Identification of pivotal genes and pathways in the synovial tissue of patients with rheumatoid arthritis and osteoarthritis through integrated bioinformatic analysis

YIRIXIATI AIHAI1*, XIADIYE TUERHONG2*, JIN-TAO YE3, XIAO-YU REN1 and PENG XU1

1Department of Joint Surgery, Xi'an Jiaotong University Affiliated Hong Hui Hospital, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi 710000; 2Department of Breast Surgery, Xinjiang Medical University Affiliated Tumor Hospital, Urumqi, Xinjiang 830011; 3Department of Orthopedics, Xi'an Jiaotong University Second Affiliated Hospital, Xi'an, Shaanxi 710004, P.R. China

Received September 23, 2019; Accepted March 30, 2020

DOI: 10.3892/mmr.2020.11406

Abstract. Rheumatoid arthritis (RA) and osteoarthritis (OA) are the two most common debilitating joint disorders and although both share similar clinical manifestations, the pathogenesis of each is different and remains relatively unclear. The present study aimed to use bioinformatic analysis to identify pivotal genes and pathways involved in the pathogenesis of RA. Microarray datasets from patients with RA and OA were obtained from the Gene Expression Omnibus (GEO) database and differentially expressed genes (DEGs) were identified using GEO2R software; Gene Ontology analysis and pathway enrichment were analyzed using the Database for Annotation, Visualization and Integrated Discovery and the Kyoto Encyclopedia for Genes and Genomes, respectively; and protein-protein interaction networks of DEGs were constructed using the Search Tool for the Retrieval of Interacting Genes database, and module analysis and pathway crosstalk of the PPI network was visualized using plugins of Cytoscape. In addition, the prediction of target mRNAs for differentially expressed microRNAs (DEMs) was performed using the starBase database and the identified pivotal genes were verified using reverse-transcription quantitative PCR in synovial tissue from patients with RA. A total of 566 DEGs were identified in GSE55457, GSE55235 while 23 DEMs were identified in the GSE72564 dataset. Upregulated DEGs were found to be mostly enriched in the ‘Cytokine-cytokine receptor interaction’ pathway, whereas downregulated DEGs were discovered to be enriched in the ‘PPAR signaling pathway’. The top 25 DEGs were mostly enriched in the ‘Chemokine signaling pathway’. In addition, six of the mRNA target genes were selected as potential biomarkers and a total of 24 genes were selected as potential hub genes. Experimental validation demonstrated that the expression levels of cytotoxic T-lymphocyte associated Protein 4 (CTLA4), Zeta-chain-associated protein kinase 70 (ZAP70) and LCK proto-oncogene (LCK) were significantly increased, whereas HGF expression levels were decreased in RA synovial tissue. In conclusion, these findings suggest that the identified DEGs and pivotal genes in the present study may further enhance our knowledge of the underlying pathways in the pathogenesis of RA. These genes may also serve as diagnostic biomarkers and therapeutic targets for RA; however, further experimental validation is necessary following the bioinformatic analysis to determine our conclusions.

Introduction

Rheumatoid arthritis (RA) and osteoarthritis (OA), which are the two most common chronic inflammatory joint diseases that culminate in joint deformity and disability worldwide, both result from persistent inflammatory disorders (1). Both RA and OA involve the destruction of articular cartilage, which eventually results in joint dysfunction (2); however, the pathogenesis of the two diseases is different and remains largely unknown. RA is a systemic autoimmune disease that involves multiple pathways and is characterized by hyperplasia of the synovium, formation of pannus, angiogenesis and T cell infiltration (3); the tumorigenic-like growth of the synovium causes progressive destruction of the bone and articular cartilage (4). Rheumatoid arthritis fibroblast-like synoviocytes (RAFLS)
are the most important components of RA synovial tissue and serve an important role in joint destruction through the secretion of various proteases and cytokines (5); the secretion of pro-inflammatory mediators is the main cause of the persistent pain in RA (6). In addition, the contributory role of rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA) in the deterioration of joints have also been investigated (7). ACPA leads to production of pro-inflammatory cytokines and differentiation of osteoclasts (8). Moreover, systemic appearance of ACPA precedes the onset of RA by a number of years (9). OA is a chronic degenerative disease, which is characterized by articular cartilage damage, osteophyte formation and subchondral bone sclerosis. OA is affected by multiple factors, such as age, gender, trauma history, obesity, heredity and joint deformity (10). It is widely accepted that synovitis is associated with symptoms such as joint pain and swelling, and may promote cartilage degradation (11).

