

RESEARCH ARTICLE

Identification of candidate sex hormone-associated genes and immune infiltration characteristics in osteoarthritis based on bioinformatics analysis and machine learning

Yishu Wang¹✉, Ling Zhu²✉, Shuna Jin^{3,4}, Yuhan Wang¹, Zhaoxiang Zeng¹, Yunzhou Zuo², Xingliang Xiang², Xugui Li^{2,5*}, Rongzeng Huang^{1*}, Chengwu Song^{1,4*} 

1 School of Pharmacy, Hubei University of Chinese Medicine, Wuhan, China, **2** The Affiliated Hospital of Wuhan Sports University, Wuhan, China, **3** School of Basic Medical Sciences, Hubei University of Chinese Medicine, Wuhan, China, **4** Hubei Shizhen Laboratory, Wuhan, China, **5** School of Acupuncture-Moxibustion and Orthopedics, Hubei University of Chinese Medicine, Wuhan, China

✉ These authors contributed equally to this work.

* lixugui2000@163.com (XL); rongzenghuang_2018@hbucom.edu.cn (RH); chengwusong_2016@hbucom.edu.cn (CS)



OPEN ACCESS

Citation: Wang Y, Zhu L, Jin S, Wang Y, Zeng Z, Zuo Y, et al. (2026) Identification of candidate sex hormone-associated genes and immune infiltration characteristics in osteoarthritis based on bioinformatics analysis and machine learning. PLoS One 21(6): e0351556. <https://doi.org/10.1371/journal.pone.0351556>

Editor: Gianpaolo Papaccio, Università degli Studi della Campania, ITALY

Received: February 28, 2026

Accepted: May 28, 2026

Published: June 12, 2026

Copyright: © 2026 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Publicly available transcriptomic datasets analyzed in this study are available from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE169077, GSE129147, and GSE55235.

Abstract

Background

Sex hormones play critical roles in the pathogenesis and progression of osteoarthritis (OA), yet the hormone-related molecular networks remain poorly defined. This study aimed to identify candidate sex hormone-associated genes in OA and to explore their potential functional enrichment and immune-related characteristics using bioinformatics analysis.

Methods

OA gene expression data were obtained from the GEO database and integrated with candidate sex hormone-associated genes retrieved from GeneCards. The R package “limma” was then used to identify differentially expressed genes (DEGs) and sex hormone-associated DEGs (SADEGs). OA-associated SADEGs, termed OA-SADEGs, were selected using weighted gene co-expression network analysis (WGCNA), and their potential biological functions and pathways were explored by GO and KEGG enrichment analyses. Hub genes were identified using three machine learning models. xCell analysis was used to estimate immune infiltration and its associations with hub genes, and hub gene expression was further evaluated in external datasets and peripheral blood samples.

Results

We identified 32 sex hormone-associated genes in OA, enriched in extracellular matrix remodeling, receptor signaling, and antigen presentation pathways. Three

Sex hormone-related gene information was obtained from the GeneCards database (<https://www.genecards.org/>). All relevant data supporting the findings of this study, including the processed differential expression results, WGCNA-related results, immune infiltration results, machine learning outputs, and clinical qRT-PCR validation data, are included within the paper and its [Supporting Information](#) files. The author-generated analysis code used in this study is available at GitHub: <https://github.com/wys436/OA-sex-hormone-OA-bioinformatics>.

Funding: This work was supported by the Natural Science Foundation of Hubei Province (Grant No. 2024AFD275, awarded to YZ) and an internal research grant from the Hubei University of Chinese Medicine (Grant No. 2022ZZXT003, awarded to SJ). No other internal funding or sources of support were received during this study. There was no additional external funding received for this study.

Competing interests: The authors have declared that no competing interests exist.

candidate hub genes (LOXL1, HLA-DRA, and CYBB) were consistently upregulated in OA and showed significant correlations with immune infiltration scores. xCell analysis identified 13 differentially enriched immune cell types, of which three were associated with hub genes. External dataset analysis and peripheral blood qRT-PCR showed upregulation of LOXL1, HLA-DRA, and CYBB in OA samples.

Conclusion

This study integrated bioinformatics and immune analyses to identify candidate sex hormone-associated genes in OA. These findings provide associative bioinformatics evidence for sex hormone-associated molecular features in OA.

Introduction

OA, the most prevalent form of degenerative joint disease, is widely recognized as a primary source of chronic pain in middle-aged and elderly individuals. It causes joint pain, structural deformities, and impaired function, and is also associated with an increased risk of cardiovascular events and fractures. Globally, OA affects more than 7% of the population, corresponding to approximately 528 million individuals, and ranks as the fifteenth leading cause of years lived with disability (YLDs), accounting for 2.2% of total global YLDs [1]. OA exhibits significant sex differences, with a notably higher prevalence and increased severity of pain among women aged over 50 years, particularly in postmenopausal populations [2]. Female patients generally experience greater pain intensity, faster disease progression, and more severe functional impairment than male patients [3]. These sex-specific disparities in OA are likely attributable to multiple factors, including sex hormone-related changes, particularly age-related declines in estrogen levels, anatomical differences, enhanced synovial inflammatory responses, and differences in central sensitization and biomechanical factors [4].

Sex hormones play a crucial role in the pathogenesis of OA, influencing its development through both direct effects on tissues and indirect effects on inflammatory and immune pathways. These hormones act on joint tissues—including cartilage, subchondral bone, and synovium—via diverse signaling mechanisms, regulating extracellular matrix metabolism, cellular apoptosis and proliferation, as well as inflammatory responses [5]. Major joint cell types, including chondrocytes, osteoblasts, and osteoclasts, express sex hormone receptors and respond to both systemic and local hormonal changes [6]. At the molecular level, sex hormones have been found to regulate key signaling pathways such as Wnt/ β -catenin and TGF- β /Smad, thereby contributing to cartilage homeostasis, immune modulation, and bone metabolism [7–9]. Despite increasing interest in the role of sex hormones in OA, current research predominantly focuses on isolated hormonal effects or individual signaling pathways. However, a comprehensive understanding of the hormone-associated molecular networks involved in OA progression remains limited.

Bioinformatics is an interdisciplinary field that integrates biological data with computational methods, enabling efficient analysis, storage, and interpretation of high-throughput biological information [10]. This capability is critical for uncovering novel insights into OA pathogenesis and identifying disease-related biomarkers. Machine learning further strengthens this strategy by enabling pattern recognition and feature prioritization for biomarker discovery and disease classification [11]. The integration of bioinformatics and machine learning provides a useful exploratory framework for identifying candidate sex hormone-associated molecular features in OA. In this study, we applied bioinformatics and immune infiltration profiling to identify differentially expressed candidate sex hormone-associated genes in OA. These candidate genes were then subjected to functional enrichment analysis, hub gene selection, and validation in external datasets. Immune cell correlation analyses were performed to evaluate their immune-related associations, and qRT-PCR further assessed hub gene expression in peripheral blood samples.

Materials and methods

Screening and processing of OA-related datasets and collection of sex hormone-associated genes

We downloaded three datasets (GSE169077, GSE129147, GSE55235) from the Gene Expression Omnibus database (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) for OA-related gene expression analysis and hub gene identification. GSE169077 was based on the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array and contained 5 normal and 6 OA samples. GSE129147 was based on the GPL15207 [PrimeView] Affymetrix Human Gene Expression Array and contained 9 normal and 10 OA samples. GSE55235 was based on the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array and contained 10 normal and 10 OA samples. GSE169077 and GSE129147 were used as the training set and GSE55235 as the test set. The dataset identifiers were converted using the “org.Hs.e.g.,db” R package. Quantile normalization was performed across the combined GEO datasets to ensure comparable expression distributions. Batch effects between the training datasets were corrected using the “sva” R package, and the effectiveness of batch correction was evaluated by visual inspection of sample distributions before and after correction (S1 Fig). Sex hormone-associated genes were initially collected from the GeneCards database (<https://www.genecards.org/>) using multiple sex hormone-related keywords, including “sex hormone”, “estradiol”, “estrogen”, “testosterone”, “progesterone”, “follicle-stimulating hormone”, and “luteinizing hormone”, combined using the Boolean operator “OR”. Genes with a relevance score > 1 were retained as preliminary candidate genes. The flowchart depicting the study design is shown in Fig 1.

Identification of SADEGs

DEGs were identified using the R package “limma”, with $|\text{Log}_2\text{FC}| > 2$ and $p < 0.05$ as the selection criteria. A heat map and volcano plot were used to visualize the DEGs, and a Venn diagram was created to identify the overlap between DEGs and sex hormone-associated genes, termed SADEGs.

WGCNA and screening for OA-SADEGs

WGCNA identifies gene modules with coordinated expression by constructing co-expression networks, providing a systems-level view of gene expression data [12]. We used the R package “WGCNA” for WGCNA analysis. Genes ranked in the top 50% by average standard deviation were selected for constructing the gene co-expression network. First, Pearson’s correlation coefficient was used to calculate gene-to-gene correlations within each group. Next, the “pickSoft-Threshold” function was used to determine the soft threshold, which converted gene correlations into weighted values to construct a weighted co-expression network. Then, hierarchical clustering was used to group genes with similar expression patterns into modules, each assigned to a unique color, with a minimum module size of 30 and deepSplit set to 3. We used the correlation coefficient and p to identify the modules most strongly associated with OA traits. Genes within these modules were filtered using the criteria $\text{MM} > 0.8$ and $\text{GS} > 0.1$ to select strongly correlated genes. The strongly correlated genes selected from the OA-associated modules were intersected with SADEGs to define OA-SADEGs.

