



Original article

Identification of differentially expressed genes and pathways crosstalk analysis in Rheumatoid and Osteoarthritis using next-generation sequencing and protein-protein networks



Shenqiang Qiu^a, Anum Munir^b, Shaukat Iqbal Malik^c, Sajid Khan^d, Amjad Hassan^{b,*}

^a Department of Hand and Foot Surgery, Shandong Provincial Hospital affiliated to Shandong First Medical University, No.324, Jingwu Road, Jinan, Shandong Province 250021, PR China

^b Department of Biotechnology, COMSATS University Islamabad, Abbottabad Campus, 22010 Abbottabad, Pakistan

^c Department of Biosciences, Faculty of Health and Life Sciences, Capital University of Science and Technology, Islamabad 44000, Pakistan

^d Department of Bioinformatics, Govt. Postgraduate College Mandian, Abbottabad 22010, Pakistan

ARTICLE INFO

Article history:

Received 17 March 2021

Revised 15 April 2021

Accepted 21 April 2021

Available online 1 May 2021

Keywords:

Cartilage
Chondrocytes
Cross talk
DEGs
NGS data

ABSTRACT

Osteoarthritis occurs when protective cartilage of bones worn out. Similarly, cartilage damage occurs mainly in the pannus cartilage in rheumatoid arthritis. It is a potentially debilitating condition, affecting women two to three times more often than men. The cause and prognosis of rheumatoid and osteoarthritis are still poorly known. However, advances in the study of disease pathogenesis have encouraged the creation of new therapeutics with improved outcomes. The purpose of this study is to investigate the differentially expressed genes potentially involved in dysregulated rheumatoid arthritis (RA) and their association to other types of arthritis, including osteoarthritis (OA). Complete RNAs were isolated for RNA expression profiling using next-generation sequencing from human primary cultured normal and RA chondrocytes. From RNA sequencing results 250 differentially expressed genes were identified using bioinformatics analysis, of which 32 were found to be significantly playing role in RA pathogenesis and its associated diseases. Molecular ontologies of the identified genes showed they are connected to Innate immune response, Protein phosphorylation, Transcription initiation from RNA polymerase II promoter, Immune response, Neoplasms of bones, as well as osteoarthritis, and Rheumatoid arthritis. Among the identified genes, TRAF1, TRAF2, BAMP, STX11, MEOX2, AES, REL, FHL3, PNMA1, SGTA, LZTS2, SIAH2, PNMA1, and TFCP2 were found to be highly enriched in the protein-protein interaction network. The significant cross talks were found in Hypertrophic cardiomyopathy, Small cell lung cancer, Proteasome, p53 signaling pathway, Arrhythmogenic right ventricular cardiomyopathy, Small cell lung cancer, SNARE interactions in vesicular transport, RIG-I-like receptor signaling pathway, and Hypertrophic cardiomyopathy pathways. The results offer new opportunities for target gene control in RA and OA cartilage destruction.

© 2021 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Rheumatoid arthritis (RA) is an inflammatory autoimmune arthritis that affects roughly 1% of the world's population. It is a

potentially crippling illness, two or three times affecting women more often than men (Essouma et al., 2020). It is marked by joint discomfort and swelling and causes permanent joint injury that adversely affects the quality of the patient's life in the absence of therapy (Coras et al., 2020). Progressive injury, systemic problems, early mortality, and socio-economic costs are linked to RA (Firestein, 2003). Osteoarthritis is a leading cause of suffering, injury, and economic loss around the world. The disorder's epidemiology is multifactorial and fluid, with genetic, biochemical, and biomechanical elements. Joint-specific aetiological conditions are also present. The origin and prognosis of RA and OA are uncertain. Advances in studying the disease's pathogenesis, however, have facilitated the development of novel therapeutics with better

* Corresponding author.

E-mail address: amjadhassan@cuatd.edu.pk (A. Hassan).

Peer review under responsibility of King Saud University.



<https://doi.org/10.1016/j.sjbs.2021.04.076>

1319-562X/© 2021 The Authors. Published by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

results. The current medical plan, which illustrates this improvement is to begin intensive therapy immediately after diagnosis and to intensify therapy in search of therapeutic remission, driven by an evaluation of disease development (McInnes and Schett, 2011). Etiopathogenetic causes that have been confirmed or at least highly hypothesized include genomic variations (Wellcome Trust Case Control Consortium, 2007), alterations in gene expressions (Tasaki et al., 2018), autoimmunity (Catrina et al., 2016), and environmental factors (Giannini et al., 2020). The first factor that contributes to the autoimmune inflammation and causes RA is still unknown, no aspect is believed to be a single cause. There is no single source of RA, no single route to progression, and no single curative method, as single treatments have reduced ultimate success rates (Ramírez et al., 2020; Platzer et al., 2019). This has contributed to the possibility that there could be RA subtypes of multiple manifestations of RA disease that differ on sex, genotype, gene expression, or microbiome composition, which will make RA an important personalized medicine case (MacGregor et al., 2000). Similarly, the pathogenesis of OA is also poorly known.

There is a complex relationship between genotype, environmental factors, and the development of RA and OA. twin experiments conducted for RA and OA include genetic factors, with concordance rates of 15–30 percent among monozygotic twins and 5 percent among dizygotic twins. Genome-wide tests indicate that the disorder is focused on immune regulatory factors (Wang et al., 2014a,b,c). The long-established correlation with human leukocyte antigen (HLA)-DRB1 locus has been verified in patients who are positive for RA. The knowledge about RA has rapidly progressed with the advancement of genomics, and various core RA-associated genes and metabolic pathways have been identified, which may have significant roles in the pathogenesis and evolution of RA. For instance, the toll-like receptor (TLR) signaling pathway has been identified as a major contributor to RA (McGovern et al., 2005). However, there is a frequent pathogenic expression of inflammatory cytokines, including TNF- α , IL-1 β , and IL-6, each of which has been effectively targeted in patients with RA and the downregulation of innate immunity in patients with RA (Wang et al., 2014a,b,c). Overexpression and functional mutations of the tumor suppressor p53 protein have been identified in RA synovial tissue, most often in patients with an advanced destructive form of the disease. Other signaling pathways involved in RA have also been documented, including the NOD-like signaling pathway (Porcelli et al., 1993), the signaling pathway of the T cell receptor (Sen, 2005), and the signaling pathway of the Wnt cell (Feldmann and Maini, 2003). Inflammatory cytokines (such as IL-6 and IL-1) (Nakano et al., 2013), human leukocyte antigen (Chen et al., 2018), and DNA methylation (Yoshizawa et al., 2008) have been reported as affecting the growth of RA and can be used as diagnostic RA markers. In RA cartilage, chondrocyte apoptosis occurs through activated p53 and c-myc and decreased Bcl-2 expression, and the degree of chondrocyte apoptosis is correlated with cartilage destruction (Proost et al., 2006). Several other pathways such as the classical mechanism of signalization of mitogen-activated protein kinase, insulin signal pathway, antigen processing, and presentation and IgA intestinal immune system. RA is also triggered by graft versus host disease and autoimmune thyroid mechanisms associated with disease (Xeu et al., 2013).