MicroRNAs (miRNAs) are a class of small non-coding RNAs of ~18–25 nucleotides in length, which are highly conserved throughout evolution and serve as important post-transcriptional regulators of gene expression. Each miRNA prevents the transcription of numerous downstream genes by incompletely binding to the 3' untranslated region of target mRNAs, whereas complete complementary base-pairing leads to the degradation of mRNAs (12). miRNAs play roles in almost all aspects of cancer biology including cell cycle, programmed cell death, tumorigenesis, angiogenesis, invasion and migration (13). Abnormal expression levels of miRNAs are also involved in the pathogenesis of RA, and their different functions in the synovium, synovial fluid and serum have been reported in past decade. MiR-16, miR-146a/b, miR-150, miR-155, and miR-223 were overexpressed in both the periphery and RA joints (14). miRNA expression levels also differ depending on the stage and activity of the disease (15). Serum miR-22 and miR-103a may predict RA development (16). Serum miR-223 levels have been associated with RA activity and disease relapse (17).

Microarray datasets for RA and OA synovial tissue samples provide a unique perspective for understanding the molecular mechanisms of the diseases (18) and facilitate the identification of potential target genes and pathways for targeted therapy. Due to the complexity of the genome, most of the genes thought to be involved in the pathogenesis of RA remain to be identified; thus, the present study aimed to investigate the molecular mechanisms and potential therapeutic targets of RA using bioinformatics analysis. Microarray data from the Gene Expression Omnibus (GEO) database was obtained and used to identify differentially expressed genes (DEGs) and differentially expressed miRNAs (DEMs) in RA compared with patients with OA. In addition, Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to predict the functional enrichment of DEGs. Selected genes obtained from the bioinformatic analysis were subsequently experimentally verified in vitro. These findings may help identify potential diagnostic biomarkers and therapeutic targets of RA.

Materials and methods

Microarray data. Microarray datasets GSE55457, GSE55235 (19) and GSE72564 were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo), a public database containing gene expression profiles, chips and microarrays. GSE55457 and GSE55235 were composed of gene expression data using Affymetrix Human Genome U133A Array platform, whereas GSE72564 was performed using the Qiagen Human miRNome miScript miRNA PCR array platform. The GSE55457 dataset comprised the gene expression data of 10 synovial tissue samples from normal donors (ND), 13 synovial tissue samples from patients with RA and 10 synovial tissue samples from patients with OA; the GSE55235 dataset comprised the gene expression data of 10 synovial tissue samples from NDs, 10 synovial tissue samples from patients with RA and 10 synovial tissue samples for patients with OA; and the GSE72564 comprised of gene expression data of FLs isolated from 4 patients with RA and 4 patients with OA, which were cultivated for 4 passages.

Identification of DEGs and DEMs. The DEGS in RA and OA samples of the three datasets were identified using GEO2R software (https://www.ncbi.nlm.nih.gov/geo/geo2r/). The cut-off values for DEG selection in the GSE55457 and GSE55235 datasets were P<0.05 and |log FCI| >1.5. The expression of non-coding RNA is different from coding RNA. The number of DEMs is too small under the cut-off criteria of DEG. Thus P<0.05 and |log FCI| >1 were used for the GSE72564 dataset. The hierarchical clustering of DEGs was performed using MeV 4.9.0 software (20). The intersections of DEGS in different microarrays were visualized using VENNY 2.1.0 software (21).

GO and KEGG pathway analysis of DEGs. To determine the biological processes, molecular functions and significantly altered metabolic pathways enriched by DEGs, DEGs were subjected to functional term enrichment analysis using The Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8; http://david.abcc.ncifcrf.gov/) for GO enrichment analysis (P<0.05) and signaling pathway enrichment analysis using the KEGG database (P<0.05).