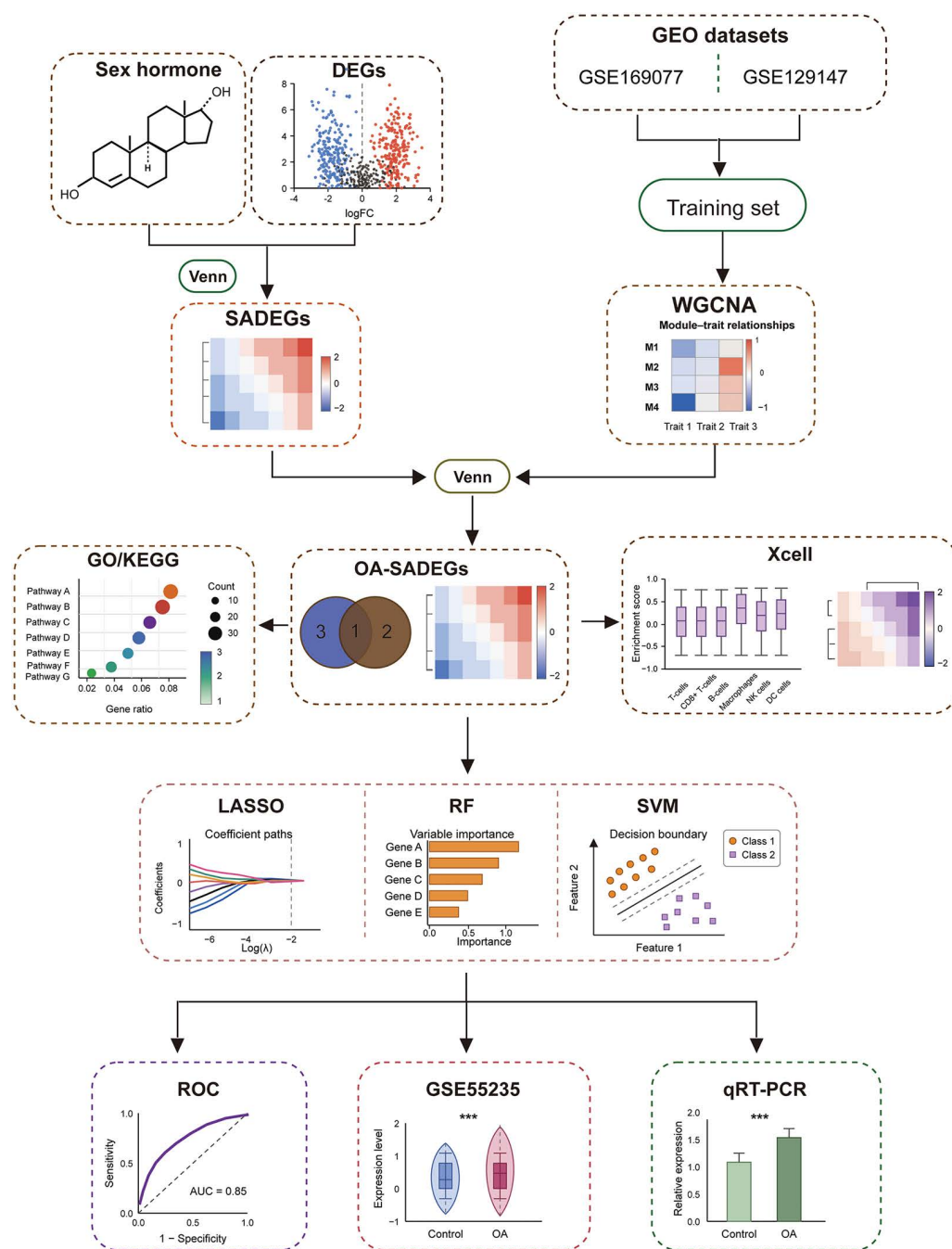


Fig 1. Flowchart showing the comprehensive analysis of sex hormone-associated genes in osteoarthritis.

<https://doi.org/10.1371/journal.pone.0351556.g001>

Functional enrichment analysis of OA-SADEGs

GO and KEGG pathway analyses of OA-SADEGs were performed using the R package “clusterProfiler”, with $p < 0.05$ considered statistically significant.

Machine learning for hub genes

Three machine learning algorithms, namely least absolute shrinkage and selection operator (LASSO) regression, support vector machine-recursive feature elimination (SVM-RFE), and random forest (RF), were used to identify hub genes from OA-SADEGs. For LASSO regression, ten-fold cross-validation was used to determine the optimal λ value with the minimum classification error. LASSO regression was performed using the R package “glmnet” and was primarily used to select feature variables from OA-SADEGs. RF, a machine learning algorithm based on decision trees, was applied to screen genes using “IncNodePurity”, implemented via the R package “randomForest”. SVM-RFE, a machine learning approach based on SVM, was employed to identify optimal variables by iteratively removing feature vectors. Ten-fold cross-validation with five replicates was performed using the R package “caret” to identify the optimal feature subset with the lowest RMSE. The genes selected by the three machine learning algorithms were intersected using a Venn diagram to identify the final hub genes.

Evaluation of the discriminative ability of hub genes

ROC curves were used to assess the discriminative ability of the hub genes, and their expression levels were compared between OA and control samples in both the training and test sets. An AUC > 0.7 in the ROC analysis was considered to indicate potential discriminative ability, while a $p < 0.05$ in the expression analysis was considered statistically significant.

xCell immune infiltration analysis

xCell is a gene signature-based method used to compute and visualize relative infiltration scores of 64 immune and stromal cell types from microarray-derived gene expression data. Statistical significance was defined as $p < 0.05$. Spearman correlation analysis was performed, with $r > 0.7$ considered to indicate a strong correlation (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). To explore immune-related associations of hub genes in OA, we analyzed the correlations between hub genes and 13 infiltrating immune cell types, using $p < 0.05$ and $r > 0.7$ as thresholds. To further examine the consistency of the immune infiltration patterns, an independent external dataset was additionally analyzed using xCell, and the major immune cell signatures identified in the training set were re-evaluated.

qRT-PCR validation

Peripheral blood samples from 20 patients with osteoarthritis (OA) and 20 individuals without OA were collected at Wuhan Orthopedics Hospital of Integrated Traditional Chinese and Western Medicine (Affiliated Hospital of Wuhan Sports University), Wuhan, China. OA was diagnosed based on clinical and radiographic criteria. Individuals without clinical or radiographic evidence of OA were included as controls. Exclusion criteria included severe systemic diseases, bone metabolic disorders, pregnancy, and other major medical conditions. All participants were adults (≥ 18 years). Female participants were postmenopausal, and male participants were age- and BMI-matched to the female participants. This study was approved by the Ethics Committee of Wuhan Orthopedics Hospital of Integrated Traditional Chinese and Western Medicine (Affiliated Hospital of Wuhan Sports University) (No. 672HREC20250310-L16). The study was conducted in accordance with the Declaration of Helsinki. All samples were collected within the ethics-approved study period. Written informed consent was obtained from all participants.

Total RNA was extracted from blood samples using the Blood RNA Extraction Kit (G3636-50T). Complementary DNA was synthesized via reverse transcription with the SweScript All-in-One RT SuperMix for qPCR (One-Step gDNA Remover) (G3337-50), and qRT-PCR was subsequently performed using 2 × Universal Blue SYBR Green qPCR Master Mix (G3326-05). Three technical replicates were maintained for each sample. GAPDH was used as the reference gene, and relative gene expression levels were quantified using the $2^{-\Delta\Delta Ct}$ method. Statistical analysis was performed using a t-test, with $p < 0.05$ considered statistically significant.

Results

Identification of DEGs and SADEGs

Differential expression analysis identified 201 genes, comprising 168 upregulated and 33 downregulated genes in osteoarthritis. Volcano plots were generated to visualize these genes, with upregulated genes shown in red and downregulated genes shown in green (Fig 2A). A heatmap highlighted the top 50 most differentially expressed genes (Fig 2B), and a total of 140 SADEGs were identified through Venn diagram analysis (Fig 2C). S1 Table provides details of the identified DEGs.

WGCNA

We performed WGCNA analysis on the training set. A soft threshold of 14 was selected as the optimal soft threshold ($R^2=0.87$) to establish the gene co-expression network (Fig 3A and 3B). An unsigned network was used, and modules were identified with a minimum module size of 30, deepSplit=3, and a merge cut height of 0.25. Eleven gene modules were obtained by hierarchical clustering (Fig 3C). In the gene-OA trait association analysis, the black module showed the strongest association with OA traits, with a correlation coefficient of 0.53 and $p=2.5e-3$ (Fig 3D and 3E). In addition, GS

