© Kamla-Raj 2018

Int J Hum Genet, 18(2): 116-128 (2018)
DOI: 10.31901/24566330.2018/18.2.695

Prediction of Key Transcription Factors-Genes Regulatory Network of Ankylosing Spondylitis Based on Transcription Factor Prognosis System

Hai-feng Song[1*], Jun-wei Gao[1], Shu-gang Li[2], Ying-peng Xia[1], Qing-feng Shen[1], Shi-bo Ma[1] and Yang Liu[1]

[1]Department of Spine Surgery, Tianjin Union Medicine Center, Tianjin, 300121, P.R. China
[2]Department of Orthopedic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing 100730, P.R. China

KEYWORDS Ankylosing Spondylitis. Different Expressed Genes. Intervertebral Joints. Transcription Factors. Patient and Public Involvements Network

ABSTRACT Ankylosing spondylitis (AS), characterized by inflammation in dorsal and intervertebral joints, is a rheumatic disease that leading to varying degrees of structural and functional impairment. Current therapeutic methods have some disadvantages, so clarifying the underlying mechanism and developing new treatment for AS are urgent. The aim of this article is to identify the key transcription factors (TFs) of AS and find new ways for AS treatment. First, the researchers used transcription factor prognosis system, which included enrichment analysis, directly impact value and indirect impact value to find the key TFs targeted different expressed genes. Then the patient and public involvements network was used to integrate the results of the above three methods for analysis. Finally, the researchers obtained the optimal TFs. In summary, this study's analysis based on bioinformatics methods discovered optimal 30 key TFs related to AS. This study may provide important potential therapeutic targets for AS.

INTRODUCTION

Ankylosing spondylitis (AS), characterized by inflammation in dorsal, thoracic and intervertebral joints, is a chronic, systemic, rheumatic disease that affects the axial bones and sacroiliac joints, leading to varying degrees of structural and functional impairment (Borse et al. 2017). Recent studies have shown that AS patients have an increased risk of heart disease, including coronary artery disease. The prevalence of AS was 0.15 percent ~ 0.86 percent (Deminger et al. 2017). The main concern in patients with AS is progression to ankylosis and abnormal fixation state (such as thoracic kyphosis) (Qu et al. 2017). The treatment for AS is usually used with repair antirheumatic drugs (DMARDs), nonsteroidal antiinflammatory drugs (NSAIDs), and nonpharmacological interventions such as physiotherapy (Wang et al. 2017). Studies have suggested that it should use anti-tumor necrosis factor (anti-TNF) drugs to treat highly disease activity AS patients (de Machado et al. 2016). Although these drugs can relieve clinical symptoms, they may still cause many adverse reactions such as relapse, disability and poor response etc. So clarifying the underlying mechanism and developing new treatment for AS are very urgent.

In recent years, some researchers begin to explore the therapeutic method of AS from the direction of gene regulation (Chen et al. 2017). And studies have reported that regulation of gene expression through transcription factors (TFs) is a basic mechanism (Joseph et al. 2017). TFs are DNA-binding proteins that regulate gene expression by binding to the promoter regions near gene transcription initiation sites (Fang et al. 2017). Studies of TF will be conducted to reveal the complex regulation mechanism of organisms (Rudnik et al. 2017). Up to now, hundreds of domains that exhibit TF activity have been identified (Sikdar et al. 2017). One of the most important reasons for phenotypic differences in organisms is regulation of gene expression achieved through transcription factor networks and their target genes (Zhang et al. 2017). TF include activators, coactivators, basic factors and some regulators (Fishwick et al. 2017). Studies have reported that putative TF-gene reg-

**Address for correspondence:*
Hai-feng Song
Department of Spine Surgery, Tianjin Union Medicine Center, No. 190 Jie Yuan Road, Hongqiao District, Tianjin, 300 121, R.P. China
Telephone: +86-022-87721989
Fac: +86-022-87721989
E-mail: drsonghf@163.com

ulatory network of AS and sarcoidosis have been identified by using biostatistics and bioinformatics tools (Xu et al. 2018). Based on different approaches or models, performance evaluation in revealing common TFs is not straightforward. In addition, TFs are an active research topic and the corresponding databases are also growing. With the development of human knowledge grows and models evolve, the available algorithms will need to be periodically re-evaluated.

In this study, the researchers presented a predictive system (transcription factor prognosis system, TFpro) to determine the AS-related TFs and discovered optimal 30 key TFs related to AS. This study may provide important potential therapeutic targets for AS.

METHODOLOGY

Screening TFs from Different Genes

Expression Data Taken From the GEO or Array Express Data Set

In the present study, the gene expression profile dataset no. GSE11886 deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/). First, the researchers combined multiple probes which correspond to the same gene, and the average value of the plurality of probes is selected as the expression value of the gene. Second, the annotation information was modified, the column name corresponding to the line, renamed "groups". The data has a total of 80 data (samples). The first group (9 samples, name IFN-Control) and the second group (7 samples, name IFN-Patient) were the first analysis. The third group (9 samples, name None-Control) and the fourth group (8 samples, name None-Patient) were the second analysis. The expression profiles of 13715 genes (first analysis) and 13715 genes (second analysis) were obtained by mapping between probes and genes.

Identification of Differentially Expressed Genes (DEGs) Using LIMMA

Firstly, according to different samples of the group (control group and experimental group), the differential expression of genes were calculated by using LIMMA package. If the number of DEGs is less than 300, the top 300 genes (bigger difference of expression value) were screened as the differential genes. Then the DEGs of 300 (the first group) and 300 (the second group) were screened out respectively. The t-test and F-test of gene expression matrix were performed by LIMMA. The lmFit function was used to linearly fit the data, eBayes statistics and used false discovery rate (FDR) to correct P values. The linear model was extracted after testing and different genes were extracted on the basis of |log fold change (FC)| >2 and FDR < 0.05.

