Identification of potential target genes for ankylosing spondylitis treatment

Yong Ni, MMa, Chengrui Jiang, MMb,∗

Abstract
This study aimed to identify the potential target genes for the treatment of ankylosing spondylitis (AS).

Dataset GSE25101 was downloaded from Gene Expression Omnibus, including 16 AS and 16 normal control blood samples. Differentially expressed genes (DEGs) were identified using unmatched t-test in limma package with adjusted P < .05. Gene ontology-biological process (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted using multifaceted analysis tool for human transcriptome. Protein–protein interaction (PPI) network was constructed using STRING and Cytoscape, and module analysis was performed using MCODE plug-in. Webgestalt was utilized to predict transcriptional factor (TF)-microRNA-target network and Comparative Toxicogenomics Database (CTD) was applied to predict chemical-target network.

A total of 334 DEGs were identified, including 136 upregulated genes and 198 downregulated genes. According to STRING, a PPI network was constructed and 1 significant clustered module was screen out with score = 6.33. MAPK7 (degree = 11) and NDUFS4 (degree = 10) were 2 important nodes in PPI network, and both of them were significantly enriched in cAMP mediated signaling pathway (P = 2.02E – 02). MAPK7 could be regulated by NFY. Both MAPK7 and NDUFS4 were 2 potential targets for Indomethacin.

MAPK7 and NDUFS4 played important roles in the pathogenesis of AS via cAMP mediated signaling pathway. Both of them could be targeted by Indomethacin.

Abbreviations: AS = ankylosing spondylitis, CI = complex I, CTD = comparative toxicogenomics database, DEGs = differentially expressed genes, GEO = Gene Expression Omnibus, GO = Gene Ontology, GO-BP = gene ontology-biological process, KEGG = Kyoto Encyclopedia of Genes and Genomes, MAPKs = mitogen-activated protein kinases, MATHT = multifaceted analysis tool for human transcriptome, miRNAs = microRNAs, NDUFS4 = NADH ubiquinone oxidoreductase iron-sulfur protein 4, ORA = overrepresentation enrichment analysis, PPI = protein-protein interaction, TF = transcriptional factor.

Keywords: ankylosing spondylitis, differentially expressed gene, inflammation, treatment target

1. Introduction
Ankylosing spondylitis (AS) is a chronic inflammatory disease and affects about 1% population worldwide.11 AS is characterized by inflammation of the spine, sacroiliac joints, facet joints, and less peripheral joints.2 It commonly presents with low-back pain and gradually develops to stiffness and disability of the spine, and creates a heavy economic and health burden on both family and society.3 Drug therapy is the most common intervention for AS treatment and several biological regents have been applied in clinical, but the selection of drugs is still confused for AS patients owing to different drugs and dose introduced by different manufactures. Lately, the Assessments in Spondylo Arthritis International Society (ASAS) recommendations have updated their knowledge, but no concrete agent and regimen has been recommended.4 Therefore, it is meaningful to have a deeper understanding of the pathogenesis of AS.

Due to the main characteristics of AS, inflammation and immune responses have been mostly revealed in recent AS researches. The primary immune response member T cells are also involved in the progression of AS.5,6 Both CD4+ and CD8+ T cells respond to collagen-derived peptides and aggrecan, which have been considered as potential targets of the autoimmune response in AS.5 Moreover, interleukin-6 (IL-6), an important cytokine in immune response, is also revealed closely related to the development of AS.7 Haroon et al8 have demonstrated that TNF-α involves in the progression of AS, and TNF-α inhibitors can significantly reduce radiographic progression of AS, especially in the initiation. Moreover, microRNAs (miRNAs) are also confirmed to participate in the progression of AS, such as the proliferation, cell death, maturation, and activation.9,10 A study also demonstrates that miR-29 performs an important role in the progression of AS, and can serve as a diagnostic marker for the process of new bone formation in AS.11 However, a clearly understand of the pathogenesis in AS is remaining limited.