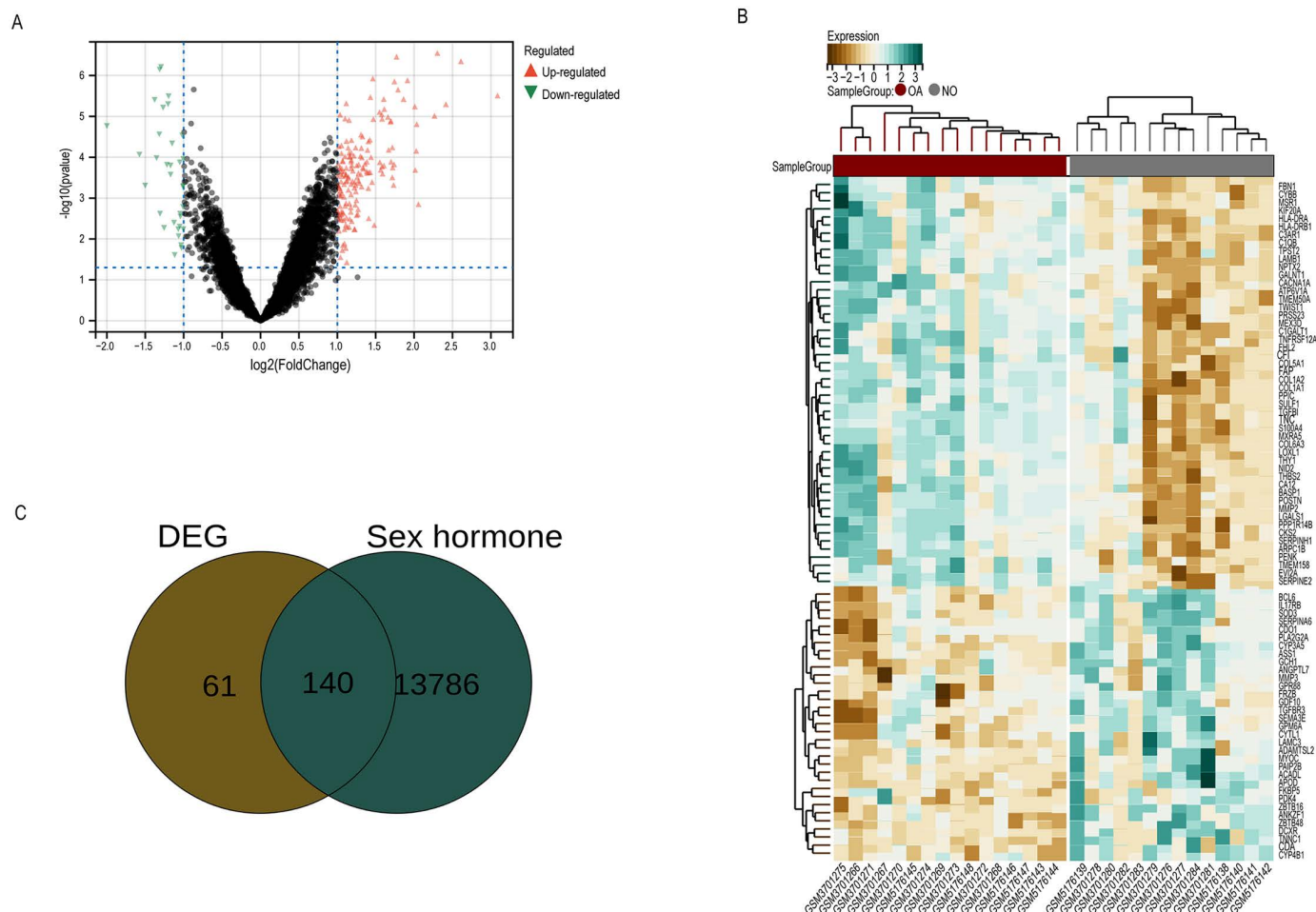


Fig 2. Identification of DEGs and SADEGs. (A) Volcano plot of DEGs. Red indicates upregulated DEGs, and green indicates downregulated DEGs. (B) Heatmap showing the top 50 DEGs. (C) Venn diagram showing the intersection of DEGs and sex hormone-associated genes.

<https://doi.org/10.1371/journal.pone.0351556.g002>

was strongly correlated with MM within the black module ($\text{cor}=0.62$; $p=6.4\text{e-}8$) (Fig 3F). A total of 216 genes strongly associated with OA were selected from the black module, and 32 OA-SADEGs were obtained by intersecting these genes with SADEGs (Fig 3G).

Functional enrichment analysis of OA-SADEGs

To explore the potential biological functions and pathways associated with OA-SADEGs, we used the R package “clusterProfiler” to perform GO and KEGG enrichment analysis. Enrichment significance was evaluated using the Benjamini–Hochberg method, and adjusted $p<0.05$ was considered statistically significant. In the biological process (BP) category, OA-SADEGs were mainly enriched in extracellular matrix organization, extracellular structure organization, ossification, and collagen fibril organization (Fig 4A). In the cellular component (CC) category, OA-SADEGs were mainly enriched in the collagen-containing extracellular matrix, collagen trimer, endoplasmic reticulum lumen, basement membrane, and collagen trimer complex (Fig 4B). In the molecular function (MF) category, OA-SADEGs were mainly enriched in extracellular matrix structural constituent, extracellular matrix structural constituent conferring tensile strength, integrin binding, and glycosaminoglycan binding (Fig 4C). The pathways in which they were mainly involved were protein digestion and absorption, ECM-receptor interaction, phagosome, staphylococcus aureus infection, focal adhesion and antigen processing and presentation (Fig 4D). Detailed results for GO and KEGG enrichment of OA-SADEGs are shown in S2–S5 Tables.

Identification of hub genes

In the LASSO analysis, four genes, LOXL1, HLA-DRA, CYBB, and ATP6V1A, were selected at the optimal λ value indicated by the vertical dashed line (Fig 5A and 5B). Gene importance was evaluated by RF using two different metrics. The top ten genes ranked by IncNodePurity were selected for further comparison (Fig 5C and 5D). LOXL1, HLA-DRA, and CYBB were selected by SVM-RFE when the RMSE reached its minimum value (Fig 5E and 5F). The three hub genes—LOXL1, HLA-DRA, and CYBB—were obtained by intersecting the genes identified by LASSO, RF, and SVM-RFE (Fig 5G). Detailed results from the three machine learning algorithms are shown in S6 Table.

Expression and discriminative ability of hub genes

The AUCs of the three hub genes were all above 0.7 in both the training and test sets (Fig 6A and 6B), suggesting that LOXL1, HLA-DRA, and CYBB had potential discriminative ability between OA and control samples. In addition, LOXL1, HLA-DRA, and CYBB showed significant expression differences between OA and control samples in both the training and test sets. All three genes were consistently upregulated in OA samples compared with control samples in both the training and test sets (Fig 6C and 6D). Consequently, LOXL1, HLA-DRA, and CYBB were identified as candidate hub genes. Hub gene expression values in the training and test sets are shown in S7 and S8 Tables.

xCell immune infiltration analysis

Pathway and hub gene analyses suggested that sex hormone-associated genes were related to immune-related pathways and features. To explore the immune-related characteristics of sex hormone-associated genes in OA, we conducted xCell immune infiltration analysis of OA-SADEGs. The relative infiltration scores of 64 immune and stromal cell types were visualized across groups, and 13 differentially enriched cell types were identified (Fig 7A): aDC, CD8+ Tcm, fibroblasts, hepatocytes, macrophages, macrophages M1, megakaryocytes, NK cells, NKT, skeletal muscle, Th1 cells, Th2 cells, and Tregs. We calculated the relative proportions of the 13 cell types to illustrate their distribution across different samples (Fig 7B). The correlation plot was used to illustrate correlations among the estimated cell infiltration scores. aDC was positively correlated with macrophages, while macrophages were positively correlated with macrophages M1. NK cells were negatively correlated with Th1 cells, and NKT cells were negatively correlated with Th2 cells (Fig 7C). Furthermore, LOXL1 was

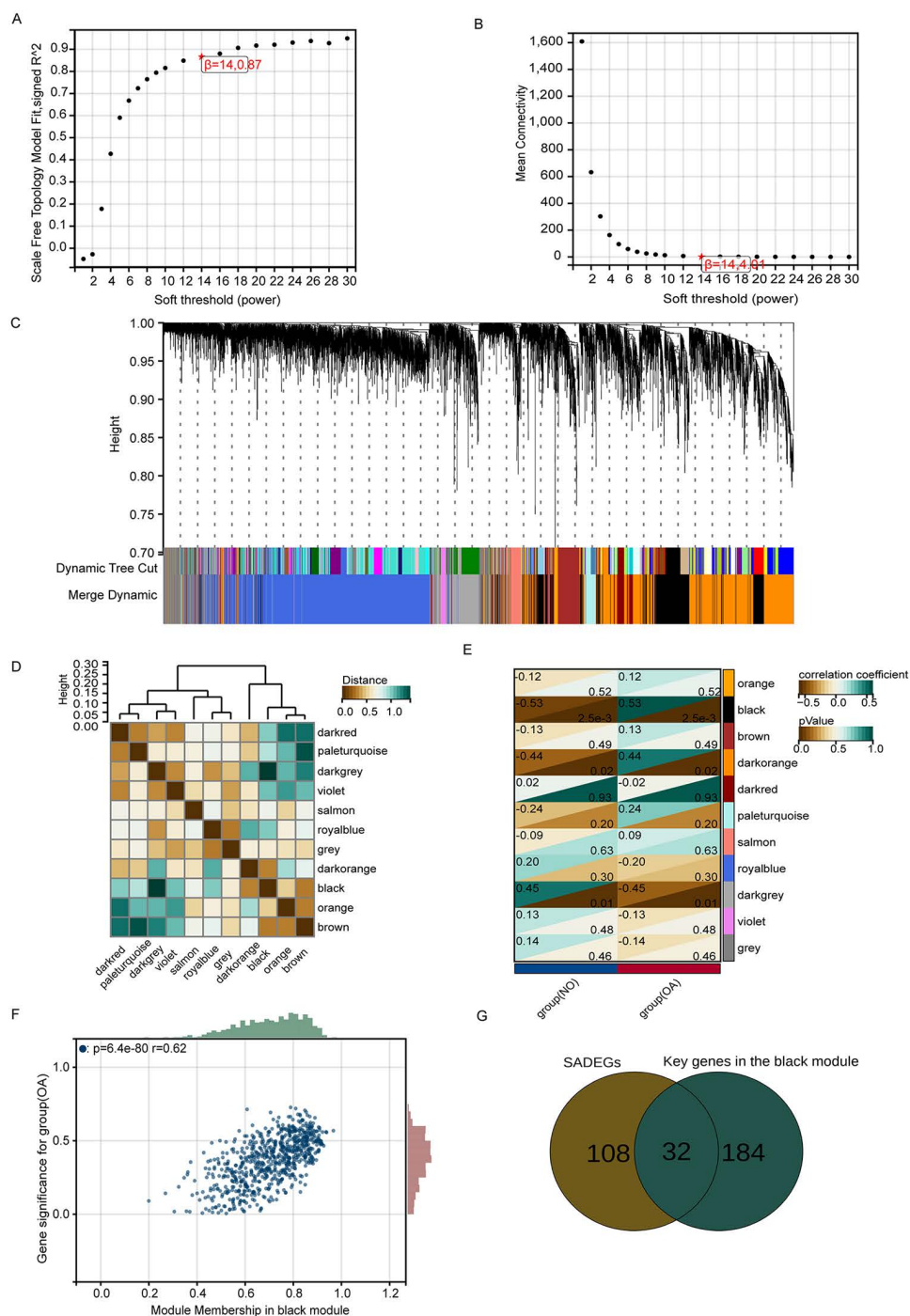


Fig 3. WGCNA. (A, B) Determination of the optimal soft threshold. (C) Eleven gene modules were obtained by hierarchical clustering. (D) Cluster plot of module feature vectors. (E) Heatmap of associations between gene modules and OA phenotypes. (F) Scatterplot showing the correlation between gene significance (GS) and module membership (MM) in the black module. (G) Venn diagram showing the intersection of SADEGs and key genes in the black module.

