

Original Article

Mutual information network-based support vector machine for identification of rheumatoid arthritis-related genes

Xin Zi¹, Yun Liu¹, Pengji Gao²

Departments of ¹Orthopedics, ²Traumatology, Linyi People's Hospital, Linyi, China

Received January 5, 2016; Accepted April 1, 2016; Epub June 15, 2016; Published June 30, 2016

Abstract: In recent years, the underlying roles of significant genes in the pathogenesis of rheumatoid arthritis (RA) have greatly aroused the interest of clinicians and researchers. In the current study, we aimed to identify the potential biomarkers of RA via mutual information network-based support vector machine (SVM) classifier. Firstly, microarray data E-GEOD-45291 was downloaded from ArrayExpress database. Next, differentially expressed genes (DEGs) identification was implemented. The differential pathways genes (DPGs) which were enriched in differential pathways were then screened through *attract* method. Subsequently, the mutual information network was constructed on the basis of the DPGs through context likelihood of relatedness (CLR) algorithm. Simultaneously, hub genes of the mutual information network were screened, followed by the extraction of the intersection of hub genes and DEGs (named as inter genes). Finally, the classification and evaluation via SVM with linear kernel was performed. Based on the results, we found that a total of 339 DEGs and 448 DPGs were screened. There were 54 hub genes. Then, 8 inter genes were identified such as mitogen-activated protein kinase kinase 2 (*MAP2K2*), tumor necrosis factor receptor (TNFR)-associated death domain (*TRADD*), BH3 interacting domain death agonist (*BID*) and so on. Accordingly, these 8 genes can classify unknown samples from patients with the highest AUC score of 1.00, MCC score of 1.00, specificity of 1.00, and sensitivity of 1.00. Our analysis might provide a novel insight into the pathogenic processes in RA. *MAP2K2*, *TRADD* and *BID* might be attractive biomarkers for the diagnosis and therapeutic intervention of RA.

Keywords: Rheumatoid arthritis, differentially expressed genes, mutual information network, support vector machines

Introduction

Rheumatoid arthritis (RA), as a common autoimmune disorder, is characterized by joint destruction, synovial proliferation, and leucocyte as well as neovascularisation extravasation. Moreover, RA is connected with disability, systemic complications, socioeconomic costs and early death [1, 2]. In 2013, RA led to 38,000 deaths, increased from 28,000 deaths in 1990 [3]. Although the outcome for RA patients has improved in recent years because of earlier and more aggressive application of nonbiologic drugs as well as the introduction of biologic therapies, unmet demands remain. Urgently, advances in understanding the pathogenesis of RA have fostered the development of new therapeutic strategy, along with improved outcomes.

The molecular mechanism of RA is still not clear. As documented, previous studies have demonstrated that there is the potential association between RA and environmental factors, for example, smoking and climate. Currently, it is generally believed that the genetic factors are closely connected with the development and progression of RA. Former studies have suggested that toll-like receptors (TLRs) including *TLR2* and *TLR4* exert important functions in the pathogenesis of RA [4, 5]. Moreover, *IL-15* is reported to be elevated in synovial fluids of patients with RA [6]. In addition, the protein tyrosine phosphatase non-receptor 22 gene (*PTPN22*) and human leucocyte antigen (*HLA*) locus have been verified to be connected with RA [7, 8]. Despite more effort to select the RA causing genes, the molecular mechanisms and