PPI network and module analysis. A protein-protein interaction (PPI) network of the DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING 11.0; https://string-db.org/) database (22), with a threshold combined score of >0.4 identified as significant, to determine the molecular mechanisms that discriminate between RA and OA pathology. Cytoscape v3.7.2 (23) was used to visualize the PPI network. In the PPI network, the edges represent the predicted functional interactions, while the network nodes represent the proteins. The degree of the node was defined as the number of interactions with other nodes. The highest degree means the most ‘popular’ gene in the PPI network. Hub genes were defined as nodes with ≥25 degrees. Interrelation analysis between the hub genes and pathways was determined using the Cytoscape plug-in ClueGO (24). The molecular complex detection (MCODE) (25) plug-in of Cytoscape was used to detect pivotal modules that may represent molecular complexes in the PPI network, using a degree cut-off value of 10 and node number >4, which was presented using Cytoscape. KEGG pathway analysis of DEGs in modules was subsequently performed using DAVID.
miRNA and target mRNA network construction. Target mRNAs of DEMs were identified using starBase v2.0 (26) and target genes were predicted using five programs: microT v5.0 (27), miRmap (28), RNA22 (29), PicTar (30) and TargetScan 7.2 (31). Genes predicted by ≥3 databases were identified as target genes for the DEMs. Subsequently, the intersection genes of MCODE network and miRNA/mRNA networks were integrated to investigate the potential interactions between DEGs and DEMs.

**Patient samples.** All experiments involving patient samples were approved by the Clinical Research Ethics Committee of Xi'an Jiaotong University (approval no. XJTULAC-2018454). Written informed consent was obtained from all participants. Human synovium biopsies were obtained from patients with RA and OA undergoing total knee arthroplasty at Xi'an Honghui Hospital from March 2018 to April 2019, who fulfilled the diagnostic criteria of the American College of Rheumatology and had a Disease Activity Score 28-joint assessment of ≥3 (32). Synovial tissue of patients with OA were obtained as the controls. The clinicopathological features of the patients are presented in Table SI.

Reverse transcription-quantitative PCR (RT-qPCR). Total RNA was extracted from synovial tissues using TRIzol® reagent (Invitrogen; Thermo Fisher Scientific, Inc.) according to the manufacturer's protocol. Total RNA (5 µg) was reverse transcribed into cDNA using the RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific, Inc.). Total RNA and primer were mixed and incubated at 65°C for 5 min. Reaction buffer RiboLock, dNTP MIX and RevertAid were mixed gently and centrifuged briefly. The mixture was incubated at 40°C for 60 min, 25°C for 5 min, finally heating at 70°C for 5 min for termination and stored at -20°C. qPCR was subsequently performed using a ChamQ Universal SYBR qPCR Master mix (Vazyme Biotech Co., Ltd.) according to the manufacturer's protocol, in a volume of 10 µl, and an Agilent StrataGene Mx3000P QPCR system (Agilent Technologies, Inc.). The following primer pairs used for the qPCR are listed in Table I. The following thermocycling conditions were used for the qPCR: Initial denaturation at 95°C for 10 min; and 45 cycles at 95°C for 10 sec, 60°C for 15 sec and 72°C for 30 sec. Expression levels were quantified using the 2-ΔΔCq method (33) and normalized to β-actin. Patient samples.

**Results**

**Identification of DEGs and DEMs.** According to the cut-off criteria (P<0.05 and log2FC>1.5), data from each microarray was separately analyzed using GEO2R to identify DEGs in RA compared with OA (RA-OA). A total of 428 DEGs were identified in the GSE55457 dataset and 289 DEGs in the GSE55235 dataset (Fig. 1B). MeV software was used for hierarchical clustering and to generate expression heat maps (Fig. 1A). To investigate the identified DEGs, DEGs shared between the two datasets and the genes found to be independently expressed were selected (Fig. 1B). These 566 genes identified in the two datasets were defined as DEGs, and consisted of 280 upregulated DEGs (uDEGs) and 286 downregulated DEGs (dDEGs). In addition, 6 miRNAs were observed to be downregulated and 17 miRNAs were upregulated in the GSE72564 dataset (Table II).

GO term enrichment analysis and KEGG pathway analysis. To determine the functions of the identified DEGs, DEGs shared between the two datasets and the genes found to be independently expressed were selected (Fig. 1B). These 566 genes identified in the two datasets were defined as DEGs, and consisted of 280 upregulated DEGs (uDEGs) and 286 downregulated DEGs (dDEGs). In addition, 6 miRNAs were observed to be downregulated and 17 miRNAs were upregulated in the GSE72564 dataset (Table II).

**Statistical analysis.** Statistical analysis was performed using SPSS 20.0 software (IBM Corp.) and GraphPad Prism 7.0 software (GraphPad Software, Inc.). The experimental data are presented as the mean ± SD. Statistical differences between two groups were determined using a Student's t-test. P<0.05 was considered to indicate a statistically significant difference.

