Send Orders for Reprints to reprints@benthamscience.net

Current Bioinformatics, 2021, *16*, 1-13

RESEARCH ARTICLE

GEREA: Prediction of Gene Expression Regulators From Transcriptome Profiling Data to Transition Networks

Min Yao[1], Caiyun Jiang[1], Chenglong Li[1], Yongxia Li[1], Shan Jiang[1], Liang He[1], Hong Xiao[1], Jima Quan[1], Xiali Huang, and Tinghua Huang[1,*]

[1]*College of Animal Science, Yangtze University, Jingzhou, Hubei 434025, China*

Abstract: ***Background*:** Mammalian genes are regulated at the transcriptional and post-transcriptional levels. These mechanisms may involve the direct promotion or inhibition of transcription *via* a regulator or post-transcriptional regulation through factors such as micro (mi)RNAs.

***Objective*:** Construct gene regulation relationships modulated by causality inference-based miRNA-(transition factor)-(target gene) networks and analysis gene expression data to identify gene expression regulators.

***Methods*:** Mouse gene expression regulation relationships were manually curated from literature using a text mining method which were then employed to generate miRNA-(transition factor)-(target gene) networks. An algorithm was then introduced to identify gene expression regulators from transcriptome profiling data by applying enrichment analysis to these networks.

***Results*:** A total of 22,271 mouse gene expression regulation relationships were curated for 4,018 genes and 242 miRNAs. GEREA software was developed to perform the integrated analyses. We applied the algorithm to transcriptome data for synthetic miR-155 oligo-treated mouse CD4$^+$ T-cells and confirmed that miR-155 is an important network regulator. The software was also tested on publicly available transcriptional profiling data for Salmonella infection, resulting in the identification of miR-125b as an important regulator.

***Conclusion*:** The causality inference-based miRNA-(transition factor)-(target gene) networks serve as a novel resource for gene expression regulation research, and GEREA is an effective and useful adjunct to the currently available methods. The regulatory networks and the algorithm implemented in the GEREA software package are available under a free academic license at http://www.thua45.cn/gerea.

ARTICLE HISTORY

Received: September 04, 2020
Revised: January 13, 2021
Accepted: January 23, 2021

DOI:

Keywords: GEREA, causality inference, gene expression regulation, transition networks, ~~mammalian genes, elongation.~~

1. INTRODUCTION

The entire process by which a protein is translated from a gene's DNA sequence is carefully controlled from the initiation of transcription, elongation, and termination to post-translational modification. These regulatory mechanisms determine the final quantity of the gene product [1]. The primary regulation of gene expression may occur at the mRNA level *via* activation or inhibition by transcription factors. These are proteins that positively or negatively coordinate with the gene transcription process by interacting with specific DNA recognition motifs located in gene promoter regions [2, 3]. Alternatively, many regulator genes interact with transcription factors to directly regulate their transcriptional activity [4]. An abundance of data has been published on the relationships between gene expression regulators and their target genes. These data may be extracted by mining the literature and performing subsequent analysis on the regulator-(target genes) sub-networks [4].

High-throughput gene expression techniques, such as microarray and mRNA sequencing, are widely adopted to elucidate the roles of genes involved in various biological conditions of interest [5, 6]. Identifying the factors responsible for regulating the expression of differentially expressed genes (DEGs) is necessary to fully elucidate the functions of these genes in specific biological contexts.

Transcription factors and micro (mi)RNAs are vital regulatory molecules functionally associated with a myriad of biological processes, including growth and development, as well as the development and pathogenesis of various diseases [7]. Transcription factors are essential to control elements

*Address correspondence to these authors at the Department of Animal Science, Yangtze University, Jingzhou, Hubei 434025, China; Tel/Fax: +86-13397215027; E-mails: thua45@yangtzeu.edu.cn and Department of Animal Science, Yangtze University, Jingzhou, Hubei 434025, China; Tel/Fax: +86-13617252063; E-mails: yuanbao5888@sina.com

1574-8936/21 $65.00+.00

required for the regulation of gene expression in specific cells and responses to particular signals [1]. MiRNAs are endogenous, short RNA molecules (19–24 nucleotides) that serve as sequence-specific, post-transcriptional regulators of protein-encoding genes [8]. Modules of miRNAs regulating miRNA target genes may be identified by pairing miRNA sequences to the 3′-UTR of the mRNAs of protein-encoding genes [9, 10]. Several bioinformatics tools have been developed to predict miRNA targets based on evolutionary seed region conservation or miRNA-to-mRNA binding energy [11-14]. For instance, databases, such as miRecords [15] and TarBase [16], compile experimentally validated miRNA targets. Meanwhile, StarBase [17] maps miRNA-to-mRNA interactions from argonaute CLIP-Seq and Degradome-Seq data.

Many miRNAs regulate target gene expression by degrading mRNA or inhibiting translation of the gene [8, 18]. Target gene expression is either positively or negatively correlated with the regulator miRNA expression level [19]. To identify the function of mRNAs involved in DEGs identified by transcriptome profiling, several bioinformatics tools were recently developed, such as DAVID [20] and GSEA [21]. Furthermore, to identify key miRNAs involved in the regulation of gene expression during biological events, several methods were recently developed, such as miReduce [22], Sylamer [23], DIANA-mirExTra [23, 24], Sigterms [25], CORNA [26], and FAME [27]. All of these platforms use sets of empirically determined, differentially expressed, protein-encoding genes and perform enrichment analyses to establish whether the DEGs are enriched for the targets of particular transcription factors or miRNAs according to sequence-based, target-predicting algorithms. However, it remains to be determined whether: (1) the regulatory effect (positive or negative interactions) between the regulators and targets should be included in the regulation network; (2) certain miRNAs inhibit miRNA target gene expression post-transcriptionally and, therefore, do not alter the mRNA level, and these target genes cannot be directly detected by transcriptome profiling. If these miRNAs regulate expression of transcription factors, regardless of the regulatory mechanism employed, then they too will downregulate the transcription factors at the protein level, ultimately altering the mRNA levels of the target genes of these transcription factors. Such changes may be detected by transcriptome profiling and analyzed by miRNA-(transition factor)-(target gene) networks. In this study,we use "transition factor" for two reasons: 1) to distinguish from the up-stream regulator of the target gene, the miRNAs; and 2) to indicate that "transition factors" are not only transcription factors.

Although considerable progress has been made, the construction of miRNA-(transition factor)-(target gene) networks for the investigation of relationships between the expression and regulation of human and mouse genes is still time-consuming. Moreover, effective tools for mining such data from literature are lacking, with the currently accepted method relying on manual extraction. There are, however, several databases available for the deposition of information regarding gene expression regulatory relationships for transcription factors, including HIRIdb [28], TRRUST [29], TFactS [30, 31], and GereDB [4]. To date, a total of 51,871 (human), 6,490 (mouse), 2,146 (mouse), and 39,000 (human) linkages have been deposited in HIRIdb, TRRUST, TFactS, and GereDB, respectively. Over 95% of regulatory gene expression relationships (49,762 out of 51,871) in HIRIdb were derived from high-throughput assays, such as deep sequencing or microarray. Meanwhile, TRRUST [29] was generated using text mining technology; however, it contains significantly fewer regulatory gene expression relationships than GereDB. TFactS [30, 31] combines data for regulatory gene expression relationships from multiple sources; however, it has the fewest interactions available compared to the other databases. Although each of these databases was constructed by applying different criteria, providing each with unique features, the overlap in data between them were under 20%, indicating that they are individually incomplete and require further development [4].

In this study, we report a further development of GereDB, curating new regulatory gene expression relationships from the abstracts of relevant studies based on mouse models. We then apply this data for the development of a new algorithm capable of identifying regulatory genes. Furthermore, we apply a network-based enrichment analysis to identify functional miRNAs and transcription factors that orchestrate a particular transcriptional profile using miR-155-treated $CD4^+$ transcriptome profile data, as well as publicly available *Salmonella* infection transcriptome data as input. This approach can assist investigators in discovering relationships between gene expression and regulation that may lead to new hypotheses and could be an effective adjunct to currently available methods.