<https://doi.org/10.1371/journal.pone.0351556.g003>

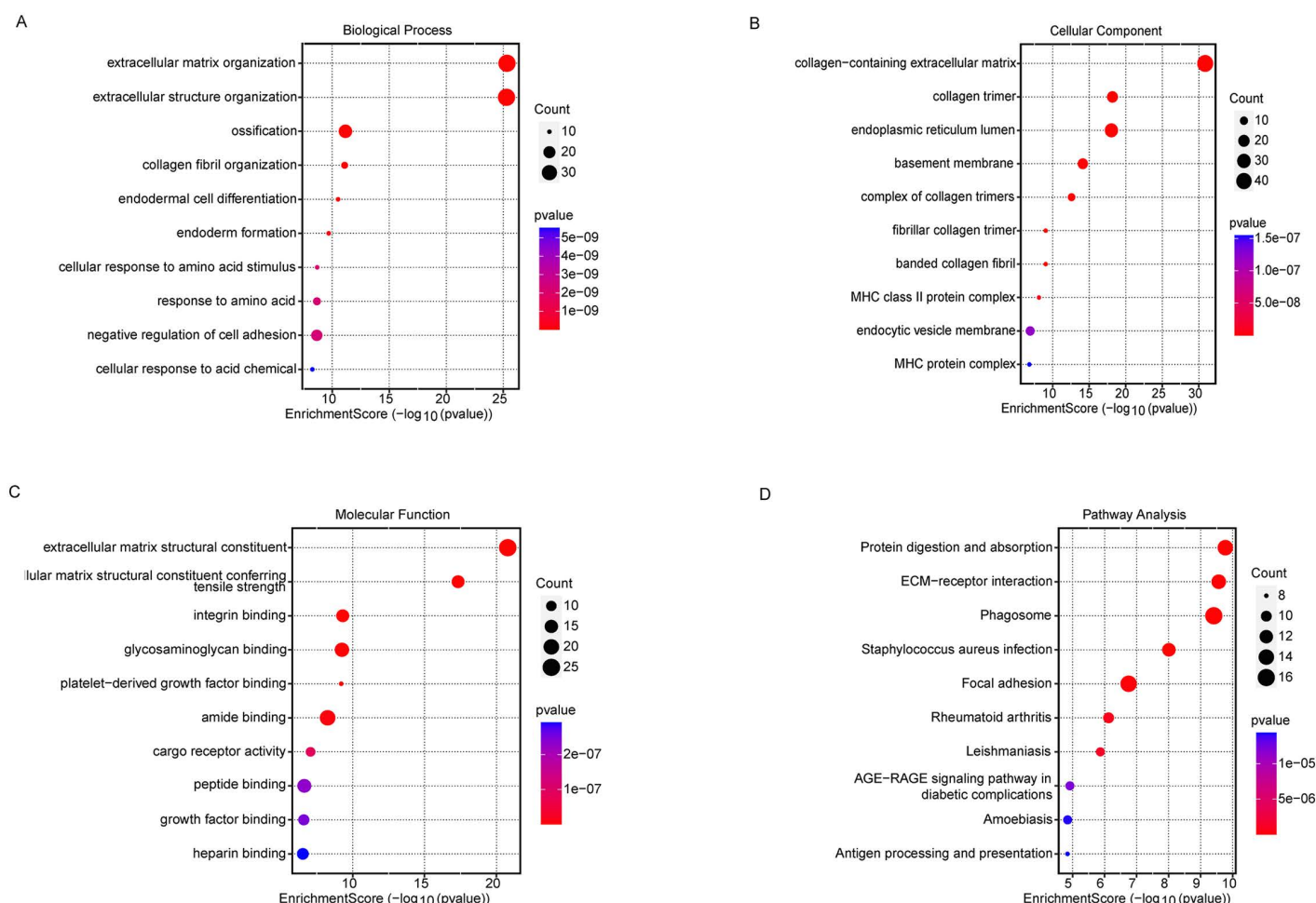


Fig 4. Functional enrichment analysis of OA-SADEGs. (A) Bubble plot of GO terms in the BP category. **(B)** Bubble plot of GO terms in the CC category. **(C)** Bubble plot of GO terms in the MF category. **(D)** Bubble plot of KEGG pathways.

<https://doi.org/10.1371/journal.pone.0351556.g004>

strongly correlated with aDC; HLA-DRA with aDC, Th1 cells, and macrophages; and CYBB with aDC and macrophages (Fig 8). Re-analysis of an independent external dataset using xCell showed generally consistent trends in the major immune infiltration patterns identified in the training set (S2 Fig and S3 Fig).

qRT-PCR validation

The qRT-PCR validation samples included 20 OA patients and 20 controls. The proportion of females was 80.0% in the OA group and 85.0% in the control group ($p = 1.000$). The mean age was 63.05 ± 10.04 years in the OA group and 63.30 ± 10.71 years in the control group ($p = 0.940$), and the mean BMI was 24.23 ± 3.28 kg/m² and 24.70 ± 4.12 kg/m², respectively ($p = 0.692$). These results indicated that the baseline characteristics were comparable between the two groups. The relative expression levels of LOXL1, HLA-DRA, and CYBB in peripheral blood samples from OA patients and controls were measured by qRT-PCR. The results indicated that LOXL1, HLA-DRA, and CYBB were significantly upregulated in OA compared to the control group (Fig 9A–C). Primer sequences for each gene are listed in S9 Table.

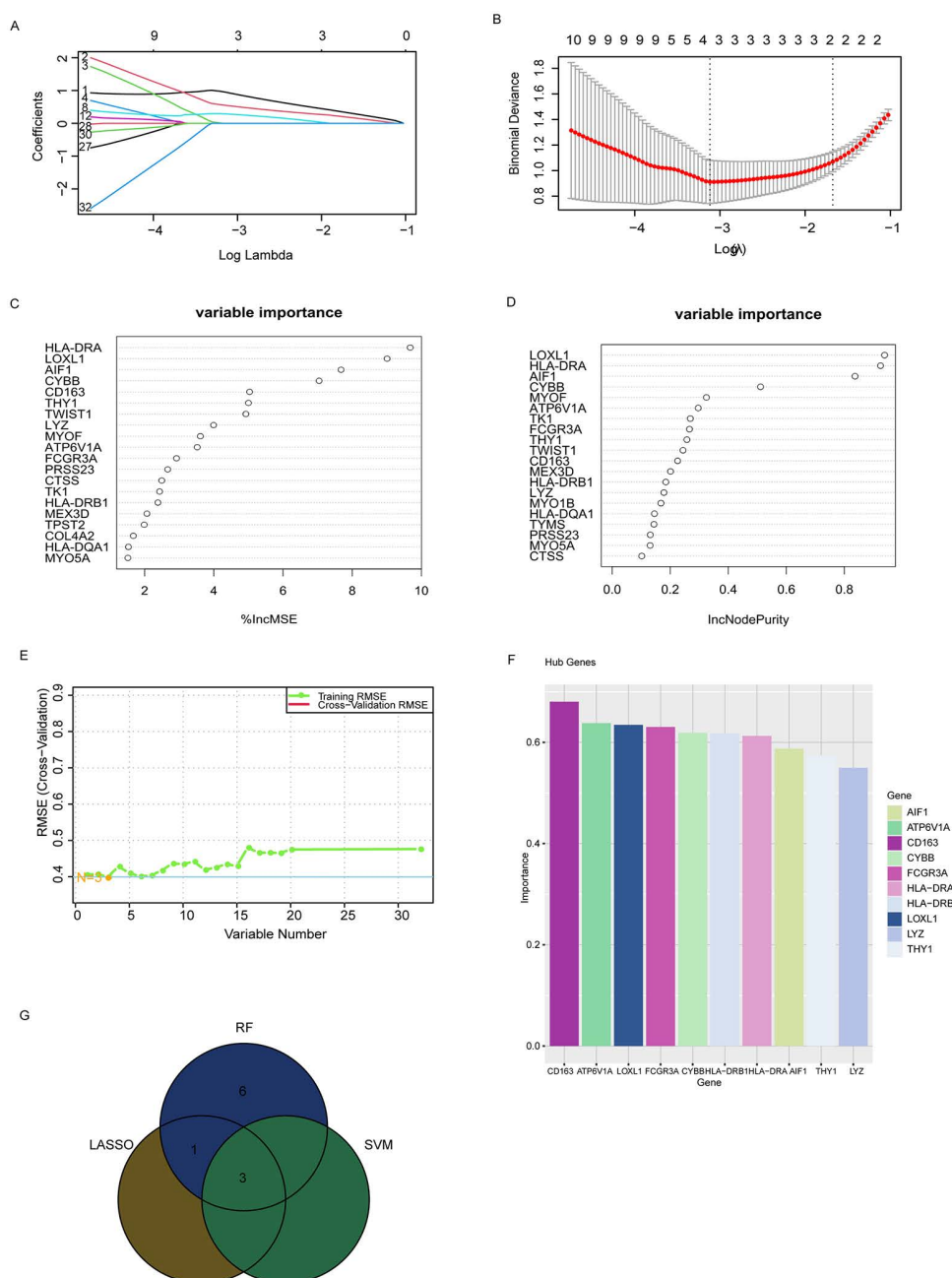


Fig 5. Identification of hub genes. (A) Ten-fold cross-validation of the LASSO model for selecting the optimal λ value. (B) LASSO coefficient analysis. (C, D) Ranking of gene importance by two RF metrics. (E, F) Ten-fold cross-validation with five replicates to identify the optimal feature subset with the lowest RMSE. (G) Venn diagram showing the intersection of genes selected by LASSO, RF, and SVM-RFE to identify the final hub genes.

