Heliyon 10 (2024) e37047

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification and experimental validation of key genes in osteoarthritis based on machine learning algorithms and single-cell sequencing analysis

Enming Yu, Mingshu Zhang, Chunyang Xi, Jinglong Yan *

Department of Orthopedics, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

ARTICLE INFO

Keywords:
Osteoarthritis
Machine learning algorithm
Differentially expressed genes
Key genes
Immune response

ABSTRACT

Purpose: Osteoarthritis (OA) is a prevalent cause of disability in older adults. Identifying diagnostic markers for OA is essential for elucidating its mechanisms and facilitating early diagnosis.
Methods: We analyzed 53 synovial tissue samples (n = 30 for OA, n = 23 for the control group) from two datasets in the Gene Express Omnibus (GEO) database. We identified differentially expressed genes (DEGs) between the groups and applied dimensionality reduction using six machine learning algorithms to pinpoint characteristic genes (key genes). We classified the OA samples into subtypes based on these key genes and explored the differences in biological functions and immune characteristics among subtypes, as well as the roles of the key genes. Additionally, we constructed a protein-protein interaction network to predict small molecules that target these genes. Further, we accessed synovial tissue sample data from the single-cell RNA dataset GSE152805, categorized the cells into various types, and examined variations in gene expression and their correlation with OA progression. Validation of key gene expression was conducted in cellular experiments using the qPCR method.
Results: Four genes *AGMAT, MAP3K8, PER1*, and *XIST*, were identified as characteristic genes of OA. All can independently predict the occurrence of OA. With these genes, the OA samples can be clustered into two subtypes, which showed significant differences in functional pathways and immune infiltration. Eight cell types were obtained by analyzing the single-cell RNA data, with synovial intimal fibroblasts (SIF) accounting for the highest proportion in each sample. The key genes were found over-expressed in SIF and significantly correlated with OA progression and the content of immune cells (ICs). We validated the relative levels of key genes in OA and normal cartilage tissue cells, which showed an expression trend consistency with the bioinformatics result except for *XIST*.
Conclusion: Four genes, *AGMAT, MAP3K8, PER1*, and *XIST* are closely related to the progression of OA, and play as diagnostic and predictive markers in early OA.

1. Introduction

Osteoarthritis (OA) manifests as a chronic condition characterized by the deterioration of joints, skeletal damage, and the degradation of articular cartilage [1]. As the condition progresses to later stages, it incapacitates approximately 53 % of those affected,

* Corresponding author.
E-mail address: 600728@hrbmu.edu.cn (J. Yan).

https://doi.org/10.1016/j.heliyon.2024.e37047
Received 11 April 2024; Received in revised form 23 August 2024; Accepted 27 August 2024
Available online 28 August 2024

profoundly influencing their mobility and overall quality of life [2]. Observations indicate a continuous rise in OA prevalence globally. Over the past two decades, the incidence of OA has escalated by 113.25 %, with an annual growth rate of 0.12 % [3]. Although pharmaceutical treatments and physical therapies can ameliorate symptoms and slow the progression of OA, there is currently no cure, according to guidelines from the American College of Rheumatology/Arthritis Foundation [4]. Recent advances in single-cell RNA sequencing (scRNA-seq) have shown that OA patients' femoral cartilage contains chondrocyte profiles that differ significantly from those of healthy individuals [5]. Additionally, new research has highlighted the importance of three disulfidptosis regulators—NCKAP1, OXSM, and SLC3A2—in the individualized assessment and treatment of OA [6]. Consequently, the importance of early detection and preventive strategies is underscored.

OA's emergence is linked to various risk factors, such as obesity, aging, excessive physical activity, inflammation, genetic predispositions, and physical trauma. Although these factors are recognized, the definitive etiological factors and pathophysiological mechanisms behind OA remain elusive. There is speculation that OA may result from a disruption in the balance between the breakdown and synthesis of molecules within the articular chondrocytes, extracellular matrix (ECM), and subchondral bone [7]. Moreover, there is a growing belief that OA may function as an immune-mediated disease, characterized by the activation of immune complexes, osteoclasts, and a dysregulation of cytokines, which are instrumental in its advancement [8]. Further elucidation of these aspects is imperative.

Synovitis is a key feature at the earliest stage of OA. Damage to the synovial tissue is an important pathological manifestation of persistent joint degeneration that accompanies the entire process of OA development [9]. Exploring the pathogenesis and diagnostic markers of OA in synovial tissue is of great significance for identifying therapeutic targets, alleviating symptoms, and improving the prognosis of this disease.

The present study examined OA-related transcriptome and single-cell sequencing data from GEO-archived synovial tissue samples. Our objectives were to identify genes indicative of OA's initiation, utilize these genes to stratify OA patients, probe the biological and immunological attributes of these clusters, and delve into the contributions of these genes to both the onset and development of OA.

2. Materials

2.1. Data preparation

We obtained the scRNA-seq dataset GSE152805 [10] along with the chip sequencing datasets GSE12021 [11] and GSE55235 [12] from the GEO database. The GSE152805 dataset, derived from *Homo sapiens,* was processed using the GPL20301 Illumina HiSeq 4000 platform. It comprised synovial cells from three knee OA patients, including three normal and six diseased cartilage samples, all included in this analysis. Similarly, datasets GSE12021 and GSE55235, also from *homo sapiens*, utilized platforms GPL96 and GPL97 respectively. Specifically, GSE12021 contained 33 synovial tissue samples—13 normal and 20 OA—while GSE55235 included 20 samples, split evenly between healthy individuals and OA patients. All these samples were integrated into our study.

2.2. Screening of DEGs in OA

The R package "limma" [13] facilitated the differential gene analysis between OA and normal samples, enabling the identification of significant DEGs in GSE12021 and GSE55235. We established thresholds of $|\text{log2Fold change (log2FC)}| > 1$ and $P < 0.05$. Genes with $\text{log2FC} > 1$ were deemed upregulated, and those with $\text{log2FC} \leq 1$ as downregulated. These DEGs were visualized using S-maps and heatmaps, and their intersections were analyzed to discern common OA-related DEGs (OADEGs). We predicted interactions between proteins linked to these OADEGs using the STRING database [14], constructing a protein-protein interaction (PPI) network visualized through Cytoscape.

2.3. Gene set variation analysis (GSVA)

The GSVA [15], an unsupervised, nonparametric method, was employed to convert gene expression data from matrices of individual genes to matrices representing specific gene sets. This included performing rank statistic calculations such as the Kolmogorov-Smirnov test on each gene set. Using the R package "gsva" (V1.42.0), we calculated the GSVA enrichment scores for each sample across both RNA sequencing datasets.

2.4. Machine learning and screening of characteristic genes in OA

A total of six machine learning algorithms were employed in the analysis, with the GSE12021 dataset serving as the training set and GSE55235 as the validation set. These algorithms included bagged decision tree [16], Bayesian [17], random forest [18], Wrapper (Boruta) [19], learning vector quantization (LQV) [20], and the least absolute shrinkage and selection operator (LASSO) [21]. The LASSO regression, a linear regression variant that introduces a penalty term (lambda × |slope|), was adopted to reduce overfitting and improve model generalizability. The logistic regression model's objective function was structured as follows:

$$min \int (\alpha_0, \alpha|X_i, Y_i + \lambda\|\alpha\|_1)$$

The penalty coefficient, represented by λ, was chosen using a 10-fold cross validation. The sum of the absolute values of each vector element is defined as $||\alpha||_1$. The regression was performed using the R package "glmnet" with 1000 simulations to identify genes with an importance greater than 0.2.

The screened genes from the six algorithms were intersected to select characteristic genes, namely Hub-OADEGs, which were defined as genes that selected by five or more algorithms. The visualization of the Hub-OADEGs was achieved through the R package "Upset" [22]. An analysis was performed on the Hub-OADEGs in the GSE12021 and GSE55235 datasets using the R package "cowplot." The results were visualized through heat maps, scatter plots, and correlation curves. Using the R package "RCircos", a chromosome localization map was successfully created [23].