Identification of RA-related genes

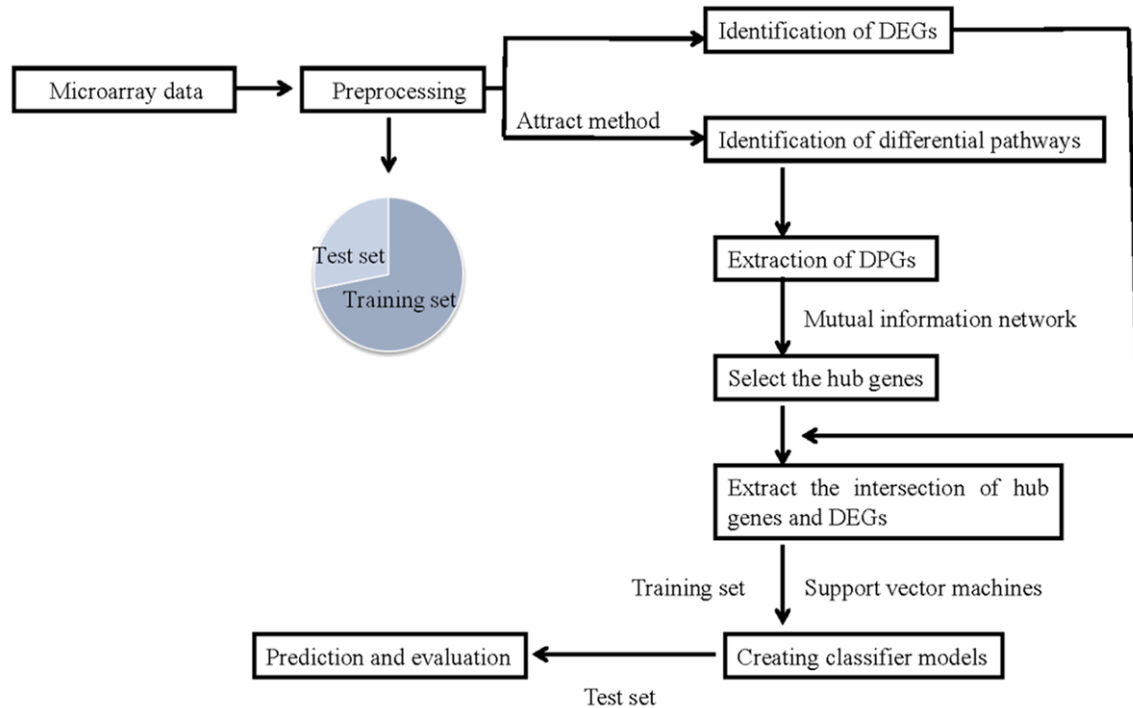


Figure 1. The flowchart of the proposed approach.

treatment of RA is difficult to establish because RA is heterogeneous in nature.

In recent years, the application of various approaches has been witnessed to identify genes that contribute to the development of RA. In general, significant genes are selected using expression-based analysis method. Kyoto encyclopedia of genes and genomes (KEGG) is then utilized to discover the potential functions of the obtained genes [9]. However, for lack of inclusion in KEGG database, several genes with highly correlated patterns to the significantly identified genes were not picked up. Excitedly, *attract* offered by Mar and coworkers [10] has been indicated to be able to expand the important genes which are not seen in traditional KEGG method. Recently, more researchers have realized that gene signatures drawn from microarrays obtained from different studies on the same disorder across parallel cohorts lack consistency [11, 12]. A potentially more effective method to solve this problem is to utilize a network-based approach [13]. As expected, constructing a classifier on the basis of network is one of the proposed novel network-based methods [14].

In the current study, we identified the biomarkers in RA using mutual information network-

based support vector machines (SVM) classifier. Firstly, differentially expressed genes (DEGs) identification was implemented. Then, differential pathways genes (DPGs) which were enriched in differential pathways were screened through *attract* method. Subsequently, the mutual information network was constructed on the basis of the DPGs through context likelihood of relatedness (CLR) algorithm. Simultaneously, the hub genes of the mutual information network were screened, followed by the extraction of the intersection of hub genes and DEGs (named as inter genes). Ultimately, the classification and evaluation was implemented via support vector machines (SVM) with linear kernel. Our analysis might provide a novel insight into the pathogenic processes in RA. Moreover, the screening of epigenetic biomarkers may allow better diagnosis and treatment of RA, and ultimately, offer opportunities to personalize the management of rheumatic disease. The flowchart of our method to predict the significant biomarkers was exhibited in **Figure 1**.

