding sites to produce only the top 2,000 binding sites per matrix by the similarity score. There were 63 matrices with greater than 2,000 perfect matches (score = 1.0), which were removed, with 494 matrices remaining.

**Target score (T-Score)**

The enrichment analysis used the binomial distribution probabilities by the gene number in the intersections of the two groups. However, the alignment score of putative microRNA target and the similarity score of the TF binding site were not considered in enrichment analysis. We defined the target score, T-Score, as follows:

\[
T - \text{Score} = \frac{\sum_i S_i}{\sum_i R_i}
\]

Where \(S_i\) is the score of the intersection target (alignment score for microRNA or similarity score for TF), \(R_i\) is each score of the target in a regulatory signature, and \(\sum R_i\) is the sum of all scores in a regulatory signature. The T-Score can evaluate the significance of the intersection targets, and the higher the T-Score, the more significant.
RESULTS

Examples for the usages of the server

For example, users can perform a sequence of queries as follows: (i) submit microarray raw data to the data pretreatment component; (ii) perform quantile normalization, data adjustment, and data filtration in the microarray data pretreatment component; (iii) go to the microarray data statistic and clustering analysis component (the selected microarray data from step 2 will be automatically transferred here); (iv) perform signal-to-noise with permutation test to obtain GESs; (v) go to the enrichment analysis component to determine significant enrichment pathways; (vi) select a significant enrichment pathway to run motif discovery (the intersected genes of both the significant pathway and the GES from step 4 will be automatically transferred to the motif discovery component); (vii) discovery of the consensus motifs in the putative promoter and 3’ UTR sequences.

Supplementary Figure S1-A shows an example of microarray data pretreatment including quantile normalization, data adjustment (clip value to minimum=0 and maximum=70,000), data filtration (keep genes with a two-fold change), and standard normalization (mean=0; standard deviation=1). To demonstrate the statistical function of CRSD in microarray analysis, the student's t-test for the group-specific marker gene selection and further analyses are used as examples. The web interface of the student's t-test is illustrated in Supplementary Figure S1-B. Users can designate the group number of samples and select the $P$ value and $Q$ value thresholds to exclude non-significant genes. The screenshot in Supplementary Figure S1-C shows a part of the student's t-test result that contains the marker genes of group-1 in the tested groups and the hyperlink of the further enrichment analysis and motif discovery for these marker genes.
DISCUSSION

The predicted TFs having significant enrichment with the pathway

Previous studies also have supported our predicted TFs having significant enrichment with the pathway (Supplementary Table S3). The $P$ value of each TRS was less than 0.05 (minimum is 1.36E-9 and maximum is 1.93E-2). Herein, we described some well-known examples with significant biological evidence, which is consistent with our prediction in the following: The NF-kappaB can regulate the transcription of IL8 through Toll-like receptor signaling pathway (10,11), and the most significant TF in this pathway is NF-kappaB ($P < 0.001$), which was found in our prediction (Supplementary Table S1). An important hypoxia responsive region of the VEGF promoter recruits the HIF-1 for up-regulating VEGF transcription which also relates to the actions of Nitric Oxide pathway (12,13). Interestingly, our results show that HIF-1 regulates VEGF transcription with significant enrichment ($P = 0.01$) in this pathway. In antigen dependent B cell activation pathway, the immunomodulatory effect involves IL4 activated by GATA-3 in T helper lymphocytes (14,15). We also predicted that GATA-3 can regulate IL4 with significant enrichment ($P < 0.001$) in this pathway. On the basis of previous literature reports, our system may provide a powerful and useful bioinformatic platform, helping researchers discover new regulatory behaviors of TFs and microRNAs.
**Supplementary Table S3.** List of the previous literature reports supporting predicted TFs having significant enrichment with the pathway