2.5. Construction of risk model for OA

Using a multivariate logistic regression model, we assessed all Hub-OADEGs. The risk-scoring formula was derived from the regression coefficients of characteristic genes.

$$\text{riskScore} = \sum_{i} \text{Coefficient}\ (\text{gene}_i) \times \text{mRNA Expression}\ (\text{gene}_i)$$

A nomograph model was developed utilizing the R package "rms" [24] to forecast OA onset in patients, based on Hub-OADEGs from the GSE12021 dataset. The model's precision and resolution were evaluated through a calibration curve from calibration analysis [25]. Diagnostic precision was further assessed by a receiver operating characteristic (ROC) curve using "pROC" [26], with an area under the curve (AUC) between 0.5 and 1 signifying enhanced diagnostic efficacy. Decision Curve Analysis (DCA) [27] was conducted to ascertain the model's predictive accuracy, facilitated by the R package "ggDCA" for generating DCA maps.

2.6. Identification of molecular subtypes of OA

Consistency clustering, a resampling-based method, was applied to delineate sub-classes and confirm clustering robustness [28]. Using the "ConsensusClusterPlus" package [29], we clustered OA samples from GSE12021 into molecular phenotypes according to Hub-OADEGs. Principal coordinate analysis (PCoA) [30] verified the clustering's accuracy and robustness. The variances among clusters were depicted through a box plot crafted with the R package "ggpubr."

2.7. Evaluation of biological characteristics of different molecular phenotypes

Gene Ontology (GO) analysis [31] was used to annotate DEGs among different molecular phenotypes using "ClusterProfiler" [32]. This analysis encompasses biological processes, molecular functions, and cellular components. In addition, the statistical variances in specific gene sets across two biological states were assessed using single-sample gene set enrichment analysis (ssGSEA) [33].

2.8. CIBERSORT

We utilized the CIBERSORT linear support vector regression tool [34], which allowed us to deconvolute the expression matrix of human immune cell subtypes. With this tool, we were able to evaluate 22 immune cell types in the GSE12021 dataset. To examine the infiltration disparities between OA and normal samples, we used the Wilcoxon test, which considered $P < 0.05$ as statistically significant.

2.9. SsGSEA immunologic infiltration analysis

Using labeling and the enrichment fraction calculation, the ssGSEA algorithm determined the relative abundance of each type of immune cell in the samples. As an example, the "ggplot2" program was used to display the pattern of immune cell infiltration across illness subtypes in the GSE12021 OA dataset [35]. Correlation heatmaps, produced using "pheatmap" [36], depicted the relationships between key characteristic genes and ICs across different OA risk groups.

2.10. Quality control on single-cell data

We utilized the "Seurat" (version 4.0) R package to import a matrix of nine samples from the single-cell dataset GSE152805 and subsequently created a Seurat object for our analyses. Quality control measures were enforced by excluding cells expressing fewer than 250 genes and those with over 20 % of unique molecular identifiers linked to mitochondrial or ribosomal genes. Doublet cells were identified and removed using the "DoubletFinder" function with default settings [37]. We also corrected for batch effects across the samples using the "Harmony" R package [38].

2.11. Cluster analysis and cell annotation

Post-quality standardization, we reanalyzed the single-cell data from GSE152805 using "Seurat". The "FindVariableFeatures" function was employed to isolate the top 2000 genes with significant expression variability. A dimensionality reduction technique called principal component analysis (PCA) was used on these genes [39]. Different kinds of cells were located by employing the

"FindNeighbors" and "FindClusters" features. Marker genes for each cluster were established with a cutoff threshold of $P < 0.05$ and a fold change greater than 0.5. Eight distinct cell types were annotated based on the literature by Chou CH et al. [10], with markers identified as: SIF- *PRG4*, synovial subintimal fibroblasts (SSF)-*WISP2*, ICs-*HLA-DRA*, smooth muscle cells (SMC)-*RGS5*, endothelial cells (EC)-*TM4SF1*, T cells-*CD3D*, mast cells-*TPSAB1*, and proliferative immune cells (PIC)-*BIRC5*.

2.12. AUCell scores

We used the "AUCell" R package (version 1.12.0) [40] to compute gene set scores for the GSE152805 dataset, passing in Hub-OADEGs as an input. These scores estimated the percentage of highly expressed genes per cell by ranking gene expression in each cell according to AUC values. For these gene sets, a threshold was set using the "AUCell_exploreThreshold" function. Then, to highlight the clusters' active characteristic genes, we used AUC scores to color-code the Uniform Manifold Approximation and Projection (UMAP) that was contained in the clusters.

2.13. Intercellular communication analysis

The "CellChat" (version 1.1.3) R package [41] was used to analyze intercellular communication networks by using the "CellChat" function to quantitatively analyze the networks from scRNA-seq data. Circle diagrams were drawn to display the cell-interaction relationship of single-cell subsets in OA, and bubble diagrams were used to count all important receptor pairs during intercellular signal transduction.

2.14. Cell differentiation of single-cell subsets

Cell differentiation in single-cell subsets was deduced using "Monocle" [42]. We generated an integrated gene expression matrix from the Seurat object to construct a cell dataset in Monocle. The "VariableFeatures" function identified genes with high variability, and the "setOrderingFilter" was used to sort cells. Dimensionality was reduced via the "DDRTree" method, and cell arrangements along developmental trajectories were estimated with "orderCells". We analyzed the trajectories of differentiation for different disease phenotypes in the dataset, examining the expression tracks of significant genes before the progression of these phenotypes. Each trajectory was analyzed following a standardized protocol with default settings.

2.15. QPCR validation of key genes in OA

C28/12 (human normal chondrocytes) were divided into two groups. One group was cultured with lipopolysaccharide (LPS) at 10 μ g/mL for 12 h to obtain OA samples, while the other group did not add LPS. Two sets of samples were subjected to fluorescence quantitative PCR detection to determine the relative expression of RNA, with primers as follows:

AGMAT (amplification fragment size 193bp)
AGMAT - F: GACCTTGGGTGGAGATCACAC.
AGMAT - R: CACCACACGCTTACAGTCCAG.
PER1 [127bp].
PER1-F: ACGGGCCGAATCGTCTACA.
PER1-R: TGGAACCATAGAACACTCCCAC.
MAP3K8 [201bp].
MAP3K8-F: CTCCCCAAAAATGGACGTTACC.
MAP3K8-R: GGATTTCCACATCAGATGGCTTA.
XIST [113bp].
XIST-F: TCTAGTCCCCCACCACCCTT.
XIST-R: GGAGGACGTGTCAAGAAGACA.

2.16. Statistical analyses

We used R (version 4.1.1) to process and analyze the data. An independent Student's t-test was used to examine continuous variables with normal distributions. The variables that did not follow a normal distribution were evaluated using the Mann-Whitney *U* test. Various genes' correlation coefficients (r) were calculated using Pearson's correlation analysis. The significance levels for all bilateral statistical tests were set at $P < 0.05$.

3. Results

3.1. The workflow of the data process and analysis

The workflow is depicted in Supplementary Fig. S1.