Material and methods

Gene expression dataset and preprocessing

Microarray dataset E-GEOD-45291 [15] was downloaded from Array Express database, whi-

Identification of RA-related genes

ch was based on the A-GEOD-13158 platform of HT_HG-U133_Plus_PM Affymetrix HT HG-U133+ PM Array Plate. Gene expression data of E-GEOD-45291 included 20 control samples, 378 RA samples with an inadequate response to disease-modifying antirheumatic drug therapy (DMARD-IR), 115 RA samples with an inadequate response to tumor necrosis factor inhibition (TNF-IR) and 292 systemic lupus erythematosus (SLE) samples. In this study, only 20 control samples and 115 RA samples (TNF-IR) were selected for further analysis. Then, repeated probes were abandoned. Next, IDs were transformed into the gene symbols. Finally, a total of 20276 genes were obtained.

Identification of DEGs

In the current study, the linear models for microarray data (LIMMA) package [16] in R language was utilized to identify DEGs between control and RA samples. The *P*-values were adjusted using false discovery rate (FDR) based on the Benjamini & Hochberg algorithm [17]. The values of $|\log FC| \geq 2$ and $FDR \leq 0.01$ were applied as the threshold.

Identification of differential pathways and DPGs

In the present study, we used the gene expression data of control and RA to identify the core pathways using attract method [10], and these core pathways showed the most differential expression changes between the two groups. Instead of examining individual genes, this attract method employed GSEA-ANOVA to test the gene sets defined by KEGG for their ability to distinguish between two groups. In briefly, the KEGG enrichment analyses were performed for the 20276 genes in both groups. Subsequently, the pathways with less than 5 genes were deleted. Then, the pathway level was examined through GSEA-ANOVA which was an ANOVA-based implementation of a gene set enrichment algorithm. On the basis of GSEA-ANOVA, an ANOVA model was fitted to each gene and the expression of one gene was regulated by a single factor standing for the groups as different levels of this class. For gene *i* as well as the corresponding expression level in every replicate sample $j = 1, \dots, r_k$ for each group $k = 1, \dots, K$, the fixed effect model was fitted:

$$y_{jk}^{(i)} = u + u_k + \varepsilon_{jk}$$

Where *u* is on the behalf of the overall mean, u_k is the effect of group *k* on the gene's expression, and ε_{jk} denotes the random normal residual error term.

Under the null hypothesis $H_0: u_1 = u_2 = \dots = u_k$, we assume that all *K* group means are equivalent. For group *k*, the mean expression was computed relying to the following formula:

$$y_{.k}^{(i)} = \frac{1}{r_k} \sum_{j=1}^{r_k} y_{jk}^{(i)}$$

According to the ANOVA model, the F-statistic value for gene *i* is counted:

$$F^{(i)} = \frac{MSS_i}{RSS_i}$$

Where MSS_i is the mean treatment sum of squares, as well as captures the amount of variation because of the group-specific effects:

$$MSS_i = \frac{1}{K-1} \sum_{k=1}^K r_k [y_{.k}^{(i)} - y_{..}^{(i)}]^2$$

RSS_i represents the residual sum of squares, and it is calculated using the following formula:

$$RSS_i = \frac{1}{N-K} \sum_{k=1}^K \sum_{j=1}^{r_k} [y_{jk}^{(i)} - y_{.k}^{(i)}]^2$$

Where *N* means the total number of samples, as well as the overall mean is counted:

$$y_{..}^{(i)} = \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{r_k} \sum_{j=1}^{r_k} y_{jk}^{(i)} \right)$$

Because large F-statistic value indicated a strong association while small F-statistic value suggested that the genes showed minimal group-specific expression changes. In an attempt to confirm this relationship more formally, T-test was employed to compare the distribution of log2-transformed F-statistics values from all the pathways to the global distribution of log2-transformed F-statistics values from all the genes with a pathway annotation. Moreover, Benjamini-Hochberg-based method was utilized to adjust the *P* values, and the pathways with FDR less than 1E-09 were considered as attractors. Then, the pathways were ranked in ascending sort of FDR values. The top 5 significant pathways were identified and designated as differential pathways. Furthermore, genes enriched in differential pathways were extracted and named as DPGs.