2. MATERIALS AND METHODS

2.1. Building microRNA-(transition-factor)-(target gene) regulatory networks

2.1.1. Training dataset preparation, word feature extraction, and model training and testing

Abstracts were retrieved using the NCBI Entrez Programming Utilities search engine (E-utilities, Esearch, and Efetch) and the query formula, "(gene regulation[MeSH Terms]) AND (mouse[MeSH Terms])" to filter related articles [32]. A total of 222,245 abstracts were obtained from the Medline 2017 database [33]. Of these, 94,102 candidate sentences were extracted manually that described gene expression regulation relationships, which were then used to establish standard sets of positive sentences. Other irrelevant sentences in these abstracts were used to establish standard sets of negative sentences (1,125,548). A training dataset and a testing dataset, consisting of 5,000 positive and 50,000 negative sentences, were created by random procedure (the ratio was set to 1:10 according to the number of positive and negative sentences in the standard sets). The datasets were available on the GEREA website (http://www.thua45.cn/gerea/data_for_manuscript.zip). The word feature selection and model training were based on: 1) Support Vector Machine (SVM) with a bag of words feature; 2) Multinomial Naive Bayes (MNB) with a bag of words feature; 3) MNB with TF-IDF (term frequency-inverse document frequency) feature; 4) SVM with TF-IDF feature; 5) SVM with averaged word vector feature (word2vec); 6) SVM with TF-IDF-weighted averaged word vector feature (word2vec), and 7)

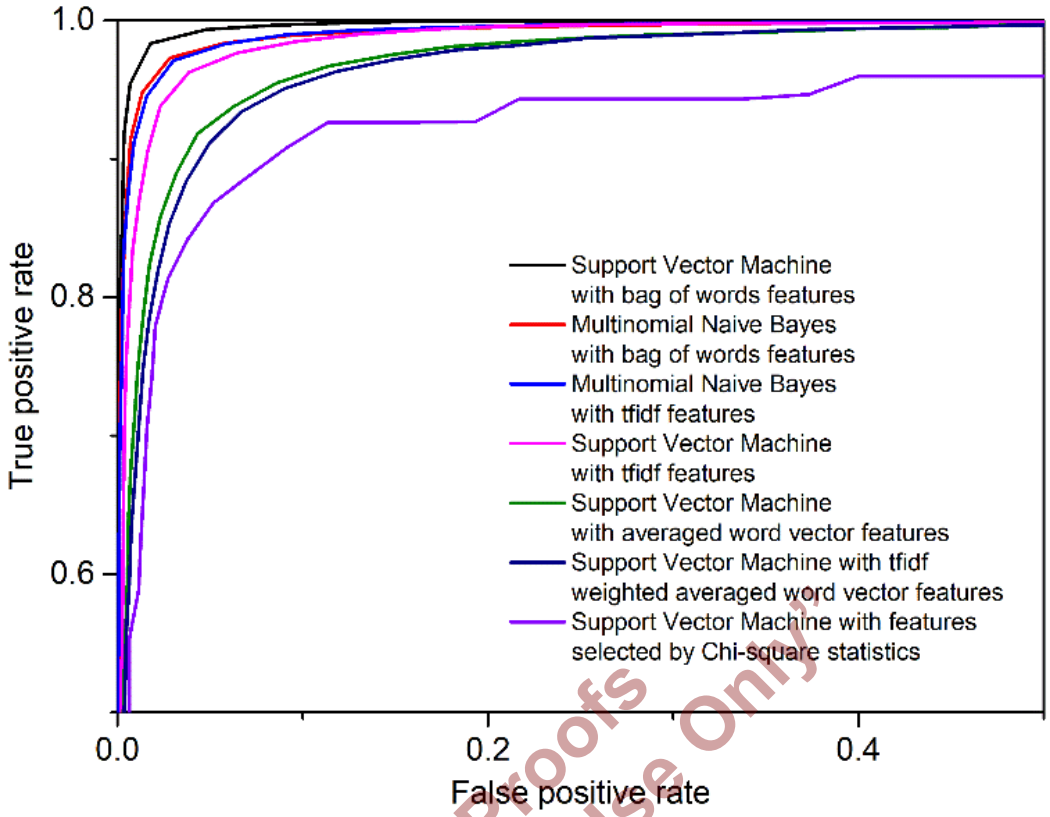

Fig. (1). Receiver operating characteristics (ROC) curve for seven-word feature selection and model training methods. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

SVM with word features selected by chi-square statistics. The sentences in the training datasets were converted to numeric attributes collectively named score datasets that were then randomly split into ten subsets for training and cross-validation. Each subset included 500 positive and 5,000 negative samples. Model training (Scikit-learn GDClassifier, loss='hinge', n_iter=100, cross-validation=10) [34] had an accuracy range from 92% (SVM with word features selected by chi-square statistics) to 95% (SVM with a bag of words features). A receiver operating characteristics (ROC) curve analysis of the model showed that the performance of the SVM with a bag of words feature was optimal, followed closely by MNB with a bag of words feature, MNB with TF-IDF feature, SVM with TF-IDF feature, SVM with averaged word vector feature (word2vec), SVM with TF-IDF-weighted averaged word vector feature (word2vec), and SVM with word features selected by chi-square statistics (Fig. **1**). Accuracy, precision, recall, and F1 scores are available in Supplemental Document 1. The testing dataset was used to assess the best model (constructed with the bag of words feature) and the results indicated that the sensitivity was 93% and specificity was 91%.

2.1.2. Manual Curation of Gene Expression Regulation Relationships

The SVM model that was constructed using the bag of words features (the best for the seven classifiers) was used to prioritize ~30,000 abstracts retrieved from Medline 2018 and before. E-utilities was run with the query formulation "mouse[MeSH Terms]" to get all mouse biological literatures [33]. A total of 29,385 sentences passed the SVM threshold of > 0.5 (classified as positive sentences containing gene expression regulation information). These sentences were then subjected to manual curation by biological undergraduates who extracted the gene symbols and established gene expression regulation relationships, and double-checked by Ph.D. scientists. The same data structure was applied to store the (gene expression regulator)-target links, as previously reported [4]. A "gene expression regulator" can directly or indirectly alter target gene expression. The regulator may be a transcription factor or other protein. Each link consisted of a gene expression regulator, a target gene, and a line connecting them representing the positive or negative effect of the regulator on the target. Finally, a total of 22,271 gene expression regulation relationships were created and (transition factor)-(target gene) links were created from 29,310 abstracts (available at the GEREA website, http://www.thua45.cn/gerea/data_for_manuscript.zip). The linkages were then annotated according to the gene information and evidence, including the descriptive sentences in literature. For the miRNA-transition factor regulatory relationships, numerous state-of-the-art miRNA target prediction software programs were available. A subset of miRTarBase targets with strong experimentally validated evidence [35] was directly applied to construct the miRNA-(transition factor) linkages; all of them were negative regulations (available in supplemental document 2). The miRNA-(transition factor) linkages and the (transition factor)-(target gene) linkages, sharing the same transition factor, were merged to generate the miRNA-(transition factor)-(target gene) networks.