**Table I. Primers used for reverse transcription-quantitative PCR.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI3CG</td>
<td>F: CACCCAAAGCATATCTCAAGC</td>
</tr>
<tr>
<td></td>
<td>R: TAAGTCAGAATACATCGTCC</td>
</tr>
<tr>
<td>LCK</td>
<td>F: GTGTTGAGAAGCTGCATTTC</td>
</tr>
<tr>
<td></td>
<td>R: GATTGGAGCCTCTGTAGTAAAC</td>
</tr>
<tr>
<td>ZAP70</td>
<td>F: AGAGCTCTGGAGTTTCTACCT</td>
</tr>
<tr>
<td></td>
<td>R: TCTCGAGCCGATTGCAAGAGA</td>
</tr>
<tr>
<td>LEP</td>
<td>F: CATCAAGACAAATTGTCCACAGG</td>
</tr>
<tr>
<td></td>
<td>R: GTCTGGTGATATTTGATACG</td>
</tr>
<tr>
<td>HGF</td>
<td>F: AATCCACTCATCTCTGGGATT</td>
</tr>
<tr>
<td></td>
<td>R: TCCCCATTACAATCGCAAAATG</td>
</tr>
<tr>
<td>CTLA4</td>
<td>F: CAGTTAGTTCGGGGTTGTTTT</td>
</tr>
<tr>
<td></td>
<td>R: TTTTCACATTCTGGCTTGTTT</td>
</tr>
<tr>
<td>SERPINE1</td>
<td>F: AACGTGGTTTCTTCCACCTAT</td>
</tr>
<tr>
<td></td>
<td>R: CAATCTTGAAATCCATAGCTG</td>
</tr>
<tr>
<td>SDC1</td>
<td>F: AAAGATATCCCTTGGTCACAGCA</td>
</tr>
<tr>
<td></td>
<td>R: GTTGGAGACGGTGGAAGATAG</td>
</tr>
<tr>
<td>NPY1R</td>
<td>F: GGACCGATGTGTAAGTTGAAATC</td>
</tr>
<tr>
<td></td>
<td>R: ACCAAAATCACAGCAATACTTA</td>
</tr>
<tr>
<td>ACTB</td>
<td>F: AAGGATCCCTATGGTGCGGAC</td>
</tr>
<tr>
<td></td>
<td>R: CAGTACAGGGATACGACAGCC</td>
</tr>
</tbody>
</table>

F, forward; R, reverse. PI3CG, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic sub-unit γ; LCK, LCK proto-oncogene Src family tyrosine kinase; ZAP70, ζ chain of T cell receptor-associated protein kinase 70; LEP, leptin; HGF, hepatocyte growth factor; CTLA4, cytotoxic T-lymphocyte associated protein 4; SERPINE1, serpin family E member 1; SDC, syndecan 2; NPY1R, neuropeptide Y receptor Y1; ACTB, β-actin.
significantly enriched signaling pathways of the uDEGs were identified as ‘Cytokine-cytokine receptor interaction’, ‘Chemokine signaling pathway’, ‘Hematopoietic cell lineage’, ‘Primary immunodeficiency’ and ‘Cell adhesion molecules (CAMs)’ (Fig. 2B). Only six downregulated pathways fulfilled the condition of P<0.05; these were ‘PPAR signaling pathway’, ‘Regulation of lipolysis in adipocytes’, ‘Dilated cardiomyopathy’, ‘Arrhythmogenic right ventricular cardiomyopathy (ARVC)’ and ‘ECM-receptor interaction’ (Fig. 2B).

**PPI network construction and module screening.** The PPI network of DEGs, constructed using STRING and visualized using Cytoscape, consisted of 481 nodes and 1,728 edges (Fig. 3). The top 25 nodes with the highest degree were selected. Pathway crosstalk analysis of the top 22 nodes identified was performed through performing KEGG pathway analysis in ClueGO (Fig. 4), with P<0.05 as the cut-off criterion. A total of 25 of these 23 genes were found to be involved in inflammatory processes (Table III), with 11 identified as chemokines and their receptors, which are abundant in the peripheral blood and in the local inflamed joints of RA (34). Thus, five different hub genes were selected for further experimental validation: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit γ (PIK3CG), lymphocyte-specific kinase (LCK), leptin (LEP), cytotoxic T-lymphocyte antigen (CTLA4) and ζ chain of T cell receptor associated protein kinase 70 (ZAP70). Module analysis was performed using the MCODE plug-in of Cytoscape and a total of 6 modules were selected using the previously mentioned criterion from the whole PPI network (Fig. 5). KEGG pathway enrichment analysis of these genes was performed using DAVID and the pathway analysis demonstrated that the selected genes in the modules were mostly enriched in the ‘Chemokine signaling pathway’, ‘Primary immunodeficiency’, ‘Proteoglycans in cancer’, ‘Wnt signaling pathway’ and ‘Asthma’ (Fig. 5).