Identification of RA-related genes

Table 1. The core Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on false discovery rate (FDR) less than 1E-09

Term	Term name	Count	FDR
KEGG	Ribosome	130	5.69E-53
KEGG	Sphingolipid signaling pathway	113	1.93E-37
KEGG	Spliceosome	112	1.33E-34
KEGG	Fc epsilon RI signaling pathway	68	2.51E-31
KEGG	Antigen processing and presentation	70	2.66E-28
KEGG	Ribosome biogenesis in eukaryotes	71	8.78E-26
KEGG	Shigellosis	61	1.37E-22
KEGG	Amino sugar and nucleotide sugar metabolism	48	2.56E-20
KEGG	Nucleotide excision repair	44	9.44E-20
KEGG	Viral myocarditis	55	1.78E-19
KEGG	Acute myeloid leukemia	57	2.51E-19
KEGG	Basal transcription factors	42	3.78E-19
KEGG	Proteasome	43	2.53E-17
KEGG	NOD-like receptor signaling pathway	56	5.10E-16
KEGG	Valine	47	3.82E-15
KEGG	DNA replication	36	5.09E-15
KEGG	Fructose and mannose metabolism	33	8.43E-14
KEGG	SNARE interactions in vesicular transport	34	1.11E-13
KEGG	Asthma	27	4.66E-12
KEGG	Base excision repair	33	5.93E-12
KEGG	Mismatch repair	23	6.78E-12
KEGG	P53 signaling pathway	67	5.17E-11
KEGG	Aminoacyl-tRNA biosynthesis	43	1.69E-10
KEGG	Glycosaminoglycan degradation	19	2.08E-10
KEGG	Thyroid cancer	29	4.14E-10
KEGG	2-Oxocarboxylic acid metabolism	17	5.16E-10
KEGG	Citrate cycle (TCA cycle)	31	6.89E-10

Mutual information network construction and hub genes identification

Mutual information network is a subgroup of network inference approach, and its theoretical foundation is to deduce a link between a couple of nodes when it has a high score on the basis of mutual information [18]. The CLR algorithm [19] is an extension of the relevance network method, which counts the mutual information for each pair of genes and acquires a score related to the empirical distribution of the mutual information values. In the current study, the DPGs were as vertices, and the microarray data were as the mutual information between vertices. Then, CLR algorithm was utilized to compute the edge score for each pair of nodes via an inference method that takes

the square weighted value as input, and then the adjacency matrix was built. Next, Igraph package was used to display the mutual information network.

Further, to investigate the biological importance of nodes in the mutual information network, we analyzed the topological centralities connected to the local scale (degree) which was employed to depict the significance of nodes. Degree gives a count of the number of interactions of a given node via totalizing the number of its neighbouring genes [20]. In our work, the nodes with the top 10% degrees were extracted as the hub genes. Significantly, the intersection of hub genes and DEGs were extracted and defined as inter genes.

Classification and evaluation via SVM

In the current study, SVM [21] with linear kernel were utilized to evaluate the classification performance of the inter genes across RA and control samples. Firstly, all samples were approximately divided into two

parts on the basis of the ratio of 6 to 4, and the 81 samples (17 of control and 64 of RA) were kept to build a train set, the rest 54 samples (3 of control and 51 of RA) constituted a test set to verify the classification models. Then, we used a 5-fold cross-validation (5-CV) to conduct on the training sets to evaluate the potential classification strength of the models. Next, the prediction power on the test sets was implemented.

To assess the classification results of inter genes, several performance measures which can offer different views were employed. Accuracy (ACC), one of the most popular measures in SVM classification, often misleads about the actual classification quality of a given classifier because it does not take into account

Identification of RA-related genes

Table 2. The top 5 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and the intersection genes between differentially expressed genes and hub genes

Category	Term name	Intersection genes	FDR
KEGG	Ribosome	RPL10, RPL22L1	5.69E-53
KEGG	Sphingolipid signaling pathway	TRADD, BID	1.93E-37
KEGG	Spliceosome	SNRPB, SF3A2, PRPF8	1.33E-34
KEGG	Fc epsilon RI signaling pathway	MAP2K2	2.51E-31
KEGG	Antigen processing and presentation	MAP2K2	2.66E-28

Note: FDR, false discovery rate.

the nature of the incorrect predictions. Consequently, the area under the receiver operating characteristics curve (AUC) is a better measure for evaluating the predictive ability of machine learners, relative to accuracy [22]. Matthews correlation coefficient (MCC) ranges from -1 to +1. The value of +1 represents the total agreement and -1 indicates total disagreement. In addition, specificity means the ratio of negative correctly classified to the practical amount of negatives as well as controls type I errors. Sensitivity is regarded as the ratio of correctly classified positives to the actual number of positives and controls type II errors. In our study, the measures of AUC, MCC, specificity as well as sensitivity were utilized to give us an adequate overview of the classification's performance.