Enrichment Analysis of TF Genes by Fisher's Test

Before enrichment analysis, a list of genes that affected by TF must be established firstly. Then TF genes were analyzed by using this list in conjunction with the classical Fisher's test. Here, Fisher's exact test was used to determine if the two overall ratios (the random ratio and the experimental ratio) were equal. The null hypothesis assumed that the two population ratios were equal (H0: p1 = p2). Alternative hypotheses might be the left tail (p1 < p2), the right tail (p1 > p2), or the two tail (p1 'd p2).

Directly Impact Value of TF

TF with a high level impact can be identified by converting the fold change in logarithmic transformation expressed and FDR adjusting P value into score of a single gene. The G value of each gene in the samples in each group was calculated by the following analytical formula.

$$G^x_s = |L^x_s| (-\log10P^x_s)$$

L is the difference logFC value of the relevant genes and P is the difference P value calculated by using limma.

This formula shows the changes of target genes for each TF. If a TF regulates many genes, and these genes changes are relatively large, the gene's G score will be higher. The more common TFs can be found by this method. At the same time, if comparing G score, the average G score and the number of genes three aspects are taken into account, a rank score is gotten. The lower the score, the greater is this gene influent in comprehensive consideration process.

Indirect Impact Value of TF

TF can not only affect the expression of changes, but also compose co-expression rela-

tionship with its target gene. To assess the importance of each TF co-expression in the network, two network sources of information (STRING database (Szklarczyk et al., 2011)and TF library) were used to calculate its effect on the local neighborhood. The TF library comes from three databases, ITFP, Marbach2016 and TRRUST. These two databases contain different types of interactions. TF library provides protein-DNA interactions for TF with known binding sites in the promoter region of the gene. This represents a low-level, targeted regulatory interaction network. STRING is an interactive metadata library that contains various interactions of protein-protein, protein-DNA and protein-RNA interactions and biological pathways. This provides a view that directly and indirectly affects the interaction of gene expression. The final STRING correlation value is obtained by multiplying the Pearson correlation coefficient of the STRING database score and the actual data.

To calculate the effect, the local network neighbourhood of TFs were weighted by the gene impact and different genes were identified by the LIMMA method. Combining these two parts, the correlation coefficient between TFs and the target genes was obtained by the following formulas.

TFCors,r = StringDatabaseCor (s,r)

TFCors,r = TFDatabaseCor (s,r) = 0.7 + N* 0.1

Transcription factor correlation coefficients were STRING database scores or TF database scores. N is the number of records of the database of s, r two genes with an interactive relationship, and the value is 0-1.

Using Patient and Public Involvement (PPI) Network Analyze All TF Results

Weight Sum

The use of distance weighting makes genes far from direct regulation less susceptible to scores. Edge-weighting compensates for highly prevalent transcription factors and prevents them from accepting artificial high scores by adjusting a large number of genes that express little difference. The researchers think that TF regulates ten Gsx genes = 100 is more important than that TF regulates 1,000 genes with Gxs = 1. The formulas for performing this weighted sum are as follows.

$$(1) \qquad N^{s}_{x,n}=\sum^{s}_{r\in Vz} Ps,r\cdot G^{s}_{r}\cdot \frac{1}{Lr,n}\cdot \frac{1}{0r,n}$$

$N_{sx,n}$ and $N_{x,ns}$ are the weight sum of TF. Ps, r is the correlation coefficient between TF and the affected genes. $L_{r,n}$ is the distance between the gene and the TF, which is the grade value. Or, n is the connectivity of the parent node.

(2) Ps,r=TFCor(s,r)* Pearson(s,r)

The correlation coefficient between TF and the affected gene is equal to the co-expression (Pearson) correlation coefficient between the two genes multiply the correlation coefficient of TF in the given background data.

PPI Network

PPI network construction generally was divided into two parts, one part is the control group and another part is the experimental group. Changes of those genes in two groups' co-expression are generally observed. The first step calculates the co-expression of two genes in control group. The second step calculates the co-expression of two genes in experimental group. The third step calculates the difference between the two genes co-expression values between the two groups and takes the difference as an absolute value. Finally, this value is first divided by the maximum value, then correct, and then multipy by the background value to correct. The background value is the relationship between the two genes given in the database. The closer the relationship, the greater the value, and the value was between 0-1.

Analysis the Coverage of TFs

TFs should be covered all different genes. Based on TF genes obtained from all results, a Venn diagram was finally obtained. The highest coverage of TF combinations needed to be found. The collection of the top 10 best genomic combinations obtained for each of these approaches was the "TFall" group.

RESULTS

The Acquisition of Differential Genes

The researchers extracted the genes that satisfy the following conditions in the tested linear model: | logFC | > 2 and p < 0.05, a total of 300 DEGs were identified. The top 10 DEGs and bar charts for the difference genes were shown in

Figure 1A. The most significant DEGs were CLEC10A, GCH1, ICOS, IFNG and CECR6. The relative venn diagrams are shown in Figure 1 B. There are 300 DEGs but 0 TF target gene. So the researchers screened that no TFs were contained in differential genes.

Enrichment Results of Gene Quantity Distribution

In the quantitative profile, it can be seen that a small number of TFs were enriched in the number of target genes, while most of the TFs were not abundant in abundance (Fig. 2A). For the first 10 genes in the data, in-depth analysis could be carried out. The target genes for these genes are abundant in differential genes. However, the presence of a large number of target genes may also be more commonly caused by genes regulated by these TFs, such as TFs associated with cell cycle regulation. Therefore, in-depth analysis of these TFs to find out the more affecting TF is more critical in disease state (Fig. 2B). At the same time, if these top 10 TFs affect most of the differential genes, it is likely that these TFs are keys to regulation. The results of TF enrichment analysis were shown in Table 1. The enriched set of transcription factors were RAD54B, NASP, DTL, CENPA, SMARCC1, NCAPD2, DEPDC1B, EXO1, RBM12, FBXO5 (the first time), PARP9, TP53, STAT4, FOXO1, HOXC10, RREB1, FOXO3, PRDM1, IFIT2, MYOD1 (the second time). The relative venn diagram is shown in Figure 2C. There were 22 transcription factor genes were obtained by enrichment analysis in 300 DEGs. This result suggests that those transcription factors rich in genes may be the focus of our attention.