<https://doi.org/10.1371/journal.pone.0351556.g005>

Discussion

Our findings identified three candidate hub genes (LOXL1, HLA-DRA, and CYBB), along with several enriched pathways and immune-related cell features that may be associated with OA. GO enrichment analysis showed that these genes were mainly enriched in biological processes such as extracellular matrix (ECM) organization and ossification. This enrichment

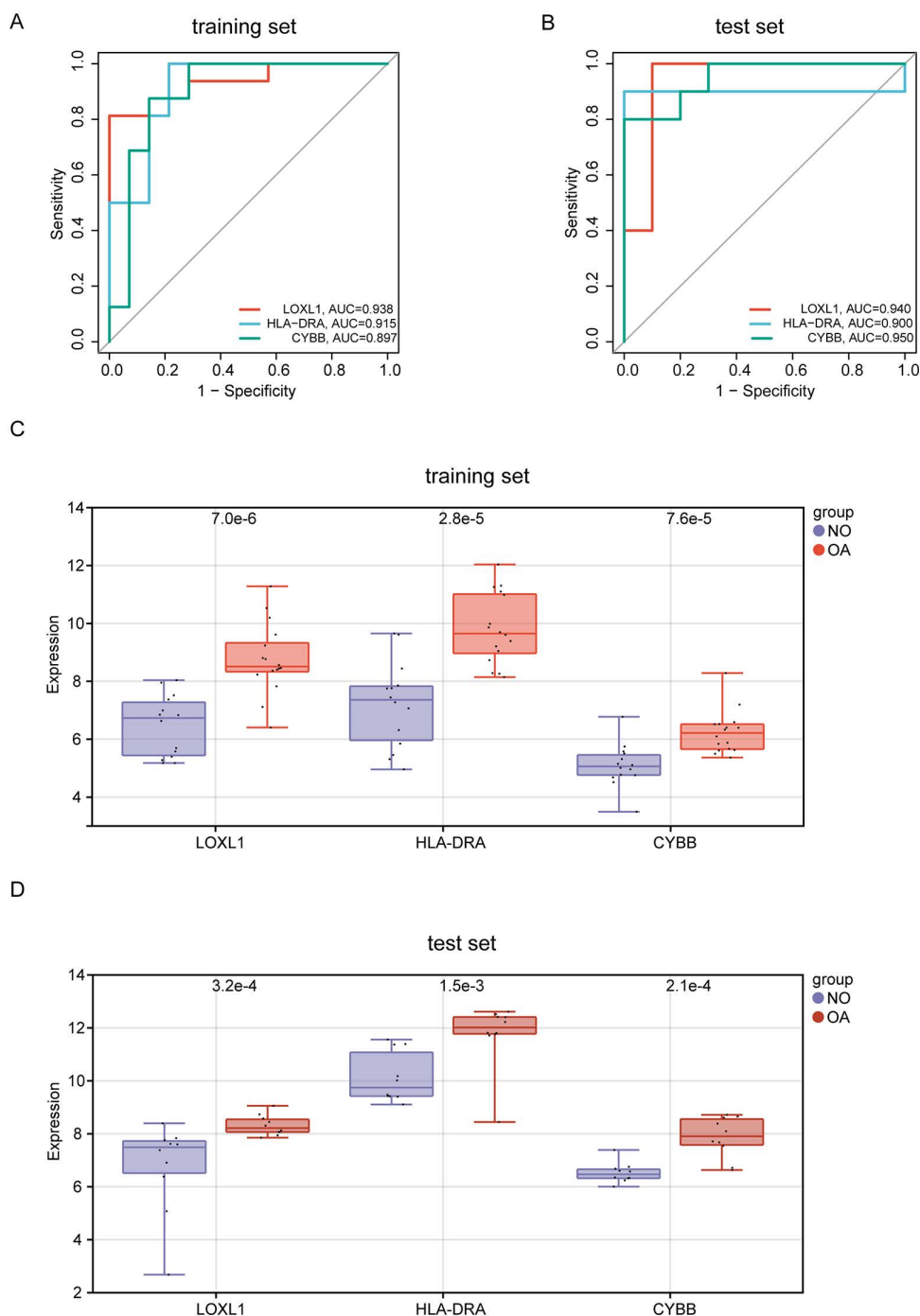


Fig 6. Evaluation of the expression and discriminative ability of hub genes. (A) ROC curve analysis of hub genes in the training set. (B) ROC curve analysis of hub genes in the test set. (C) Box plot of hub gene expression in Normal and OA samples in the training set. (D) Box plot of hub gene expression in Normal and OA samples in the test set. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

<https://doi.org/10.1371/journal.pone.0351556.g006>

is notable because ECM homeostasis is central to cartilage integrity and joint remodeling in OA [13]. Ossification contributes to the formation of osteophytes, which in turn leads to joint stiffness and pain in patients with OA [14]. Among the

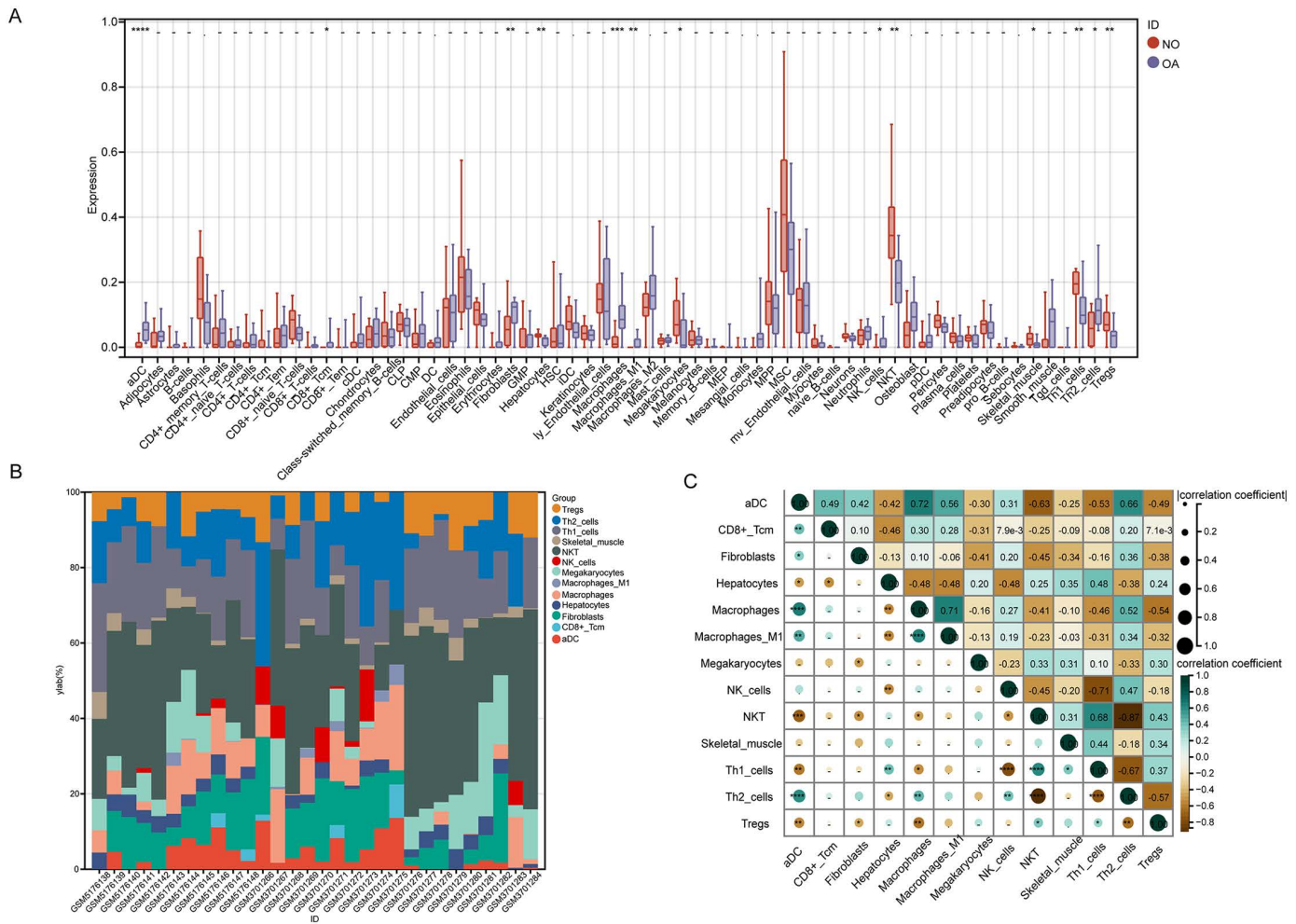


Fig 7. xCell immune infiltration analysis of OA-SEDEGs. (A) Box plot of the 64 immune cell types. (B) Relative proportions of 13 differentially enriched immune cell types. (C) Correlation heatmap showing correlations among the 13 differentially enriched immune cell types.