**miRNA/mRNA network construction.** The target genes were predicted using starBase and compared with DEGs to identify pivotal target mRNAs as potential biomarkers. Only the intersection was selected and visualized using Cytoscape (Fig. 6). The MCODE results and the miRNA/mRNA network were analyzed and a total of 6 genes were chosen as candidate genes (Table II), consisting of two dDEGs [hepatocyte growth factor (HGF) and neuropeptide Y receptor Y1 (NPY1R)] and four uDEGs [CD69, serpin family E member 1 (SERPINE1), syndecan 1 (SDC1) and C-X-C motif chemokine ligand 3 (CXCL3)].

**Experimental validation.** RT-qPCR was performed to verify the expression levels of the 9 candidate genes in OA and RA synovial tissue. The expression levels of CTLA4, ZAP70, LCK, PIK3CG, SERPINE1, SDC1, NPY1R in RA tissue compared with OA tissue were observed to be consistent with the predictions (Fig. 7). In addition, the experimental validation found that CTLA4, ZAP70 and LCK expression levels were significantly increased, whereas HGF expression levels were significantly decreased in RA synovial tissue compared with OA synovial tissue (P<0.05; Fig. 7).

**Discussion**

Microarray and bioinformatic analysis is widely used to investigate the causes and underlying mechanisms of different types of disease (35), and gene expression profiling of rheumatoid arthritis (RA) has been conducted in numerous previous studies. Orange et al (36) Used RNA-seq of synovial tissue along with histologic analysis identified three distinct
Table II. Identification of the differentially expressed microRNAs in the GSE72564 dataset obtained from the Gene Expression Omnibus database.

<table>
<thead>
<tr>
<th>miRNA ID</th>
<th>P-value</th>
<th>Log. fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-670</td>
<td>0.00181</td>
<td>1.5375</td>
</tr>
<tr>
<td>hsa-miR-26a</td>
<td>0.00365</td>
<td>2.0325</td>
</tr>
<tr>
<td>hsa-miR-2116</td>
<td>0.00745</td>
<td>1.8950</td>
</tr>
<tr>
<td>hsa-miR-653</td>
<td>0.01371</td>
<td>-1.2775</td>
</tr>
<tr>
<td>hsa-miR-190a</td>
<td>0.02222</td>
<td>1.1625</td>
</tr>
<tr>
<td>hsa-miR-2276</td>
<td>0.02301</td>
<td>-1.8875</td>
</tr>
<tr>
<td>hsa-miR-548b</td>
<td>0.02336</td>
<td>1.5025</td>
</tr>
<tr>
<td>hsa-miR-579</td>
<td>0.02551</td>
<td>1.2850</td>
</tr>
<tr>
<td>hsa-miR-496</td>
<td>0.02800</td>
<td>-1.0725</td>
</tr>
<tr>
<td>hsa-miR-1305</td>
<td>0.02961</td>
<td>1.7775</td>
</tr>
<tr>
<td>hsa-miR-30c</td>
<td>0.02995</td>
<td>1.6525</td>
</tr>
<tr>
<td>hsa-miR-20b</td>
<td>0.03323</td>
<td>1.1150</td>
</tr>
<tr>
<td>hsa-miR-4262</td>
<td>0.03394</td>
<td>1.1850</td>
</tr>
<tr>
<td>hsa-miR-4263</td>
<td>0.03396</td>
<td>-1.0450</td>
</tr>
<tr>
<td>hsa-miR-502</td>
<td>0.03450</td>
<td>1.1650</td>
</tr>
<tr>
<td>hsa-miR-1258</td>
<td>0.03540</td>
<td>1.4250</td>
</tr>
<tr>
<td>hsa-miR-708</td>
<td>0.03672</td>
<td>2.1575</td>
</tr>
<tr>
<td>hsa-miR-1193</td>
<td>0.04084</td>
<td>-1.4750</td>
</tr>
<tr>
<td>hsa-miR-1299</td>
<td>0.04256</td>
<td>1.6225</td>
</tr>
<tr>
<td>hsa-miR-218</td>
<td>0.04671</td>
<td>2.7450</td>
</tr>
<tr>
<td>hsa-miR-346</td>
<td>0.04710</td>
<td>-1.0200</td>
</tr>
<tr>
<td>hsa-miR-3116</td>
<td>0.04784</td>
<td>1.1275</td>
</tr>
<tr>
<td>hsa-miR-499a</td>
<td>0.04941</td>
<td>1.4275</td>
</tr>
</tbody>
</table>