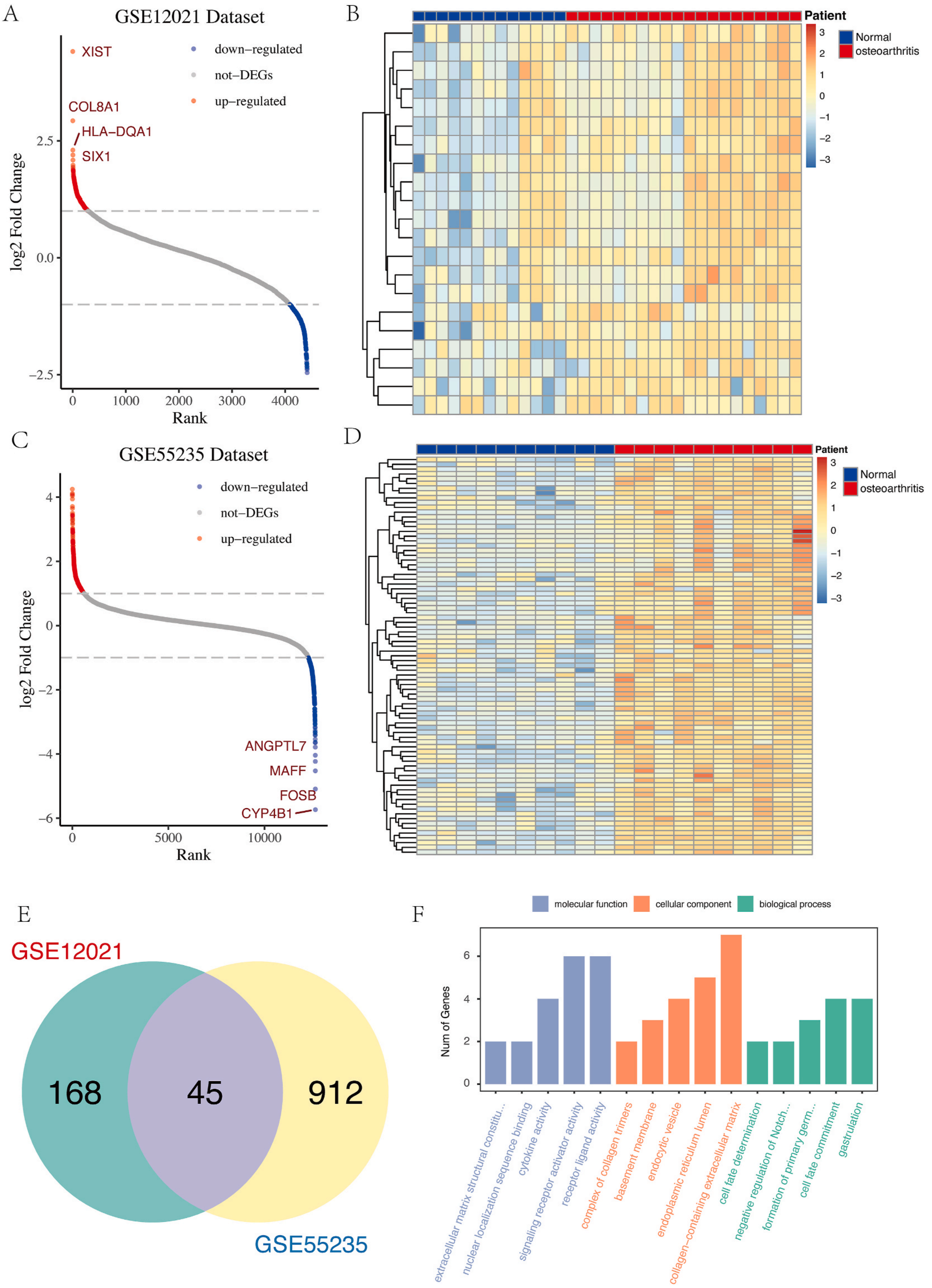


Fig. 1. Differential gene expression analysis for OA. (A) GSE12021 dataset; comparison of DEGs between OA patients with healthy individuals; red, high expression in patients with OA; blue, high expression in healthy individuals. (B) Correlation heatmap of DEGs in the GSE12021 dataset; darker color represents higher correlation. (C) GSE55235 dataset; comparison of the DEGs between patients with OA and healthy individuals, red, high expression in patients with OA; blue; high expression in healthy individuals. (D) Correlation heatmap of DEGs in the GSE55235 dataset. (E) Venn plots of DEGs in the GSE12021 and GSE55235 datasets. (F) GO enrichment analysis of DEGs.

3.2. Screening of OADEGs and their basic biological functions

This analysis explored genetic variances between OA and normal samples employing the GSE12021 and GSE55235 datasets (Supplementary Table S1). In the GSE12021 dataset, we identified 4415 DEGs, consisting of 291 upregulated and 327 downregulated DEGs (Fig. 1A and B, Supplementary Table S2). Conversely, the GSE55235 dataset yielded 1596 DEGs, with 484 upregulated and 299 downregulated (Fig. 1C and D, Supplementary Table S3). An intersection of these datasets pinpointed 45 overlapping OADEGs (Fig. 1E). GO enrichment analysis was performed to define the biological functions of these OADEGs, significantly involving biological processes such as gastrulation, cell fate commitment, and cell fate determination (Supplementary Table S4); cellular components like collagen-containing ECM, collagen trimer complexes, and basement membranes (Supplementary Table S5); and molecular functions

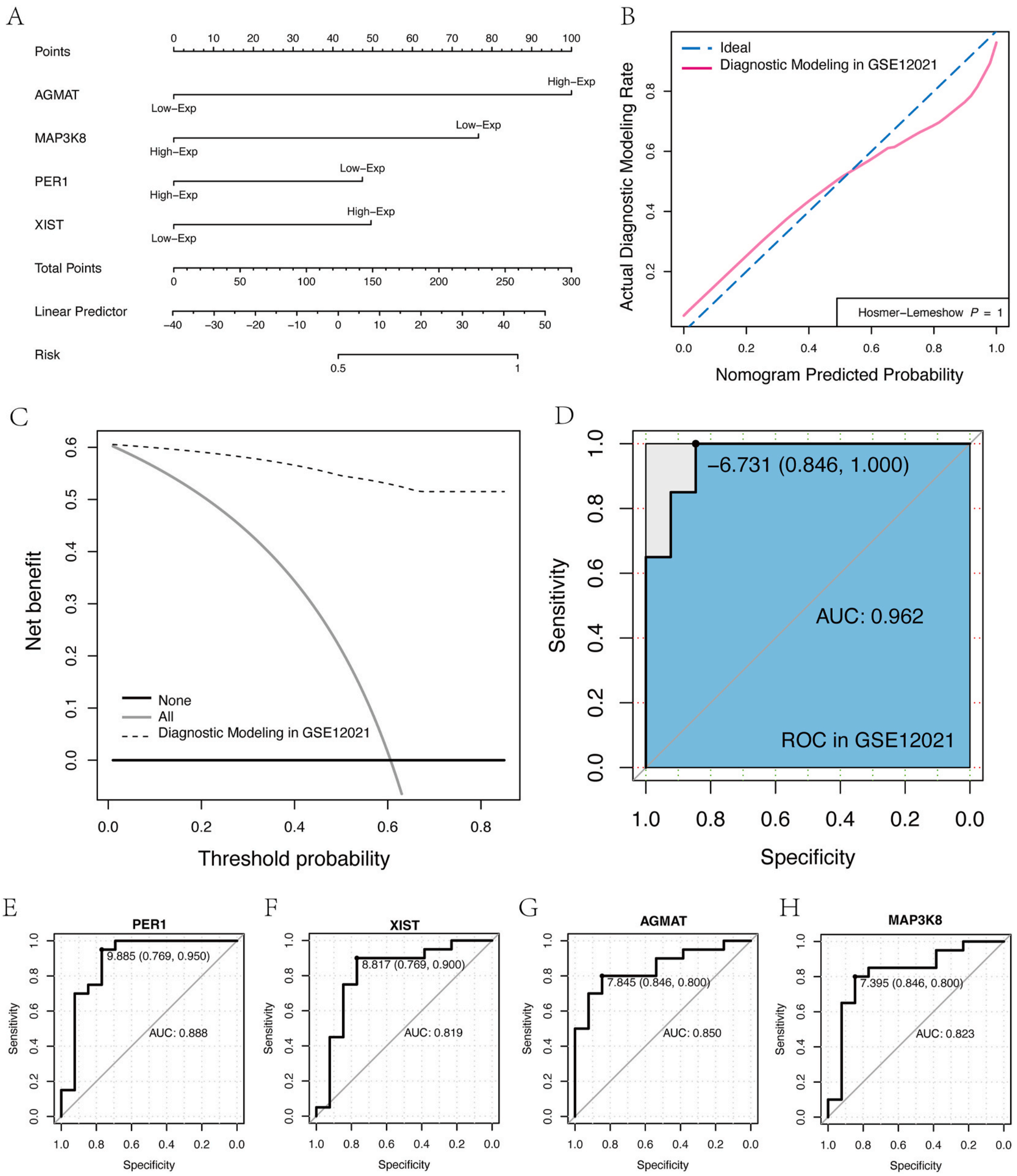


Fig. 2. Nomogram model for OA in GSE12021. (A) Nomogram, with the score of the important 7 genes in the middle, and the Total score representing the total score obtained by the patient based on the 7 gene score. The total score corresponds to the incidence probability below; (B) The calibration of the Homer Lemeshow evaluation model; (C) The clinical decision curve of the diagnostic model for OA. The vertical axis represents the net income, whereas the horizontal axis represents probability threshold or threshold probability; (D) ROC curve of the predictive model for OA occurrence; (E) ROC curve of PER1 gene expression predicting the onset of OA; (F) ROC curve of XIST gene expression predicting the onset of OA; (G) ROC curve of AGMAT gene expression predicting the onset of OA; (H) ROC curve of MAP3K8 gene expression in predicting the onset of OA.

