Screening Key IncRNAs of Ankylosing Spondylitis Using Bioinformatics Analysis

Jian-Xiong Wang*, Xu Zhao*, Sheng-Qian Xu

Department of Rheumatology & Immunology, The First Affiliated Hospital of Anhui Medical University, Hefei, People's Republic of China

*These authors contributed equally to this work

Correspondence: Sheng-Qian Xu, Email xsqian-1112@163.com

Background: Long non-coding RNAs (lncRNAs) are important regulators in ankylosing spondylitis (AS). Few studies have examined the lncRNA-RNA binding protein (RBP) interaction in AS. This study performed bioinformatics analysis and clinical verification to identify key lncRNAs and propose their RBP interaction.

Methods: Three GEO datasets of AS were analyzed by differential expression analysis. The differentially expressed lncRNAs between the AS and control groups were screened out, and the intersecting lncRNAs were regarded as target lncRNAs. Functional was performed to identify target lncRNAs by enrichment analysis, co-expressed RNA analysis, and lncRNA-RBP interaction analysis. Finally, this study analyzed the differential expression level and clinical value of lncRNAs between the AS and control groups.

Results: Linc00304, linc00926, and MIAT were differentially expressed and upregulated. Enrichment analysis indicated that the key KEGG terms were the T-cell receptor signaling pathway and B-cell receptor signaling pathway. The key molecular function term was protein binding, and the key biological process term was adaptive immune response. In qRT-PCR results, 44 samples were validated. linc00304 expression was positively correlated with bath ankylosing spondylitis disease activity index (BASDAI), bath ankylosing spondylitis functional index (BASFI), erythrocyte sedimentation rate (ESR), and c-reactive protein (CRP). linc00926 expression was only positively correlated with ESR, whereas MIAT expression was positively correlated with BASFI, ESR, and CRP. Logistic regression revealed that linc00304, ESR, and CRP were the independent risk factors for BASDAI activation. The area under the curve (AUC) of serum linc00304 level in the diagnosis of AS was 0.687 (cutoff value: 0.413, specificity: 0.423, sensitivity: 0.900). AUC of linc00926 was 0.664 (cutoff value: 0.299, sensitivity: 0.882, specificity: 0.417). AUC of MIAT was 0.623 (cutoff value: 0.432, specificity: 0.443, sensitivity: 0.890) (all P <0.05).

Conclusion: Overall, this study uncovered three novel lncRNAs, which were upregulated in AS, and proposed a new lncRNA-RBP-mRNA interaction that might regulate adaptive immune response.

Keywords: ankylosing spondylitis, long non-coding RNA, RNA binding protein, RNA sequencing

Introduction

Ankylosing spondylitis (AS) is a long-term inflammatory arthritis caused by autoimmune imbalance. It is often accompanied by inflammatory back pain, three peripheral joint manifestations (arthritis, enthesitis, and dactylitis), three extrarticular manifestations (uveitis, psoriasis, and crohn’s disease), two genetic signatures (family history of AS and HLA-B27), and two inflammatory signatures (good response to NSAIDs and elevated CRP level). When the trigger point of structural damage is activated (magnetic resonance imaging showed post-inflammatory fat deposition or bone marrow edema),1 the process of structural damage will not cease, even if the activity is controlled. The mechanism underlying the cessation of the process remains unclear. In recent years, several studies have shown that long noncoding RNAs (lncRNAs) are pervasively dysregulated in different cell types or tissues. lncRNAs in whole blood cells, peripheral blood mononuclear cells (PBMCs), T cells, hip joint cells, and osteogenically differentiated bone marrow mesenchymal stem cells participate in the regulation of key proinflammatory cytokines, such as IL-1β, TNF-α, IL-6, and IL-23/IL-17.2–5 More mechanistic studies are needed to identify more important lncRNAs and their
downstream pathways that are critical to the pathogenesis of AS. lncRNAs are longer than other RNAs; they contain multiple functional domains or several RNA binding protein (RBP) binding sites. Most lncRNAs have been identified to have RBPs that bind to their binding sites. On the one hand, lncRNAs can regulate the bioactivity and stability of RBPs in ribonucleoprotein complexes, thereby affecting the interaction between RBPs and mRNAs, and directly bind with RBPs to regulate other target mRNAs. On the other hand, RBP can regulate the expression, stability, and function of lncRNAs. Therefore, the interaction between RBPs and lncRNAs plays an important role in the regulation of other target genes, thereby participating in various disease processes. However, the interaction mechanism in AS has been rarely studied.