There has been a great deal of work to find or test genes and pathways of relevance for RA progression (Lee et al., 2004). So far, a lot of work has been done to gain a more detailed view of the detailed genome and expression data (Giannopoulou et al., 2015), but still, there is a space to find more. For the study of gene expression data, there are currently various traditional approaches, including comparing differentially expressed genes and clustering. However, a growing number of studies indicate that only looking for genes with substantially different levels of expression under

various conditions lacks adequate specificity to represent the relationship between genes (Goulielmos et al., 2016). In conjunction with the use of bioinformatics methods, modern sequencing technologies allow for wider investigation and a greater understanding of pathogenesis and disease variants in RA (Ammari et al., 2013). Interactions between synovial fibroblasts and chondrocytes and the suggested autocrine and paracrine effects of chondrocytes in the degradation of RA cartilage indicate that chondrocytes play a significant role in RA pathogenesis (Chhangawala et al., 2015). Therefore, the current research uses chondrocytes of the hip joint to classify differentially expressed genes (DEGs) in RA. Furthermore, Protein-protein Interactions (PPI) and Kyoto encyclopedia of genes and genomes (KEGG) pathways data, the pathway cross-talk network were also developed between RA linked pathways. This study would also shed light on deeper knowledge of RA's molecular pathways. In the meantime, this may provide fresh insight into the cure for it.

2. Material and methods

2.1. Cell culture of human primary chondrocytes

3 normal (N) from healthy individuals having a catalog number C-12710 and 3 RA (D) pre-mature chondrocytes of human hip joint isolates (catalog number; C-12750) were purchased from Promo Cell company (Heidelberg, Germany). The N chondrocytes were cryopreserved in nature whereas, the D were proliferating. Both types were grown to mature Chondrocytes by keeping them in a Growth Medium containing 5% CO₂ at 37 °C in the humidified incubator until they get matured. Then, the mature N and D cells were collected for RNA extraction and expression profiling.

3. Sequencing of RNA and its expression profiling

The same methodology was used for RNA sequencing as reported in (Yoshizawa et al., 2008), Trizol[®] Reagent was used to collect RNA from both N and D samples. OD260 / OD280 1.91 for N and D 1.92 absorption ratio detection was used to check the quality of collected RNAs, RNA integrity was checked by the Nano Drop-1000 spectrophotometer 8.9 integrity value was observed for N and 7.8 for D. Once the RNAs were extracted, RNA sequencing was performed using Illumina sequencing, adapters were then added to each end of the fragments and were labeled with several probe IDs for gene identification, short read length sequences of 75 nucleotides were generated, which was sufficient for the study of differential expression (Kaya et al., 2019). For each sample, approximately 28 million reads were produced, which were reasonable for stable detection of protein-coding genes (Russo et al., 2018).

3.1. Preprocessing of raw sequences

Once the raw sequences were obtained, they were trimmed to generate proper read libraries using the Cuff link tool. The Cuff link tool provided results containing probe IDs instead of gene symbols, therefore, GEO2R server was used to convert the probe IDs into their respective gene symbols. probe samples with faulty or unfavorable values were removed. Once the transcripts were generated, the data was mapped to human genome sequencing by Picard software. The alignment was done to find the presence of splice variants and modifications. Expression values were estimated using the Bitseq tool. The gene expression values obtained through the Bitseq tool were varying largely, therefore min-max normalization was performed within the range 0–1. The data were then pre-processed in the R studio (Li and Barber, 2019) using the R language Preprocess Core kit.

3.2. Differentially expressed genes

On preprocessed results, the Limma package (Li and Barber, 2019) in the R language was used to discern DEGs between the N and D cells. The cutoff condition for DEGs was ranging from 0.06 to 0.33 as the p-value, fold shift (B): –7.5 to 3.3, and logFC to –3.7 to 5.11. The raw p values were then modified to acceptable p values using the Benjamini & Hochberg process (Sherman et al., 2007). A total of 250 genes was identified as differentially expressed

3.3. Gene ontology (GO) of DEGs

The DEGs were clustered on a functional basis using the DAVID Bioinformatics tool that also offers enrichment values for each cluster. For each gene ontology (GO) term, a p-value was calculated by the hypergeometric distribution associated with a group of genes and then updated using the Benjamini method for multiple testing. Enrichment scores were then measured for each cluster as the negative logarithm of the individual GO p-values using geometric calculation (Sherman et al., 2007). Of those 250 genes, 31 were discovered to be functionally enriched in RA.

3.4. Pathways enrichment analysis of DEGs

GenCLIP 2.0 server was used to identified enriched pathways of DEGs, it is a web-based text-mining tool for gene clustering and creation of molecular networks (Wang et al., 2014a,b,c). For GO classifications, the cutoff conditions were set as $p = 0.05$ and a hit = 20. The probability that a set of randomly picked genes contain more than 30 genes or pair of 19 genes related to the disease keywords is calculated as P-value, which means the distribution of the number of related gene and gene pairs is similar to the normal distribution. To classify enriched pathways, the gene list obtained from GenCLIP 2.0 was then sent to the KEGG Pathways database. For this function, as a cutoff value, the p-value was set to 0.03. The enriched pathways obtained by KEGG were further cross-checked at a cutoff p-value (0.03 to 0.07) using the Network analyst tool (Xia et al., 2014).