Results

Identification of DEGs and DPGs

Based on $|\log_{2}FC| \geq 2$ and $FDR \leq 0.01$, a total of 339 DEGs including 4 down- and 335 up-regulated genes were identified. Overall, 27 core pathways were extracted based on $FDR < 1E-09$, as shown in **Table 1**. We defined the top 5 pathways as differential pathways, and the genes originated from the differential pathways were named after DPGs. The top five pathways were the ribosome, sphingolipid signaling pathway, spliceosome, Fc epsilon RI signaling pathway and antigen processing and presentation. These 5 pathways were significantly different between control and RA group tested at the $1E-09$ level. Additionally, these were the most "representative pathways" in the sense that these pathways contained larger numbers of genes. Overall, there were 448 DPGs in these 5 differential pathways.

Mutual information network construction and hub genes identification

We applied the mutual information to identify the hub genes on the basis of the 448 DPGs in 5 differential pathways. In general, there were 44 hub genes. Then, the intersection of hub genes and DEGs were extracted. Overall, 8 inter genes such as small nuclear ribonucleo protein polypeptides B and B1 (SNRPB, degree = 429), mitogen-activated protein kinase kinase 2 (MAP2K2, degree = 429), tumor necrosis factor receptor (TNFR)-associated death domain (TRADD, degree = 420), splicing factor 3a, subunit 2, 66 kDa (SF3A2, degree = 420), ribosomal protein L10 (RPL10, degree = 416), BH3 interacting domain death agonist (BID, degree = 410), pre-mRNA processing factor 8 (PRPF8, degree = 405), and ribosomal protein L22-like 1 (RPL22L1, degree = 399) were identified. Based on the differential pathways (**Table 2**), we found that MAP2K2 was involved in the two pathways of Fc epsilon RI signaling pathway and antigen processing and presentation.

Classification and evaluation via SVM

SVM was used in our study to assess the classification performance of the 8 inter genes between RA and control subjects. During the 5-CV evaluation, the classification results of inter genes were ideal to distinguish all samples correctly. Moreover, our results achieved the highest AUC score of 1.00, MCC score of 1.00, specificity of 1.00, and sensitivity of 1.00. Accordingly, these 8 genes can classify unknown samples from RA patients with high accuracy.

Discussion

To our knowledge, RA is the most frequently observed type of arthritis. Worriedly, the molec-

Identification of RA-related genes

ular mechanisms of RA have not been elucidated clearly. Thus, advances in understanding the pathogenesis of RA have significant implications for the development of new therapeutics and drug research. Our study demonstrates that several promising genes including *MAP2K2*, *TRADD* and *BID* might provide important clues to the development process of RA.

In the current study, one immune pathway of antigen processing and presentation, involved in *MAP2K2* gene, was identified in RA samples, suggesting that immune response is participated in this disorder. This finding was consistent with former studies [23, 24]. Upon stimulation, TLR2 activates the *MAP2K1* and *MAP2K2*, and in turn they switch on the extracellular signaling regulated ERK1/2 [25]. Of note, ERK1/2 has been demonstrated to play an important part in mediating the TNF- α secretion [26]. Moreover, the main inflammatory cytokine, TNF- α is a key player in inflammatory disorders including RA [23]. Inhibition of TNF- α expression can effectively suppress the occurrence of inflammation and anti-TNF- α therapy improves the clinical symptoms of RA [27]. Thus, *MAP2K2* might be an attractive biomarker for RA diagnosis and therapeutic intervention.