Consequently, this study attempted to find reliable lncRNAs that were highly expressed in multiple RNA-sequencing (RNA-seq) database. Additionally, the study explored the lncRNA–RBP interaction and evaluated and clinically validated the value of the target lncRNAs.

**Materials and Methods**

**AS Data Collection**

The flow chart and overall design of this case–control study are shown in Figure 1. Data were obtained from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo). The GSE181364 dataset included lncRNA expression profiles of whole blood from three healthy individuals and five patients with AS and associated bone bridge formation. The GSE18781 dataset included array-based lncRNA expression profiles of peripheral blood from 18 patients with AS and 24 control human neonates. The GSE39340 dataset included lncRNA expression profiles of peripheral blood from five patients with AS, 10 patients with rheumatoid arthritis (RA), and 10 patients with osteoarthritis (OA).

---

**Figure 1** The overall design and flowchart.
Differential Analysis

IncRNA profiles were established for each sample of patients with AS, OA, or healthy individuals for control groups. Differential expression analyses of IncRNA expression profiles of patients with AS and control groups in the GSE181364, GSE18781, or GSE39340 dataset were performed using the limma package of the R software and significance criterion of $P < 0.05$. Thereafter, the differentially expressed IncRNAs (DEIncRNAs) between patients with AS and control groups were identified. DEIncRNAs were screened for intersection analysis with three GEO databases to obtain the intersecting IncRNAs.

Enrichment Analysis and Coexpression Analysis and RBP Network Analysis

The IncRNAs and predicted target genes of mRNAs were annotated, in terms of their GO categories and KEGG pathways, using the online DAVID tool\(^\text{11}\) (https://david.ncifcrf.gov). COXPRESdb (https://coxpresdb.jp) could provide co-expressed genes of IncRNAs, and enrichment analysis of co-expressed genes was helpful in estimating IncRNA function.\(^{12}\) ENCORI (https://starbase.sysu.edu.cn) is a database of RBP expression and IncRNA information.\(^{13}\) The IncRNA–RBP interaction network was constructed using the Cytoscape software (version 3.6.1).

qRT-PCR

During the clinical validation stage, RNA was extracted from PBMCs of 24 patients with AS and 20 healthy controls. Real-time quantitative polymerase chain reaction (qRT-PCR) was performed using the Trizol method. cDNA was transcribed using PrimeScript™RT Master Mix. PCR was performed on the ABI7500 PCR platform. The RT-qPCR thermocycling protocol was performed as follows: 50°C for 2 min, 95°C for 10 min, 94°C for 30 s, and 60°C for 60 s in 40 cycles. qPCR data were relatively quantified using $2^{-\Delta\Delta C_t}(\Delta C_t = C_{\text{target}} - C_{\text{reference}}, -\Delta\Delta C_t = \text{sample } \Delta C_t - \beta\text{-actin } \Delta C_t)$.\(^{14}\) The primer information are shown in Supplementary Material 1.

Clinical Verification

This study enrolled 24 patients with AS and 20 age-sex matched healthy controls from the Department of Rheumatology and the Physical Examination Center of the First Affiliated Hospital of Anhui Medical University from May 2021 to May 2022 (Clinical details available in Supplementary Material 2). All patients with AS met the modified New York 1984 criteria.\(^{15}\) All clinical data and participant information were obtained by questionnaire administration. The indicators included general patient information, disease course, laboratory index [erythrocyte sedimentation rate (ESR, mm/h) and C-reactive protein (CRP, mg/l)], disease activity score [Ankylosing Spondylitis Disease Activity Score with C-reactive protein (ASDAScrp)], Bath ankylosing spondylitis disease activity index (BASDAI), Bath ankylosing spondylitis functional index (BASFI), and modified stoke ankylosing spondylitis spine score (mSASSS)].\(^{19}\) BASDAI score >4 was defined as AS disease activity.\(^{16}\) All recruiters were first examined. The exclusion criteria included incomplete clinical data, history of any acute or other chronic diseases, and chronic medication use. This study was approved by the Ethics Committee (No. 20200740) and was performed in accordance with the Declaration of Helsinki. All participants provided informed consent.