3.5. Construction of protein-protein interactions networks

To identify human protein-protein interactions (PPIs), PICKLE (Protein InterAction KnowLedgEbase) 2.0 was used (Gioutlakis et al., 2017). It is a protein database that is accessible through the internet. A total of 14,774 PPI pairs were derived from PICKLE. Using the Network Analyst, the PPI network was constructed. The edges and nodes of the generated PPI network were so complicated that further research was needed to see the enriched functional modules of the PPI network, so the important subnetworks were also developed. Finally, using the EnrichNet tool (Glaab et al., 2012), a study of the modules for GO and pathway enrichment was carried out.

3.6. Pathways cross talk generation

KEGG database, EnrichNet, and GeneCLIP 2.0 enriched pathways were screened and major RA pathways were identified with at least 1 protein overlap between any given pair of DEGs. To calculate the similarity of gene expression by weight, a Pearson correlation coefficient was used, the nodes and edge weights were calculated in the networks. To calculate the statistical significance of functional interactions, the following formula was used.

$$S(e) = f(\text{diff}(x), \text{corr}(x, y), \text{diff}(y)) = -2 + \sum_{i=1}^n \log_e(p_i) \quad (2)$$

where $\text{diff}(x)$ and $\text{diff}(y)$ represent the quantitative measurement of the differential expressions of gene x and gene y , $\text{corr}(x,y)$ shows the frequency of interaction between gene x and gene y based on the rates of expression; f illustrates a generalized method of data aggregation across a variety of statistical tools taking into account various data sets, and Pathways cross-talk network was developed in a Gephi Tool.

3.7. Disease ontology (DO) of DEGs

The TargetMine tool was used to evaluate the direct relation between DEGs and RA. It is an advanced data warehouse for identified disease-related genes and recovery of candidate target genes and proteins for RA from large-scale studies

4. Results

The expression profiling of normal and diseased chondrocytes using the NGS technique was obtained from RNA sequencing. Using the Bioinformatics technique, the differently expressed genes in N and D were identified. The findings showed that 250 DEGs were identified after preprocessing, including 121 down-regulated genes and 129 up-regulated genes. Dual specificity phosphatase 6 (DUSP6) had the lowest statistical significance (p-value) of 0.0934 and had the highest LogFC = 5.113 and CUB and Sushi multiple domains 1 (CSMD1) had the highest $p = 0.23$ and the lowest fold changes of logFC = –0.374 among all DEGs, respectively. Limma offers several alternatives for P-value modification. These modifications, also known as numerous tests, seek to correct false positive outcomes. The false rate discovery approach of Benjamini & Hochberg was chosen as it is the most frequently used correction for microarray data and offers a strong balance between statistically important gene detection and false-positive limitation. The p-value between 0 and 1 shows significant discovery rates, whereas, positive LogFC values demonstrate upregulated genes and negative values show down-regulated genes.

It was found that only 32 of the 250 DEGs were specifically involved in arthritis, osteoarthritis, and associated diseases. So, for further study, a total of 32 genes was chosen. The GO terms of DEGs (Table 1) obtained by DAVID shows that the highest number of genes (10 out of 32) were playing role in protein phosphorylation and innate immune response mechanisms, showing 31.25% of DEGs, whereas, inflammation, RA, neoplasms, and osteoarthritis include 9.3% of total DEGs. Fig. 1 shows the heat map created for the DEGs.

From table 1, it was observed that TNFRSF1B gene was playing common role in all types of arthritis including, osteoarthritis, osteoporosis, rheumatoid arthritis, and arthritis. Gene clusters and their molecular networks were discovered using Gene Clip 2.0. The p-value was set to ≤ 0.05 and Hits to ≥ 10 as the cutoff criterion for GO divisions. The highest number of DEGs were substantially enriched in processes such as Inflammatory Response Pathway (P-value; 1.013e-06, O value; 5.041e-05), NF- κ B activation by non typeable Hemophilus influenza having p-value = 5.92e-07 and q value = 3.927e-05 with enrichment score of 4.74, signal transduction through il1r (P-value; 2.221e-08, Q value; 4.419e-06) with enrichment score 4.69, and NHR having p-value; 5.109e-06, and q value; 0.0001 (enrichment score 3.87). The detailed Gene clip 2.0 results are shown in Table 2. The gene interaction network containing interactions of only selected DEGs was also generated to determine the type of interactions among the DEGs (Fig. 2). The interactions were identified from the PICKLE tool. Approximately 14,474 interactions were obtained. The Network Analyst was used to obtain the PPI network of DEGs with interacting proteins to define the relation between DEGs and dif-

Table 1
Gene ontologies of DEGs involved in Rheumatoid Arthritis.

Clusters	Gene Ontologies	Genes Names	P-Value	Benjamani Test
Enrichment Score 2.93				
Cluster 1	Innate immune response	DEFB126, IRGM, IGHA1, IGHA2, IGHG1, IGHG2, IGHG3, IGHM, IFNA8, TRIM8	1.3E-2	6.2E-1
	Transcription initiation from RNA polymerase II promoter	MED30, NR1D1, NR4A3, PGR, THRA	4.5E-2	1.0E0
	Protein phosphorylation	FAM20C, NIM1K, NEK5, WNK4, BMP2, CCND1, HIPK1, IRAK2, MAP4K4, PRKAB2	1.9E-2	7.6E-1
	Immune response	TNFRSF10A, TNFRSF1B, IGHA1, IGHA2, IL1RN, LIF	3.1E-1	1.0E0
	Arthritis, Rheumatoid arthritis	TNFRSF1B, DUSP1, IL1RN, MAP2K3, MAP2K6	4.0E-1	1.0E0
Cluster 2	Neoplasms	TNFRSF10A, CCND1, IL1RN	4.1E-1	1.0E0
	Rheumatoid arthritis	TNFRSF1B, IL1RN, PGR	5.6E-1	1.0E0
	Osteoarthritis	TNFRSF1B, BMP2, IL1RN	1.4E-1	1.0E0
Cluster 3	Osteoporosis	TNFRSF1B, BMP2, IL1RN, PGR	2.3E-1	1.0E0
	Arthritis, Rheumatoid/Rheumatoid Arthritis	NFKBIE, TNFRSF1B, IL1RN	9.4E-2	1.0E0