The Direct Impact Value of TFs

The specific influence of these TFs in the TF set isshown in Table 2. There were 9 TF genes. G score, Ave score, Num genes and rank_p value of every TF gene were shown in the table. The highest G score was 30.43, which was SP1, and the number genes of it is 8324. Firstly, the relationship between G value and the affected genes was shown Figure 3A. The influence of TFs increases with the number of genes that can be regulated. However, if the influence deviates from the trajectory and appears above the line, the TF is more likely to be specifically regulated in this study. Then, the relationship between influence and average is shown in Figure 3B. The average G value was about 0.13. The middle area is the point that should be paid more attention. The influential TFs generally did not have large

Table 1: Transcription Factors (TFs) enrichment analysis

TF name	*TF p value*	*False discovery rate (FDR)*	*Num genes*
PARP9	4.37E-08	4.74E-05	6
TP53	1.13E-05	0.006130333	102
STAT4	4.17E-05	0.015092466	4
FOXO1	6.20E-05	0.016858599	115
HOXC10	9.39E-05	0.018333167	70
RREB1	0.000110559	0.018333167	146
FOXO3	0.000118061	0.018333167	138
PRDM1	0.000144592	0.019646394	113
IFIT2	0.000179388	0.021666126	3
MYOD1	0.000267012	0.029024249	89

Table 2: The specific influence of Transcription Factors (TFs)

TF gene	*G score*	*Ave score*	*Num genes*	*rank_p*
SP1	30.42915848	0.003655593	8324	0
TCFL5	0.047013508	0.047013508	1	0
SLA2	0.038881172	0.038881172	1	0.02988
PPP1R13L	0.038822993	0.038822993	1	0.03031
SUGP1	0.023919907	0.023919907	1	0.2411
FLI1	21.63437033	0.003710233	5831	0.24521
HF1H3B	21.6422333	0.003633076	5957	0.24586
EGR	21.63549025	0.003628898	5962	0.2461
TRIM31	0.022983612	0.022983612	1	0.26105

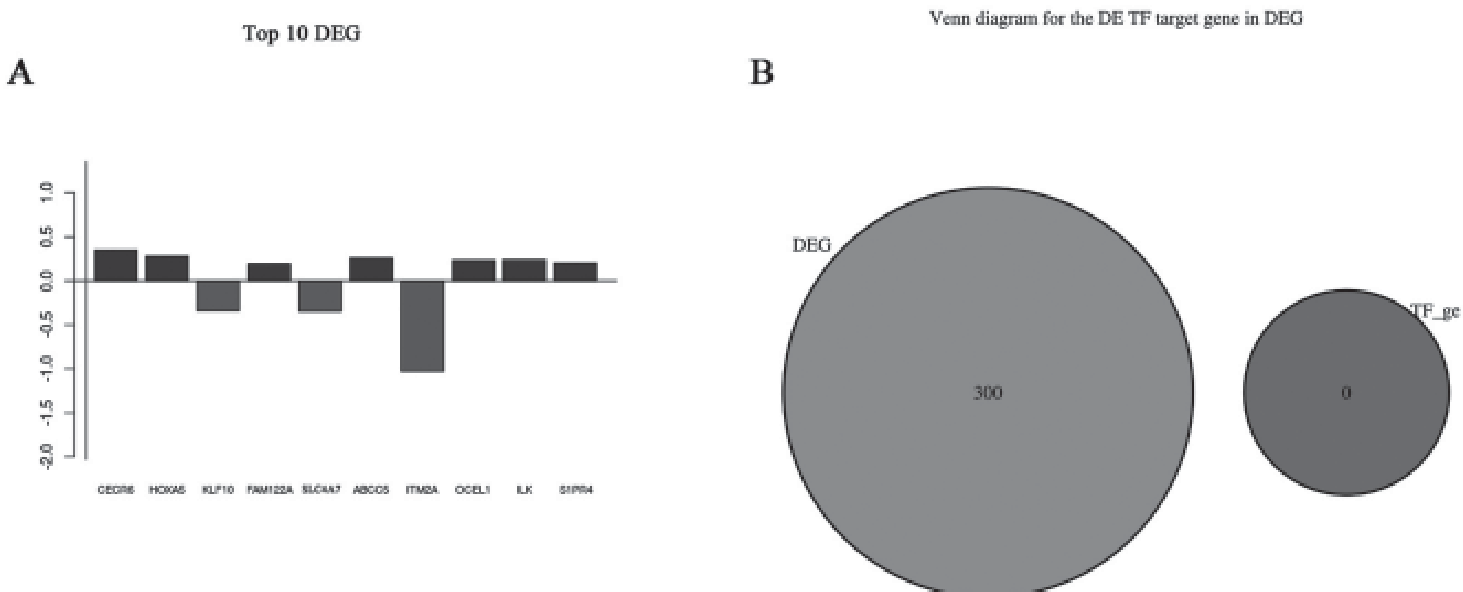

Fig. 1. The enrichment analysis of ankylosing spondylitis differentially expressed genes. The abscissa is top 10 differentially expressed genes. The ordinate is | log FC | for these 10 genes, respectively.
Source: Author