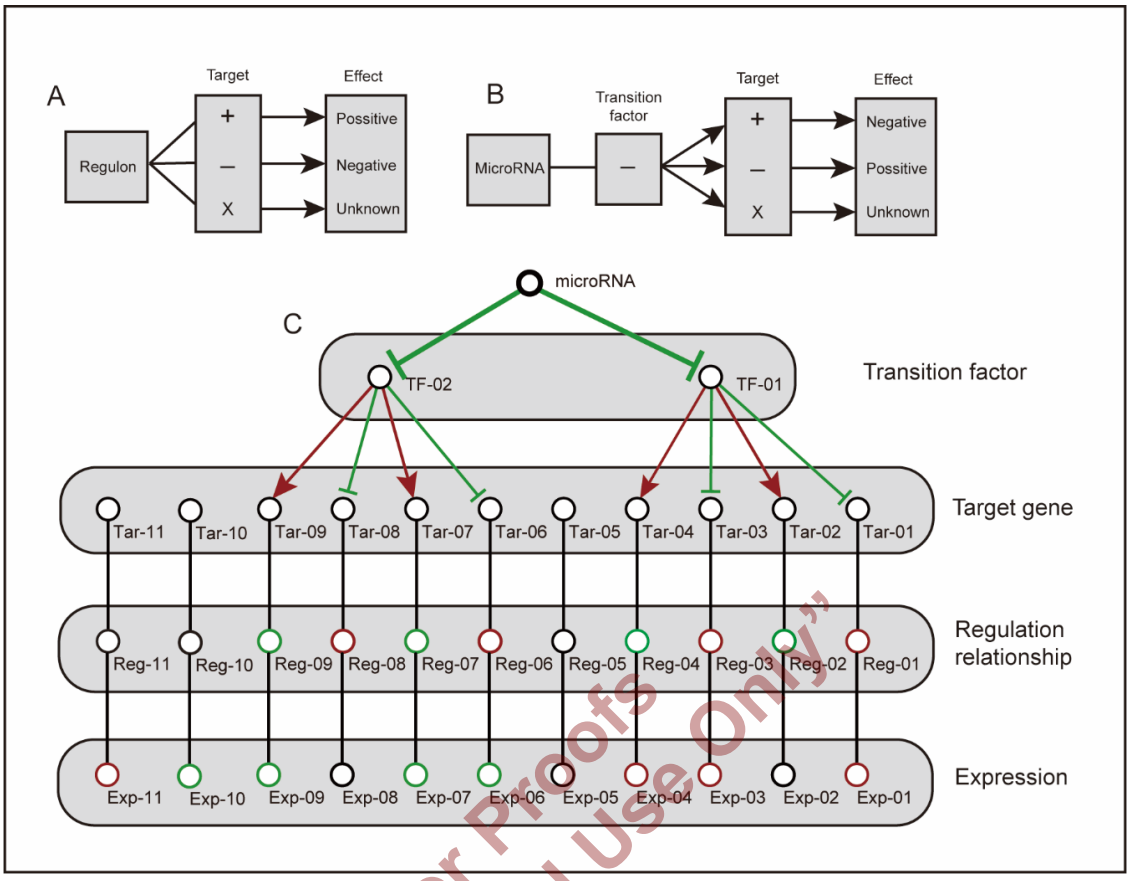

Fig. (2). Regulation relationship inference machine. **A**. Regulation relationship between a target gene and its regulator. **B**. Regulation relationship between a target gene and its regulator via an intermediate transition factor. **C**. Example of networks comprising miRNA, two transitions (transcription) factors (TF-01 and TF-02), and 12 target genes (Tar-01 to Tar-11). Expression directions of the 12 targets (Exp-01 to Exp-11) are represented as red circles (up-regulated), green circles (down-regulated), and black circles (unchanged). Relationship between the miRNA and target gene are represented as red circles (positive), green circles (negative), and black circles (unknown). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The network file is available on the GEREA website (http://www.thua45.cn/gerea/data_for_manuscript.zip).

2.2. Network-based Enrichment Analysis

2.2.1. Regulation Relationship Inference Machine

If a target gene can be controlled by a regulator gene, the relationship between the target and regulator may be positive (+), negative (–), or unknown (×) (Fig. **2A**). When a regulator up-regulates a target gene, the relationship between them is "positive,"; whereas when a regulator downregulates a target gene, the relationship between them is "negative." In certain cases, the relationship between regulator and target is not defined, resulting in an "unknown" relationship. A target gene may also be regulated by a transition factor, such as a transcription factor in a positive (+), negative (-), or unknown (×) manner. Moreover, the transition factor may be inhibited (-) by a miRNA. Hence, through the intermediate transition factor, the relationship between the miRNA and target could be: 1) (-)–(+)–negative, 2) (-)–(-)–positive, or 3) (-)–(×)–unknown (Fig. **2B**). A total of 22,271 manually curated mouse gene expression regulation relationships for 4,018 genes and 242 miRNAs created in section 2.1.2 were used to populate the miRNA-(transition factor)-(target gene) networks. The network is downloadable from the GEREA website along with the GEREA software package.

2.2.2. Enrichment Analysis by Fisher's Exact Test

Significantly over-represented regulatory networks were defined as network patterns consisting of target genes that occur with significantly higher frequency in the DEG lists. After the miRNA-(transition factor)-(target gene) regulation networks were constructed, the gene expression profiling data ~~were~~was loaded ~~with~~to the targets of transcription factors (Fig. **2C** expression row). When a total of N (N = 11 in Fig. **2C**) genes were measured in the transcription profiling ex-periment, K (K = 8 in Fig. 2C, red and green circles) genes were differentially expressed, and O (O = N - K = 3 in Fig. **2C**, black circles) genes were not differentially expressed. For the K genes, U genes were up-regulated (U = 4 in Fig. **2C**, red circles), and W genes were down-regulated (W = 4 in Fig. (**2C**), green circles). For the U genes, A genes were positively regulated by their miRNAs via transition factors TF-01 (A = 2, Reg-03 and Reg-01 in Fig. **2C**), B genes were negatively regulated by their miRNAs via transition factors TF-01 (B = 1, Reg-04 in Fig. 2C), and C represents an un-known relationship between them and the miRNAs and transition factors (C = 1, Reg-11 in Fig. **2C**). For the W genes, D

A The three by three contingency table for overall test

Relationship / Expression	Positive	Negative	Unknown	Total
Diff (Up)	A	B	C	A+B+C
Diff (Down)	D	E	F	D+E+F
No diff	G	H	I	G+H+I
Total	A+D+G	B+E+H	C+F+I	N

B The three by four contingency table (full information)

Expression	Relationship	Positive	Negative	Unknown	Total
Diff	Up	A	B	C	A+B+C
	Down	D	E	F	D+E+F
No diff	Up	J	K	L	J+K+L
	Down	M	Q	R	M+Q+R
Total		A+D+J+M	B+E+K+Q	C+F+L+R	N

C The two by two contingency table for up-driven test

Expression / Casualty	Upregulation	Other	Total
Diff	A+E	B+C+D+F	A+B+C+D+E+F
No diff	J+Q	K+L+M+R	J+K+L+M+Q+R
Total	A+E+J+Q	B+D+C+F+K+L+M+R	N

D The two by two contingency table for down-driven test

Expression / Casualty	Downregulation	Other	Total
Diff	B+D	A+C+E+F	A+B+C+D+E+F
No diff	K+M	J+L+Q+R	J+K+L+M+Q+R
Total	B+D+K+M	A+C+E+F+J+L+Q+R	N

Fig. (3). Contingency tables deduced from the inference machine. **A**: contingency table for P0. **B**: All information deduced from the inference machine. **C** and **D**: breakdown contingency tables (from B) for the up-regulation and down-regulation tests, respectively.

genes were positively regulated by their miRNAs via transition factors TF-02 (D = 1, Reg-06 in Fig. **2C**), E genes were negatively regulated by their miRNAs via transition factors TF-02 (E = 2, Reg-07 and Reg-09 in Fig. **2C**), and F represents an unknown relationship between them and the miRNAs and transition factors (F = 1, Reg-10 in Fig. **2C**). For the O genes, G genes were positively regulated by their miRNAs via transition factors TF-02 (G = 1, Reg-08 in Fig. **2C**), H genes were negatively regulated by their miRNAs via transition factors TF-01 (H = 1, Reg-02 in Fig. **2C**), and I represents an unknown relationship between their miRNAs and transition factors (I = 1, Reg-05 in Fig. **2C**). If the targets of TF-01 and TF-02 are overlapped, they are counted only once. The values for A–I were used to build a 3 × 3 contingency table (Fig. **3A**).