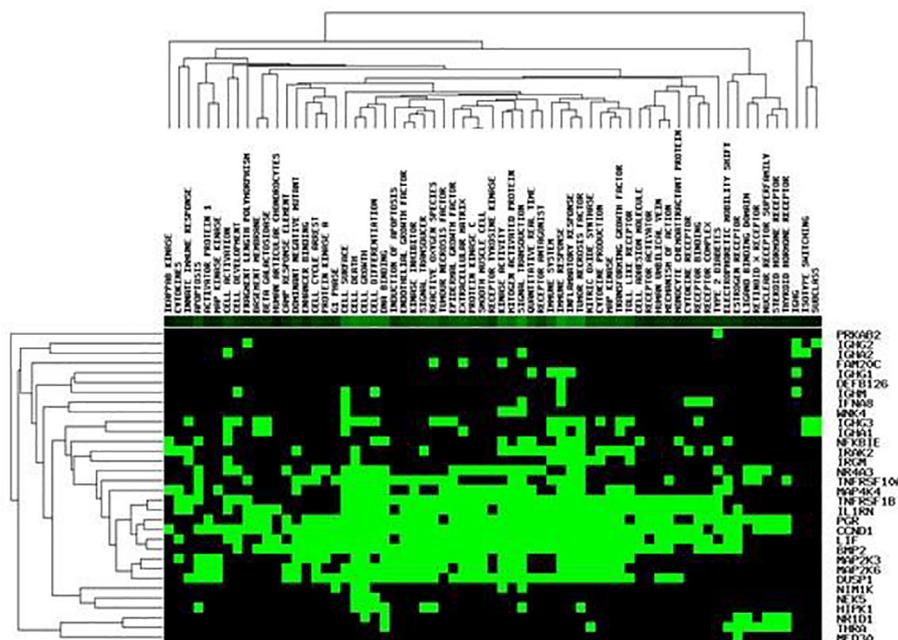


Fig. 1. Heat map of RA DEGs. The relative levels of expression of a single gene for all samples are expressed in each row; each column represents the expression levels for a single sample. The black color reflects favorably reported corresponding gene-term association, green color displays corresponding gene-term association not yet reported.

ferent functional modules. The network was so intricate that sub-functional networks were also created that would be more important in the network (Figs. 3 and 4). The density of the modules was between 0.51 and 0.68. Degrees of interacting nodes in a network and their betweenness centrality was calculated to determine the influence of each node in a network

To classify the significant pathways that have been altered in the RA and linked to other types of arthritis, a mathematical approach on the route level was used. The study of the importance of pathway crosstalk effects was focused on the pathways acquired by Network Analyst and KEGG (Fig. 5). Numerous diseases caused by DEGs were found from the Target mine database. Several signaling pathways, cancer pathways, and transcription pathways were seen as significant by the molecular crosstalk between host and candidate pathways. The disease ontologies of selected DEGs are shown in Table 3.

Current research has identified differentially expressed genes in RA chondrocytes that are connected to altered cell cycle processes,

inflammatory processes, Innate immune responses. arthritis, and protein phosphorylation, with bioinformatics methods by the systemic study. TRAF1, TRAF2, BAMP, STX11, MEOX2, AES, REL, FHL3, PNMA1, SGTA, LZTS2, SIAH2, PNMA1, and TFCP2 were found to be highly enriched in the PPI network (Fig. 3).

5. Discussions

Rheumatoid arthritis is a chronic autoimmune condition that includes inflammatory cell activation, synovial hyperplasia, and cartilage and bone destruction (Silman and Pearson, 2002). And the introduction of sequencing for next-generation technologies such as RNA sequencing has become feasible and inexpensive, offering more systematic and precise transcriptome analysis. Short fragments of complementary DNA are sequenced in RNA sequencing and mapped into the reference genome. RNA sequencing facilitates both the discovery of differentially expressed genes and

Table 2
Enriched pathways of DEGs involved in Rheumatoid Arthritis.