All statistical analyses were performed using SPSS (version 24.0). The Chi-square test was used for qualitative data. Differences in normally distributed data and skewed data were assessed using t-test and nonparametric test, respectively. Pearson and Spearman correlation analyses were used to assess the relationship between different parameters. Multivariate logistic regression analysis was performed to identify independent predictors of disease activity (BASDAI) by analyzing relevant clinical indicators and target IncRNA levels. Findings from this analysis were used to establish a forest map. $P < 0.05$ indicated statistical significance in all analyses. To further indicate the predictive performance of target IncRNA levels, we plotted receiver operating characteristic (ROC) curves for the diagnosis of AS dependence and calculated the area under the curve (AUC) for each dataset value.
Results

Identification and Value of Target IncRNAs

Comparing the profiles of DElncRNA in three AS GEO databases, we found that linc00304, linc00926, and MIAT in the GSE181364 database were differentially expressed in both GSE18781 and GSE39340 databases (Figure 2A–C). The Heatmap showed the association between the target lncRNA expression levels and these AS-related biomarkers (Figure 2D). Target IncRNAs expressed significantly differently in patients with AS than in healthy controls, and they were significantly correlated with such AS biomarkers.

Functional Analysis of Target IncRNAs

Target prediction of co-expressed mRNAs was performed using COXPRESdb programs, and the top 50 co-expressed mRNAs were predicted for three IncRNAs. Thereafter, GO and KEGG analyses were performed on the co-expressed mRNAs. The results showed that 532 terms were annotated for co-expression of linc00304, 564 terms were annotated for co-expression of linc00926, and 670 terms were annotated for co-expression of MIAT. The top 10 items of statistical significance in the GO and KEGG analyses are shown in Figure 3. According to our findings, binding proteins, immune system, T-cell receptor signaling pathway (TCR signaling pathway), B-cell receptor signaling pathway (BCR signaling pathway), integral component of membrane, and adaptive immune system were frequently enriched, which might be the core factors involved in the systemic inflammatory process of AS.

![Figure 2](https://doi.org/10.2147/JIR.S387258)

**Figure 2** (A) The expression level of linc00304 in two AS RNA-sequencing data. Upregulated expression of linc00304 in patients with AS. (B) The expression level of linc00926 in two AS RNA-sequencing data. Upregulated expression of linc00926 in patients with AS. (C) The expression level of MIAT in two AS RNA-sequencing data. Upregulated expression of MIAT in patients with AS. (D) Heatmap showing the association between the expression level of target IncRNAs and AS-related biomarkers.
Exploration of the Potential lncRNA-RBP Interaction

Using the ENCORI website, this study predicted 13 RBPs that were associated with target lncRNAs and 16 key co-expressed mRNAs (Figure 4). Therefore, we considered linc00304, linc00926, and MIAT to be related with mRNAs through those RBPs and involved in the pathogenesis of AS by regulating cellular immunity.

Validation by qRT-PCR

Three target lncRNA expression levels were verified by qRT-PCR in 20 healthy individuals and 24 patients with AS. The results indicated that linc00304, linc00926, and MIAT were upregulated and showed differentially expressed levels between patients with AS and healthy controls (Figure 5A). The results were consistent with those of the GEO data, confirming reliability.