miR, microRNA.

molecular subtypes of RA that correlated with specific clinical phenotypes.

In the present study, genome-wide transcriptomic datasets from the GEO database were used to identify DEGs in patients with RA or osteoarthritis (OA). Combining the data from the GSE55457 and GSE55235 dataset revealed that 566 genes were differentially expressed between OA and RA samples (280 genes were upregulated and 286 genes were downregulated). In addition, six miRNAs were found to be downregulated and 17 miRNAs were upregulated in the GSE72564 dataset.

GO functional term enrichment analysis indicated that upregulated (u)DEGs were mostly enriched in BP such as ‘Immune response’, CC such as ‘External side of plasma membrane’ and MF such as ‘Chemokine receptor activity’, whereas downregulated (d)DEGs were significantly enriched in ‘cell adhesion’ at the BP level, ‘Extracellular region’ at CC level and ‘heparin binding’ at the MF level. These findings are consistent with the previous studies that reported that immune responses and chemokine activity were vital processes for RA development and progression (37,38). KEGG signaling pathway enrichment analysis indicated that uDEGs were mostly enriched in ‘Cytokine-cytokine receptor interaction’ and ‘Chemokine signaling pathway’. A diverse range of cytokines have been observed to have important roles in the pathophysiological processes of RA, such as interleukin (IL)-1, IL-17 and tumor necrosis factor (TNF)-α (39). In fact, the development of therapeutic agents, such as rituximab and infliximab, which are targeted agents against cytokines or their receptors, have significantly increased the success rate for the treatment of the disease (40). The dDEGs were found to be enriched in the ‘PPAR signaling pathway’ and ‘Regulation of lipolysis in adipocytes’. Peroxisome proliferator activated receptors (PPARs) are transcription factors that belong to the nuclear receptor family (41) and they have been reported to inhibit the expression of matrix metalloproteinases (MMP) and the release of pro-inflammatory cytokines when exposed to IL-1β (42). Decreased PPAR activation may promote FLSs proliferation and expression levels of c-Myc, Cyclin D1, MMP-1 and MMP-9 (43).

In addition, based on the PPI network constructed of the DEGs, the top 25 nodes with the highest degree were selected and pathway interactions of these genes were analyzed using ClueGO; this revealed that 17 of these genes were involved in the inflammatory process. Some of these genes identified in the present study have previously been attributed to the pathogenesis of RA; for example, 11 of them were chemokines and their receptors, including C-X-C motif chemokine ligand 1 (CXCL1), CXCL2, CXCL3, C-X-C motif chemokine receptor 4 (CXCR4), C-C motif chemokine ligand 5 (CCL5), CCL8, CXCL9, CXCL10, CXCL11, CXCL13 and CCL20. Chemokines are divided into four classes depending on the location of the conserved cysteine in the protein molecule: C-X-C, C-C, C-X-C and C-X-C (44). Chemokines and their receptors control hemostasis during recirculation and promote the recruitment of immune cells during inflammation (45). Of note, the role of multiple chemokines and growth factors, such as vascular endothelial growth factor, in angiogenesis and the inflammatory response has been reported in RA (46). CXCL8, also known as IL-8, is the most commonly studied chemokine of the CXC subfamily and the expression levels of CXCL8 have been closely associated to the symptoms and disease activity of patients with RA (47). CCL20, also known as liver and activation-regulated chemokine, is one of the few known chemokine ligands that only pairs with a sole receptor, CCR6 (45). CCL20 is able to recruit CCR6+ mononuclear cells to the synovial fluid of patients with RA and has been found to be significantly neutralized with the addition of an anti-CCL20 antibody (48). CCL20 has also been reported to activate osteoblast proliferation and osteoclast differentiation, which may be maintain bone homeostasis in RA bone destruction (49). Furthermore, the blockade of inflammatory cytokines through biological agents, such as infliximab, inhibited the production of CCL20 (50). CCL5 is found to serve a positive role in leukocyte recruitment during inflammation (51). In addition, human RAFLS treated with CCL5 demonstrated significantly increased expression levels of MMP-1 and MMP-13, which resulted in the degradation of type I collagenase (52). The CXC subfamily are the predominant chemotactic cytokines of the neutrophils; CXCL10 and its receptor CXCR3, have been found to promote FLS invasion and joint erosion by increasing the expressing levels of receptor activator of nuclear factor-κ B ligand (RANKL) in CD4+ T cells in RA (53,54). Notably, it has been suggested that serum concentrations of CXCL9 and CXCL10 may also serve as sensitive biomarkers for RA (55).
PIK3CG has been demonstrated to serve an important role in proliferation, apoptosis and adhesion (56). In hTNF-α transgenic mice, reduced MMP-3 expression levels and invasive properties were observed in synovial fibroblasts in the absence of PIK3γ (57). In line with this observation, the pharmacological inhibition of PIK3CG was discovered to effectively
alleviate chronic inflammation in RA (58). Thus, it is thought that the full therapeutic potential of selective PIK3γ inhibitors remains to be investigated.