Genes not differentially expressed also exhibited fold changes in known directions. Suppose that J + M genes were not differentially expressed and were positively controlled by the regulator (J was up-regulated and M was down-regulated); K + Q genes not differentially expressed, were negatively controlled by the regulator (K was up-regulated and Q was down-regulated); and L + R genes not differentially expressed, had unknown relationships to their regulator (L was up-regulated and R was down-regulated). The values for A–R could be used to construct a 3 × 4 contingency table (full information, Fig. **3B**) by which the P0 was calculated and from which up-regulation/down-regulation analyses were derived. In the DEG list, the expression changes of the A + E genes may be caused by up-regulation of the regulator, while those of the B + D genes could be caused by down-regulation of the regulator, accounting for the casualty inference of the analysis. A 2 × 2 contingency table was constructed for the Fig. **3B** data (Figs. **3C** and **3D**), by which the P1 and P2 were calculated (P1 for up-regulation and P2 for down-regulation analysis). The contingency tables were analyzed using a statistical method developed by Mehta et al. [36]. The computer code for Mehta and Patel's algorithm for Fisher's exact test on unordered R × C contingency tables was imported into GEREA. The code, written in double precision FORTRAN 77, currently provides the fastest available method for executing Fisher's exact test. P values were corrected by the Benjamini and Hochberg method (false discovery rate: FDR). Corrected $P < 0.05$ was regarded as statistically significant. If corrected $P0 < 0.05$ and corrected $P1 < 0.05$, the changes in the network were caused by increased regulator activity, whereas if corrected $P0 < 0.05$ and corrected $P2 < 0.05$, the changes in the network were caused by decreased regulator activity.

2.3. Development of the GEREA Software

The GEREA software was developed using networks to analyze the miRNA function with gene expression profiling data as input. Three major steps are built into the program: 1) The regulatory network is constructed based on the linkages in the network file. 2) The transcriptome data is loaded into the data file. The program matches each gene ID in the transcriptome to the target gene ID in the network. Both are offi-

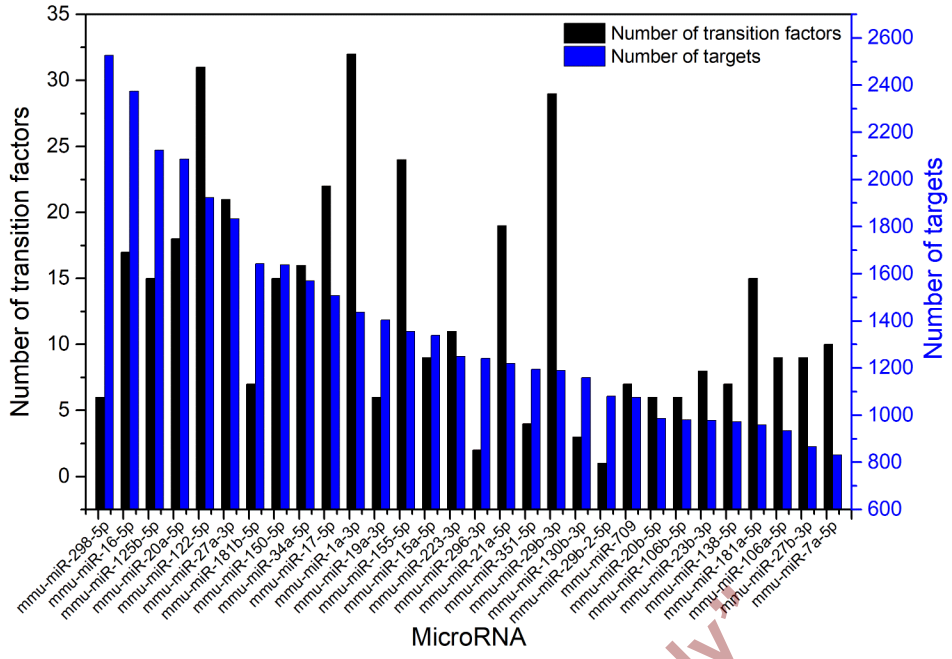

Fig. (4). Number of transition factors and targets for each miRNA regulator (top 30). Black bars represent the number of transition factors for each miRNA, and green bars represent the number of targets. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

cial NCBI gene symbols. If they are detected, then the expression data is mapped to the target gene. If not, the data are omitted. 3) A statistical analysis is performed on the network. The A–R values in the contingency table are counted. Statistical analysis is performed using Fisher's exact test calculated in the FEXACT FORTRAN subroutine [37]. The statistical data and the network are printed into files. The former is formatted as a tab-delimited text file. The web-based interface (www.thua45.cn/gerea) supports online GEREA analysis using custom data. Demo data may be visualized on and downloaded from the GEREA demo data page. Data may be uploaded to the database on the GEREA submit page. A GEREA job can be started in the GEREA run page using the specified data and link database. The GEREA website runs the jobs sequentially, and each job requires ~5 min on average. The results page automatically refreshes every 15 s. For very large jobs, the open-source GEREA program may be downloaded and run to analyze the data locally (offline) (https://sourceforge.net/projects/gerea/).

3. RESULTS

3.1. Overview of the miRNA-(transition factor)-(target gene) Regulation Network

The manually curated links included 242 miRNAs, 524 transition factors, and 4,018 target genes, for which a total of 31,948 evidence sentences were assigned. Each link is usually manually curated from > 1 evidence of abstract candidates. A total of 980 miRNA-(transition factor) linkages and 22,271 (transition factor)-target linkages were generated in the regulatory network. Fig. (**4**) presents the number of transition factors and targets for the top 30 miRNA regulators with the highest number of targets. The results indicate that mmu-miR-1a-3p can regulate the largest number of transition factors (a total of 32), followed closely by mmu-miR-122-5p (31), mmu-miR-29b-3p (29), and mmu-miR-155-5p (24). Meanwhile, mmu-miR-298-5p can regulate the largest number of target genes (a total of 2,526), followed closely by mmu-miR-16-5p, mmu-miR-125b-5p, and mmu-miR-20a-5p. This regulatory gene expression information can be used to construct miRNA-(transition factor)-target networks, representing our current understanding of which network motif is the most important "regulator" for gene expression when dividing a large network into smaller blocks [38].

A side-by-side comparison among the GereDB mouse section and to that of HIRIdb [28], TRRUST [29], and TFactS [30, 31] was performed. However, since HIRIdb is designed specifically for human gene expression regulation, it does not contain any information related to mouse and was, therefore, excluded. The results showed that there were 1957 overlaps with TRRUST, and 552 overlaps with TFactS. Interestingly, 97% of the gene expression regulation relationships deposited in GereDB did not overlap with TRRUST and TFactS, indicating that GereDB is a unique resource for gene expression regulation relationships in the community (Fig. **5**).

3.2. Case Study 1: Transcriptome Analysis of Synthetic miR-155 Oligo-treated Mouse $CD4^+$ T-cells

A case study was performed to determine if GEREA is capable of detecting gene expression regulators in real-world transcriptome datasets. Mouse $CD4^+$ T-cells were purified with a $CD4^+$ T-cell isolation kit (Miltenyi Biotec, Shanghai, China) and by magnetic bead separation. Cell purity was confirmed by flow cytometry. A total of 3×10^4 cells with three replicates were transfected by synthetic miR-155 oligo using a previously described protocol (three replicates) [39]. At 8 h post-transfection, cells were collected, and total RNA

was isolated. Sequencing libraries were prepared with an Illumina Truseq RNA sample preparation kit according to the manufacturer's protocol (Illumina, San Diego, CA, USA). Single-read sequencing (read length: 50 bp) was conducted on an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA). Bowtie 2 [40] mapped the clean reads to the reference gene set extracted from the NCBI reference sequence database [41]. The reads per transcript were counted, quantile normalized, and DEGs were identified using the limma R package [42]. All data discussed in the present study were deposited in the NCBI GEO database [43] under accession number GSE138715.