Validation of Clinical Value

This study evaluated the correlation between the expression of target lncRNAs and patient disease status. Specifically, we performed Pearson correlation analyses comparing the expression levels of target lncRNAs (linc00304, linc00926, and MIAT) and AS clinical features (ESR, CRP, ASDAScrp, BASDAI, and BASFI). Results confirmed that linc00304 expression levels were positively correlated with BASDAI, BASFI, ESR, and CRP (Figure 5B–E), linc00926 levels were positively associated with ESR (Figure 5F), and MIAT levels were positively correlated with BASFI, ESR, and CRP (Figure 5G–I) (all P < 0.05).
Additionally, multivariate logistic regression analysis was performed to test whether the prognostic performances of target lncRNAs were independently associated with disease activity (BASDAI score >4). The analysis revealed that the HR of linc00304 was 7.633 ($P = 0.046$, 95% confidence interval (CI) = 1.140−56.014). ESR and CRP levels were closely

Figure 4 lncRNA-RBP-mRNA interaction in AS. Blue represents lncRNAs, brown represents miRNAs, green represents mRNAs, and red represents the key functional enrichment terms.

Figure 5 (A) qRT-PCR-based validation of the expression of three differentially expressed lncRNAs in PBMCs of control group and patients with AS. (B–I) Correlations between target lncRNAs and AS clinical disease activity.

Additionally, multivariate logistic regression analysis was performed to test whether the prognostic performances of target lncRNAs were independently associated with disease activity (BASDAI score >4). The analysis revealed that the HR of linc00304 was 7.633 ($P = 0.046$, 95% confidence interval (CI) = 1.140−56.014). ESR and CRP levels were closely
related to the onset of AS and were risk factors, indicating that linc00304, ESR, and CRP could independently predict the onset of AS (Figure 6A).

Lastly, this study assessed the specificities and sensitivities of lncRNAs in predicting the onset of AS by calculating the area under the curve (AUC) values and cutoff values (best Youden Index). The ROC curve for serum target lncRNAs is presented in Figure 6B–D. The AUC of the ROC for the use of serum linc00304 in the diagnosis of AS was 0.687 (95% CI = 0.532–0.842), and the cutoff value was 0.413 (specificity = 0.423, sensitivity = 0.900). The AUC of the ROC for the use of serum linc00926 in the diagnosis of AS was 0.664 (95% CI = 0.490–0.838), and the cutoff value was 0.299 (sensitivity = 0.882, specificity = 0.417). The AUC of the ROC for the use of serum MIAT in the diagnosis of AS was 0.623 (95% CI = 0.423–0.822), and the cutoff value was 0.432 (specificity = 0.443, sensitivity = 0.890) (all P <0.05).

**Discussion**

Currently, AS is an incurable and debilitating autoimmune disease, and controlling disease activity cannot completely inhibit structural damage. Over the past decade, the discovery and characterization of the function of lncRNAs have revealed multiple regulatory roles. One recent study showed that peripheral blood lncRNA NONHSAT118801.2 and NONHSAT183847.1 levels were reduced in patients with AS and negatively correlated with CRP and ESR levels in these individuals, suggesting a potential protective role of lncRNAs in this pathological context.

RNA-sequencing data revealed that lncRNA contributed to disease progression of AS, as lncRNA–mRNA interaction was common. This study identified three upregulated DElncRNAs (linc00304, linc00926, and MIAT) in three AS GEO databases and validated the differential expression of those lncRNAs by qRT-PCR with 44 samples. The correlation analysis between the target lncRNA expression levels and AS activity indicators showed that linc00304 expression was positively correlated with BASDAI, BASFI, ESR, and CRP in AS (Figure 5B–E). linc00926 expression was only positively correlated with ESR (Figure 5F), whereas MIAT expression was positively correlated with BASFI, ESR, and CRP (Figure 5G–I). Linc00304, ESR, and CRP were the independent risk factors for AS activity. Although three lncRNAs
could partially predict the onset of AS, further studies are needed to understand the mechanisms underlying the effect of these target lncRNAs on disease progression.

The enrichment analysis results indicated that three novel lncRNAs (linc00304, linc00926, and MIAT) were involved in specific immune cell regulations pertinent to AS pathogenesis, such as TCR signaling pathway and BCR signaling pathway. The important molecular function term was protein binding, and the important biological process term was adaptive immune response (Figure 3). Bioinformatics analysis identified those essential genes in terms, including cytotoxic T-lymphocyte-associated protein 4 (CTLA4), CD79A, heme oxygenase 1 (HMOX1), toll-like receptor 10 (TLR10), complement component receptor 2 (CR2), kinesin family member 5A (KIF5A), inducible T-cell kinase (ITK), CD22, Fc Epsilon Receptor II (FCER2), UL16-binding protein 3 (ULBP3), B lymphoid tyrosine kinase (BLK), CD180, CD79B, CD3G, lymphocyte activation gene 3 (LAG3), and CD8A.