LCR is a member of the Src family of protein tyrosine kinases (59), which is only expressed in T cells and natural killer (NK) cells, and is involved in T cell receptor-mediated T cell proliferation and differentiation (60). In the present study, LCK was observed to be significantly over-expressed in RA synovial tissue and the overexpression of LCK has been found to contribute to numerous autoimmune diseases, such as inflammatory bowel diseases and type I diabetes (59). Selective LCK inhibitors have been developed and have reported potential upon their application in inflammatory disorders, including RA (60). In addition, previous bioinformatical analysis indicated that LCK expression was increased in RA (61), which is consistent with what was observed in the present study, in which the aberrant expression of LCK was experimentally validated.

Leptin is one of the most important hormones and cytokines secreted by adipose tissue and the past decade of research has proven that leptin controls body weight by inhibiting food intake (62). As a cytokine, Leptin modulates inflammatory responses and the immune system by promoting Th1 cell activation and the production of pro-inflammatory cytokines, such as IL-6, IL-2 and TNF-α (63,64); however, the expression levels and role of Leptin in the pathogenesis of RA remains unclear. Some previous studies have observed higher expression levels of Leptin in the serum, synovial fluid and synovial tissue from patients with RA, which suggested that Leptin expression levels are related to disease activity (65,66). In contrast, other studies have reported no change in the expression levels of leptin and no correlation with disease activity (67), because the individual characteristics of patients such as age, race and body weight were inconsistent (66). Thus, further investigations on Leptin will help determine its mechanism of action and its potential as a treatment option in RA.

ZAP70, also known as SRK, encodes an enzyme belonging to the protein tyrosine kinase family. In a previous study, ZAP70+B cells were identified as a biomarker of response to B cell depletion therapy in RA (68) and partial ZAP-70 deficiency was found to alter the balance between the activation and apoptotic processes of T cells (69). In the present study, ZAP70 was found to be significantly enriched in the NF-κB signaling pathway, which is a widely researched signaling pathway in RA. Meanwhile, CTLA-4 is a T cell surface glycoprotein, which helps prevent the development of autoimmune diseases through the inhibition of T cell activation (70).

The intersected genes in MCODE and target genes of miRNAs were selected for further experiment. In the present study, submodules in MCODE analysis were found to be associated
with the ‘Chemokine signaling pathway’, ‘Primary immunodeficiency’, ‘Wnt signaling pathway’, ‘Asthma’ and ‘Proteoglycans in cancer’. *HGF, SERPINE1, SDC1, NPY1R, CTLA4 and CXCL3* were selected as potential biomarkers. *HGF* has been discovered to stimulate hepatocyte proliferation, in addition to promoting proliferation, angiogenesis, anti-inflammatory and apoptotic
Figure 6. miRNA/mRNA network. Differentially expressed miRNAs are presented as blue diamonds. Upregulated DEGs are presented in red and downregulated DEGs are marked in green. DEGS, differentially expressed genes; miR/miRNA, microRNA.