| TF^ * | Pathway | Target gene | P value | Q value† | T-score‡ | Reference |
|-------|---------|-------------|---------|----------|----------|-----------|
| NF-kappaB (p65) | Toll-like receptor signaling pathway | IL8 (Hs.624) | 1.36E-09 | 6.51E-07 | 1.10E-02 | (10,11) |
| AP-1 | Inhibition of Matrix Metalloproteinases | MMP2 (Hs.513617) | 1.56E-04 | 1.11E-02 | 2.50E-03 | (16) |
| E2F | Sonic Hedgehog (SHH) Receptor Ptc1 Regulates cell cycle | CDC2 (Hs.334562) | 3.08E-04 | 1.44E-02 | 3.05E-03 | (17) |
| GATA-1 | NO2-dependent IL 12 Pathway in NK cells | CCR5 (Hs.450802) | 1.93E-02 | 1.57E-01 | 1.49E-03 | (18) |
| HIF-1 | Actions of Nitric Oxide in the Heart | VEGF (Hs.73793) | 1.47E-02 | 1.81E-01 | 2.01E-03 | (12,13) |
| ETS | Hematopoietic cell lineage | GP9 (Hs.1144) | 4.19E-05 | 2.84E-03 | 6.51E-03 | (19) |
| Sp3 | NFAT and Hypertrophy of the heart (Transcription in the broken heart) | HAND1 (Hs.152531) | 6.14E-05 | 4.73E-03 | 5.00E-03 | (20) |
| GATA-3 | Antigen Dependent B Cell Activation | IL4 (Hs.73917) | 5.59E-04 | 2.47E-02 | 2.01E-03 | (14,15) |
| Oct-1 | p53 Signaling Pathway | GADD45A (Hs.80409) | 1.30E-02 | 1.63E-01 | 1.54E-03 | (21) |
| AP-2 | Urea cycle and metabolism of amino groups | CKB (Hs.173724) | 7.33E-04 | 2.16E-02 | 3.50E-03 | (22) |

^ TF, Transcription factor. † Q value (false discovery rate estimation) is obtained by adjusting P value for multiple hypothesis test. ‡ T-Score is calculated by using the alignment scores for microRNA or similarity scores for TF.
Supplementary Figure S1. The screenshots of the microarray data pretreatment and statistic analysis. (A) Users can perform quantile normalization, rescale normalization, data adjustment, data filtration, and standard normalization, which are implemented in the microarray data pretreatment component within the integrated environment. (B) The P value and Q value thresholds can be specified, and group membership can be assigned in the student's t-test component to discover group-specific marker genes. (C) The P value and Q value will be calculated by the student's t-test and multiple hypothesis test, respectively. The screenshot shows the group-specific marker genes and the hyperlink of the further enrichment analysis and motif discovery for these genes.
Supplementary Figure S2. A group of genes ($G_A$) was assessed for the significant enrichment of another group of genes ($G_B$). $G_A$ or $G_B$ could be the genes that belong to a GO annotation, pathway, MRS, TRS, GES, or user input data. The possible set $P_A$ is defined as the set of all of the possible genes of $G_A$. $E$ is the intersection between $G_A$ and $G_B$. 
REFERENCES

1. Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A. et al. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, 31, 28-33.

2. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, 34, D140-144.

3. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res.*, 32, D109-111.

4. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, 5, R1.

5. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, 2, e363.

6. Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T.R., Ghosh, D. and Chinnaiyan, A.M. (2005) Mining for regulatory programs in the cancer transcriptome. *Nat. Genet.*, 37, 579-583.

7. Halees, A.S. and Weng, Z. (2004) PromoSer: improvements to the algorithm, visualization and accessibility. *Nucleic Acids Res.*, 32, W191-194.

8. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, 31, 374-378.