Adaptive immune response is an immunomodulatory process by T/B lymphocytes. Previous studies have indicated that T cell molecular associations, reception, and infections contribute to AS progression. In a 2019 RNA sequencing study, the TCR signaling pathway was a significantly enriched term of the DEmRNAs co-expressed with DElncRNAs in patients with AS. However, CTLA-4 costimulatory signal molecules were already known participants in the pathogenesis of inflammatory arthritis. The CTLA-4 drug could inhibit costimulatory signals from antigen presenting cells and prevent activation of T cells, which could improve clinical signs and symptoms of patients with inflammatory arthritis and retard the radiological progression of structural damage of affected joints. A case–control study comprising 38 patients with AS found that these patients expressed the highest CTLA-4 levels (p < 0.001). Aberrant activation of TLRs can produce a wide range of immune-stimulating cytokines and chemokines, suggesting that the TLR family may play a key role in the occurrence and development of autoimmune diseases. New evidence suggests that lncRNAs are involved in the regulation of TLR in the pathogenesis of spondyloarthropathies, including AS. For example, lncRNA-PCAT1 negatively regulates miR-145-5p, promotes the expression of TLR4, and stimulates osteogenic differentiation by activating the TLR signaling pathway of human adipose-derived stem cells.

According to the mediatory immunopathogenetic process of AS, aberrant-activated B cells were correlated with AS risk. In a study, phosphoproteomic analysis of PBMCs from 20 patients with AS and 20 controls revealed that TCR and BCR might be closely related to AS. DNA microarray analysis showed that those differential genes were involved in the BCR and TCR signaling pathways. Given these evidence, lncRNAs were speculated to trigger BCR signaling activation and further lead to B cell activation by binding of the ligand to BCR, which contributes to the cascade of inflammation to enhance AS risk.

Another key term is protein binding; most lncRNAs have been reported to have the ability to interact with the corresponding RBP. Luo et al. showed that the mRNA expression of RBP(YTHDF2) in PBMCs was significantly decreased, and RBP was positively correlated with m6A demethylase in AS. Logistic regression analysis showed that RBP expression level was a risk factor for AS. AUC for RNA binding protein in AS and control groups was 0.692, indicating that RBP expression levels in PBMCs may be involved in the pathogenesis of AS.

In conclusion, this study found three novel lncRNAs (linc00304, linc00926, and MIAT) that were significantly expressed in patients with AS. These lncRNA-RBP-mRNA interactions may participate in the development or progression of AS by adaptive immune response. Importantly, using the combination of lncRNA expression levels with conventional clinical risk factors, such as CRP, ESR, BASDAI, and BASFI, this study found these three lncRNAs to be closely related with disease activity of AS. These data suggest that linc00304, linc00926, and MIAT may be the key regulators of AS.

Data Sharing Statement
The datasets for this study can be found in the GEO database [Number: GSE181364, GSE18781, and GSE39340].

Contribution to the Field Statement
AS is an incurable and debilitating autoimmune disease, and if not controlled, the rate of disability is high. Based on GEO public database, we propose that linc00304, linc00926, and MIAT probably regulate the adaptive immune response in AS by lncRNA-RBP-mRNA interactions. TCR and BCR signaling pathways were the most critical pathways in the development and progression of AS. According to reports, most lncRNAs interact with their corresponding RBPs.
However, the RBP–lncRNA interaction in AS has not been studied. Overall, this study proposed and clinically verified that lncRNA-RBP-mRNA interaction explains a new vital regulation mechanism in AS.

**Acknowledgments**

This project was supported by the Wu Jieping Medical Foundation (Grant No. 320.6750) and National Natural Science Foundation of China (Grant No. 8157060087).

**Author Contributions**

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

**Funding**

This project was supported by the Wu Jieping Medical Foundation (Grant No. 320.6750) and National Natural Science Foundation of China (Grant No. 8157060087).

**Disclosure**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**References**