Figure 7. Reverse transcription-quantitative PCR results of the expression levels of identified candidate genes in samples from OA vs. RA. Expression levels of each gene was normalized to ACTB. *P<0.05, **P<0.01 vs. OA. OA, osteoarthritis; RA, rheumatoid arthritis. PI3KCG, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic sub-unit γ; LCK, LCK proto-oncogene Src family tyrosine kinase; ZAP70, ζ chain of T cell receptor-associated protein kinase 70; LEP, leptin; HGF, hepatocyte growth factor; CTLA4, cytotoxic T-lymphocyte-associated protein 4; SERPINE1, serpin family E member 1; SDC1, syndecan 1; NPYIR, neuropeptide Y receptor Y1.
processes in cells other than hepatocytes (71). SERPINE1 is a member of urokinase plasminogen activating system; increased expression levels of SERPINE1 have also been related to invasion and migration in head and neck cancer cells (72). Previous KEGG pathway analysis has indicated that SERPINE may be related to the HIF-1 signaling pathway, which alongside its downstream receptors, is activated in a hypoxic environment during RA (73); however, to the best of our knowledge, there are currently no studies published on the differential expression and function of SERPINE1 in the pathogenesis of RA. SDC1 is expressed in hepatocytes, epithelial cells and the endothelium, where it interacts with multiple chemokines and cytokines to regulate differentiation, migration and proliferation (74). In addition, SDC is a phenotypic marker and regulator of NKT17; in a previous study, SDC1 deficiency led to a significant expansion of NKT17 cells, which was detected by the increased expression levels of IL-17 (75). NPY1 is mediated by the NPY receptor subtypes Y1-R and Y2-R (76). NPY accompanied with NPY1 receptor (NPY1R) and NPY2 receptor (NPY2R) are overexpressed in human OA cartilage, which results in chondrocyte hypertrophy and cartilage degradation (77). Increased expression levels of NPY1R have been found to regulate proliferation and metastasis in breast cancer (78). In the present study, NPY1R was discovered to be involved in the ‘regulation of lipolysis in adipocytes’ pathway.

The present study experimentally validated that PIK3CG, LCK, ZAP70, HGF, CTLA4, SERPINE1 exhibited consistent expression trends with the microarray datasets GSE55457 and GSE55235, whereas LEP, SDC-1 and NPY1R were downregulated, but PCR demonstrated them to be upregulated., which may be caused by disease activity and ethnic differences. LCK, ZAP70 and CTLA4 expression levels were observed to be overexpressed, whereas expression levels of HGF were significantly decreased in the RA synovium. Conversely, previous studies have reported significant increases in the expression levels of serum HGF in patients with RA (79), which is inconsistent with our findings. Therefore, further research is required to determine the role of HGF.

To conclude, transcriptome sequencing has been widely used for research into both cancer and autoimmune diseases. In the present study, bioinformatics methods were used to analyze the presence of DEGs in the synovial tissue of patients with RA compared with OA. The expression levels of LCK, ZAP70 and CTLA4 were found to be overexpressed, whereas HGF expression levels were significantly decreased in RA. These genes and their related signaling pathways may provide diagnostic biomarkers and future therapeutic targets for RA; however, only CTLA4, ZAP70, LCK and HGF were validated to be significantly differentially expressed in the synovial tissue of patients with RA, which indicated that experimental validation is necessary following bioinformatics analysis.

Acknowledgements

Not applicable.

Funding

The present study was supported by a grant from the National Natural Science Foundation of China (grant no. 81772410).

Availability of data and materials

The datasets analyzed during the current study are available from the GEO database (www.ncbi.nlm.nih.gov/gds/) with the accession nos. GSE55457, GSE55235 and GSE72564.

Authors’ contributions

YA, XT and PX designed the experiments and wrote the manuscript; JTY collected the data and XYR performed the experimental validation and statistical analysis. YA wrote the manuscript All authors read and approved the final version and agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

The present study was approved by the Clinical Research Ethics Committee of Xi’an Jiaotong University (approval no. XJTULAC-2018454). Clinical information in this study was collected after obtaining written informed consent from all patients.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


60. Kumar Singh P, Kashyap A and Silakari O: Exploration of the
61. Camps M, Rückle T, Ji H, Ardissone V, Rintelen F, Shaw J,
62. Procaccini C, Pucino V, Mantzoros CS and Matarese G: Leptin
63. la Cava A: Leptin in inflammation and autoimmunity.