To further analyze the pathogenesis of AS, Pimentel-Santos et al12 have created a whole transcriptional microarray using the whole blood samples. Basing on this dataset, the differentially expressed genes (DEGs) between AS patients and normal control were screened, and found that SPOCK2 and EP300 might modulate cartilage and bone metabolism in AS. Moreover, this dataset is also utilized for further sub-pathway
identifications,[13] protein–protein interaction network (PPI) analysis,[14] and meta-analysis.[15] In the current study, this microarray dataset was utilized to identify the potential target genes for the treatment of AS.

2. Methods

2.1. Data sourcing

Dataset of GSE25101[12] was downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), including 16 AS blood samples and 16 healthy blood samples. All the samples included in this dataset were sequenced on the platform of GPL6947 Illumina Human HT-12 V3.0 expression beadchip.

2.2. Identification DEG

Basing on the annotation files, probe IDs in the standardized series matrix files, which were downloaded from GEO, were converted into gene symbols, and non-matched gene symbols were rejected. For different probes matching the same gene, mean value was used as the final expression level of the gene. Then, DEGs between AS and health control patients were screened using non-matching t test in Limma package (version 3.26.9, http://www.bioconductor.org/packages/3.2/bioc/html/limma.html) in R 3.4.1 (http://CRAN.R-project.org/bin/windows/base), and P value was calculated. Benjamini–Hochberg method was used to adjust P value. DEGs were considered as the adjusted P value <.05.

2.3. DEGs enrichment analyses

Gene Ontology (GO, http://geneontology.org/) is a free research database which defines the concepts relating to gene function in 3 respects: biological process, molecular function, and cellular component.[16] Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg) is a public database for linking genes/proteins to high-levels function of biological system.[17] Multifaceted Analysis Tool for Human Transcriptome (MATHT, http://www.biocloudservice.com) was utilized to perform enrichment analyses for both up- and down-regulated DEGs, including GO-BP and KEGG pathway enrichment analyses. In MATHT, mRNA enrichment, which is based on the Fisher test, in Functional Enrichment of Gene Set Function was applied to perform the enrichment analysis, and P <.05 was considered significantly enriched.

2.4. PPI network construction

According to the PPI intersections of 4 databases, including STRING (version 10.0, http://string-db.org/),[18] mintche (http://mentha.uniroma2.it/about.php), BioGRID (version 3.4, https://wiki.thebiogrid.org/),[19] and HRRD (Release 9, https://wiki.thebiogrid.org/),[20] PPI interactions of up- and down-regulated DEGs were extracted. Then, Cytoscape (version 3.2.0) was utilized to construct PPI network, and degree centrality of topology property was applied to estimate the score of a node in the network. The higher score node obtained, the more potential hub gene it would be. Moreover, the plug-in MCODE[22] in Cytoscape was used to conduct sub-network analysis of PPI. Meanwhile, GO-BP and KEGG enrichment analyses of sub-network were also carried out.

2.5. Disease related regulation network analyses

Overrepresentation Enrichment Analysis (ORA) in webgestalt (http://www.webgestalt.org/option.php)[23] was applied to predict the regulatory relationships between transcriptional factor (TF), miRNA, and DEGs. Then, Cytoscape was used to construct regulatory networks. Gene sequenced on illumina humanht 12 V3 platform was utilized as the reference, and enriched gene number ≥5 and P value adjusted by BH method <.05 were set as the cutoff for regulatory network analysis.

2.6. Construction of network between DEGs and small chemical molecule

The Comparative Toxicogenomics Database (CTD)[24] is a free research tool describing relationships between drugs, diseases, genes/proteins, phenotypes, and interaction modules. In the current study, chemical–gene interaction of AS in CTD was applied to analyze interactive between DEGs and small chemical molecules. Then, Cytoscape was used to construct a network based on the identified interactions.