GEREA was assessed using the dataset for synthetic miR-155 oligo-treated mouse $CD4^+$ T-cells. The GEREA input files were created and run using the GEREA option "-f 1.5 –q 0.05". The GEREA run took 395 s and generated two output files. A total of 37 miRNAs and their targets were significantly enriched in the DEGs when FDR ≤ 10^{-6} (Table **1** listed top 25, a full list is available in Supplemental Document 3). The network regulated by miR-155 was significant in up-regulation analysis (FDR = 5.09E-16, ranked 2ed). Thus, the GEREA test indicated that the transcriptional changes in the targets of the miR-155 networks were caused by miR-155 up-regulation. This result corresponds to the experimental design (miR-155 overexpression).

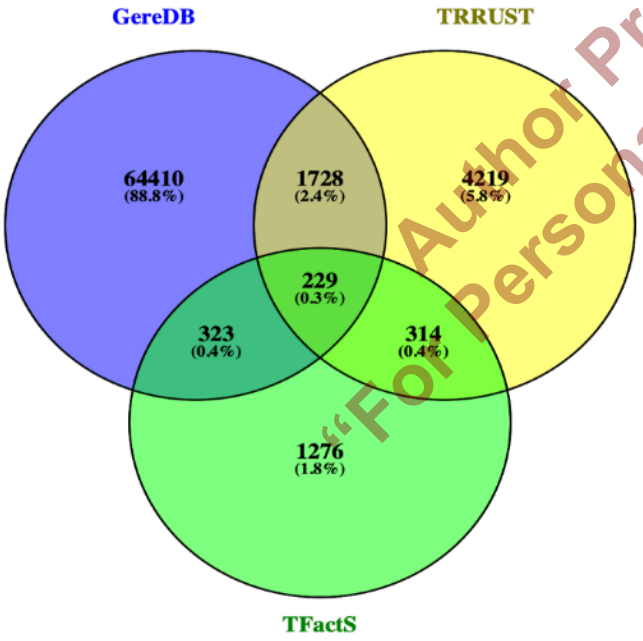

Fig. (5). Comparison of GereDB, TRRUST, and TFactS datasets. Venn diagram depicting the number of gene expression regulation relationships deposited in GereDB (blue), TRRUST (yellow), and TFactS (green), and their intersection. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The dataset for the synthetic miR-155 oligo-treated $CD4^+$ T-cells was then assessed by GEREA for transition factor enrichment analysis (method in Fig. **2A**, full list of results in Supplemental Document 4). The GEREA input file was the same as that for the miRNA enrichment analysis. It was run with the option "-f 1.5 –q 0.05". The GEREA run took 475 s and generated two output files. A total of 38 regulators and their targets were significantly enriched in the up- or down-regulation analyses when FDR ≤ 10^{-3} (Table **2** listed top 25).

3.3. Case Study 2: Analysis of GEREA Using Publicly Available Transcriptome Data

GEREA was then assessed using a publicly available dataset of mice peripheral blood infected with *Salmonella* bacteria. The raw transcriptome dataset was obtained from the NCBI GEO database with accession number GSE115164. The data were normalized using the quantile method in Bioconductor [44]. Comparisons were made between the samples collected before infection (day 0) and two days post-infection (day 2). The GEREA run took 362 seconds. The results indicated that targets of 38 miRNA regulators were significantly enriched in the DEGs between day 2 versus day 0 samples (FDR < 10E-6, Table **3** listed top 25, full list available in Supplemental document 5). The top-ranked miRNA, mmu-miR-298-5p, was significantly down-regulated, leading to the hypothesis that many DEGs (mmu-miR-298-5p targets) were induced via altering of mmu-miR-298-5p expression. The second-ranked miRNA, mmu-miR-16-5p, was significant in both the up- and down-regulation analyses (2.01E-05 and 3.28E-28), indicating that the regulation of mmu-miR-16-5p target genes is complex and requires further investigation. The possible reasons for this paradox could be: 1) The regulatory relationships deposited in GEREDB were established by particular experimental conditions, and these relationships may differ when the biological conditions are altered. For example, most miRNAs down-regulate target gene expression, while in certain cellular conditions, some miRNAs up-regulate target gene expression; 2) the regulation of gene expression is complex, and multiple factors are involved at different levels of the regulation cascade (in this study, the miRNA level and the transition factor level were considered). Hence, uncertainties remain regarding which factor contributes to the major effect in a particular biological experiment. If the miRNA serves as the major effector, the GEREA result may fulfill the gene expression regulation mechanism in the experiment. Otherwise, if the transition factors account for the major effectors, the GEREA result may (if all of them change according to the rules of the regulatory networks) or may not (if they were changed in different directions) fulfill the gene expression regulation mechanism in the experiment. The third-ranked miRNA is mmu-miR-125b-5p, which was significantly down-regulated, leading to the hypothesis that many DEGs (mmu-miR-125b-5p targets) were induced by attenuating the effect of mmu-miR-125b-5p [45]. Interestingly, it has been reported that two of the top-ranked miRNAs, miR-16 and miR-125b, were differentially expressed in the peripheral blood of animals infected by *Salmonella*. Meanwhile, ssc-miR-16 was down-regulated by at least four-fold in the peripheral blood of *Salmonella*-infected piglets [46]. Through expression analyses, differences were identified between pre- and postnatal stages of salmonellosis for miR-125b, which were suppressed two days after *Salmonella* inoculation in pigs [45].

3.4. Analysis of GEREA Using Randomized Datasets

To determine if the regulators identified as significant in up- or down-regulation test were caused by random effects, a

Table 1. Significantly enriched miRNA-(transition factor)-(target gene) regulatory networks in the transcriptome data for the synthetic miR-155 oligo-treated mouse CD4$^+$ T-cells (Top 25).

miRNA Name	Number of Genes in the Contingency Table									P values		
	A	B	C	D	E	F	G	H	I	P0	P1	P2
mmu-miR-16-5p	26	82	514	26	121	480	169	510	6824	1.69E-29	3.50E-16	1.06E-08
mmu-miR-155-5p	40	29	532	19	46	505	176	165	6964	3.81E-19	5.09E-16	0.0124139
mmu-miR-223-3p	15	36	530	17	68	490	72	227	6924	1.57E-23	1.16E-14	4.28E-06
mmu-miR-20a-5p	24	49	522	19	92	479	107	416	6822	1.48E-21	1.46E-14	0.00053855
mmu-miR-298-5p	25	64	515	37	120	474	149	531	6781	2.79E-29	1.56E-14	2.56E-08
mmu-miR-181b-5p	25	28	517	32	78	463	133	304	6777	7.00E-25	1.65E-14	0.00026721
mmu-miR-21a-5p	53	33	541	47	36	523	181	148	7052	2.88E-29	2.64E-14	3.60E-13
mmu-miR-27a-3p	28	49	523	30	73	477	171	281	6807	9.26E-24	4.02E-14	3.52E-07
mmu-miR-122-5p	22	85	501	19	85	477	135	419	6749	2.21E-24	6.35E-13	3.31E-09
mmu-miR-125b-5p	18	52	512	22	92	464	129	399	6718	4.50E-21	1.55E-12	4.30E-05
mmu-miR-34a-5p	19	45	532	28	35	522	176	198	7004	3.57E-13	0.0196141	8.68E-12
mmu-miR-106b-5p	12	27	529	3	44	509	42	161	6970	1.32E-14	2.18E-11	0.015037
mmu-miR-15a-5p	17	45	524	14	66	497	114	271	6915	3.02E-17	2.39E-10	3.69E-05
mmu-miR-130b-3p	22	12	520	19	48	477	102	199	6841	4.13E-14	2.92E-10	0.114678
mmu-miR-429-3p	4	12	550	15	14	539	26	25	7132	2.88E-16	7.01E-05	3.29E-10
mmu-miR-29b-2-5p	21	12	520	19	46	479	97	179	6860	2.05E-14	3.90E-10	0.0428807
mmu-miR-17-5p	21	38	524	13	56	495	88	271	6879	1.29E-14	4.69E-10	0.00097524
mmu-miR-150-5p	21	41	527	24	64	483	93	300	6871	1.99E-19	8.28E-10	5.37E-07
mmu-miR-19a-3p	19	38	514	25	64	476	107	282	6789	8.05E-19	1.74E-09	5.98E-07
mmu-miR-296-3p	11	33	515	18	62	476	73	267	6795	1.46E-16	9.74E-09	4.08E-05
mmu-miR-182-5p	15	24	532	21	22	501	64	139	6955	2.18E-12	0.00199847	9.84E-09
mmu-miR-181a-5p	16	24	530	9	39	506	86	146	6970	2.15E-12	2.10E-08	0.00469923
mmu-miR-1a-3p	13	43	530	17	59	504	100	282	6861	2.30E-13	2.82E-08	0.0002393
mmu-miR-29a-3p	11	29	533	15	35	511	64	130	6965	5.61E-15	3.90E-08	2.00E-06
mmu-miR-23b-3p	14	10	546	17	38	526	91	138	7059	9.98E-12	3.91E-08	0.130397