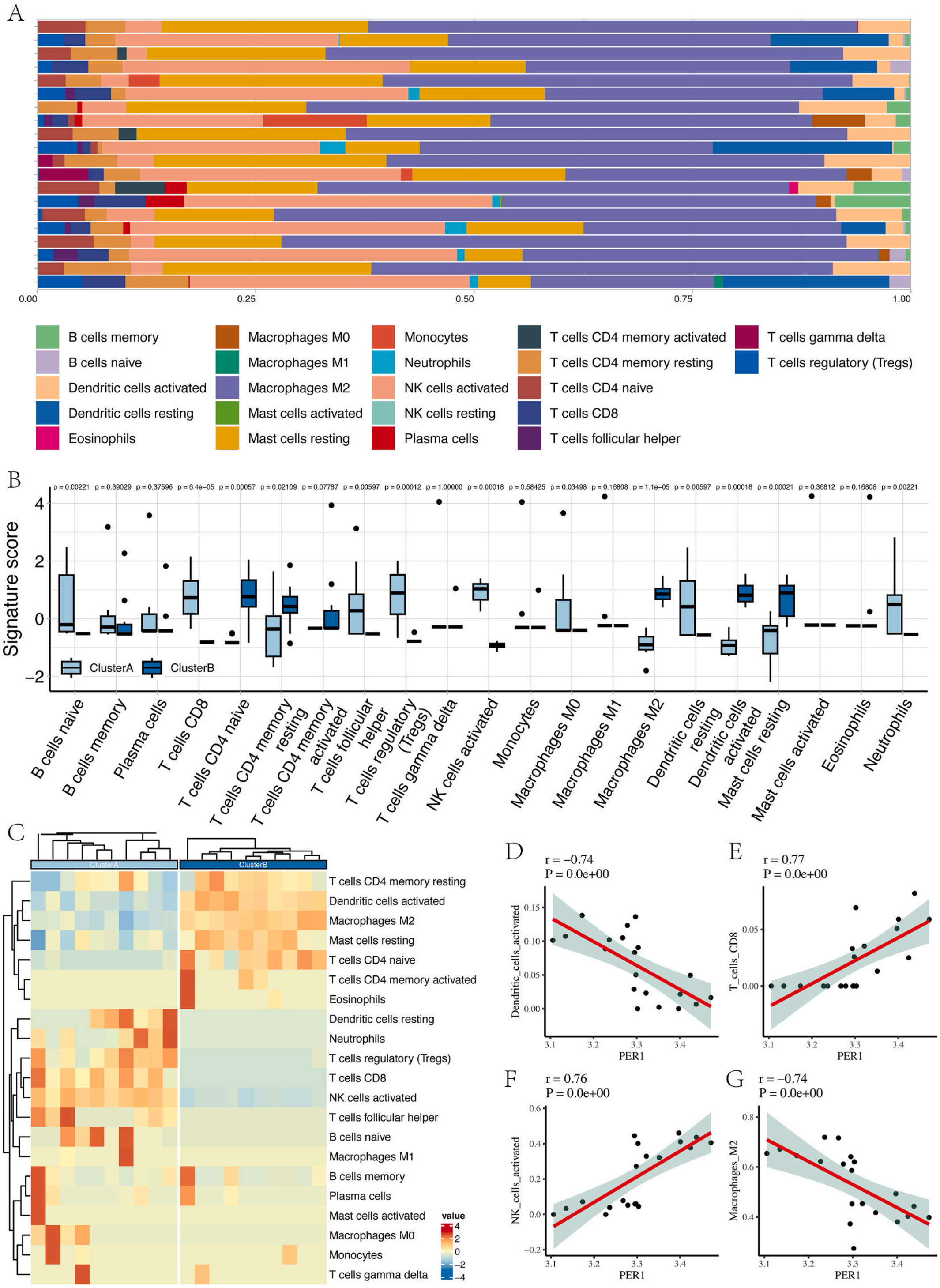


Fig. 3. Immune infiltration analysis conducted with the CIBERSORT algorithm on the GSE12021 dataset. (A) Histogram depicting the proportion of ICs. (B) Box plot comparing the abundance of immune cell infiltration across OA subtypes. (C) Heatmap illustrating the correlation of immune cell infiltration abundance, displayed in red and blue gradients. (D) Correlation analysis between the expression level of the PER1 gene and the infiltration abundance of activated dendritic cells. (E) Correlation analysis linking the expression level of the PER1 gene with CD8T cell infiltration. (F) Analysis of the correlation between the PER1 gene expression level and the infiltration of activated NK cells. (G) Examination of the correlation between PER1 gene expression and the infiltration abundance of M2 macrophages.

such as receptor ligand activity and nuclear localization sequence binding (Supplementary Table S6, Fig. 1F).

3.3. PPI network of OADEGs

A PPI network of the 45 OADEGs was established using the STRING database and visualized with Cytoscape software (Supplementary Fig. S2A). This network comprised 43 nodes and 12 interaction pairs, with node weights calculated using the Cluster-Coefficient algorithm in the "CytoHubba" Cytoscape plug-in. The results showed that NFKBIA, CXCL12, and TLR7 had the highest weights and highest correlations with other genes (Supplementary Fig. S2B).

Using the GSE12021 dataset for training and the GSE55235 dataset for validation, this study constructed a diagnostic model for OA. Using the bagged tree (Supplementary Fig. S3A), Bayesian (Supplementary Fig. S3B), random forest (Supplementary Fig. S3C), Boruta (Supplementary Fig. S3D), LQV (Supplementary Fig. S3F) and LASSO-logistic (Supplementary Fig. S3E) algorithms, the relationship between the 45 OADEGs and the incidence of OA in the GSE12021 dataset was analyzed, and the characteristic genes of different algorithms were intersected (Supplementary Fig. S3G). Four important characteristic genes of OA (Hub-OADEGs) were found to be closely related to the pathogenesis of OA: AGMAT, MAP3K8, PER1, and XIST.

The GSE12021 dataset was used to create a diagnostic model for OA (Fig. 2A), which was verified across both the training and validation cohorts. A P-value of 1 in the training set (GSE12021) indicated a minimal risk of Type I error, demonstrating that the model's predictions were well-aligned with actual data, meaning that the model's calibration was tested by the Hosmer-Lemeshow test (Fig. 2B). DCA assessed the model's clinical utility in the training set (GSE12021), demonstrating significant clinical value and benefit (Fig. 2C). Model diagnostics revealed an AUC of 0.962 for the training set (Fig. 2D). The four Hub-OADEGs had excellent diagnostic ability independent of the onset of OA, with AUC value of PER1 0.888 (Fig. 2E), XIST 0.819 (Fig. 2F), AGMAT 0.850 (Fig. 2G), MAP3K8 0.823 (Fig. 2H).

The Hosmer–Lemeshow dataset showed the same result in GSE55235 as in GSE12021 (Supplementary Fig. S2A) and DCA analysis (Supplementary Fig. S2B). The C-index of the diagnostic model constructed in this study in the training set (GSE12021) and the validation set (GSE55235) is 0.99 and 1, respectively (Supplementary Fig. S2C). We also verified the diagnostic ability of the four Hub-OADEGs in diagnosing the onset of OA in the validation set (GSE55235). The results showed that the AUC values of AGMAT (Supplementary Fig. S2D), MAP3K8 (Supplementary Fig. S2E), PER1 (Supplementary Fig. S2F), and XIST (Supplementary Fig. S2G) were 0.890, 0.980, 1, and 0.815, respectively, indicating that all four genes had excellent diagnostic ability.

3.4. Correlation between the Hub-OADEGs

In the GSE12021 dataset, correlation coefficients were recorded as 0.39 between PER1 and MAP3K8, -0.59 between XIST and MAP3K8, and 0.42 between XIST and AGMAT (Supplementary Fig. S5A). For the validation set GSE55235, the coefficients were 0.69 for PER1 and MAP3K8, -0.73 for XIST and MAP3K8, and -0.77 for MAP3K8 and AGMAT (Supplementary Fig. S5B). The positions of the four hub genes on human chromosomes are described with a circle map in (Supplementary Fig. S5C), in which MAP3K8 is located on chromosome 10, PER1 is located on chromosome 17, XIST is located on chromosome X, and AGMAT is located on chromosome 1.