3. Results

3.1. Identification of DEGs

After preprocessing, a total of 11,505 genes were screened out from 18,168 probes. Followed with t test in limma, total 334 DEGs were identified in this study, including 136 upregulated genes and 198 down-regulated genes. The heatmap of DEGs is showed in Fig. 1.

3.2. GO-BP and KEGG pathway enrichment analyses

Based on the variations of DEGs expression, GO-BP and KEGG pathway enrichment analyses were carried out for up- and down-regulated DEGs, respectively. For GO-BP enrichment analysis, the upregulated DEGs were mainly concentrated in terms of hydrogen ion transmembrane transport (P = 9.45E-04, muscle...
filament sliding (2.50E–03), mitochondrial translational elongation (P = 3.22E–03), mitochondrial translational termination (P = 3.36E–03), and cell cycle arrest (P = 3.42E–03); however, the down-regulated DEGs were significantly enriched in the terms of viral process (P = 8.06E–05), nucleotide-excision repair, DNA duplex unwinding (P = 1.53E–03), T cell migration (P = 3.24E–03), nucleotide-excision repair, preincision complex assembly (P = 3.44E–03) and global genome nucleotide-excision repair (P = 4.56E–03) (Table 1). Meanwhile, for the KEGG enrichment analysis, the upregulated DEGs were significantly enriched in the pathways of oxidative phosphorylation (P = 8.68E–06), Parkinson disease (P = 1.41E–05), non-alcoholic fatty liver disease (P = 2.20E–05), Alzheimer disease (P = 4.75E–05), and Huntington disease (P = 1.22E–04), while the downregulated DEGs were remarkably enriched in the pathways of primary immunodeficiency (P = 3.37E–04), 2-oxocarboxylic acid metabolism (P = 5.86E–04), cysteine and methionine metabolism (P = 6.26E–03), nucleotide excision repair (1.13E–02), and biosynthesis of antibiotics (1.87E–02) (Table 2).

3.3. PPI network and module analyses

Depended on the interactive relationships, a PPI network was constructed. As the Fig. 2A presents, there were 215 nodes with 363 edges contained in the PPI network. The top 10 nodes in PPI were EP300 (degree = 11), NEDD8 (degree = 11), PPP2R1A (degree = 11), ACTR1B (degree = 11), MAPK7 (degree = 11), PRPF8 (degree = 11), RNPS1 (degree = 10), NDUF5 (degree = 10), RPL26l-1 (degree = 10), and RPS7 (degree = 10). Meanwhile, based on the threshold, a significant enriched module was screened from the PPI network (enriched score = 6.33), including 7 nodes with 19 edges (Fig. 2B).

Moreover, GO-BP and KEGG analyses were performed for the top 10 nodes and selected module. As the enriched shows, the top 10 nodes were significantly enriched in nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (P = 2.83E–05), splicing (P = 3.34E–03), cAMP-mediated signaling (P = 2.02E–02), and signal recognition particle-dependent cotranslational protein targeting to membrane (P = 4.93E–02). The selected module was remarkably enriched in GO-BP terms of hydrogen ion transmembrane transport (P = 9.05E–07), mitochondrial electron transport, cytochrome c to oxygen (P = 2.02E–05), aerobic respiration (P = 5.59E–05), electron transport chain (P = 2.50E–03), and oxidative phosphorylation (P = 4.64E–03), and in KEGG pathways of oxidative phosphorylation (P = 4.54E–11), Parkinson’s disease (P = 6.78E–11), non-alcoholic fatty liver disease (NAFLD) (P = 9.87E–11), Alzheimer disease (P = 1.89E–10), and Huntington disease (P = 4.26E–10) (Table 3).

3.4. TF-miRNA-target regulation network

Based on the webgestal database, the TF-miRNA-target regulation network of DEGs was predicted and constructed for. Finally, a total of 119 nodes with 181 edges were contained in this regulation network, including 69 downregulated genes, 29 upregulated genes, 7 TPs, and 14 miRNAs (Fig. 3).