random test was performed using shuffled datasets. The two datasets used were from case 1 and case 2, which were shuffled five times (random exchange between the gene IDs and the expression value) and run with GEREA using the same parameters as those applied for the case study. For the miR-155 oligo-treated mouse CD4$^+$ T-cells datasets used in case study 1, only one miRNA was significant (FDR < 0.05) in the up-regulation test (0.019), and none were significant in the down-regulation test. For the Salmonella infection dataset used in case study 2, one miRNA was significant (FDR < 0.05) in the up-regulation test (0.03), and two were significant (FDR < 0.05) in the down-regulation test (0.002 and 0.02). These results indicate that the significantly over-represented miRNAs reported in case 1 and case 2 are not random events.

4. DISCUSSION

Gene expression regulation is a crucial molecular mechanism in eukaryotic organisms that is essential for the normal development and maintenance of healthy cells. Hence, deviation from standard coordination programs may lead to severe disease [47]. Transcription factors bind to short DNA sequences or motifs in genes with specific positive or negative regulatory patterns. Hence, specific genes will be transcribed into the primary RNA transcript [48]. MiRNAs control target gene expression by post-transcriptional inhibition, which is one of the numerous events occurring between the gene DNA sequence and its corresponding protein [49]. While previous studies often dealt with individual regulatory interactions, high-throughput experiments have dramatically adv-

Table 2. Significantly enriched (transition factor)-(target gene) regulatory networks in transcriptome data for synthetic miR-155 oligo-treated mouse CD4$^+$ T-cells (Top 25).

Regulator	Gene Numbers in the Contingency Table									P values		
	A	B	C	D	E	F	G	H	I	P0	P1	P2
Il2	13	1	547	32	5	514	77	11	7071	1.31E-14	0.0200996	7.44E-11
Tgfb1	12	21	520	46	19	479	179	97	6860	8.60E-15	0.0505541	8.23E-11
Il1b	23	4	533	45	7	498	148	35	6975	2.08E-14	0.00312185	9.95E-10
Tnf	30	10	517	56	17	479	254	68	6805	9.12E-15	3.29E-05	9.55E-08
Cd28	10	2	547	22	2	527	53	10	7100	1.43E-09	0.145938	6.17E-07
Il6	14	4	540	32	11	506	112	28	7028	1.88E-12	0.00148039	1.61E-06
Mapk1	16	1	534	28	9	505	99	19	7002	1.61E-10	0.00077478	4.79E-06
Trp53	3	22	529	11	13	521	72	63	6973	2.81E-08	0.247933	4.80E-06
Mapk8	11	1	545	22	4	518	60	16	7060	3.63E-09	0.0171536	7.62E-06
Il21	3	2	554	13	3	539	15	9	7148	1.15E-09	0.0516756	9.88E-06
Tbx21	2	1	556	4	5	546	2	2	7166	1.68E-09	1.88E-05	0.00553668
Il25	0	1	558	7	0	548	4	4	7167	0	1	2.16E-05
Il15	11	0	549	18	3	533	42	6	7117	1.41E-10	0.00105096	2.39E-05
Ifng	17	7	534	42	17	484	198	69	6878	1.39E-10	0.023602	2.73E-05
Il4	9	5	545	27	13	507	94	42	7015	9.81E-11	0.00996519	2.88E-05
Il10	7	7	547	15	11	524	39	46	7073	1.88E-08	0.00886346	3.87E-05
Il7	11	1	547	12	4	535	39	4	7121	2.58E-08	4.36E-05	0.00493448
Tlr3	5	0	555	5	4	545	16	3	7149	3.90E-06	4.97E-05	0.365043
Crp	6	0	554	3	3	548	16	8	7149	0.00058817	7.45E-05	1
Csf3	3	4	552	11	2	540	33	10	7131	3.74E-05	1	8.64E-05
Rb1	0	11	547	2	1	548	10	13	7142	3.74E-07	1	8.71E-05
Tnfsf10	3	3	553	8	1	544	22	11	7136	0.00025708	1	0.00013717
Mapk14	12	5	536	20	1	523	82	15	7036	1.20E-06	0.180003	0.00015442
Igf1	15	2	541	24	1	529	99	28	7027	5.98E-06	0.421865	0.0001718
Fasl	1	1	558	6	3	546	6	0	7165	0	0.0981048	0.00018104

Table 3. Significantly enriched miRNA-(transition factor)-(target gene) regulatory networks in the transcriptome data of *Salmonella*-infected mouse peripheral blood samples (Top 25).

Regulator miRNA	Gene Numbers in the Contingency Table									P values		
	A	B	C	D	E	F	G	H	I	P0	P1	P2
mmu-miR-298-5p	55	290	1515	48	85	1722	485	1048	20008	4.89E-60	0.262848	3.08E-35
mmu-miR-16-5p	54	242	1567	46	129	1719	327	1076	20098	2.56E-41	2.01E-05	3.28E-28
mmu-miR-125b-5p	52	205	1500	36	66	1718	344	811	19874	1.91E-41	0.0870758	1.86E-24
mmu-miR-122-5p	45	180	1545	37	102	1710	304	788	19980	2.89E-32	8.22E-05	1.10E-21