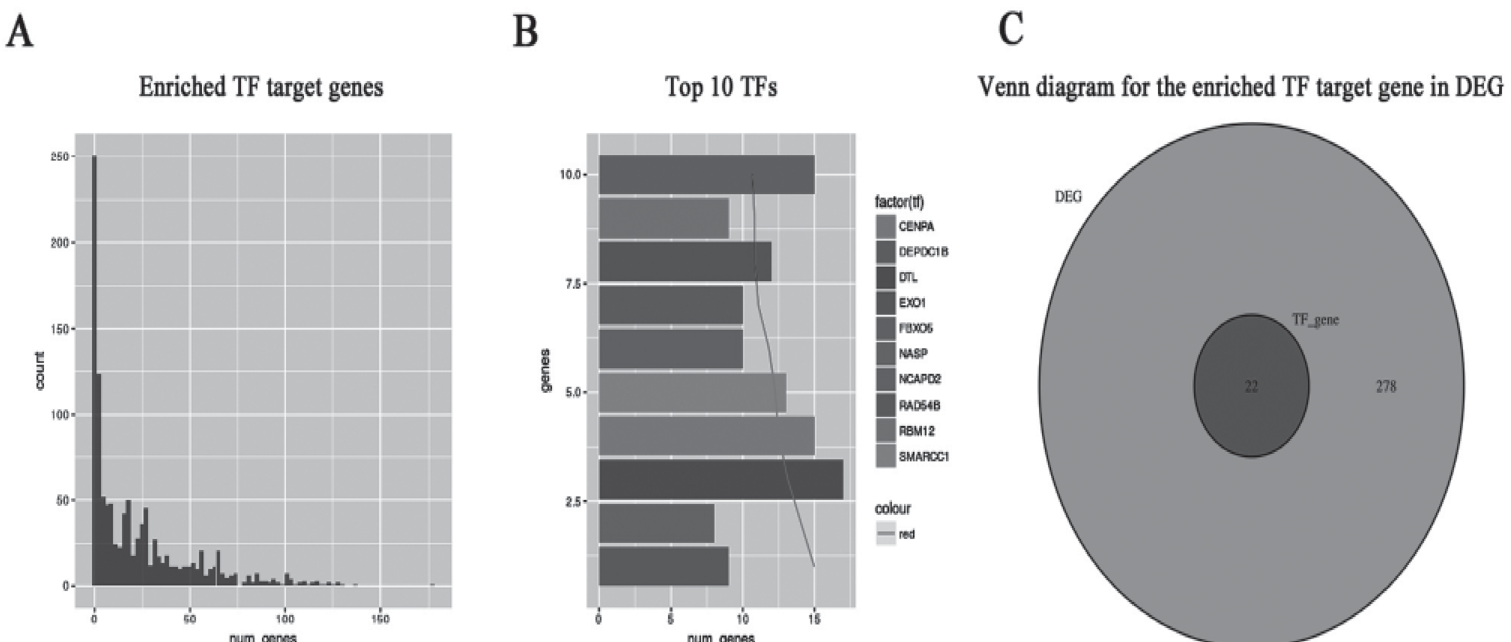

Fig. 2. Enrichment of gene quantity distribution. (A) A small number of transcription factors are enriched in the number of target genes, while most of the transcription factors are not abundant in abundance. So those transcription factors rich in genes may be the focus of our attention. (B) The top 10 TFs, which located to the left of Panel A, are our focus. (C) Venn diagram for the enriched TF target gene in DEG
Source: Author

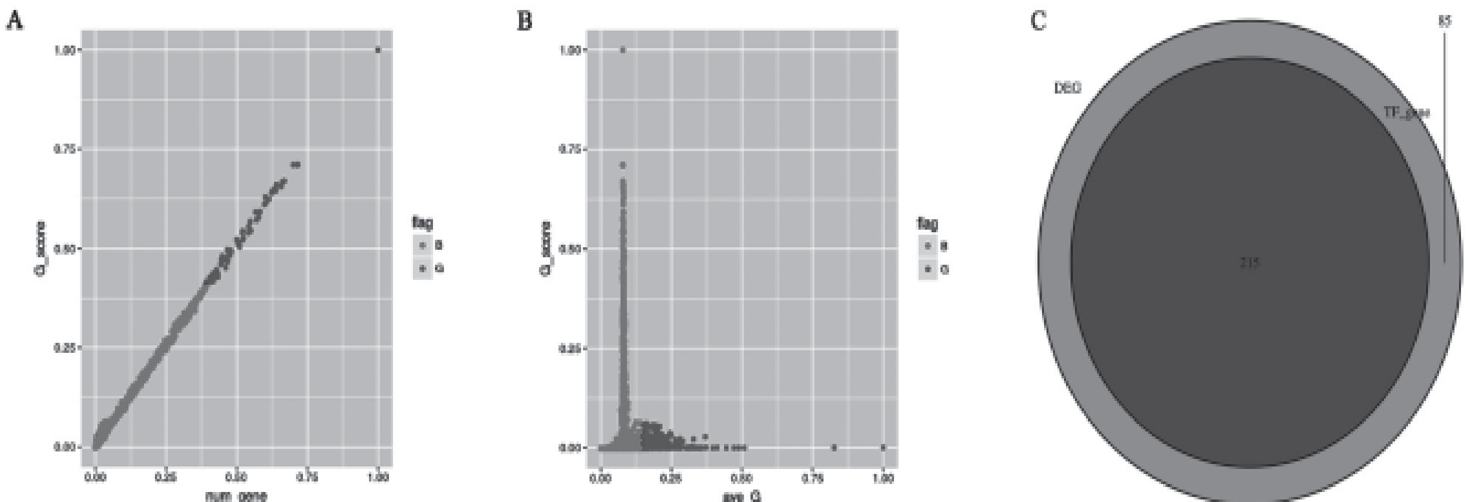

Fig. 3. The relationship between G value and the affected genes. (A) The light color points indicate the rank value is less than 0.5, while the deep color points indicate the rank value is above 0.5. (B) The light color points indicate the average of transcription factors below the overall mean plus standard deviation. The deep color points indicate the average of transcription factors above the overall mean plus standard deviation. We should focus on the middle area. (C) Target genes of the TF with top G in the expression profile in DEG.
Source: Author

average values, while the average large TFs generally gathered in the G value smaller area. At the same time, according to the influence of TFs, the top 10 TFs affected most of the difference genes. If so, it was possible that these TFs might be included in the regulatory genes. These genes were SP1, TCFL5, SLA2, PPP1R13L, SUGP1, FLI1, HF1H3B, EGR, TRIM31, HR (the first time) and SP1, SLA2, HF1H3B, TCFL5, EGR, FLI1, SP4, AP1, SP3, TFAP2 (the second time). The relative venn diagram is shown in Figure 3C, there were 215 transcription factor genes in 300 DEGs were obtained by using indirect influence value method.