<https://doi.org/10.1371/journal.pone.0351556.g007>

cellular components, the identified genes were mainly enriched in the collagen-containing extracellular matrix and collagen trimer, suggesting a potential association with cartilage structural components. Regarding molecular functions, these genes were mainly enriched in extracellular matrix structural constituents and integrin binding. Integrins mediate interactions between cells and the ECM and regulate cell signaling, adhesion, and migration, which may contribute to OA progression [15]. KEGG analysis showed enrichment in pathways related to protein digestion and absorption, ECM–receptor interaction, phagosome, and antigen processing and presentation, suggesting potential links with matrix remodeling and immune-related processes in OA. Enrichment of the protein digestion and absorption pathway may reflect protease-related extracellular matrix degradation in OA. Collagen, a major extracellular matrix component, undergoes degradation, which is a key feature of OA [16]. Matrix metalloproteinase-13 (MMP-13) is a key collagenase whose overexpression contributes to cartilage erosion and degeneration and may thereby exacerbate OA [17]. Phagocytosis is crucial for removing apoptotic cells and pathogen-related molecules from the joints [18]. In OA, macrophages and other immune cells phagocytose apoptotic chondrocytes, influencing disease progression [19]. Enrichment of the antigen processing and presentation pathway suggests a potential association between immune-related processes and OA progression. An imbalance in

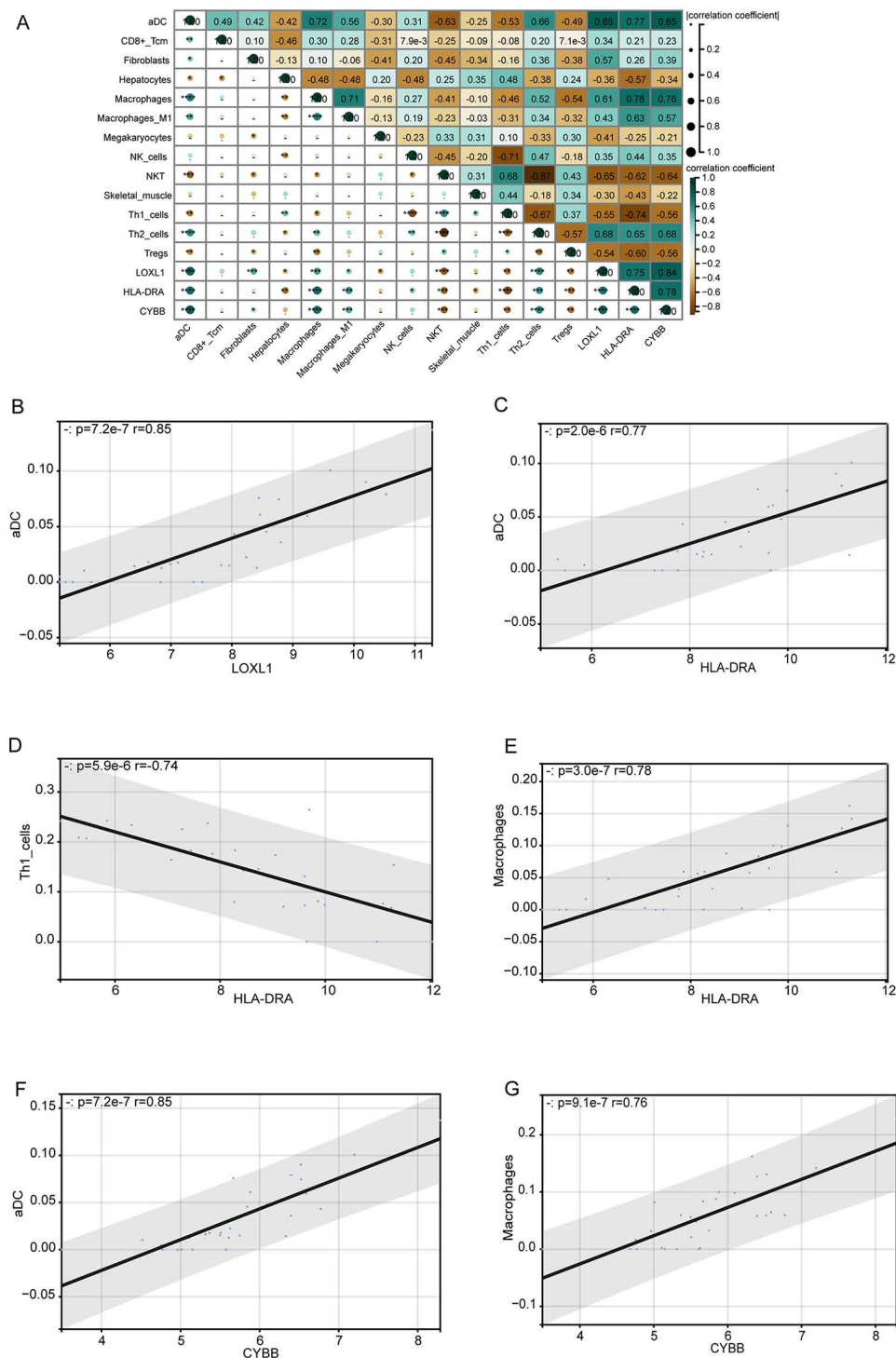


Fig 8. Correlation analysis between hub genes and immune cell infiltration scores. (A) Heatmap of the association between hub genes and 13 immune cell infiltration scores. (B) LOXL1 was positively correlated with aDC ($p=7.2e-7$, $r=0.85$). (C) HLA-DRA was positively correlated with aDC ($p=2.0e-6$, $r=0.77$). (D) HLA-DRA was negatively correlated with Th1 cells ($p=5.9e-6$, $r=-0.74$). (E) HLA-DRA was positively correlated with macrophages ($p=3.0e-7$, $r=0.78$). (F) CYBB was positively correlated with aDC ($p=7.2e-7$, $r=0.85$). (G) CYBB was positively correlated with macrophages ($p=9.1e-7$, $r=0.76$).

<https://doi.org/10.1371/journal.pone.0351556.g008>

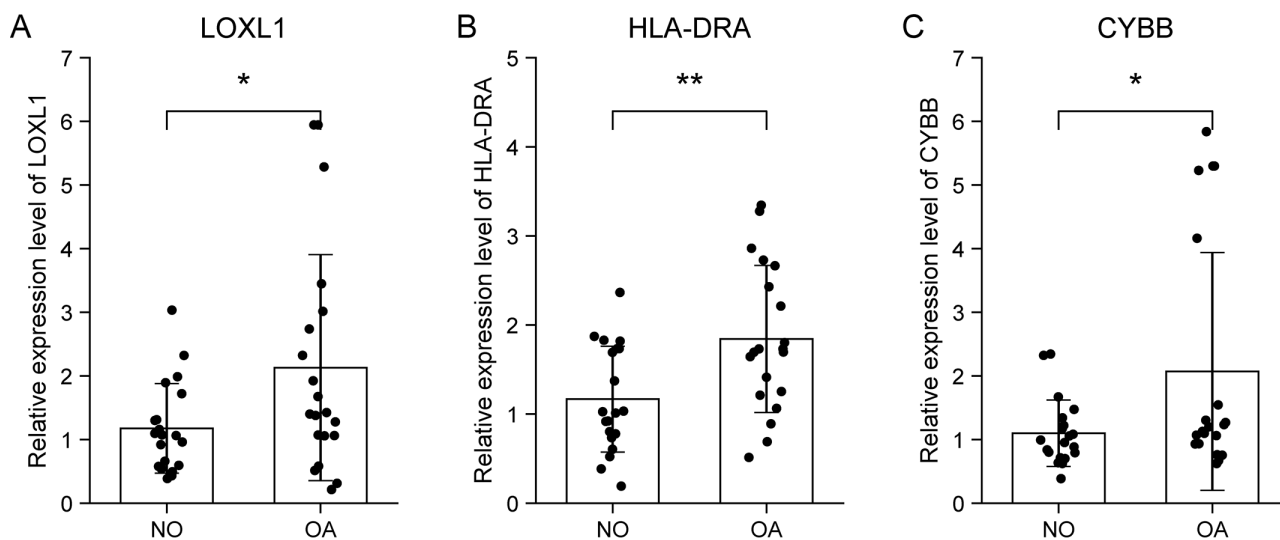


Fig 9. qRT-PCR. (A-C) LOXL1, HLA-DRA, and CYBB were significantly upregulated in peripheral blood samples from OA patients compared with the NO group. * $p < 0.05$, ** $p < 0.01$.

<https://doi.org/10.1371/journal.pone.0351556.g009>

immune responses may contribute to chronic inflammation and cartilage destruction [20]. Reduced estrogen levels may promote macrophage secretion of osteoclast-related cytokines and osteoclast activation, thereby potentially contributing to OA progression [21].

LOXL1 (lysyl oxidase-like 1), a member of the lysyl oxidase family, plays a critical role in connective tissue biogenesis. The prototypic member of this family is essential for ECM formation. LOXL1 has been reported to promote the cross-linking of collagen and elastin through its lysyl oxidase activity, thereby enhancing ECM strength and stability. In OA, LOXL1 has been reported to be upregulated and associated with cartilage degeneration and increased inflammation. Specifically, LOXL1 may participate in synovial inflammation through the PI3K/AKT signaling pathway [22]. Treatment with inhibitors such as ZSTK474 has been reported to reduce inflammatory responses associated with LOXL1 and PI3K/AKT signaling [23]. Although direct evidence that sex hormones regulate LOXL1 in OA is lacking, sex hormones may be indirectly linked to LOXL1-associated ECM remodeling in OA. In fibrotic contexts, E2 has been shown to inhibit the TGF- β 1/Smad3 pathway and reduce LOXL1-mediated ECM hypercross-linking, thereby delaying fibrosis [24,25]. HLA-DRA (major histocompatibility complex, class II, DR alpha) encodes the alpha chain of the HLA-DR class II molecule. Proteins encoded by HLA-DRA play a crucial role in antigen presentation, primarily by presenting antigens to CD4+ T cells and thereby initiating immune responses. Elevated HLA-DRA expression in OA has been reported to be associated with increased immune cell infiltration, including macrophages and T cells, and higher levels of proinflammatory cytokines, such as IL-1 β and TNF- α [26]. Sex hormones, such as estrogen, have been shown to modulate the expression of HLA class II molecules in immune cells. Changes in sex hormone levels may indirectly influence HLA-DRA-mediated inflammation in OA via the antigen presentation pathway, though this hypothesis requires experimental validation [27]. CYBB (cytochrome b-245 β chain), also referred to as NADPH oxidase 2 (NOX2), functions as a key subunit of the NADPH oxidase complex, generating reactive oxygen species (ROS), and its overexpression contributes to oxidative stress. Oxidative stress is involved in OA pathogenesis by inducing chondrocyte apoptosis, promoting matrix degradation, enhancing inflammation, impairing cellular function, and contributing to matrix sclerosis [28,29]. NOX2-derived oxidative stress may exacerbate MMP-mediated cartilage degradation, potentially contributing to tissue destruction in OA [30]. Oxidative stress is associated with homocysteine, a well-established biomarker linked to ROS production and redox imbalance.