Table 3. contd…

Regulator miRNA	Gene Numbers in the Contingency Table									P values		
	A	B	C	D	E	F	G	H	I	P0	P1	P2
mmu-miR-155-5p	109	73	1617	43	41	1741	283	350	20386	1.91E-41	1.47E-21	1.05E-08
mmu-let-7e-5p	46	80	1597	15	19	1768	113	154	20611	8.08E-45	1.58E-10	1.39E-19
mmu-miR-19a-3p	37	149	1518	36	45	1722	307	548	20035	1.52E-32	0.198886	1.64E-19
mmu-miR-296-3p	25	146	1516	23	43	1726	235	525	20047	8.56E-32	0.472754	3.79E-19
mmu-miR-351-5p	24	140	1521	24	46	1722	230	505	20055	4.20E-30	0.26861	6.44E-19
mmu-miR-15a-5p	37	134	1594	33	74	1738	210	571	20276	2.68E-27	8.47E-05	1.14E-18
mmu-miR-20a-5p	42	182	1573	25	79	1732	234	878	20087	1.87E-27	0.0729133	2.72E-18
mmu-miR-27a-3p	88	144	1555	41	69	1721	343	635	20078	6.11E-38	5.61E-09	3.12E-17
mmu-miR-181a-5p	38	95	1598	17	35	1752	158	314	20406	4.81E-28	0.00040952	7.96E-16
mmu-miR-17-5p	32	118	1594	21	48	1743	184	567	20230	9.09E-19	0.211181	8.76E-13
mmu-miR-203-3p	12	22	1657	5	7	1784	17	40	20788	9.68E-16	0.00416845	3.44E-11
mmu-miR-1a-3p	36	108	1620	33	71	1736	179	610	20197	2.50E-17	0.0004939	5.23E-11
mmu-miR-106b-5p	18	84	1607	7	33	1760	80	339	20433	9.90E-18	0.0681212	5.23E-11
mmu-miR-182-5p	25	72	1601	14	28	1756	137	246	20427	9.02E-19	0.00865317	6.34E-11
mmu-miR-130a-3p	5	23	1663	5	3	1790	11	29	20796	3.67E-15	0.0900785	1.11E-10
mmu-miR-466l-3p	38	34	1621	6	13	1777	83	84	20690	1.54E-24	1.12E-10	8.50E-06
mmu-miR-21a-5p	54	71	1656	58	21	1768	439	274	20610	1.67E-16	0.169895	1.51E-10
mmu-miR-150-5p	35	127	1594	20	65	1735	194	648	20237	1.64E-18	0.00665409	1.63E-10
mmu-miR-223-3p	35	113	1585	13	47	1747	168	516	20299	2.86E-21	0.00916381	2.91E-10
mmu-miR-206-3p	19	43	1652	34	37	1752	151	216	20579	1.26E-12	0.00657037	4.26E-10
mmu-miR-181b-5p	64	123	1555	34	69	1704	319	694	19918	7.62E-20	0.00050595	7.36E-10

anced the perspectives on gene regulation. It is now evident that the only way to clarify gene regulatory activity is to address the complex interaction network in the entire ensemble of regulatory gene expression elements [50].

Accurate network regulator discovery plays a pivotal role in the study of complex networks as it provides a systematic approach toward disclosing the major effectors in various gene expression patterns across multiple types of interactions. However, detection of the regulators in complex networks is challenging. Numerous methods can elucidate the transcription factors or miRNA-related regulatory networks; however, complete information required to connect them is lacking [47, 49-51]. In principle, miRNAs and transition factor regulatory networks may be connected by identifying the miRNA binding sites in the 3′-UTR regions of the transcription factors and identifying the regulatory relationships between the transition factors and target genes. Computational biologists have sought to predict miRNA and transcription factor binding sites [52, 53]. However, current methods have only proven effective at forecasting direct target relationships between miRNAs and mRNAs or between transcription factors and target genes. Hence, several questions remained to be addressed. 1) The coverage of the gene expression relationships deposited in the databases. 2) The integrity of gene expression relationships deposited in the database must be assessed. For example, the regulatory effect (positive or negative interactions) between the regulators and targets should be included in the regulation network. 3) The structure of the data storage and analyzing methods must be improved. For example, in some cases, miRNA inhibits the expression of certain target genes post-translationally without altering mRNA expression levels, and these target genes cannot be directly detected by transcriptome profiling.

In the current work, we have addressed the first question by extracting gene expression regulation information from literature abstracts. The extracted 64,410 unique records from GereDB were compared to TRRUST [29] and TFactS [30, 31], demonstrating the value of our work. We further addressed the second question by manually extracting additional information from the sentence descriptions in the literature regarding the gene expression regulation relationships, resulting in the identification of the regulation effect (up or down-regulation) for over 70% of the gene regulation relationships. Finally, we addressed the third question by introducing miRNA-(transition factor)-(target gene) networks using an inference machine and Fisher's exact test-based analysis methods. If miRNAs regulate transition factors post-

transcriptionally without impacting the mRNA levels, they can be detected with our tool via analysis of target gene transcriptome data.

CONCLUSION

Here, we developed a new algorithm, designated GEREA, capable of detecting genetic regulators. This platform identifies regulatory networks involving miRNAs, transition factors, and target genes. This new method was developed to analyze the functions of regulatory genes using transcription data as input and effectively constructed reliable miRNA-(transition factor)-(target gene) regulation networks based on publicly accessible literature. Moreover, this algorithm established how miRNAs and transition factors orchestrate particular transcriptional profiles. Specifically, the generated miRNA-(transition factor)-target network confirmed cooperation between miRNAs and transition factors in cellular systems. It further determined that regulator genes participate in transcriptional and post-transcriptional transcriptome organization in miR-155-stimulated mouse $CD4^+$ T-cells and peripheral blood collected from *Salmonella*-infected mice. However, the miRNA-(transition factor)-(target gene) regulatory networks generated in this study are incomplete and must be expanded to account for additional transcription factors and miRNAs.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All procedures involving animals and collecting the samples were adhered to ethical guidelines and approved by the Animal Care and Use Committee of Yangtze University (China, YZU-2018-0031).

HUMAN AND ANIMAL RIGHTS

No humans were used in the studies that is the basis of this research. All animal procedures were in accordance with the US “Public Health Service's Policy on Humane Care and Use of Laboratory Animals" and "Guide for the Care and Use of Laboratory Animals."

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

This project was funded by the National Natural Science Foundation of China [NSFC Grant No. 31902231 and 31402055], the College Students' Innovation and Entrepreneurship Training Program of Yangtze University [Grant No. 2019110]. The funding body played no role in the design of the study, or in the collection, analysis, or interpretation of data, or in writing the manuscript.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors declare that they have no competing interests.

ACKNOWLEDGMENTS

Declared none.

REFERENCES

[1] Goodbourn S, King P. Eukaryotic Gene Transcription. New York: Oxford University Press 1996.

[2] Elnitski L, Jin VX, Farnham PJ, Jones SJ. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res 2006; 16(12): 1455-64. http://dx.doi.org/10.1101/gr.4140006 PMID: 17053094

[3] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet 2009; 10(4): 252-63. http://dx.doi.org/10.1038/nrg2538 PMID: 19274049

[4] Huang T, Huang X, Shi B, Yao M. GEREDB: Gene expression regulation database curated by mining abstracts from literature. J Bioinform Comput Biol 2019; 17(4): 1950024. http://dx.doi.org/10.1142/S0219720019500240 PMID: 31617460

[5] Torres TT, Metta M, Ottenwälder B, Schlötterer C. Gene expression profiling by massively parallel sequencing. Genome Res 2008; 18(1): 172-7. http://dx.doi.org/10.1101/gr.6984908 PMID: 18032722

[6] Blohm DH, Guiseppi-Elie A. New developments in microarray technology. Curr Opin Biotechnol 2001; 12(1): 41-7. http://dx.doi.org/10.1016/S0958-1669(00)00175-0 PMID: 11167071

[7] Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. Development 2005; 132(21): 4653-62. http://dx.doi.org/10.1242/dev.02073 PMID: 16224045

[8] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004; 116(2): 281-97. http://dx.doi.org/10.1016/S0092-8674(04)00045-5 PMID: 14744438

[9] Fu J, Tang W, Du P, *et al.* Identifying microRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. BMC Syst Biol 2012; 6(1): 68. http://dx.doi.org/10.1186/1752-0509-6-68 PMID: 22703586

[10] Peng X, Li Y, Walters KA, *et al.* Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. BMC Genomics 2009; 10(1): 373. http://dx.doi.org/10.1186/1471-2164-10-373 PMID: 19671175

[11] Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA 2004; 10(10): 1507-17. http://dx.doi.org/10.1261/rna.5248604 PMID: 15383676

[12] Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet 2007; 39(10): 1278-84. http://dx.doi.org/10.1038/ng2135 PMID: 17893677

[13] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res 2008; 36(Database issue): D149-53. PMID: 18158296

[14] Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003; 115(7): 787-98. http://dx.doi.org/10.1016/S0092-8674(03)01018-3 PMID: 14697198

[15] Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 2009; 37(Database issue): D105-10. http://dx.doi.org/10.1093/nar/gkn851 PMID: 18996891