Pathway	Hit DEGs	Total Genes	P-Value	Q-Value
<i>cluster1 Enrichment Score: 4.74</i>				
nflkb activation by nontypeable hemophilus influenzae (BioCarta)	4	29	5.92e-07	3.927e-05
Vitamin D in inflammatory diseases (Wikipathways)	3	22	1.809e-05	0.0003
Regulation of p38-alpha and p38-beta (PID)	3	31	5.205e-05	0.0006
tnf/stress related signaling (BioCarta)	3	23	2.077e-05	0.0003
TNF receptor signaling pathway (PID)	3	46	0.0002	0.0013
Interleukin-10 signaling (Reactome)	3	6	2.411e-07	2.399e-05
Inflammatory Response Pathway (Wikipathways)	4	33	1.013e-06	5.041e-05
<i>cluster2 Enrichment Score: 4.69</i>				
signal transduction through il1r (BioCarta)	5	37	2.221e-08	4.419e-06
activated TAK1 mediates p38 MAPK activation (Reactome)	3	21	1.565e-05	0.0003
Structural Pathway of Interleukin 1 (IL-1) (Wikipathways)	3	44	0.0002	0.0011
IL1-mediated signaling events (PID)	2	36	0.0033	0.0064
<i>cluster3 Enrichment Score: 4.62</i>				
Signaling mediated by p38-gamma and p38-delta (PID)	3	12	2.627e-06	0.0001
Photodynamic therapy-induced AP-1 survival signaling. (Wikipathways)	3	50	0.0002	0.0015
<i>cluster4 Enrichment Score: 4.12</i>				
BMP2 signaling TAK1 (INOH)	3	15	5.406e-06	0.0001
Fc Epsilon Receptor I Signaling in Mast Cells (SMPDB)	2	42	0.0044	0.0078
Cellular roles of Anthrax toxin (PID)	3	22	1.809e-05	0.0003
<i>cluster5 Enrichment Score: 3.87</i>				
NHR (Signalink)	4	49	5.109e-06	0.0001
Nuclear Receptors (Wikipathways)	2	38	0.0036	0.0067
<i>cluster6 Enrichment Score: 2.80</i>				
Uptake and function of anthrax toxins (Reactome)	2	10	0.0002	0.0016
Uptake and actions of bacterial toxins (Reactome)	2	25	0.0016	0.0044
the 41bb-dependent immune response (BioCarta)	2	13	0.0004	0.0021
human cytomegalovirus and map kinase pathways (BioCarta)	2	16	0.0006	0.0028
4-hydroxytamoxifen, Dexamethasone, and Retinoic Acids Regulation of p27 Expression (Wikipathways)	2	17	0.0007	0.0031
gata3 participate in activating the th2 cytokine genes expression (BioCarta)	2	19	0.0009	0.0036
TGF-beta signaling TAK1 (INOH)	2	20	0.0001	0.0037
TLR ECSIT MEKK1 p38 (INOH)	2	23	0.0013	0.0042
CXCR3-mediated signaling events (PID)	2	38	0.0036	0.0067
TSH (NetPath)	2	26	0.0017	0.0045
TLR ECSIT MEKK1 JNK (INOH)	2	26	0.0017	0.0045
p38 MAPK signaling pathway (PID)	2	29	0.0021	0.005
MAPK Cascade (Wikipathways)	2	29	0.0021	0.005
fmlp induced chemokine gene expression in hmc-1 cells (BioCarta)	2	30	0.0023	0.0052
Trk receptor signaling mediated by the MAPK pathway (PID)	2	33	0.0028	0.0056
Serotonin HTR1 Group and FOS Pathway (Wikipathways)	2	33	0.0028	0.0056
toll-like receptor pathway (BioCarta)	2	35	0.0031	0.0061
Tacrolimus/Cyclosporine Pathway, Pharmacodynamics (PharmGKB)	2	44	0.0048	0.0085
RhoA signaling pathway (PID)	2	45	0.0051	0.0087
<i>cluster7 Enrichment Score: 2.80</i>				
Mammary gland development pathway - Puberty (Stage 2 of 4) (Wikipathways)	2	13	0.0004	0.0021
Mammary gland development pathway - Pregnancy and lactation (Stage 3 of 4) (Wikipathways)	2	33	0.0028	0.0056
Validated transcriptional targets of AP1 family members Fra1 and Fra2 (PID)	2	37	0.0034	0.0066
Circadian rhythm - Homo sapiens (human) (KEGG)	2	31	0.0024	0.0054
<i>cluster8 Enrichment Score: 2.41</i>				
NOD1/2 Signaling Pathway (Reactome)	2	33	0.0028	0.0056
Nucleotide-binding domain, leucine-rich repeat containing receptor (NLR) signaling pathways (Reactome)	2	47	0.0055	0.0094
Oxidative Damage (Wikipathways)	2	40	0.004	0.0073

accurate quantitative determination of exon and isoform expression (Grassi et al., 1998; Hart et al., 2013; Tarazona et al., 2011) along with the localization of transcription initiation sites and novel splicing variations (Grassi et al., 1998). In the altered joint microenvironment of RA, highly enriched genes were found to be potentially impaired, having identical expression profiles in primary RA chondrocytes and arrays of clinical RA synovial tissues. Features of the damaged articular joints in RA (Grassi et al., 1998) are synovitis with joint fragmentation. The development of pannus inside the joint results in direct interaction of synovial fibroblasts with bone and cartilage tissues, contributing to cartilage degradation and bone erosion (Lipsky, 2007). At the pre-arthritis stage, the synovium is a relevant trait of RA and causes inflammation, angiogenesis, and cartilage injury, affecting the entire structure of the joint (Jeon et al., 2008; Wang et al., 2014a,b,c).

Chen et al (2018) used knee chondrocyte samples and identified 463 differentially expressed genes in RA chondrocytes, enriched in biological functions related to altered cell cycle process, inflammatory response, and hypoxic stimulation. We used a similar methodology as that of Chen et al., (2018) but used chondrocytes of the hip joint instead of the knee. Chen and colleagues found fibroblast growth factor 9 (FGF9), kynureninase (KYNU), and regulator of the cell cycle (RGCC) as the top dysregulated genes identified to be potentially affected in the RA joint microenvironment. Li et al. (2019) found several DEGs from synovial tissues including CXCL13, CD247, CCL5, GZMB, IGKC, IL7R, UBD//GABBR1, ADAMDEC1, BTC, AIM2, SHANK2, CCL18, LAMP3, CR1, and IL32 significantly involved in RA. However, in our study, the most significant genes involved in RA were found to be TRAF1, TRAF2, BAMP, STX11, MEOX2, AES, REL, FHL3, PNMA1, SGTA, LZTS2, SIAH2, PNMA1, and TFPC2, con-

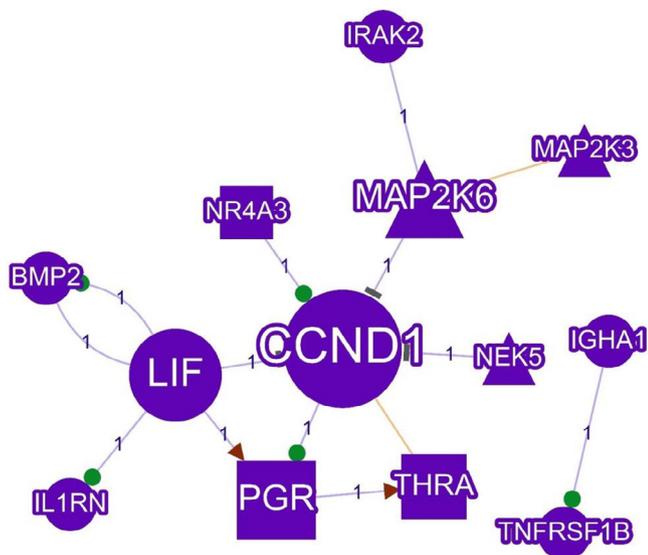


Fig. 2. The Interaction network of DEGs obtained through Gene Clip 2.0 server, the weight of the edges is written in numbers, the circles represent gene, triangles are displaying the enzymes, whereas transcription factors are shown by squares. The arrowheads represent interactions, whereas green circles are showing the gene regulations, and horizontal lines in front of arrows represent inhibitions.