The Indirect Impact Value of TFs

This local network was limited to a maximum of three edges. The effect node at each edge was further reduced from the seed TFs, which was located on and depends on the extent of its parental prominence. The relative venn diagram is shown in Figure 4, there were 162 transcription factor genes in 300 DEGs were obtained using the direct impact value method.

PPI Final Network Edge Value Distribution

Information about the network edge distribution was shown in PPI distribution (Fig. 5A). In Figure 5B, the density of border values mostly converges in the middle. Those have a high degree of co-expression, that the collection to the right of the figure, need to be attention. From Figure 5B (the distribution graph of G network scores), it can be seen that the G value is not evenly distributed, and most of the G network score of TFs is not high. Therefore, those well-performing TFs need to be screened. The network diagram of the three TFs (TP53, NR3C1 and MAX) of the highest network access G value is shown in Figure 6. Finally, according to the influence of TFs, the top 10 TFs affected most of the difference genes. If so, it is possible that these TFs have to be included in the regulatory genes. These genes were TP53, NR3C1, MAX, RELA, EP300, CREBBP, ELK1, SPI1, SRF, RREB1 (the first time) and SPI1, NR3C1, TP53, POU2F1, RELA, RREB1, ELK1, EP300, TCF3, CREBBP (the second time).

Analysis of TF Coverage

As shown in Figure 7, we can find the optimal coverage of differential genes associated with AS, then find the optimal combination. The enriched-TFt represents TFs-related differential genes that are screened by enrichment. The Gnet-TFt represents TFs-related differential genes that are screened by indirect impact value. The G-TFt represents TFs-related differential genes that are screened by direct impact value. The TFt-DEG represents TFs-related differential genes that are screened by the above three methods. The first 10 combinations of the best genotypes obtained for each method were the "TFall" group. There were three useful methods, enrichment analysis, direct and indirect impact value. So, There were 30 genes, which are shown in Table 3 for details.

Table 3: The best genotypes "TFall" group

	Gene		*Gene*		*Gene*
1	RAD54B	11	SP1	21	TP53
2	NASP	12	TCFL5	22	NR3C1
3	DTL	13	SLA2	23	MAX
4	CENPA	14	PPP1R13L	24	RELA
5	SMARCC1	15	SUGP1	25	EP300
6	NCAPD2	16	FLI1	26	CREBBP
7	DEPDC1B	17	HF1H3B	27	ELK1
8	EXO1	18	EGR	28	SPI
9	RBM12	19	TRIM31	29	SRF
10	FBXO5	20	HR	30	RREB1

DISCUSSION

In this study's report, potential TFs target genes related to AS were screened by enrichment, direct and indirect impact value. At last, the optimal 30 TFs, including FLI1, EP300 and so on, were predicted.

Regulation of eukaryotic gene expression is one of the most advanced fields in molecular biology. And regulation of transcription level is the most important in gene expression (Du et al. 2017). Due to the interaction between protein and protein, protein and DNA, and the formation of complex macromolecular complexes, the regulation of eukaryotic transcription level is a multi-level complex process (Majewska et al. 2017). TFs, trans-acting factors, can directly or indirectly recognize or bind to the 8 bp to 112 bp core sequence of cis-acting elements in transcriptional regulation. Besides, TFs are involved in regulating the transcriptional efficiency of target genes (Carrillo et al. 2017). Thus, the interaction between TFs and regulatory sequences is central to determine a gene expressed or not (Lozano et al. 2017). With the in-depth study of TFs

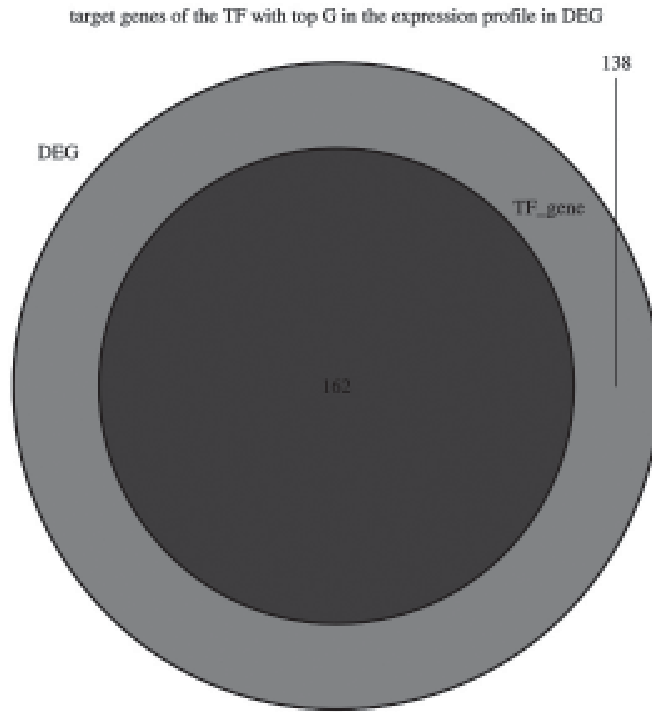

Fig. 4 target genes of the TF with top G in the network in DEG.