TRADD was another hub gene in our study. *TRADD* is a central adaptor in the TNFR1 signaling complex which regulates cell death as well as inflammatory signals [28]. Moreover, *TRADD* interacts with TNF receptor associated factor 2 (TRAF2) which then induces the activation of NF- κ B signaling [29]. Moreover, it is well established that NF- κ B pathway exerts key roles in macrophage production of TNF- α and destructive metal matrix proteinase (MMPs) [30]. Remarkably, Houseman and colleagues have indicated that determination of baseline MMP-3 can provide additional prognostic information for RA patients [31]. Accordingly, these results indicate that *TRADD* might be applied to monitor the NF- κ B signaling pathway in RA.

In the current study, *BID* was identified as a hub gene. *BID* is a member of B-cell lymphoma 2 (Bcl-2) family which is made up of anti-apoptotic proteins and pro-apoptotic proteins. Moreover, Bcl-2 has been indicated to be more highly expressed in RA synovial tissue [32]. To the best knowledge, apoptosis plays a crucial role in the development of RA [33], which is initiated

via death receptors or mitochondria-related pathways. Fas is a major death receptor and Fas-associated via death domain can lead to the activation of caspase-8 [34]. Moreover, caspase-8 results in apoptosis through cleaving *BID*, which lead to the mitochondrial dysfunction. The local modulation of mitochondrial pathway or inhibition of the expression of anti-apoptotic molecules, could be therapeutically beneficial in RA [33]. Thus, *BID* might be a therapeutic target for the diagnosis and treatment of RA.

In summary, *MAP2K2*, *TRADD* and *BID* might be used as genetic markers for RA diagnosis and treatment in the future. Our study might offer bioinformatics support to further study the mechanism underlying RA. Nevertheless, the drawbacks were taken into consideration. The identified genes in the present study were predicted through the bioinformatics approaches, yet detailed analyses of their expression have not been conducted by experiments. Related experimental data are required to verify the effects and mechanisms of *MAP2K2*, *TRADD* and *BID* in RA in later work.

Disclosure of conflict of interest

None.

Address correspondence to: Pengji Gao, Department of Traumatology, Linyi People's Hospital, Linyi 276-000, Shandong Province, China. E-mail: gaopengjisd@163.com

References

- [1] Strand V and Singh JA. Improved health-related quality of life with effective disease-modifying antirheumatic drugs: evidence from randomized controlled trials. *Am J Manag Care* 2008; 14: 234-254.
- [2] Firestein GS. Evolving concepts of rheumatoid arthritis. *Nature* 2003; 423: 356-361.
- [3] GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; 385: 117-171.
- [4] Iwahashi M, Yamamura M, Aita T, Okamoto A, Ueno A, Ogawa N, Akashi S, Miyake K, Godowski PJ, Makino H. Expression of Toll-like receptor 2 on CD16+ blood monocytes and synovial tis-