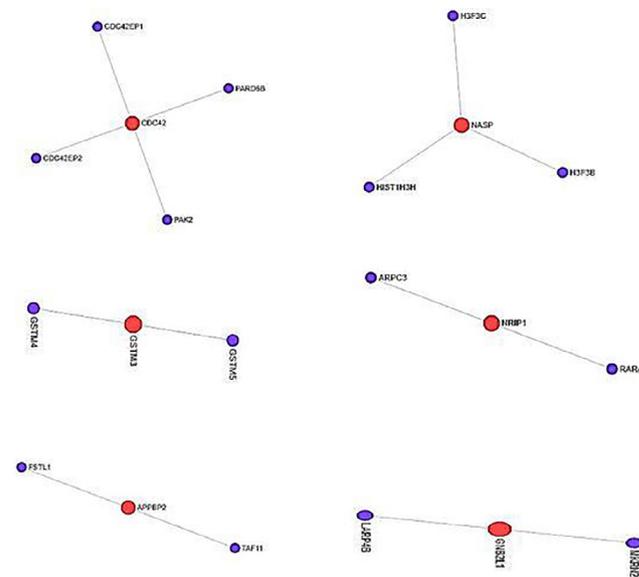


Fig. 4. Sub-networks functional modules obtained from the PPI shown in Fig. 3.

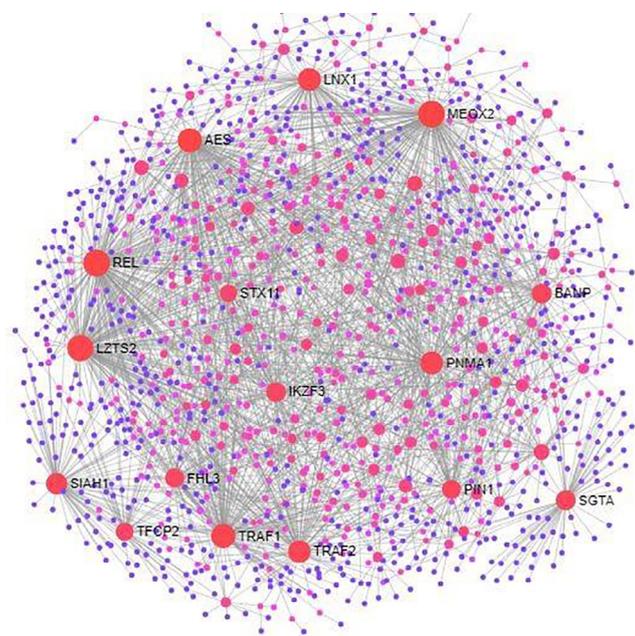


Fig. 3. The disease gene interaction network of DEGs, blue squares represent associated diseases of DEGs and Red nodes show DEGs, DEGs for which no disease association was observed are shown by light pink color.

nected to Innate immune response, Protein phosphorylation, Transcription initiation from RNA polymerase II promoter, Immune response, Neoplasms of bones, as well as Arthritis, and Rheumatoid arthritis. Similarly, [Heruth et al \(2012\)](#) found that other novel dysregulated networks and pathways such as cell morphology, cell-to-cell signaling and interaction, cellular motion, cell growth and proliferation, and cellular development may contribute to the pathogenesis of RA, in addition to known inflammatory and immune responses. 120 new genes and isoforms not previously associated with rheumatoid arthritis were also identified in their study. Whereas, our study identified Hypertrophic cardiomyopathy

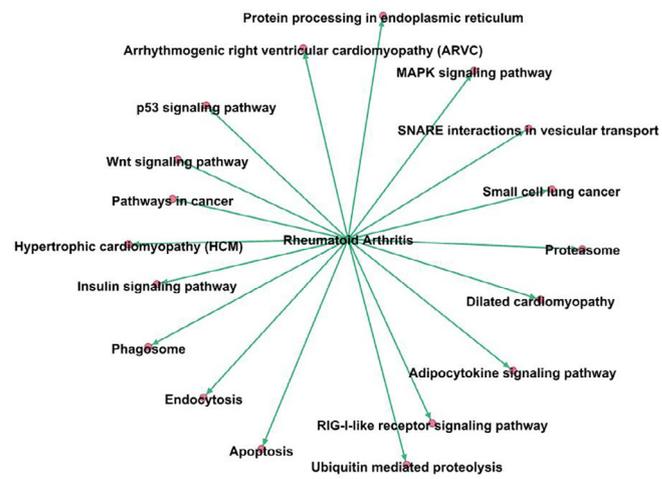


Fig. 5. Crosstalk network of Rheumatoid Arthritis related DEGs and their associated pathways.

(HCM), Small cell lung cancer, Proteasome, p53 signaling pathway, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Small cell lung cancer, SNARE interactions in vesicular transport, RIG-I-like receptor signaling pathway, Hypertrophic cardiomyopathy (HCM) as most significant in RA.

A statistical test was used to classify the significant pathways altered in RA. The generated crosstalk was based on the PPI database and KEGG database. The molecular crosstalk between host and candidate pathways shows that Hypertrophic cardiomyopathy (HCM), Small cell lung cancer, Proteasome, p53 signaling pathway, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Small cell lung cancer, SNARE interactions in vesicular transport, RIG-I-like receptor signaling pathway, Hypertrophic cardiomyopathy (HCM) were most significant among all the cross-talk pathways. Besides, DO term annotation found that multiple DEGs played roles in different bone disorders and neoplasms. This could validate the precision of the results to some extent. Also, a significant number of bone abnormalities have been proposed to be developed by certain roles of the BMP2 gene. Besides, [Wang et al. \(2014a,b,c\)](#) stated that BMP2 is a morphogenic protein responsible for the formation

Table 3
Disease Gene Ontologies of selected DEGs of rheumatoid arthritis.