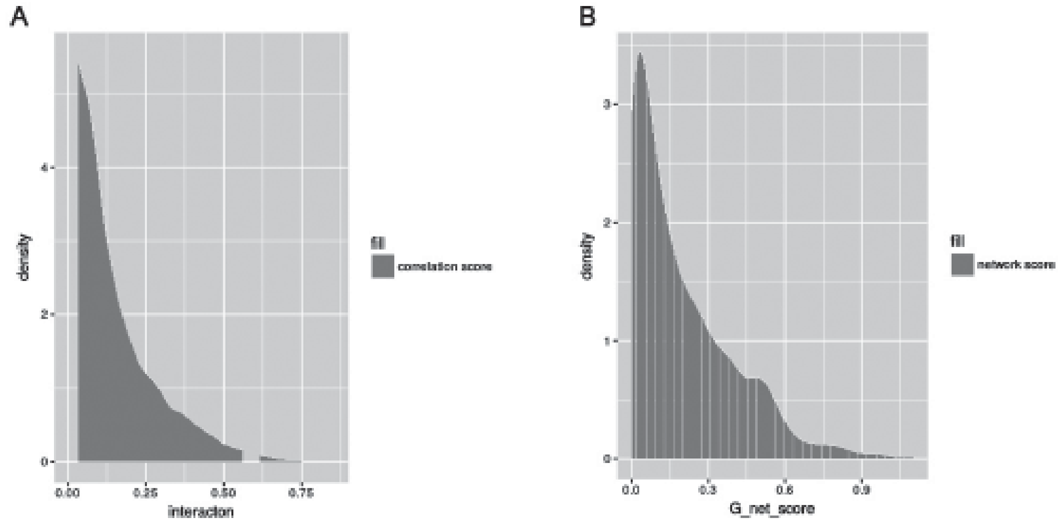

Fig. 5. Distribution of PPI final network edge value. (A) G values are not evenly distributed, and most of the transcription factors have a low G-score. (B) As can be seen from the figure, the density of edge values mostly converges in the middle. Therefore, we should pay attention to those who have a higher value, that is, the figure on the right side of the collection.
Source:Author

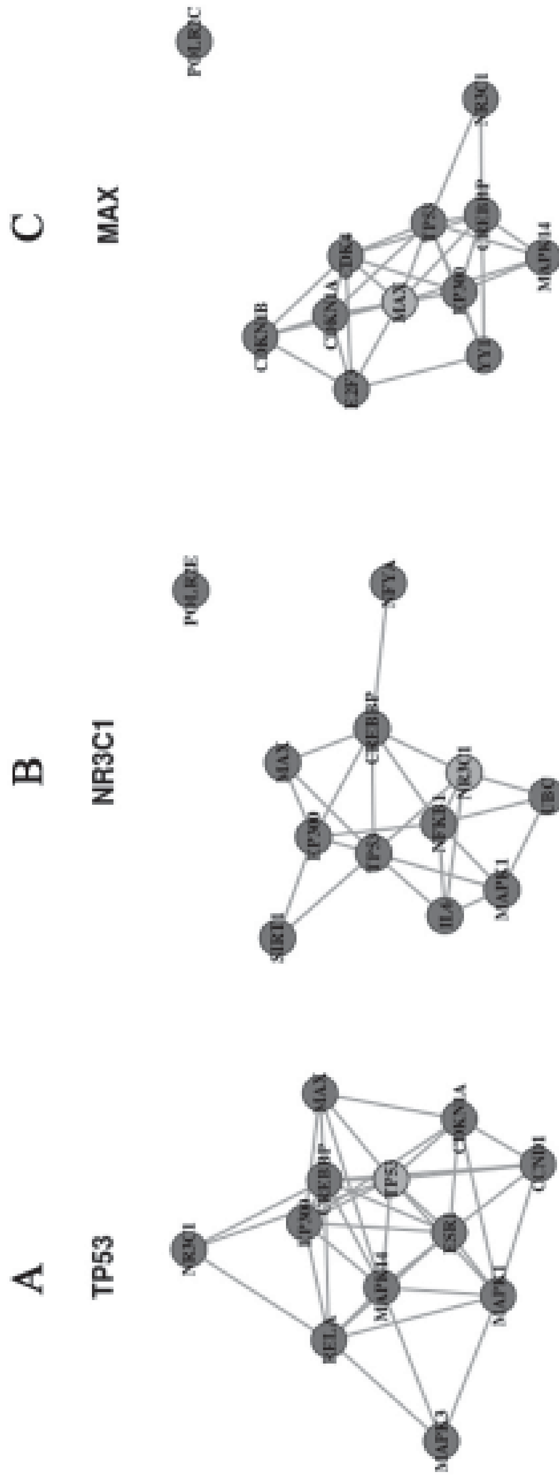

Fig. 6. The network graph of the three transcription factors for the highest network pathway G value. The light color dots (TP53, NR3C1, MAX) indicate the transcription factors and the deep color dots are the regulatory genes.
Source: Author