Clusters A and B were identified using a consistent clustering method based on the four Hub-OADEGs. Supplementary Fig. S6A shows the grouping results. Supplementary Fig. S6B shows the consistency index for the different categories. As shown in the figure, the rising slope dropped the lowest when the samples were divided into two categories, indicating that the two categories are more reasonable. Supplementary Fig. S6C is a gravel map for the best classification parameters. To demonstrate the stability of the consistent clustering, we validated the two subtypes of OA samples based on the GSE12021 dataset using PCoA, which showed relatively high stability (Supplementary Fig. S6D).

Analysis of differential gene expression between Clusters A and B revealed 3560 genes, with 508 genes upregulated in Cluster A and 869 in Cluster B (Supplementary Fig. S6E). GO enrichment analysis demonstrated that Cluster A predominantly featured enrichment in biological processes such as regulation of protein aggregation and oxidative phosphorylation, cellular components including *tertiary granule*, and molecular functions like NADH dehydrogenase (ubiquinone) activity (Supplementary Table S7, Supplementary Fig. S6F).

3.5. Differences in immune characteristics between subtypes of OA

Utilizing the GSE12021 dataset, we assessed 22 types of immune cell infiltrations across the two OA subtypes through CIBERSORT analysis (Fig. 3A). The results indicated significantly elevated levels of immature CD8T cells, CD8T cells, regulatory T cells, and resting dendritic cells in Cluster A. In contrast, Cluster B showed higher quantities of naive CD4 T cells, memory CD4 T cells, activated dendritic cells, and resting mast cells (Fig. 3B). The thermograph displaying these findings is shown in Fig. 3C. Further correlation analysis linked the Hub-OADEGs with various immune cell types; notably, PER1 was negatively correlated with activated dendritic cells ($r = -0.74$, $p < 0.001$) (Fig. 3D) and positively correlated with CD8T cells ($r = 0.77$, $p < 0.001$) (Fig. 3E) and activated NK cells ($r = 0.76$, $p < 0.001$) (Fig. 3F). Additionally, a negative correlation existed between PER1 and M2 macrophage content ($r = -0.74$, $p < 0.001$) (Fig. 3G).

Results showed that Cluster A had more activated CD8 T cells, central memory CD8 T cells, effector memory CD4 T cells, and effector memory CD8 T cells than Cluster B ($P < 0.05$) (Supplementary Fig. S7A), when comparing the levels of immune cell infiltration between the two OA subtypes. Significant associations were found between the content of ICs and all four Hub-OADEGs ($P < 0.05$) (Supplementary Fig. S7B).

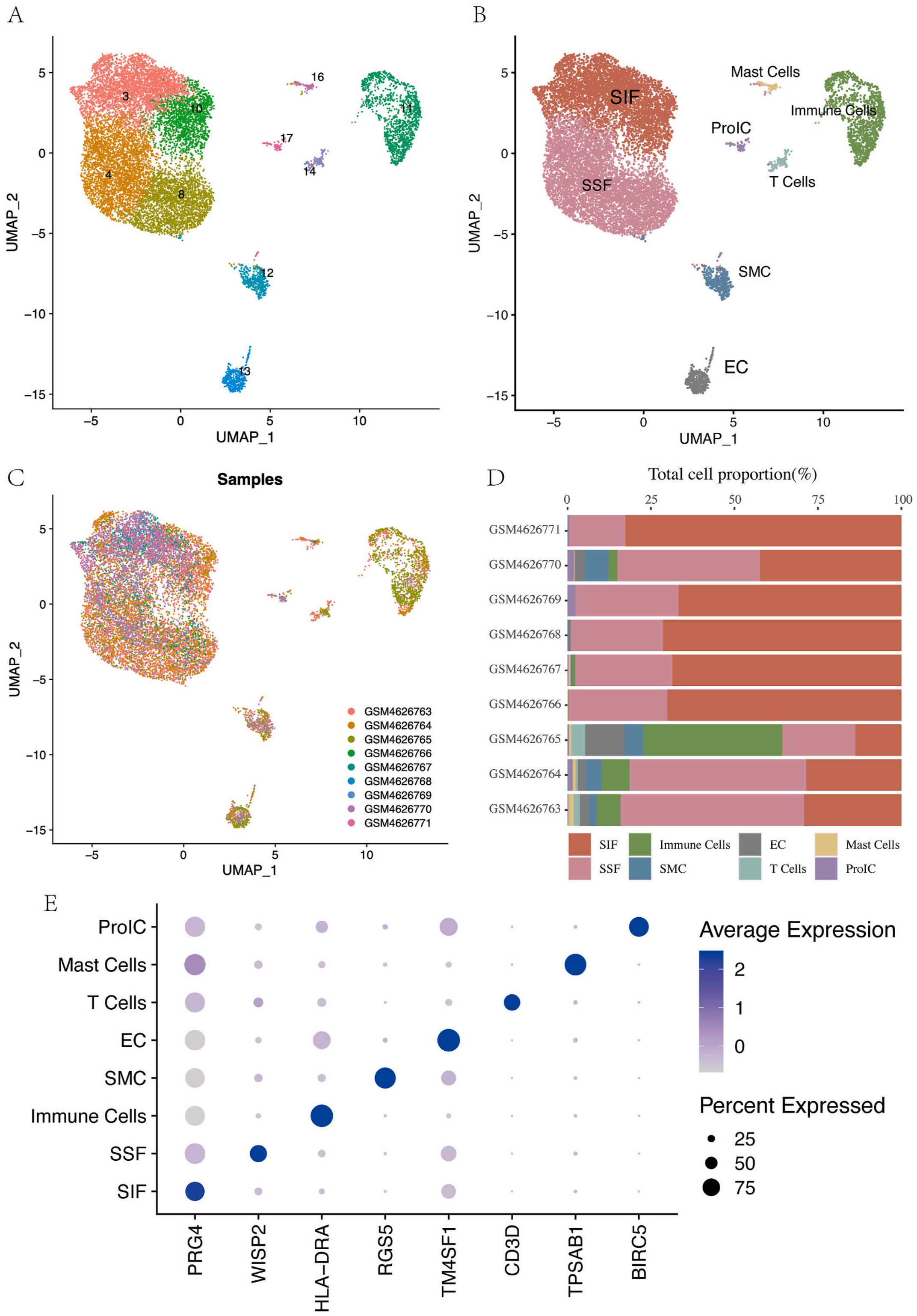
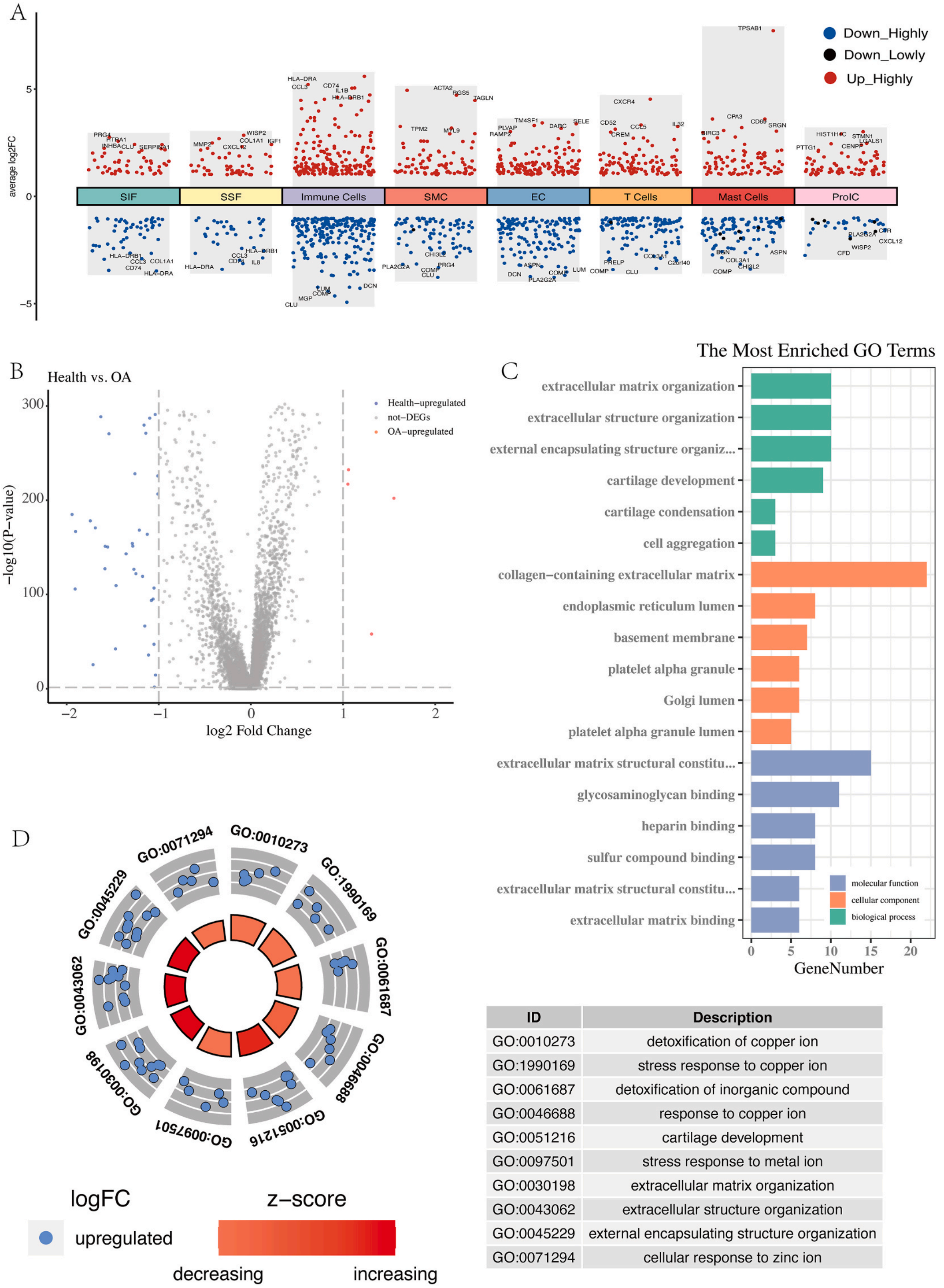