3.5. Small chemical molecule-target network

According to CTD, a total of 10,668 results were screened out involved in AS. After deleted the duplicated ones, 7014

---

### Table 1

| DEGs | Terms | Gene | P value |
|------|-------|------|---------|
| Up-regulated | GO:0190260: hydrogen ion transmembrane transport | COX7A2, UQCR9H, COX7B, COX6A1, UQCRB | 9.45E–06 |
| Down-regulated | GO:0003040: muscle filament sliding | MYL6, TNNC2, MYL6B, MYH4 | 2.50E–03 |

**BP** = biological process, **DEG** = differentially expressed gene, **GO** = gene ontology.

---

### Table 2

| DEGs | Terms | Gene | P value |
|------|-------|------|---------|
| Up-regulated | hsa00190: oxidative phosphorylation | NDUF5B3, COX7A2, NDUF5A, UQCR9H, COX7B | 8.68E–06 |
| Down-regulated | hsa00401: Parkinson disease | NDUF5B3, COX7A2, NDUF5A, UQCR9H, COX7B | 1.41E–05 |

**DEGs** = differentially expressed gene, **KEGG** = Kyoto Encyclopedia of Genes and Genomes.
chemical–gene interaction pairs were obtained. Then, 334 DEGs were mapped to the chemical–gene interaction pairs, and 124 chemical–target interactions were matched and utilized to construct network. A total of 107 nodes with 124 edges were included in this network, including 40 up-regulated genes, 61 down-regulated genes, and 6 chemical molecules (Fig. 4).

4. Discussion

In the current study, a total of DEGs were screened from microarray dataset GSE25101, including 136 upregulated genes and 198 downregulated genes. Based on these identified DEGs, a PPI network was constructed, and a significant module was screened out \((P=6.33)\). GO-BP and KEGG enrichment analyses for the

**Table 3**
The top 5 results of GO-BP and KEGG enrichment analysis for the top 10 nodes and selected module involved in PPI.

| Enrichment analysis | Term | Gene | \(P\) value |
|---------------------|------|------|-------------|
| Top 10 nodes        | GO-BP| GO:0000184→nuclear-transcribed mRNA catabolic process, nonsense-mediated decay | PPP2R1A, RPL26L1, RNPS1, RPS7 | 2.83E-05 |
|                     |      | GO:0003830→RNA splicing | PPP2R1A, PRPF8, RNPS1 | 3.34E-03 |
|                     |      | GO:0019033→cAMP-mediated signaling | NDUFS4, MAPK7 | 2.02E-02 |
|                     |      | GO:0006614→SRP-dependent cotranslational protein targeting to membrane | RPL26L1, RPS7 | 4.93E-02 |
| Module              | GO-BP| GO:1902600→hydrogen ion transmembrane transport | UQCRH, COX7B, COX6A1, UQCRB | 9.05E-07 |
|                     |      | GO:0006123→mitochondrial electron transport, cytochrome c to oxygen | COX7B, COX6A1, COX5B | 2.02E-05 |
|                     |      | GO:0009060→aerobic respiration | UQCRH, COX6A1, UQCRB | 5.59E-05 |
|                     |      | GO:0022900→electron transport chain | NDUFB3, NDUFS4 | 2.50E-03 |
|                     |      | GO:0006119→oxidative phosphorylation | UQCRH | 4.64E-03 |
| PATHWAY             |       | hsa05012:Parkinson’s disease | NDUFB3, NDUFS4, UQCRH, COX7B, COX6A1 | 4.54E-11 |
|                     |       | hsa00902:Neurodegenerative diseases | NDUFB3, NDUFS4, UQCRH, COX7B, COX6A1 | 6.78E-11 |
|                     |       | hsa00903:Non-alcoholic fatty liver disease (NAFLD) | NDUFB3, NDUFS4, UQCRH, COX7B, COX6A1 | 9.87E-11 |
|                     |       | hsa05010:Alzheimer’s disease | NDUFB3, NDUFS4, UQCRH, COX7B, COX6A1 | 1.80E-10 |
|                     |       | hsa05016:Huntington’s disease | NDUFB3, NDUFS4, UQCRH, COX7B, COX6A1 | 4.26E-10 |

BP=biological process, GO=gene ontology, KEGG=Kyoto Encyclopedia of Genes and Genomes.
selected module were also carried out. Meanwhile, a TF-miRNA-target regulatory network was predicted and constructed, and 7 TFs and 14 miRNAs were obtained. Moreover, a chemical-target network was predicted and created, and 6 small chemical molecules were identified.