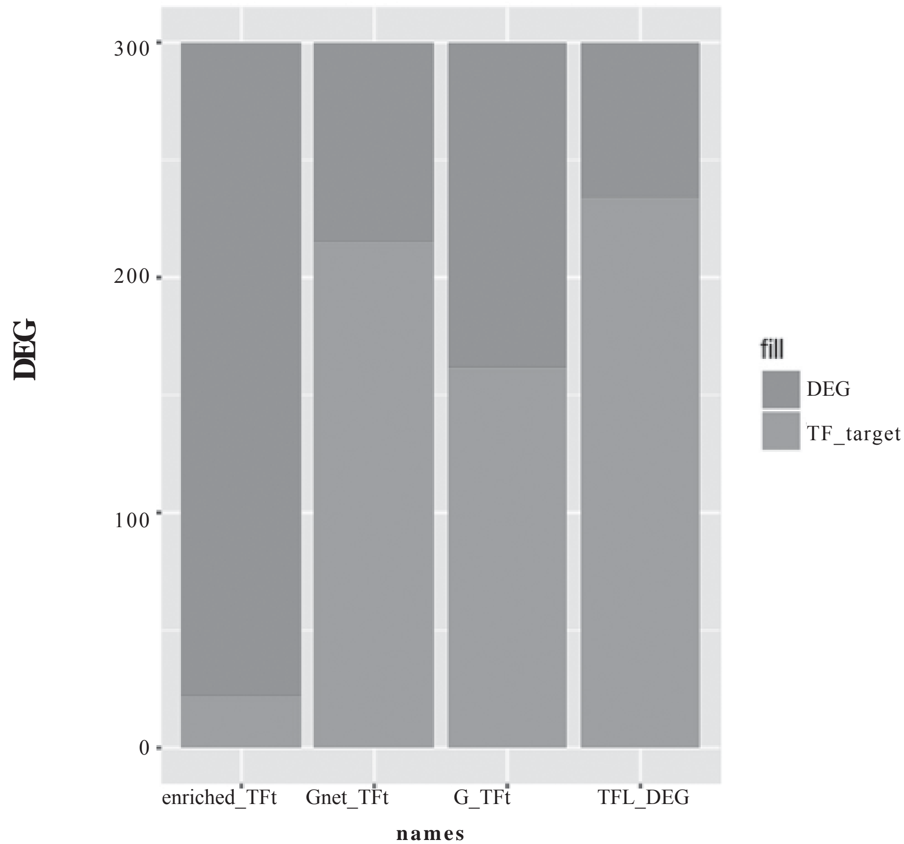

Fig. 7. Transcription factor coverage. The deep color represents DEG not regulated by TFs, and light color represents DEG regulated by TFs.
Source: Author

and their research methods, especially the development of bioinformatics in recent years, the mechanism of transcriptional regulation and diseases contain AS will be further study. TFs generally include three major functional domains, DNA-specific sequence binding domains, transcriptional activation domains, and regulatory domains between proteins and proteins (van der Does et al. 2016). Some small chemical molecules can affect the binding of TFs to DNA molecules, so they can regulate TFs-mediated regulation of gene expression (Alizadeh et al. 2017). Studies have found 14 DEGs in AS patients, among which EP300 was identified as interesting candidate genes that might be able to provide insight into AS progression (Kim et al. 2015). Studies have reported that selective endothelial-cell deletion of FLI1 in mice led to down-regulation of classic endothelial-cell markers, which was also known as down-regulated in vasculature of patients with systemic sclerosis, such as AS patients (van Bon et al. 2014). This study was consistent with these studies, showing that the researchers' method was effective and practical. Of course, those TFs target genes, such as RAD54B, NASP, DTL and so forth, have not been reported on AS, which will be the focus of the future research, providing a theoretical basis for the treatment of AS.

Bioinformatics methods can analyze a large number of DNA sequence data and gene expression products, more in-depth analysis of the functional relationship between genes and proteins (TFs are a type of DNA-binding proteins) (van Bon et al. 2014). A long time ago, using bioinformatics methods to analyze genomic DNA sequence information as a starting point, structur-

al modeling and functional prediction of proteins and RNAs, and consequent drug design, were a common and sophisticated routine (Liu et al. 2017). Studies constructed a database, which allows the identification of target genes for each transcription factor and serves as a starting point for further studies on the regulation of transcription factor combinations (Does et al. 2016). The researchers' approach used TFpro to establish AS-specific transcription factor regulation models through expression profiling and PPI networks. The results obtained by these models are useful for clarifying the mechanism of action of specific transcription factors and ultimately for understanding which pathways of these transcription factors play in the organism. Finally, three TFs (TP53, NR3C1, MAX) related with AS for the highest network pathway G value were obtained. The tumor suppressor TP53 functions multiply affect in cells and regulate growth negatively via many mechanisms, including increased senescence (Z et al. 2017). Recently, it has been recognized that TP53 plays a vital role in the differentiation and proliferation of progenitor/stem cells (Gao et al. 2016). It has been found that TP53 mutation is an independent predictor of tumor recurrence in lower and intermediate cancers (Fagerholm et al. 2017). NR3C1 is a glucocorticoid receptor (GR) and ligand-activated TF (Wu et al. 2017). Specifically, the NR3C1 gene encodes a GR that is expressed in brain and many peripheral tissues and is participated in binding cortisol and modulating cortisol response and levels (Sheinkopf et al., 2016). Studies have shown that epigenetic mechanism regulating the expression of NR3C1 promoter may affect the biological and behavioral aspects of the human infant's stress response (Conradt et al. 2015). These findings were also providing a theoretical basis for the treatment of AS in the future.

CONCLUSION

In summary, based on bioinformatics methods, our analysis discovered optimal 30 key TFs related to AS. This study may provide important potential therapeutic targets for AS.

RECOMMENDATIONS

However, there were few limitations in this paper. Firstly, the analysis results need to be confirmed by the clinical data of large samples. Secondly, the dataset used in this work were downloaded from the GEO database, not produced by us. So, the database may have some limitations. Additionally, the results based on bioinformatics prediction were obtained from this work, but were not confirmed relying on experiments. Therefore, to reveal new insights into the role of TFs in AS, further researches need to be study.

REFERENCES

Alizadeh F, Tavakkoly-Bazzaz J, Bozorgmehr A, Azarnezhad A, Tabrizi M et al. 2017. Association of transcription factor 4 (TCF4) gene mRNA level with schizophrenia, its psychopathology, intelligence and cognitive impairments. *J Neurogenet,* 32(1): 51

Andrysik Z, Galbraith MD, Guarnieri AL et al. 2017. Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. *Genome Research*.

Borse RH, Brown C, Muszbek N, Chaudhary MA, Kachroo S 2017. Cost-effectiveness of golimumab in ankylosing spondylitis from the uk payer perspective. *Rheumatology and Therapy,* 4: 1-17.

Carrillo AJ, Schacht P, Cabrera IE, Blahut J, Prudhomme L et al. 2017. Functional profiling of transcription factor genes in neurospora crassa. *G3 (Bethesda),* 7: 2945.