Disease	Genes Involved	P-value
Osteomyelitis (disease)	IGHM,IL1RN	0,029,443
Osteopenia	CCND1	0,029,453
Lethal osteosclerotic bone dysplasia	FAM20C	1,850,106
Osteomyelitis, sterile multifocal, with periostitis and pustulosis	IL1RN	2,748,507
Arthritis	IGHM	0,003,864
Rheumatoid arthritis	IL1RN, NFKBIE, NR4A3	0,003,873
Juvenile arthritis due to defect in LACC1	DUSP1, IL1RN, NR4A3	3,495,559
Bone disease	BMP2	0,005,940
Bone marrow disorder	CCND1	0,005,956
Bone pain	CCND1	0,151,825
Lethal osteosclerotic bone dysplasia	FAM20C	1,850,106
Infection of bone	IGHM,IL1RN	2,242,472
Abnormality of bone marrow cell morphology	CCND1	4,021,634
Retrusion of upper jaw bones	BMP2	4,280,640
Deficiency of upper jaw bones	BMP2	4,280,642
Hypertrophic malar bone	BMP2	4,280,651
Increased size of the cranium	BMP2	4,280,663
Big calvaria	BMP2	4,280,664
Short stature, facial dysmorphism, and skeletal anomalies with or without cardiac anomalies	BMP2	4,693,481

and growth of bones In many cases of arthritis, the IL1RN gene has also been found to play a part. These observations inferred that dysregulation of these genes may be essential for RA progression.

6. Conclusion

In conclusion, identified DEGs, specifically TRAF1, TRAF2, BAMP, STX11, MEOX2, AES, REL, FHL3, PNMA1, SGTA, LZTS2, SIAH2, PNMA1, and TFCP2, may be pivotal genes for RA and may prove useful markers for the prediction of RA, neoplasms, and therapeutic targets for the treatment of RA and OA patients. Moreover, these genes are associated with bone formation and growth. Also, Hypertrophic Cardiomyopathy, Small Cell Lung Cancer, Proteasome, p53 signaling pathway, Arrhythmogenic Right Ventricular Cardiomyopathy, Small Cell Lung Cancer, SNARE vesicular transport interactions, RIG-I receptor signaling pathway, Hypertrophic Cardiomyopathy have been shown to play important roles in RA-related procedures in the crosstalk network. Our analysis sheds new light on the RA process and treatment. However, in the future, the result mentioned above, not just genes, but also RA-related pathways, can be tested and confirmed jointly in subsequent work by animal and clinical trials.

Significance

Focus

- Identification of Differentially expressed genes involved in Rheumatoid and Osteo Arthritis.
- Generation of Pathways cross talk to determine the mechanism of disease.

Key message

- TRAF1, TRAF2, BAMP, STX11, MEOX2, AES, REL, FHL3, PNMA1, SGTA, LZTS2, SIAH2, PNMA1, and TFCP2 are responsible for RA and OA pathogenesis.
- These genes can be used as a target for controlling RA.

Limitations and strength

- The study further requires clinical testing and validation.
- Cross-talk pathways will help understand the disease occurrence mechanism and the role of several genes in arthritis.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to the Research Innoscience China for providing a platform to collaborate and conduct research.

References

- Ammari, M., Jorgensen, C., Apparailly, F., 2013. Impact of microRNAs on the understanding and treatment of rheumatoid arthritis. *Curr. Opin Rheumatol.* 25, 225–233.
- Catrina, A.L., Joshua, V., Klareskog, L., Malmström, V., 2016. Mechanisms involved in triggering rheumatoid arthritis. *Immunol. Rev.* 269 (1), 162–174.
- Chen, Y.J., Chang, W.A., Wu, L.Y., Hsu, Y.L., Chen, C.H., Kuo, P.L., 2018. Systematic analysis of differential expression profile in rheumatoid arthritis chondrocytes using next-generation sequencing and bioinformatics approaches. *Int. J. Med. Sci.* 15 (11), 1129.
- Chhangawala, S., Rudy, G., Mason, C.E., et al., 2015. The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol.* 16, 131.
- Coras, R., D Murillo-Saich, J., Guma, M., 2020. Circulating Pro-and Anti-Inflammatory Metabolites and Its Potential Role in Rheumatoid Arthritis Pathogenesis. *Cells* 9(4), 827.
- Essouma, M., Nkeck, J.R., Endomba, F.T., Bigna, J.J., Ralandison, S., 2020. Epidemiology of rheumatoid arthritis in sub-Saharan Africa: a systematic review and meta-analysis protocol. *System. Rev.* 9, 1–4.
- Feldmann, M., Maini, R.N., 2003. Lasker Clinical Medical Research Award. TNF defined as a therapeutic target for rheumatoid arthritis and other autoimmune diseases. *Nat. Med.* 9, 1245–1250.
- Firestein, G.S., 2003. Evolving concepts of rheumatoid arthritis. *Nature* 423, 356–361.
- Giannini, D., Antonucci, M., Petrelli, F., Bilia, S., Alunno, A., Puxeddu, I., 2020. One year in review 2020: Pathogenesis of rheumatoid arthritis. *Clin. Exp. Rheumatol.* 38, 387–397.
- Giannopoulou, E.G., Elemento, O., Ivashkiv, L.B., 2015. Use of RNA sequencing to evaluate rheumatic disease patients. *Arthritis Res Ther.* 17, 167.
- Gioutlakis, A., Klapa, M.I., Moschonas, N.K., 2017. PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS One* 12, (10) e0186039.
- Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., Valencia, A., 2012. EnrichNet: network-based gene set enrichment analysis. *Bioinformatics* 28 (18), i451–i457.
- Goulielmos, G.N., Zervou, M.I., Myrthianou, E., et al., 2016. Genetic data: The new challenge of personalized medicine, insights for rheumatoid arthritis patients. *Gene* 583, 90–101.
- Grassi, W., De Angelis, R., Lamanna, G., et al., 1998. The clinical features of rheumatoid arthritis. *Eur. J. Radiol.* 27 (Suppl 1), S18–S24.
- Hart, T., Komori, H.K., LaMere, S., et al., 2013. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* 14, 778.
- Heruth, D.P., Gibson, M., Grigoryev, D.N., Zhang, L.Q., Ye, S.Q., 2012. RNA-seq analysis of synovial fibroblasts brings new insights into rheumatoid arthritis. *Cell Biosci.* 2 (1), 1–6.
- Jeon, C.H., Ahn, J.K., Chai, J.Y., et al., 2008. Hypoxia appears at pre-arthritis stage and shows co-localization with early synovial inflammation in collagen induced arthritis. *Clin. Exp. Rheumatol.* 26, 646–648.
- Kaya, E., Agca, M., Adiguzel, F., Cetin, M., 2019. Spatial data analysis with R programming for environment. *Human Ecol. Risk Assessment: Int. J.* 25 (6), 1521–1530.
- Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J., Pavlidis, P., 2004. Coexpression analysis of human genes across many microarray data sets. *Genome Res.* 14, 1085–1094.