Fig. 4. scRNA-seq revealed the complexity of OA. (A) Based on the single cell dataset (GSE152805), cells were clustered into 10 clusters through UMAP. (B) Six cell types were annotated through singleR and sample sources. (C) The UMAP diagram shows the distribution of 9 samples. (D) The bar chart displays the proportion of each cell type in 9 samples. (E) The heat map displays the expression of genes specifically expressed for each cell type in the cell, with circle size representing the average expression value of characteristic genes.



ID	Description
GO:0010273	detoxification of copper ion
GO:1990169	stress response to copper ion
GO:0061687	detoxification of inorganic compound
GO:0046688	response to copper ion
GO:0051216	cartilage development
GO:0097501	stress response to metal ion
GO:0030198	extracellular matrix organization
GO:0043062	extracellular structure organization
GO:0045229	external encapsulating structure organization
GO:0071294	cellular response to zinc ion

Fig. 5. Heterogeneity analysis for OA in GSE152805. (A) Analysis of differences in cell subpopulations related to different OA; red, high expression genes in this cell subpopulation; blue, low expression genes in this cell subpopulation; (B) Comparing the DEGs of patients with OA in healthy samples; (C) GO enrichment analysis of highly expressed genes in OA; (D) Analysis of differences in GO enrichment scores related to OA. GO, Gene Ontology.

3.6. Interaction network of mRNA-miRNA or mRNA-TF for Hub-OADEGs

Networks for mRNA-miRNA and mRNA-TF were established based on the four Hub-OADEGs. The mRNA-TF network encompassed 198 interactions involving 156 TFs, with MAP4K8 and PER1 engaging with 79 and 58 TFs, respectively (Supplementary Fig. S8A). The network of mRNA-miRNA included 192 interactions with 188 miRNAs, where PER1 was linked to 111 miRNAs and AGMAT to 64 miRNAs (Supplementary Fig. S8B).

3.7. Quality control of single-cell data using seurat

Using the Seurat R package (version 4.0.2), we created objects for additional analysis by importing a count matrix for nine samples from the single-cell dataset. Following initial quality control by Cell Ranger, we conducted further quality assessment, retaining cells with over 250 genes, more than 500 UMIs, and a log10GenesPerUMI above 0.8, mitochondrial UMI ratio below 20 %, and red blood cell gene ratio below 5 % as high-quality cells. The “DoubletFinder” R package was used to eliminate doublets, resulting in 35,737 cells. The “NormalizeData” function with the “LogNormalize” method standardized the sequencing depth for dataset GSE152805. We utilized the “vst” method to identify 2000 variable features by activating the “FindVariableFeatures” function. To reduce the effect of sequencing depth, the data was scaled using the “ScaleData” function. The ElbowPlot function was used by PCA to identify important principal components. The top 20 components were then selected for Uniform Manifold Approximation and Projection based on their statistical significance. The variables’ genes were used as inputs.

3.8. Cluster analysis and cell type annotation of single-cell data

After rigorous quality control, 35,737 cells from the single-cell OA dataset were successfully categorized into ten distinct clusters using UMAP for dimensionality reduction and visualization (Fig. 4A). Cells were annotated into eight cell types as shown in (Fig. 4B): Cluster 3 and Cluster 10 were annotated as SIFs (5032,36.93 %); Clusters 7 and 19 were annotated as B cells (2109, 5.37 %); Clusters 10 and 21 were annotated as EC (1306, 3.33 %); Clusters 4 and 8 were annotated as SSF (5966, 43.78 %); Cluster 11 was annotated as ICs (1307, 9.59 %); Cluster 12 was annotated as SMC (465,3.41 %); Cluster 13 was annotated as EC (449, 3.29 %); Cluster 14 was annotated as T cells (169, 1.24 %); Cluster 16 was annotated as mast cells (122, 0.90 %); Cluster 17 was annotated as ProIC (117, 0.86 %). Fig. 4C shows the distribution of cell subtypes across the samples. The proportion of each cell type per sample was determined; SIFs constituted a significant proportion in each sample (Fig. 4D). Key marker genes for the various cell types, identified through the Cellmarker database and publications by Chou et al. [7] are displayed in the form of dot graphs: SIF (PRG4), (WISP2), ICs (HLA-DRA), SMC (RGS5), EC (TM4SF1), T cells (CD3D), mast cells (TPSAB1), and ProIC (BIRC5) (Fig. 4E).

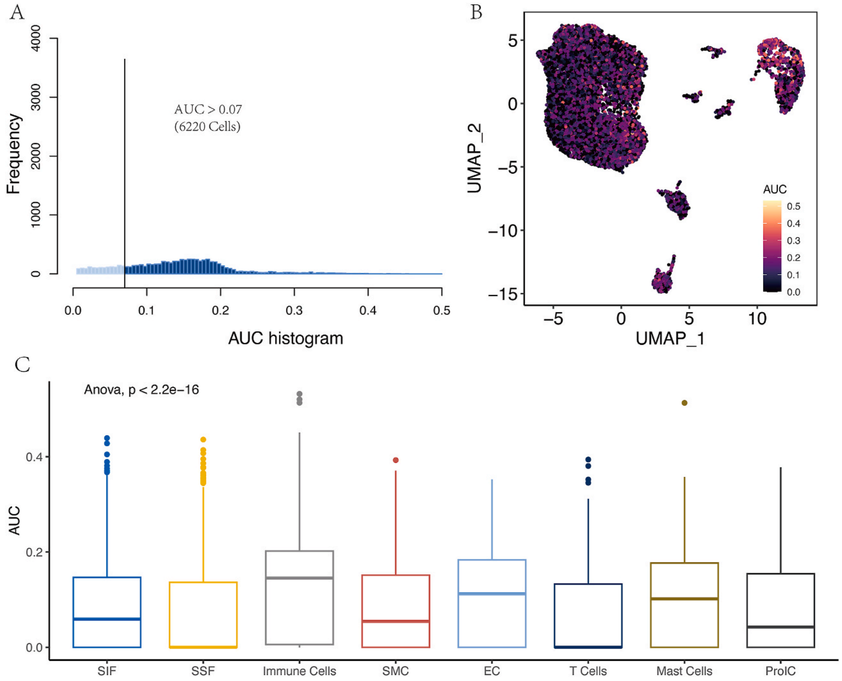


Fig. 6. Hub OA-related gene score in GSE152805. (A) Score of four important OA characteristic genes: AGMAT, MAP3K8, PER1, and XIST. 6220 cells exceed the threshold of 0.07. (B) UMAP is plotted based on the scores of four important OA characteristic genes, namely AGMAT, MAP3K8, PER1, and XIST, for each cell; (C) Area under curve (AUC) box plot of the scores of four important OA characteristic genes, AGMAT, MAP3K8, PER1, and XIST, in different cell subtypes.