MAPK7, also known as ERK5, is a member of mitogen-activated protein kinases (MAPKs) with a unique C-terminal domain which determines its location and activity in cells. According to the previous studies, MAPK7 is considered to be a critical mediator of inflammation, but the role of MAPK7 is still controversial. The studies of Wilhelmsen et al. and Mackesy et al. have documented that MAPK7 is a proinflammatory factor, but the researches of Le et al. and Wu et al. have revealed that MAPK7 is an anti-inflammatory factor in various cells. In the current study, MAPK7 was detected as significantly downregulated in AS patients compared with normal controls, which indicated that MAPK7 might perform as an anti-inflammatory role in the progression of AS, and this hypothesis might be involved in TGF-β-Smad pathway, but this hypothesis was still needed to be confirmed.

NADH ubiquinone oxidoreductase iron-sulfur protein 4 (NDUFS4), the important component of the mitochondrial complex I (CI), was another protein that enriched in CAMP signaling pathway with the top 9 degree in PPI network in this study. CI is the largest complex in mitochondrial respiratory chain, and its abnormal expression also relates to neurological and oxidative phosphorylation disorder, which will result in the imbalance of autophagy, differentiation, metabolism, and immune response. However, the mechanism of NDUFS4 in AS is still rarely reported. Jin et al. have demonstrated that NDUFS4 is critical for osteoclastogenesis, and deficiency of NDUFS4 will lead to a disorder of osteoclast to macrophage. In the current study, NDUFS4 was significantly upregulated in AS patients while compared with normal controls, and GO-BP analysis result showed that NDUFS4 was significantly enriched in mitochondrial respiratory related biological process. Therefore, we deduced that high-expression of NDUFS4 might induce a dysfunction of mitochondrial respiratory and result in a high-level of oxidative damage and inflammatory response to promote the progression of AS. These evidences indicated that NDUFS4 might a positive role in the development of AS, and this effect might be performed via cAMP mediated signaling pathway.

Indomethacin is a common anti-inflammatory drug in clinical treatment of AS, but the effect of Indomethacin in AS treatment is still not determined owing to several reasons, including the disease degree, side-effects, and health condition of patients. In this study, the chemical-target predication had revealed that both MAPK7 and NUDFS4 could be targeted by Indomethacin. These identifications might indicate that MAPK7 and NUDFS4 might be potential targets for Indomethacin. Thus, it was of significant to further analyze the interactive loci of target genes and Indomethacin, so that could develop a more effective structure of Indomethacin to improve the interactions between drug and target proteins.

There were still some limitation should be strengthened in this study. First, because of the limited sample size, there might be some false positive results involved in this study. Thus, further analysis with larger sample should be considered. Next, lacking in the enough clinical samples, the experimental verification was deficiency.
In conclusion, AS was a chronic inflammatory disease in sine, and mitochondrial respiratory disorder might be an important characteristic of its pathogenesis. MAPK7 and NUDFS4 were 2 important participators involved in the regulation of inflammation via CAMP mediated signaling pathway to TGF-β pathway, and could serve as potential target genes for Indomethacin.