- Li, W.C., Bai, L., Xu, Y., Chen, H., Ma, R., Hou, W.B., Xu, R.J., 2019. Identification of differentially expressed genes in synovial tissue of rheumatoid arthritis and osteoarthritis in patients. *J. Cell Biochem.* 120 (3), 4533–4544. <https://doi.org/10.1002/jcb.27741>. Epub 2018 Sep 27 PMID: 30260019.
- Li, A., Barber, R.F., 2019. Multiple testing with the structure-adaptive Benjamini-Hochberg algorithm. *J. Roy. Statist. Soc.: Series B (Statistical Methodology)* 81 (1), 45–74.
- Lipsky, P.E., 2007. Why does rheumatoid arthritis involve the joints?. *N Engl. J. Med.* 356, 2419–2420.
- MacGregor, A.J., Snieder, H., Rigby, A.S., et al., 2000. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43, 30–37.
- McGovern, D.P., Hysi, P., Ahmad, T., et al., 2005. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. *Hum Mol Genet.* 14, 1245–1250.
- McInnes, I.B., Schett, G., 2011. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* 365 (23), 2205–2219.
- Nakano, K., Whitaker, J.W., Boyle, D.L., Wang, W., Firestein, G.S., 2013. DNA methylome signature in rheumatoid arthritis. *Ann. Rheum. Dis.* 72, 110–117.
- Platzer, A., Nussbaumer, T., Karonitsch, T., Smolen, J.S., Aletaha, D., 2019 Jul 25. Analysis of gene expression in rheumatoid arthritis and related conditions offers insights into sex-bias, gene biotypes and co-expression patterns. *PLoS one.* 14, (7) e0219698.
- Porcelli, S., Yockey, C.E., Brenner, M.B., Balk, S.P., 1993. Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4-8-alpha/beta T cells demonstrates preferential use of several V beta genes and an invariant TCR alpha chain. *J. Exp. Med.* 178, 1–16.
- Proost, P., Struyf, S., Loos, T., Gouwy, M., Schutyser, E., Conings, R., Ronsse, I., Parmentier, M., Grillet, B., Opdenakker, G., Balzarini, J., 2006. Coexpression and interaction of CXCL10 and CD26 in mesenchymal cells by synergising inflammatory cytokines: CXCL8 and CXCL10 are discriminative markers for autoimmune arthropathies. *Arthritis Res. Therapy* 8 (4), 1–4.
- Ramírez, J., Inciarte-Mundo, J., Cuervo, A., Celis, R., Ruiz-Esquide, V., Castellanos-Moreira, R., Ponce, A., Gómez-Puerta, J., Sanmartí, R., Cañete, J.D., 2020. Comparable Clinical and Radiographic Outcomes among DAS28-ESR-based Remission Criteria and SDAI, CDAI, and ACR/EULAR Boolean definitions in patients with Established Rheumatoid Arthritis in Clinical Practice. *Observational Study of 5 years of Follow-Up.*
- Russo, P.S., Ferreira, G.R., Cardozo, L.E., Bürger, M.C., Arias-Carrasco, R., Maruyama, S.R., Hirata, T.D., Lima, D.S., Passos, F.M., Fukutani, K.F., Lever, M., 2018. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *Bmc Bioinformatics.* 19 (1), 1–13.
- Sen, M., 2005. Wnt signalling in rheumatoid arthritis. *Rheumatology.* 2005, 1;44 (6):708–13. Deon D. Ahmed S, Tai K, et al: Cross-talk between IL-1 and IL-6 signaling pathways in rheumatoid arthritis synovial fibroblasts. *J Immunol.* 2001;167: 5395-5403.
- Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H.C., Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8 (9), R183.
- Silman, A.J., Pearson, J.E., 2002. Epidemiology and genetics of rheumatoid arthritis. *Arthritis Res. Therapy* 4 (S3), S265.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., et al., 2011. Differential expression in RNA-seq: A matter of depth. *Genome Res.* 21, 2213–2223.
- Tasaki, S., Suzuki, K., Kassai, Y., Takeshita, M., Murota, A., Kondo, Y., Ando, T., Nakayama, Y., Okuzono, Y., Takiguchi, M., Kurisu, R., 2018. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat. Commun.* 9 (1), 1–2.
- Wang, K., Zhao, L., Liu, X., Hao, Z., Zhou, Y., Yang, C., Li, H., 2014 Nov 1a. Differential co-expression analysis of rheumatoid arthritis with microarray data. *Mol. Med. Rep.* 10 (5), 2421–2426.
- Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., Idowu, O., 2014b. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* 1 (1), 87–105.
- Wang, J.H., Zhao, L.F., Lin, P., Su, X.R., Chen, S.J., Huang, L.Q., Wang, H.F., Zhang, H., Hu, Z.F., Yao, K.T., Huang, Z.X., 2014c. GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms. *Bioinformatics* 30 (17), 2534–2536.
- Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447 (7145), 661.
- Spandidos Publications style Xue, F., Zhang, C., He, Z., Ding, L., Xiao, H., 2013. Analysis of critical molecules and signaling pathways in osteoarthritis and rheumatoid arthritis. *Mol. Med. Rep.* 7, 603–607.
- Xia, J., Benner, M.J., Hancock, R.E., 2014. NetworkAnalyst-integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* 42 (W1), W167–W174.
- Yoshizawa, T., Hammaker, D., Sweeney, S.E., Boyle, D.L., Firestein, G.S., 2008. Synovial innate immune responses: I. Differential regulation of interferon responses and the JNK pathway by MAPK kinases. *J. Immunol.* 181 (5), 3252–3258.