9. Wingender, E. (2004) TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. *In Silico Biol.*, 4, 55-61.

10. Hoffmann, E., Thiefes, A., Buhrow, D., Dittrich-Breiholz, O., Schneider, H., Resch, K. and Kracht, M. (2005) MEK1-dependent delayed expression of Fos-related antigen-1 counteracts c-Fos and p65 NF-kappaB-mediated interleukin-8 transcription in response to cytokines or growth factors. *J. Biol. Chem.*, 280, 9706-9718.

11. Akira, S. (2003) Toll-like receptor signaling. *J. Biol. Chem.*, 278, 38105-38108.

12. Fukumura, D., Gohongi, T., Kadambi, A., Izumi, Y., Ang, J., Yun, C.O., Buerk, D.G., Huang, P.L. and Jain, R.K. (2001) Predominant role of endothelial nitric oxide synthase in vascular endothelial growth factor-induced angiogenesis and vascular permeability. *Proc. Natl. Acad. Sci. USA*, 98, 2604-2609.

13. Coles, L.S., Diamond, P., Lambrusco, L., Hunter, J., Burrows, J., Vadas, M.A. and Goodall, G.J. (2002) A novel mechanism of repression of the vascular...
endothelial growth factor promoter, by single strand DNA binding cold shock domain (Y-box) proteins in normoxic fibroblasts. *Nucleic Acids Res.*, **30**, 4845-4854.

14. Srivastava, K., Teper, A.A., Zhang, T.F., Li, S., Walsh, M.J., Huang, C.K., Kattan, M., Schofield, B.H., Sampson, H.A. and Li, X.M. (2004) Immunomodulatory effect of the antiasthma Chinese herbal formula MSSM-002 on TH2 cells. *J. Allergy Clin. Immunol.*, **113**, 268-276.

15. Tykocinski, L.O., Hajkova, P., Chang, H.D., Stamm, T., Sozeri, O., Lohning, M., Hu-Li, J., Niesner, U., Kreher, S., Friedrich, B. *et al.* (2005) A critical control element for interleukin-4 memory expression in T helper lymphocytes. *J. Biol. Chem.*, **280**, 28177-28185.

16. Bergman, M.R., Cheng, S., Honbo, N., Piacentini, L., Karliner, J.S. and Lovett, D.H. (2003) A functional activating protein 1 (AP-1) site regulates matrix metalloproteinase 2 (MMP-2) transcription by cardiac cells through interactions with JunB-Fra1 and JunB-FosB heterodimers. *Biochem. J.*, **369**, 485-496.

17. Dalton, S. (1992) Cell cycle regulation of the human cdc2 gene. *EMBO J.*, **11**, 1797-1804.

18. Moriuchi, M., Moriuchi, H. and Fauci, A.S. (1999) GATA-1 transcription factor transactivates the promoter for CCR5, a coreceptor for human immunodeficiency virus type 1 entry. *Blood*, **93**, 1433-1435.

19. Bastian, L.S., Kwiatkowski, B.A., Breininger, J., Danner, S. and Roth, G. (1999) Regulation of the megakaryocytic glycoprotein IX promoter by the oncogenic Ets transcription factor Fli-1. *Blood*, **93**, 2637-2644.

20. Vasichek, R., Meinhardt, G., Haidweger, E., Rotheneder, H., Husslein, P. and Knoffler, M. (2003) Expression of the human Hand1 gene in trophoblastic cells is transcriptionally regulated by activating and repressing specificity protein (Sp)-elements. *Gene*, **302**, 115-127.

21. Fan, W., Jin, S., Tong, T., Zhao, H., Fan, F., Antinore, M.J., Rajasekaran, B., Wu, M. and Zhan, Q. (2002) BRCA1 regulates GADD45 through its interactions with the OCT-1 and CAAT motifs. *J. Biol. Chem.*, **277**, 8061-8067.

22. Willis, D., Zhang, Y. and Molloy, G.R. (2005) Transcription of brain creatine kinase in U87-MG glioblastoma is modulated by factor AP2. *Biochim. Biophys. Acta*, **1728**, 18-33.