Previous studies have shown that sex hormones may modulate homocysteine levels, which may in turn influence NOX-mediated ROS production [31]. Additionally, evidence suggests that the antioxidant properties of E2 help protect cartilage against ROS-induced damage [32].

In this study, we identified 13 differentially enriched immune cell types and observed significant associations between hub genes and selected immune cells. These findings provide insights into potential associations among sex hormone-associated genes, immune-related features, and OA. Macrophages, particularly the M1 subtype, have been implicated in synovial inflammation and cartilage degradation by producing pro-inflammatory cytokines, such as TNF- α and IL-1 β , and ROS via NOX2 activity [33]. Several studies have shown that the M1/M2 macrophage ratio is significantly higher in patients with knee OA than in controls, and this ratio is positively correlated with Kellgren-Lawrence grade [34]. The positive correlation between macrophages and aDC may suggest a coordinated pro-inflammatory immune pattern in OA. An imbalance between Th1 and Th2 cells has been reported in OA, with Th1 cells secreting IFN- γ and TNF- α and potentially promoting inflammation and joint degeneration, whereas Th2 cells are generally associated with anti-inflammatory responses [35]. Natural killer (NK) cells are innate immune cells that have been implicated in OA-related inflammation, partly through the production of IL-6 and TNF- α and recruitment via the CXCL10/CXCR3 axis [36,37]. They may also eliminate activated T cells through LFA-1-, NKG2D-, and TRAIL-mediated mechanisms and modulate adaptive immunity by regulating T cell activation [36]. The interaction between NK and T cells may be involved in immune infiltration and OA progression. Sex hormones may also be involved in OA-related immune regulation by influencing key immune cells. Estrogen has been reported to enhance anti-inflammatory macrophage responses, including cytokine production, chemotaxis, and phagocytosis [38]. It has also been reported to modulate NK cell activity by reducing NK cell infiltration, cytotoxicity, and proliferation, which may contribute to reduced inflammation [39]. Additionally, estrogen may modulate the Th1/Th2 balance by promoting Th2 cell cytokine secretion, thereby contributing to immune regulation [40]. Androgen has been shown to promote macrophage migration and the production of anti-inflammatory cytokines [41]. Correlation analysis showed that all three hub genes were strongly associated with immune cell infiltration, suggesting potential links between sex hormone-associated genes and immune alterations in OA. Furthermore, the potential role of sex hormones in modulating immune responses may provide a theoretical context for understanding OA-related immune alterations, but the underlying mechanisms require further experimental validation.

Although this study identified candidate sex hormone-associated genes in OA through bioinformatics analysis and qRT-PCR validation, several limitations remain. First, the links between LOXL1, HLA-DRA, and CYBB and sex hormone regulation in OA remain indirect and associative. The initial sex hormone-associated gene list was derived from the broadly annotated GeneCards database without gene-by-gene manual curation and may include genes with indirect or weak sex hormone-related associations. Therefore, the findings should be interpreted as candidate associations rather than direct evidence of sex hormone-mediated regulation. Second, clinical validation still has limitations. Direct estrogen and testosterone measurements were not performed, and detailed endocrine-related characteristics were not fully available. Therefore, the relationship between hormone status and candidate gene expression could not be directly assessed. Future studies should include larger validation samples with direct hormone measurements and more detailed endocrine information. Third, qRT-PCR validation was performed using peripheral blood samples rather than joint tissues. Therefore, the results should be interpreted as supportive biomarker evidence, rather than biological confirmation of tissue-level mechanisms or direct validation of cartilage or synovial transcriptomic findings. Future studies using joint tissues, synovial fluid, or paired blood–tissue samples are needed. Fourth, xCell immune infiltration analysis was based on computational estimation rather than direct biological validation, and experimental validation of the inferred immune cell patterns was not performed. In addition, functional experiments, such as gene knockdown or overexpression, chondrocyte or synoviocyte assays, ECM degradation analysis, inflammatory signaling assays, oxidative stress assays, or in vivo validation, were not performed. Therefore, LOXL1, HLA-DRA, and CYBB should be regarded as candidate associated genes or candidate markers rather than functionally validated regulators.

Conclusion

In summary, by integrating bioinformatics analysis with qRT-PCR validation, this study systematically explored candidate sex hormone-associated genes in OA and identified three candidate hub genes (LOXL1, HLA-DRA, and CYBB), which exhibited significant upregulation in both external datasets and peripheral blood samples. Furthermore, immune infiltration analysis suggested immune-related features associated with OA, providing associative evidence for the potential links among sex hormone-associated genes, immune alterations, and OA.

Supporting information

S1 Fig. Evaluation of batch effect removal before and after batch correction. (A) UMAP plot before batch correction. (B) UMAP plot after batch correction. The separation between datasets was markedly reduced after batch correction, indicating that batch effects were effectively alleviated.

(TIF)

S2 Fig. External assessment of immune infiltration patterns and hub gene–immune cell associations using xCell. (A) Differential immune cell infiltration patterns between the NO and OA groups in the independent external dataset, as estimated by xCell. (B) Correlation heatmap showing the associations between hub genes and major immune cell populations in the independent external dataset.

(TIF)

S3 Fig. External assessment of representative hub gene–immune cell correlations. (A–F) Representative correlations between hub genes and major immune cell populations in the independent external dataset, including CYBB–macrophages, CYBB–aDC, HLA-DRA–macrophages, HLA-DRA–Th1 cells, HLA-DRA–aDC, and LOXL1–aDC.

(TIF)

S1 Table. List of differentially expressed genes.

(XLSX)

S2 Table. GO biological process enrichment analysis of OA-SADEGs.

(XLSX)

S3 Table. GO cellular component enrichment analysis of OA-SADEGs.

(XLSX)

S4 Table. GO molecular function enrichment analysis of OA-SADEGs.

(XLSX)

S5 Table. KEGG pathway enrichment analysis of OA-SADEGs.

(XLSX)

S6 Table. Hub gene identification results from LASSO, SVM-RFE, and random forest analyses.

(XLSX)

S7 Table. Expression values of hub genes in the training set.

(XLSX)

S8 Table. Expression values of hub genes in the test set.

(XLSX)

S9 Table. Primer sequences used for qRT-PCR validation.

(XLSX)

Author contributions

Conceptualization: Yishu Wang, Shuna Jin, Rongzeng Huang, Chengwu Song.

Data curation: Yishu Wang, Ling Zhu, Yunzhou Zuo, Xugui Li.

Funding acquisition: Shuna Jin, Yunzhou Zuo.

Investigation: Yishu Wang, Yuhan Wang, Zhaoxiang Zeng, Xingliang Xiang.

Project administration: Shuna Jin, Rongzeng Huang, Chengwu Song.

Resources: Ling Zhu, Yunzhou Zuo, Xingliang Xiang, Xugui Li.

Software: Yishu Wang, Yuhan Wang, Zhaoxiang Zeng.

Supervision: Shuna Jin, Xingliang Xiang, Rongzeng Huang, Chengwu Song.

Validation: Yishu Wang, Shuna Jin, Yuhan Wang, Zhaoxiang Zeng.

Writing – original draft: Yishu Wang.

Writing – review & editing: Shuna Jin, Rongzeng Huang, Chengwu Song.