References

[1] Ghasemirad M, Attaya H, Lesha E, et al. Ankylosing spondylitis: a state of the art factual backbone. World J Radiol 2015;7:236–52.
[2] Nordström D, Kauppi M. [Axial spondyloarthritis and ankylosing spondylitis]. Duodecim 2010;126:1467–74.
[3] Ravani A, Vincenzi F, Bortoluzzi A, et al. Role and function of A2A and A3 adenosine receptors in patients with ankylosing spondylitis, psoriatic arthritis and rheumatoid arthritis. Int J Mol Sci 2017;18:pi: E697.
[4] Heijde DVD, Sieper J, Maksymowych WP, et al. 2010 update of the international ASAS recommendations for the use of anti-TNF agents in patients with axial spondyloarthritis. Ann Rheum Dis 2011; 70:903–8.
[5] Duchmann R, Lambert C, May E, et al. CD4+ and CD8+ clonal T cell expansions indicate a role of antigens in ankylosing spondylitis: a study in HLA-B27+ monozygotic twins. Clin Exp Immunol 2001;123:315–22.
[6] Bollow M, Fischer T, Reihnhauer H, et al. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active spondylitis-cellularity correlates with the degree of enhancement detected by magnetic resonance imaging. Ann Rheum Dis 2000;59:135–40.
[7] Gratacós J, Collado A, Fillola X, et al. Serum cytokines (IL-6, TNF-alpha, IL-1 beta and IFN-gamma) in ankylosing spondylitis: a close correlation between serum IL-6 and disease activity and severity. Br J Rheumatol 1994;33:927–31.
[8] Haroon N, Ioman RD, Leach Tj, et al. The impact of tumor necrosis factor α inhibitors on radiographic progression in ankylosing spondylitis. Arthritis Rheum 2013;65:2645–54.
[9] Lai NS, Yu HC, Chen HC, et al. Aberrant expression of microRNAs in T cells from patients with ankylosing spondylitis contributes to the immunopathogenesis. Clin Exp Immunol 2013;173:47–57.
[10] Huang CH, Wei JC, Chang WC, et al. Higher expression of whole blood microRNA-21 in patients with ankylosing spondylitis associated with programmed cell death 4 mRNA expression and collagen cross-linked C-telopeptide concentration. J Rheumatol 2014;41:1104–11.
[11] Huang J, Song G, Yin Z, et al. Elevated miR-29a expression is not correlated with disease activity index in PBMCs of patients with ankylosing spondylitis. Mod Rheumatolat 2013;24:331–4.

[12] Pimentel-Santos FM, Ligeiro D, Matos M, et al. Whole blood transcriptional profiling in ankylosing spondylitis identifies novel candidate genes that might contribute to the inflammatory and tissue-destructive disease aspects. Arthritis Res Ther 2011;13:R27.

[13] Chen K, Zhao Y, Chen Y, et al. A sub-pathway based method to identify candidate agents for ankylosing spondylitis. Molecules 2012;17:12460–8.

[14] Zhang C, Shen L. Functional modules analysis based on protein-protein network analysis in ankylosing spondylitis. Eur Rev Med Pharmacol Sci 2012;16:1821–7.

[15] Park R, Kim T-H, Ji JD. Gene expression profiles in patients with axial spondyloarthritis: meta-analysis of publicly accessible microarray datasets. J Rheum Dis 2016;23:365–72.

[16] Consortium TGO. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res 2017;45:D331–8.

[17] Kanetska M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44:D457–62.

[18] Szkarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39:D561–8.

[19] Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res 2015;43:D470–8.

[20] Prasad TS, Goel R, Kandasamy K, et al. Human protein reference database—2009 update. Nucleic Acids Res 2009;37:D767–72.

[21] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[22] Bandettini WP, Kellman P, Mancini C, et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson 2012;14:83.

[23] Wang J, Duncan D, Shi Z, et al. WEB-based Gene Set Enrichment Analysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013;41:W77–83.

[24] Davis AP, King BL, Mockus S, et al. The comparative toxicogenomics database: update 2011. Nucleic Acids Res 2013;39:D1067–72.

[25] Dale M, Higgins A, Carolanrees G. Sherlock 3CG: tip confirmation system for placement of peripherally inserted central catheters: a nice medical technology guidance. Appl Health Econ Health Policy 2016;14:41–9.