Chen WC, Wei JC, Lu HF, Wong HS, Woon PY et al. 2017. rs657075 (CSF2) is associated with the disease phenotype (BAS-G) of ankylosing spondylitis. *International Journal of Molecular Sciences,* 18: 83.

Conradt E, Fei M, LaGasse L, Tronick E, Guerin D et al. 2015. Prenatal predictors of infant self-regulation: The contributions of placental DNA methylation of NR3C1 and neuroendocrine activity. *Front Behav Neurosci,* 9: 130.

de Machado MA, Almeida AM, Kakehasi AM, Acurcio FA 2016. Real life experience of first course of anti-TNF treatment in ankylosing spondylitis patients in Brazil. *Rheumatol Ther,* 3: 143-154.

Deminger A, Klingberg E, Lorentzon M, Geijer M, Göthlin J et al. 2017. Which measuring site in ankylosing spondylitis is best to detect bone loss and what predicts the decline: Results from a 5-year prospective study. *Arthritis Research and Therapy,* 19: 273.

Does HCVD, Fokkens L, Yang A, Schmidt SM, Langereis L et al. 2016. Transcription factors encoded on core and accessory chromosomes of fusarium oxysporuminduce expression of effector genes. *Plos Genetics,* 12: e1006401.

Du M, Zhang Q, Bai L 2017. Three distinct mechanisms of long-distance modulation of gene expression in yeast. *Plos Genetics,* 13: e1006736.

Fagerholm R, Khan S, Schmidt MK, Garcíaclosas M, Heikkilä P et al. 2017. TP53-based interaction analysis identifies cis-eQTL variants for TP53BP2, FBXO28, and FAM53A that associate with survival and treatment outcome in breast cancer. *Oncotarget,* 8: 18381-18398.

Fang D, Zhu J 2017. Dynamic balance between master transcription factors determines the fates and func-

tions of CD4 T cell and innate lymphoid cell subsets. *Journal of Experimental Medicine,* 214: 1861.
Fishwick C, Higgins J, Percivalalwyn L, Hustler A, Pearson J et al. 2017. Heterarchy of transcription factors driving basal and luminal cell phenotypes in human urothelium. *Cell Death and Differentiation,* 24: 809-818.
Gao L, Hong X, Guo X, Cao D, Gao X et al. 2016. Targeted next-generation sequencing of dedifferentiated chondrosarcoma in the skull base reveals combined TP53 and PTEN mutations with increased proliferation index, an implication for pathogenesis. *Oncotarget,* 7: 43557-43569.
Joseph SR, Pálfy M, Hilbert L, Kumar M, Karschau J et al. 2017. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife,* 6: e23326
Kim KJ, Lee S, Kim WU 2015. Applications of systems approaches in the study of rheumatic diseases. *Korean J Intern Med,* 30: 148-160.
Liu Y, Irie T, Yada T, Suzuki Y 2017. A new computational method to predict transcriptional activity of a DNA sequence from diverse datasets of massively parallel reporter assays. *Nucleic Acids Research,* 45(13): e124.
Lozano T, Gorraiz M, Lasarte-Cia A, Ruiz M, Rabal O et al. 2017. Blockage of FOXP3 transcription factor dimerization and FOXP3/AML1 interaction inhibits T regulatory cell activity: Sequence optimization of a peptide inhibitor. *Oncotarget,* 8: 71709-71724.
Majewska M, Wysokinska H, Kuzma L, Szymczyk P 2017. Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: A comprehensive overview. *Gene,*644: 38-48.
Rudnik R, Bulcha JT, Reifschneider E, Ellersiek U, Baier M 2017. Specificity versus redundancy in the RAP2.4 transcription factor family of Arabidopsis thaliana: Transcriptional regulation of genes for chloroplast peroxidases. *Bmc Plant Biology,* 17: 144.
Sheinkopf SJ, Righi G, Marsit CJ, Lester BM 2016. Methylation of the glucocorticoid receptor (NR3C1) in placenta is associated with infant cry acoustics. *Front Behav Neurosci,* 10: 100.
Sikdar S, Datta S 2017. A novel statistical approach for identification of the master regulator transcription factor. *BMC Bioinformatics,* 18: 79.
Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A et al. 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res,* 39: D561-D568.
Qu Z, Qian B, Qiu Y, Zhang Y, Hu J et al. 2017. Does the position of conus medullaris change with increased thoracolumbar kyphosis in ankylosing spondylitis patients? *Medicine,* 96: e5963.
van Bon L, Affandi AJ, Broen J, Christmann RB, Marijnissen RJ et al. 2014. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. *N Engl J Med,* 370: 433-443.
van der Does HC, Fokkens L, Yang A, Schmidt SM, Langereis L et al. 2016. Transcription factors encoded on core and accessory chromosomes of fusarium oxysporum induce expression of effector genes. *PLoS Genet,* 12: e1006401.
Wang D, Hou Z, Gong Y, Chen S, Lin L et al. 2017. Bone edema on magnetic resonance imaging is highly associated with low bone mineral density in patients with ankylosing spondylitis. *PLoS One,* 12: e0189569.
Wu LH, Cheng W, Yu M, He BM, Sun H et al. 2017. Nr3C1-Bhlhb2 axis dysregulation is involved in the development of attention deficit hyperactivity. *Mol Neurobiol,* 54: 1196-1212.
Xu Z, Wang X, Zheng Y 2018. Screening for key genes and transcription factors in ankylosing spondylitis by RNA-Seq. *Experimental and Therapeutic Medicine,* 15: 1394-1402.
Zhang X, Liu X, Zhang D, Tang H, Sun B et al. 2017. Genome-wide identification of gene expression in contrasting maize inbred lines under field drought conditions reveals the significance of transcription factors in drought tolerance. *PLoS One,* 12: e0179477.

Paper received for publication on March 2018
Paper accepted for publication on May 2018