[16] Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Res 2009; 37(Database issue): D155-8. http://dx.doi.org/10.1093/nar/gkn809 PMID: 18957447

[17] Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res 2011; 39(Database issue): D202-9. http://dx.doi.org/10.1093/nar/gkq1056 PMID: 21037263

[18] Jackson RJ, Standart N. How do microRNAs regulate gene expression? Sci STKE 2007; 2007(367): re1. http://dx.doi.org/10.1126/stke.3672007re1 PMID: 17200520

[19] Most D, Leiter C, Blednov YA, Harris RA, Mayfield RD. Synaptic microRNAs Coordinately Regulate Synaptic mRNAs: Perturbation by Chronic Alcohol Consumption. Neuropsychopharmacology 2016; 41(2): 538-48. http://dx.doi.org/10.1038/npp.2015.179 PMID: 26105134

[20] Huang DW, Sherman BT, Tan Q, *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res 2007; 35(Web Server issue): W169-75. http://dx.doi.org/10.1093/nar/gkm415

[21] Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005; 102(43): 15545-50. http://dx.doi.org/10.1073/pnas.0506580102 PMID: 16199517

[22] Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. Cell-type-specific signatures of microRNAs on target mRNA expression. Proc Natl Acad Sci USA 2006; 103(8): 2746-51. http://dx.doi.org/10.1073/pnas.0511045103 PMID: 16477010

[23] van Dongen S, Abreu-Goodger C, Enright AJ. Detecting microRNA binding and siRNA off-target effects from expression data. Nat Methods 2008; 5(12): 1023-5. http://dx.doi.org/10.1038/nmeth.1267 PMID: 18978784

[24] Alexiou P, Maragkakis M, Papadopoulos GL, Simmosis VA, Zhang L, Hatzigeorgiou AG. The DIANA-mirExTra web server: from gene expression data to microRNA function. PLoS One 2010; 5(2): e9171. http://dx.doi.org/10.1371/journal.pone.0009171 PMID: 20161787

[25] Creighton CJ, Nagaraja AK, Hanash SM, Matzuk MM, Gunaratne PH. A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. RNA 2008; 14(11): 2290-6. http://dx.doi.org/10.1261/rna.1188208 PMID: 18812437

[26] Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. Bioinformatics 2009; 25(6): 832-3. http://dx.doi.org/10.1093/bioinformatics/btp059 PMID: 19181683

[27] Ulitsky I, Laurent LC, Shamir R. Towards computational prediction of microRNA function and activity. Nucleic Acids Res 2010; 38(15): e160. http://dx.doi.org/10.1093/nar/gkq570 PMID: 20576699

[28] Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics 2012; 13: 405.

[29] Han H, Cho JW, Lee S, *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 2018; 46(D1): D380-6. http://dx.doi.org/10.1093/nar/gkx1013 PMID: 29087512

[30] Essaghir A, Demoulin JB. A minimal connected network of transcription factors regulated in human tumors and its application to the quest for universal cancer biomarkers. PLoS One 2012; 7(6): e39666. http://dx.doi.org/10.1371/journal.pone.0039666 PMID: 22761861

[31] Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin JB. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. Nucleic Acids Res 2010; 38(11): e120. http://dx.doi.org/10.1093/nar/gkq149 PMID: 20215436

[32] Sayers EW, Beck J, Brister JR, *et al.* Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2020; 48(D1): D9-D16. http://dx.doi.org/10.1093/nar/gkz899 PMID: 31602479

[33] Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 3. advanced searching, MeSH and My NCBI. J Telemed Telecare 2014; 20(2): 102-12. http://dx.doi.org/10.1177/1357633X13519036 PMID: 24614997

[34] Pedregosa F, Varoquaux Gl, Gramfort A, *et al.* Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011; 126(85): 2825-30.

[35] Chou CH, Shrestha S, Yang CD, *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018; 46(D1): D296-302. http://dx.doi.org/10.1093/nar/gkx1067 PMID: 29126174

[36] Mehta CR, Patel NR. ALGORITHM 643: FEXACT: a FORTRAN subroutine for Fisher's exact test on unordered times contingency tables. ACM Trans Math Softw 1986; 12(2): 154-61. http://dx.doi.org/10.1145/6497.214326

[37] Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR. FlowRepository: a resource of annotated flow cytometry datasets associated with peer-reviewed publications. Cytometry A 2012; 81(9): 727-31. http://dx.doi.org/10.1002/cyto.a.22106 PMID: 22887982

[38] Boyle AP, Araya CL, Brdlik C, *et al.* Comparative analysis of regulatory information and circuits across distant species. Nature 2014; 512(7515): 453-6. http://dx.doi.org/10.1038/nature13668 PMID: 25164757

[39] Yao M, Gao W, Tao H, Yang J, Liu G, Huang T. Regulation signature of miR-143 and miR-26 in porcine Salmonella infection identified by binding site enrichment analysis. Mol Genet Genomics 2016; 291(2): 789-99. http://dx.doi.org/10.1007/s00438-015-1146-z PMID: 26589421

[40] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10(3): R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25 PMID: 19261174

[41] O'Leary NA, Wright MW, Brister JR, *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 2016; 44(D1): D733-45. http://dx.doi.org/10.1093/nar/gkv1189 PMID: 26553804

[42] Ritchie ME, Phipson B, Wu D, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7): e47. http://dx.doi.org/10.1093/nar/gkv007 PMID: 25605792

[43] Barrett T, Wilhite SE, Ledoux P, *et al.* NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013; 41(Database issue): D991-5. PMID: 23193258

[44] Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003; 19(2): 185-93. http://dx.doi.org/10.1093/bioinformatics/19.2.185 PMID: 12538238

[45] Yao M, Gao W, Yang J, Liang X, Luo J, Huang T. The regulation roles of miR-125b, miR-221 and miR-27b in porcine Salmonella infection signalling pathway. Biosci Rep 2016; 36(4): 52-8. http://dx.doi.org/10.1042/BSR20160243 PMID: 27474500

[46] Huang T, Huang X, Chen W, *et al.* MicroRNA responses associated with Salmonella enterica serovar typhimurium challenge in peripheral blood: effects of miR-146a and IFN-γ in regulation of fecal bacteria shedding counts in pig. BMC Vet Res 2019; 15(1): 195. http://dx.doi.org/10.1186/s12917-019-1951-4 PMID: 31186019

[47] Friard O, Re A, Taverna D, De Bortoli M, Corá D. CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. BMC Bioinformatics 2010; 11(1): 435. http://dx.doi.org/10.1186/1471-2105-11-435 PMID: 20731828

[48] Latchman DS. Transcription factors: an overview. Int J Biochem Cell Biol 1997; 29(12): 1305-12. http://dx.doi.org/10.1016/S1357-2725(97)00085-X PMID: 9570129

[49] Wang J, Lu M, Qiu C, Cui Q. TransmiR: a transcription factor-microRNA regulation database. Nucleic Acids Res 2010; 38(Database issue): D119-22. http://dx.doi.org/10.1093/nar/gkp803 PMID: 19786497

[50] Zacher B, Abnaof K, Gade S, Younesi E, Tresch A, Fröhlich H. Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data. Bioinformatics 2012; 28(13): 1714-20. http://dx.doi.org/10.1093/bioinformatics/bts257 PMID: 22563068

[51] Zhang S, Li Q, Liu J, Zhou XJ. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. Bioinformatics 2011; 27(13): i401-9. http://dx.doi.org/10.1093/bioinformatics/btr206 PMID: 21685098

[52] Tompa M, Li N, Bailey TL, *et al.* Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 2005; 23(1): 137-44. http://dx.doi.org/10.1038/nbt1053 PMID: 15637633

[53] Mazière P, Enright AJ. Prediction of microRNA targets. Drug Discov Today 2007; 12(11-12): 452-8. http://dx.doi.org/10.1016/j.drudis.2007.04.002 PMID: 17532529