3.9. Heterogeneity analysis of cell subsets in OA

We analyzed the differential genes among various subsets of patients with OA in the GSE152805 data set using the "Findallmarkers" function, the results of which showed that genes PRG4 and HTRA1 were overexpressed, HLA-DRB1 and CCL3 low-expressed in SIF; Genes MMP2 and WISP2 were overexpressed, whereas HLA-DRA and CCL3 were low-expressed in SSF; HLA-DRA and CCL3 were

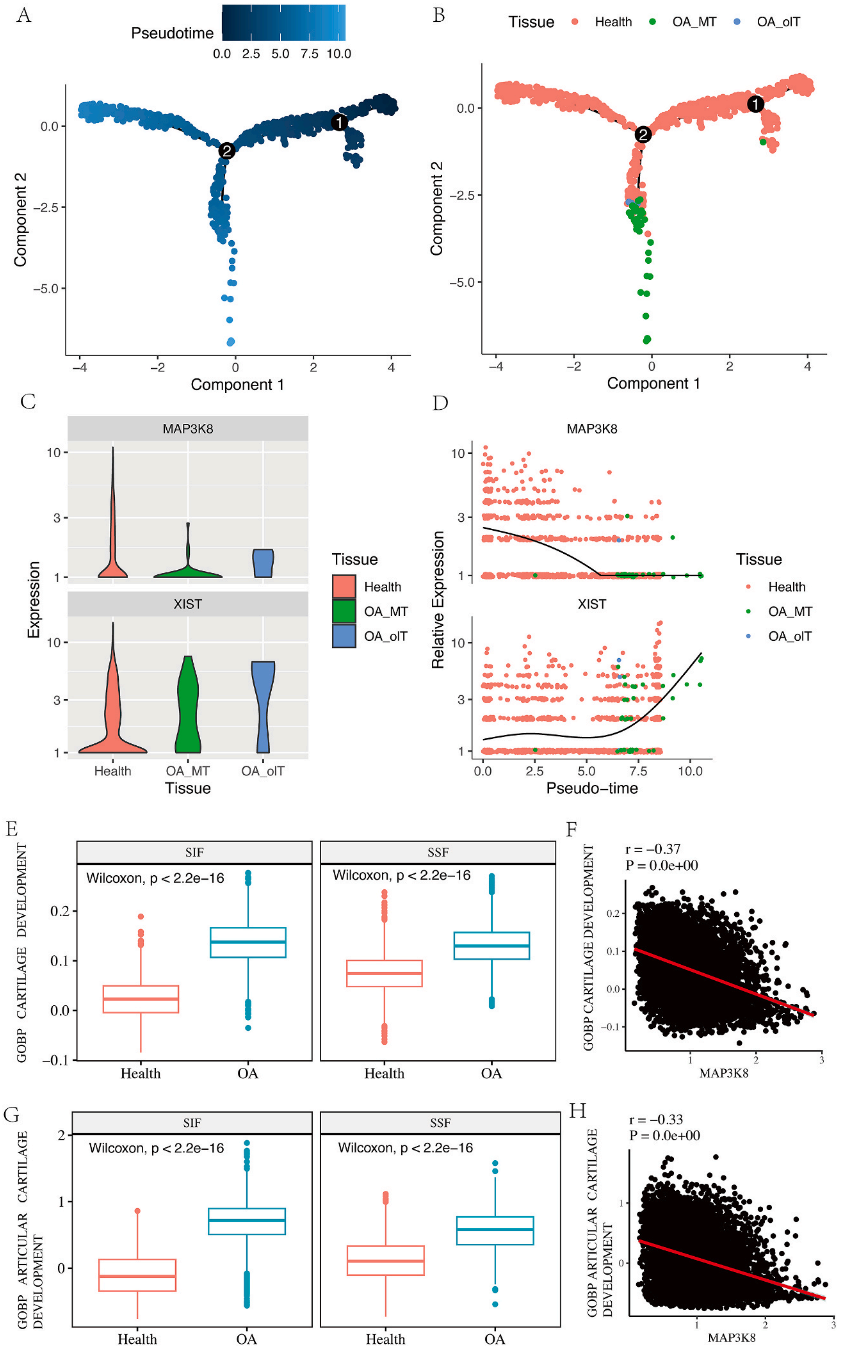


Fig. 7. Monocle analysis in GSE152805. (A) Pseudo time series analysis (time process); (B) Quasi temporal analysis (different disease processes); (C) The expression levels of MAP3K8 and XIST in the quasi temporal process; (D) Changes in the expression levels of MAP3K8 and XIST over time; (E) The enrichment differences of the cartilage development pathway in different disease types of OA and Health groups; Green color represents OA; Red color represents Health. (F). The relationship between MAP3K8 expression level and cartilage development pathway score; (G) Enrichment differences of the articular cartilage development pathway in different disease types of OA and normal samples; Geen color represents OA; Red color represents Health. (H). The relationship between MAP3K8 expression level and articular cartilage development pathway score; OA, osteoarthritis.

overexpressed, where CLU and DCN were low-expressed in ICs; ACTA2 and RGS5 were overexpressed, whereas CLU and COMP were low-expressed in SMC; TM4SF1 and SELE were overexpressed, whereas PLA2G2A and DCN were low-expressed in EC; CXCR4 and CD52 were overexpressed, whereas CLU and C2orf40 were low-expressed in T cells; CPA3 and CD69 were overexpressed, whereas CHI3L2 and ASPN were low-expressed in mast cells; PTTG1 and STMN1 were overexpressed, whereas CFD and WISP2 were low-expressed in ProIC (Fig. 5A, Supplementary Table S8). We analyzed the heterogeneity between OA and healthy samples from a single-cell perspective in the GSE12021 dataset and found that TIMP1, COL3A1, MSMP, and IL11 were highly expressed in OA samples (Fig. 5B). The GO enrichment analysis showed that the ECM pathway (Fig. 5C) was highly enriched in OA, whereas pathways such as detoxification of copper ion and stress response to copper ion were also activated in patients with OA (Fig. 5D).

3.10. Expression score of Hub-OADEGs in OA cell subsets

To explore autophagic expression in OA patients, we utilized the “AUCell” R package to analyze the functional activity and high expression gene proportion in each cell using four key Hub-OADEGs. This analysis revealed two prominent AUC peaks; by applying a threshold of 0.07, we observed that 6220 cells displayed relatively elevated AUC values (Fig. 6A). These cells were mainly distributed in synovial membrane fibroblasts and ICs (Fig. 6B). Compared with subsynovial fibroblasts, AGMAT, MAP3K8, PER1, and XIST were mainly highly expressed in synovial fibroblasts (Fig. 6C, $P < 0.001$).

3.11. Mesocellular evolution and pseudotemporal analysis of OA cells

In this study, a pseudotemporal analysis was conducted using the GSE152805 dataset to construct developmental trajectory maps of healthy samples (health), medial tibial (MT), platform cartilage damage (OA MT), and lateral tibial (oLT) platform cartilage damage (OA OIT) in pseudotime. We colored the pseudo-time series diagram from two aspects: the pseudo-time process (Fig. 7A) and sample phenotype (Fig. 7B). We found that the pseudo-time trajectory developed from a healthy phenotype along the medial tibial (MT) plateau cartilage injury (OA MT) to the lateral tibial (oLT) plateau cartilage injury (OA oIT). Next, we analyzed the relationships between four genes (AGMAT, MAP3K8, PER1, and XIST) and OA progression. Comparative analyses demonstrated that MAP3K8 expression was significantly higher in healthy versus OA-affected samples (Fig. 7C, top), and this expression decreased progressively as OA progressed (Fig. 7D, bottom). In a similar pattern, XIST expression was also higher in healthy samples than in those affected by OA (Fig. 7C, bottom). As OA progressed, XIST expression gradually decreased (Fig. 7D). Next, when comparing healthy samples with OA samples, cartilage development was highly enriched in both synovial and subsynovial fibroblasts ($P < 0.001$, Fig. 7E). Articular cartilage development was highly enriched in OA samples, both in synovial and subsynovial fibroblasts ($P < 0.001$, Fig. 7G). There was a significant negative correlation between MAP3K8 expression and both general cartilage ($r = -0.37$, $P < 0.001$, Fig. 7F) and articular