References

1. Leifer VP, Katz JN, Losina E. The burden of OA-health services and economics. *Osteoarthritis Cartilage*. 2022;30(1):10–6. <https://doi.org/10.1016/j.joca.2021.05.007> PMID: 34023527
2. Tang S, Zhang C, Oo WM, Fu K, Risberg MA, Bierma-Zeinstra SM, et al. Osteoarthritis. *Nat Rev Dis Primers*. 2025;11(1):10. <https://doi.org/10.1038/s41572-025-00594-6> PMID: 39948092
3. Segal NA, Nilges JM, Oo WM. Sex differences in osteoarthritis prevalence, pain perception, physical function and therapeutics. *Osteoarthritis Cartilage*. 2024;32(9):1045–53. <https://doi.org/10.1016/j.joca.2024.04.002> PMID: 38588890
4. Peshkova M, Lychagin A, Lipina M, Di Matteo B, Anzillotti G, Ronzoni F, et al. Gender-related aspects in osteoarthritis development and progression: A review. *Int J Mol Sci*. 2022;23(5):2767. <https://doi.org/10.3390/ijms23052767> PMID: 35269906
5. Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*. 2021;184(18):4784–818.e17. <https://doi.org/10.1016/j.cell.2021.07.038> PMID: 34450027
6. Gulati M, Dursun E, Vincent K, Watt FE. The influence of sex hormones on musculoskeletal pain and osteoarthritis. *Lancet Rheumatol*. 2023;5(4):e225–38. [https://doi.org/10.1016/S2665-9913\(23\)00060-7](https://doi.org/10.1016/S2665-9913(23)00060-7) PMID: 38251525
7. Tang J, Liu T, Wen X, Zhou Z, Yan J, Gao J, et al. Estrogen-related receptors: Novel potential regulators of osteoarthritis pathogenesis. *Mol Med*. 2021;27(1):5. <https://doi.org/10.1186/s10020-021-00270-x> PMID: 33446092
8. Yan Y-S, Qu Z, Yu D-Q, Wang W, Yan S, Huang H-F. Sex steroids and osteoarthritis: A Mendelian randomization study. *Front Endocrinol (Lausanne)*. 2021;12:683226. <https://doi.org/10.3389/fendo.2021.683226> PMID: 34248845
9. Ma N, Gao F. Correlation between low testosterone levels and the risk of osteoarthritis: A cross-sectional analysis of NHANES data (2011–2016). *BMC Musculoskelet Disord*. 2025;26(1):23. <https://doi.org/10.1186/s12891-024-08272-6> PMID: 39773699
10. Kumar N, Narayan Das N, Gupta D, Gupta K, Bindra J. Efficient automated disease diagnosis using machine learning models. *J Healthc Eng*. 2021;2021:9983652. <https://doi.org/10.1155/2021/9983652> PMID: 34035886
11. Auslander N, Gussow AB, Koonin EV. Incorporating machine learning into established bioinformatics frameworks. *Int J Mol Sci*. 2021;22(6):2903. <https://doi.org/10.3390/ijms22062903> PMID: 33809353
12. Liu Y. CWGCNA: An R package to perform causal inference from the WGCNA framework. *NAR Genom Bioinform*. 2024;6(2):lqae042. <https://doi.org/10.1093/nargab/lqae042> PMID: 38666214
13. Hodgkinson T, Kelly DC, Curtin CM, O'Brien FJ. Mechanosignalling in cartilage: An emerging target for the treatment of osteoarthritis. *Nat Rev Rheumatol*. 2022;18(2):67–84. <https://doi.org/10.1038/s41584-021-00724-w> PMID: 34934171
14. Fujii Y, Liu L, Yagasaki L, Inotsume M, Chiba T, Asahara H. Cartilage homeostasis and osteoarthritis. *Int J Mol Sci*. 2022;23(11):6316. <https://doi.org/10.3390/ijms23116316> PMID: 35682994
15. Goldring MB. Integrin-dependent recruitment of Src to ROS-producing endosomes in osteoarthritic cartilage. *Sci Signal*. 2023;16(809):eadj9760. <https://doi.org/10.1126/scisignal.adj9760> PMID: 37906627
16. Xu R, Wu J, Zheng L, Zhao M. Undenatured type II collagen and its role in improving osteoarthritis. *Ageing Res Rev*. 2023;91:102080. <https://doi.org/10.1016/j.arr.2023.102080> PMID: 37774932

17. Hu Q, Ecker M. Overview of MMP-13 as a promising target for the treatment of osteoarthritis. *Int J Mol Sci.* 2021;22(4):1742. <https://doi.org/10.3390/ijms22041742> PMID: 33572320
18. Zheng S, Li Y, Yin L, Lu M. Identification of sulfur metabolism-related gene signature in osteoarthritis and TM9SF2's sustenance effect on M2 macrophages' phagocytic activity. *J Orthop Surg Res.* 2024;19(1):62. <https://doi.org/10.1186/s13018-023-04384-2> PMID: 38218914
19. Zhao K, Ruan J, Nie L, Ye X, Li J. Effects of synovial macrophages in osteoarthritis. *Front Immunol.* 2023;14:1164137. <https://doi.org/10.3389/fimmu.2023.1164137> PMID: 37492583
20. Woodell-May JE, Sommerfeld SD. Role of inflammation and the immune system in the progression of osteoarthritis. *J Orthop Res.* 2020;38(2):253–7. <https://doi.org/10.1002/jor.24457> PMID: 31469192
21. Emmanuelle N-E, Marie-Cécile V, Florence T, Jean-François A, Françoise L, Coralie F, et al. Critical role of estrogens on bone homeostasis in both male and female: From physiology to medical implications. *Int J Mol Sci.* 2021;22(4):1568. <https://doi.org/10.3390/ijms22041568> PMID: 33557249
22. Morici L, Allémann E, Jordan O, Nikolić I. Promising LOX proteins for cartilage-targeting osteoarthritis therapy. *Pharmacol Res.* 2025;212:107627. <https://doi.org/10.1016/j.phrs.2025.107627> PMID: 39875019
23. Hu J, Liu X, Xu Q, Zhu M, Wang S, Quan K, et al. Mechanism of lysine oxidase-like 1 promoting synovial inflammation mediating rheumatoid arthritis development. *Aging (Albany NY).* 2024;16(1):928–47. <https://doi.org/10.18632/aging.205429> PMID: 38217541
24. Sethi A, Mao W, Wordinger RJ, Clark AF. Transforming growth factor-beta induces extracellular matrix protein cross-linking lysyl oxidase (LOX) genes in human trabecular meshwork cells. *Invest Ophthalmol Vis Sci.* 2011;52(8):5240–50. <https://doi.org/10.1167/iovs.11-7287> PMID: 21546528
25. Smith LC, Moreno S, Robertson L, Robinson S, Gant K, Bryant AJ, et al. Transforming growth factor beta1 targets estrogen receptor signaling in bronchial epithelial cells. *Respir Res.* 2018;19(1):160. <https://doi.org/10.1186/s12931-018-0861-5> PMID: 30165855
26. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol.* 2019;20(7):928–42. <https://doi.org/10.1038/s41590-019-0378-1> PMID: 31061532
27. Luckey D, Medina K, Taneja V. B cells as effectors and regulators of sex-biased arthritis. *Autoimmunity.* 2012;45(5):364–76. <https://doi.org/10.3109/08916934.2012.665528> PMID: 22429183
28. Ansari MY, Ahmad N, Haqqi TM. Oxidative stress and inflammation in osteoarthritis pathogenesis: Role of polyphenols. *Biomed Pharmacother.* 2020;129:110452. <https://doi.org/10.1016/j.biopha.2020.110452> PMID: 32768946
29. Yao X, Sun K, Yu S, Luo J, Guo J, Lin J, et al. Chondrocyte ferroptosis contribute to the progression of osteoarthritis. *J Orthop Translat.* 2020;27:33–43. <https://doi.org/10.1016/j.jot.2020.09.006> PMID: 33376672
30. van Dalen SCM, Kruisbergen NNL, Walgreen B, Helsen MMA, Slöetjes AW, Cremers NAJ, et al. The role of NOX2-derived reactive oxygen species in collagenase-induced osteoarthritis. *Osteoarthritis Cartilage.* 2018;26(12):1722–32. <https://doi.org/10.1016/j.joca.2018.08.014> PMID: 30195046
31. Tenkorang MA, Snyder B, Cunningham RL. Sex-related differences in oxidative stress and neurodegeneration. *Steroids.* 2018;133:21–7. <https://doi.org/10.1016/j.steroids.2017.12.010> PMID: 29274405
32. Cagnacci A, Gazzo I, Stigliani S, Paoletti AM, Anserini P, Londero AP, et al. Oxidative Stress: The Role of Estrogen and Progesterone. *J Clin Med.* 2023;12(23):7304. <https://doi.org/10.3390/jcm12237304> PMID: 38068356
33. Zhang H, Cai D, Bai X. Macrophages regulate the progression of osteoarthritis. *Osteoarthritis Cartilage.* 2020;28(5):555–61. <https://doi.org/10.1016/j.joca.2020.01.007> PMID: 31982565
34. Liu B, Zhang M, Zhao J, Zheng M, Yang H. Imbalance of M1/M2 macrophages is linked to severity level of knee osteoarthritis. *Exp Ther Med.* 2018;16(6):5009–14. <https://doi.org/10.3892/etm.2018.6852> PMID: 30546406
35. Wen Z, Qiu L, Ye Z, Tan X, Xu X, Lu M, et al. The role of Th/Treg immune cells in osteoarthritis. *Front Immunol.* 2024;15:1393418. <https://doi.org/10.3389/fimmu.2024.1393418> PMID: 39364408
36. Zheng C-Q, Zeng L-J, Liu Z-H, Miao C-F, Yao L-Y, Song H-T, et al. Insights into the roles of natural killer cells in osteoarthritis. *Immunol Invest.* 2024;53(5):766–87. <https://doi.org/10.1080/08820139.2024.2337025> PMID: 38622991
37. Benigni G, Dimitrova P, Antonangeli F, Sansevero E, Milanova V, Blom A, et al. CXCR3/CXCL10 axis regulates neutrophil-NK cell cross-talk determining the severity of experimental osteoarthritis. *J Immunol.* 2017;198(5):2115–24. <https://doi.org/10.4049/jimmunol.1601359> PMID: 28108560
38. Barcena ML, Christiansen-Mensch C, Aslam M, Haritonow N, Ladilov Y, Regitz-Zagrosek V. Upregulation of mitochondrial Sirt3 and alleviation of the inflammatory phenotype in macrophages by estrogen. *Cells.* 2024;13(17):1420. <https://doi.org/10.3390/cells13171420> PMID: 39272992
39. Baker AE, Brautigam VM, Watters JJ. Estrogen modulates microglial inflammatory mediator production via interactions with estrogen receptor beta. *Endocrinology.* 2004;145(11):5021–32. <https://doi.org/10.1210/en.2004-0619> PMID: 15256495
40. González DA, Díaz BB, Rodríguez Pérez M del C, Hernández AG, Chico BND, de León AC. Sex hormones and autoimmunity. *Immunol Lett.* 2010;133(1):6–13. <https://doi.org/10.1016/j.imlet.2010.07.001> PMID: 20637236
41. Martin CS, Crastin A, Sagmeister MS, Kalirai MS, Turner JD, MacDonald L, et al. Inflammation dynamically regulates steroid hormone metabolism and action within macrophages in rheumatoid arthritis. *J Autoimmun.* 2024;147:103263. <https://doi.org/10.1016/j.jaut.2024.103263> PMID: 38851089