Identification of RA-related genes

- sue macrophages in rheumatoid arthritis. *Arthritis Rheum* 2004; 50: 1457-1467.
- [5] McGarry T, Veale DJ, Gao W, Orr C, Fearon U and Connolly M. Toll like receptor 2 (TLR2) induces migration and invasive mechanisms in rheumatoid arthritis. *Arthritis Res Ther* 2015; 17: 153.
- [6] McInnes IB, Leung BP, Sturrock RD, Field M and Liew FY. Interleukin-15 mediates T-Cell dependent regulation of tumor necrosis factor production in rheumatoid arthritis. *Nat Med* 1997; 3: 189-95.
- [7] Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM and Spoecker JM. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004; 75: 330-337.
- [8] Weyand CM and Goronzy JJ. Association of MHC and rheumatoid arthritis. HLA polymorphisms in phenotypic variants of rheumatoid arthritis. *Arthritis Res* 2000; 2: 212-216.
- [9] Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; 28: 27-30.
- [10] Mar JC, Matigian NA, Quackenbush J and Wells CA. attract: A method for identifying core pathways that define cellular phenotypes. *PLoS One* 2011; 6: e25445.
- [11] Ein-Dor L, Zuk O and Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *P Nat Acad Sci* 2006; 103: 5923-5928.
- [12] Ein-Dor L, Kela I, Getz G, Givol D and Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 2005; 21: 171-178.
- [13] Chuang HY, Lee E, Liu YT, Lee D and Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007; 3: 140.
- [14] Zhu Y, Shen X and Pan W. Network-based support vector machine for classification of microarray samples. *BMC Bioinformatics* 2009; 10: Suppl 1: S21.
- [15] Bienkowska J, Allaire N, Thai A, Goyal J, Plavina T, Nirula A, Weaver M, Newman C, Petri M and Beckman E. Lymphotoxin-LIGHT Pathway Regulates the Interferon Signature in Rheumatoid Arthritis. *PLoS One* 2014; 9: e112545.
- [16] Smyth GK. Limma: linear models for microarray data. In: *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer 2005. pp. 397-420.
- [17] Benjamini Y, Drai D, Elmer G, Kafkafi N and Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001; 125: 279-284.
- [18] Meyer PE, Kontos K, Lafitte F and Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J Bioinform Syst Biol* 2007; 2007: 8-8.
- [19] Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ and Gardner TS. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007; 5: e8.
- [20] Haythornthwaite C. Social network analysis: An approach and technique for the study of information exchange. *Library & information science research* 1996; 18: 323-342.
- [21] Chang CC and Lin CJ. LIBSVM: A library for support vector machines. *ACM T Intel Syst Tec* 2011; 2: 27.
- [22] Huang J and Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE TKnowl Data En* 2005; 17: 299-310.
- [23] Russo C and Polosa R. TNF- α as a promising therapeutic target in chronic asthma: a lesson from rheumatoid arthritis. *Clin Sci* 2005; 109: 135-142.
- [24] Kaneko K, Miyabe Y, Takayasu A, Fukuda S, Miyabe C, Ebisawa M, Yokoyama W, Watanabe K, Imai T and Muramoto K. Chemerin activates fibroblast-like synoviocytes in patients with rheumatoid arthritis. *Arthritis Res Ther* 2011; 13: R158.
- [25] Eliopoulos AG, Wang CC, Dumitru CD and Tsi-chlis PN. Tpl2 transduces CD40 and TNF signals that activate ERK and regulates IgE induction by CD40. *EMBO J* 2003; 22: 3855-3864.
- [26] Rousseau S, Papoutsopoulou M, Symons A, Cook D, Lucocq JM, Prescott AR, O'Garra A, Ley SC and Cohen P. TPL2-mediated activation of ERK1 and ERK2 regulates the processing of pre-TNF α in LPS-stimulated macrophages. *J Cell Sci* 2008; 121: 149-154.
- [27] Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, Smolen JS, Weisman M, Emery P, Feldmann M, Harriman GR, Maini RN; Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *New Engl J Med* 2000; 343: 1594-1602.
- [28] Aggarwal BB. Signalling pathways of the TNF superfamily: a double-edged sword. *Nat Rev Immunol* 2003; 3: 745-756.
- [29] Walczak H. TNF and ubiquitin at the crossroads of gene activation, cell death, inflammation, and cancer. *Immunol Rev* 2011; 244: 9-28.
- [30] Tas SW, Vervoordeldonk MJ, Hajji N, May MJ, Ghosh S and Tak PP. Local treatment with the selective I κ B kinase β inhibitor NEMO-binding domain peptide ameliorates synovial inflammation. *Arthritis Res Ther* 2006; 8: R86.

Identification of RA-related genes

- [31] Houseman M, Potter C, Marshall N, Lakey R, Cawston T, Griffiths I, Young-Min S and Isaacs JD. Baseline serum MMP-3 levels in patients with Rheumatoid Arthritis are still independently predictive of radiographic progression in a longitudinal observational cohort at 8 years follow up. *Arthritis Res Ther* 2012; 14: R30.
- [32] Perlman H, Georganas C, Pagliari LJ, Koch AE, Haines K and Pope RM. Bcl-2 expression in synovial fibroblasts is essential for maintaining mitochondrial homeostasis and cell viability. *J Immunol* 2000; 164: 5227-5235.
- [33] Peng S. Fas (CD95)-related apoptosis and rheumatoid arthritis. *Rheumatology* 2006; 45: 26-30.
- [34] Liu H and Pope RM. The role of apoptosis in rheumatoid arthritis. *Curr Opin Pharmacol* 2003; 3: 317-322.