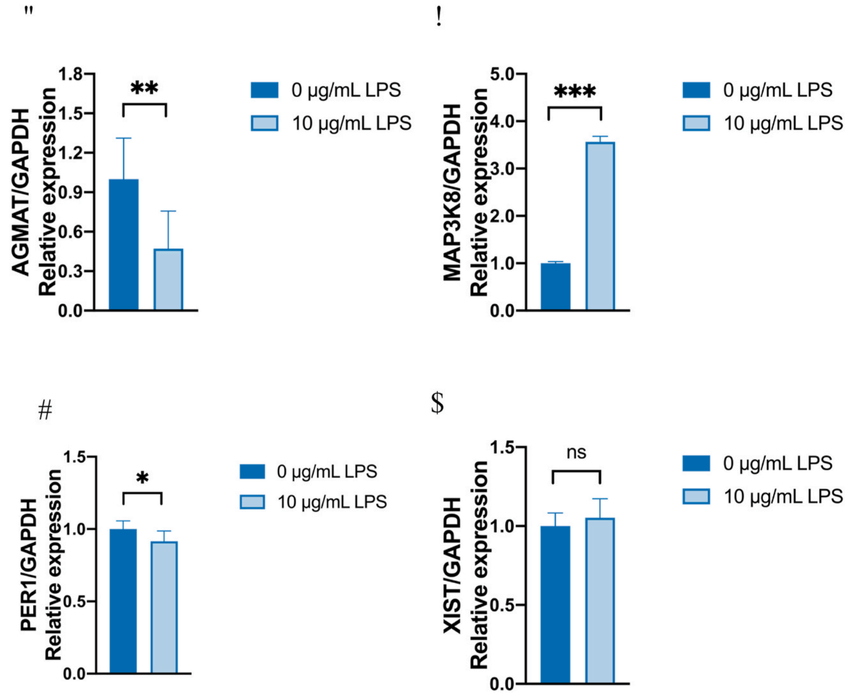


Fig. 8. qRT-PCR results show that the expression levels of key genes in OA cells and normal cartilage tissue cells （A）AGMAT was significantly reduced($P < 0.05$) （B）The total expression of MAP3K8 was significantly increased in OA cells ($P < 0.05$) (C)PER1was significantly reduced($P < 0.05$) (D)XIST was no significant expression differences in cells.

cartilage development (r = −0.33, P < 0.001, Fig. 7H).

3.12. Analysis of communication between different cell types

Utilizing “CellChat” for quantitative and visual analysis of cellular communication among eight cell types, we mapped the interaction intensity and frequency using heat maps and circular charts (Supplementary Figs. S9A and S9B). Notably, interactions between ProIC and cell types such as SMC and EC were markedly intense. Through an in-depth analysis of the cell-cell communication network, we pinpointed the collagen and FN1 pathways as the foremost influences on the reception and emission signals of SIF in synovial fibroblasts, respectively (Supplementary Fig. S9C). Therefore, we used a bubble diagram to illustrate the cellular communication received (Supplementary Fig. S9D) or emitted by SIFs via the collagen and FN1 pathways. Among the cellular communications received by SIF, the number of fibroblasts communicating through the collagen pathway is the highest, and they communicated with CD44 receptor ligands through COL6A2. Among the communications emitted by SIF, they mainly communicate with each other through FN1 and CD44 receptor ligand pairs.

3.13. Laboratory validation of key genes in OA

We utilized the qPCR technique to assess the relative expression of pivotal genes in OA cells compared to normal cartilage tissue cells (Fig. 8A–D). Expression of MAP3K8 markedly escalated in OA cells (P < 0.05), whereas levels of AGMAT/PER1 substantially declined (P < 0.05).

4. Discussion

OA is a widespread joint disorder predominantly affecting the knees, severely impairing quality of life and potentially resulting in disability [43]. Owing to the lack of effective interventions to prevent or reverse its progression, individuals with advanced disease frequently necessitate joint replacement surgery [44]. Therefore, identifying characteristic biological diagnostic markers is crucial for early diagnosis of the disease as well as for understanding its pathogenesis. Synovial tissue is the connective tissue membrane of bones and joints that nourishes, maintains, and protects the normal function of joints, together with the synovial fluid. Synovitis is a key factor in the initiation and pathological processes of OA and leads to pain and clinical symptoms [45]. Therefore, genetic research on synovial tissue samples is important for exploring the pathogenesis of OA.

In our study, aiming to pinpoint genes for the early detection of OA, we evaluated synovial tissue samples from two datasets, GSE12021 and GSE55235, encompassing 53 samples (30 OA and 23 normal). This analysis led to the identification of 45 DEGs linked to OA onset. We applied six machine-learning algorithms to perform dimensionality reduction on the 45 DEGs mentioned above. By obtaining the intersection of the results of the six algorithms, we identified four Hub-OADEGs: *AGMAT, MAP3K8, PER1,* and *XIST*. Diagnostic analysis found that in both the training and validation sets, these four genes had high diagnostic value and could independently predict OA, serving as molecular markers for the early diagnosis of OA. We also validated the expression of the four genes in cell experiment, which showed difference in OA comparing to control, except *XIST.* That could be a performance bias due to the small sample of our cell experiment.

Yang et al. [46] uncovered the circRSU1-miR-93-5p-MAP3K8 axis, a regulator of OA progression via oxidative stress modulation, posing as a viable target for OA treatment. In parallel, Kanbe et al. [47] reported that circadian genes such as Per1 and Per 2 exhibit differential expression in chondrocytes, significantly influencing their metabolic rhythm in OA contexts. These findings emphasize the critical role of these four genes in OA, marking their diagnostic relevance as a novel discovery in our study.

To delve deeper into the role of four pivotal Hub-OADEGs in OA progression, we applied an unsupervised clustering approach to organize 33 synovial tissue samples into two groups (Cluster A and Cluster B) based on these genetic markers. Examination of functional enrichment within these clusters indicated that Cluster A was primarily associated with biological processes like protein catabolism and the regulation of oxidative phosphorylation, as well as with respiratory chain complexes. It also showed enrichment in cellular components such as the inner mitochondrial membrane protein complex and tertiary granules, along with enzymatic functions including NADH dehydrogenase (ubiquinone), oxidoreductase, and NADH dehydrogenase (quinone) activities acting on NAD(P)H. Therefore, we speculate that Clusters A and B represent two different stages of OA development. Patients in Cluster A were in the early stage of OA, whereas those in Cluster B were in the intermediate or later stages. In the initial stages, OA is characterized by the progressive degradation and destruction of cartilage tissue [48], with inflammatory responses heightening the activity of proteolytic enzymes that expedite the breakdown of cartilage cells’ ECM. These characteristics are consistent with the biological functions prevalent in Cluster A.

Immune mechanisms play a vital role in both the initiation and progression of OA. Our comparative analysis of immune profiles between the clusters showed that Cluster A exhibited enhanced levels of innate immune responses, including higher numbers of neutrophils, activated NK cells, resting dendritic cells, and naive B cells, which are indicative of early immune activity in OA progression. The primary pathological alterations in the OA synovium are driven largely by a mix of inflammatory and anti-inflammatory cytokines. During this stage, the innate immune system takes charge as the first line of defense against pathogens. It helps eliminate damaged and aging cells and coordinates the activation and control of targeted immune responses [49]. The innate immune system is mainly regulated by a barrier composed of monocyte macrophages, natural killer cells, T lymphocytes, complement cells, and a series of cytokines that are involved in local inflammatory reactions and tissue repair. In later stages of OA, as depicted by Cluster B, the immune landscape is dominated by the activation of dendritic cells, M2 type macrophages, and resting mast cells, which contribute to the autoimmune aspects of OA. During this phase, the body releases inflammatory chemicals that disrupt immunological homeostasis and speed up the breakdown of connective tissues including bone and cartilage [50]. Cluster B's higher concentrations of CD4 naïve and CD4 memory resting T cells suggest an active adaptive immune response. This response is marked by the permeabilization of the ECM immune barrier and the presentation of surface antigens specific to chondrocytes, which in turn activate related T cells within the immune repertoire. During this phase, T cells emit lymphokines to neutralize antigens, which concurrently inflicts damage and intensifies OA progression [47]. These observations corroborate the aforementioned insights.

ssGSEA identified disparities in the distribution of ICs between Clusters A and B, noting particularly that γδ T cells were less abundant in Cluster A than in Cluster B. Recent research suggests that γδ T cells play roles in age-related conditions, such as OA [51, 52]. Faust et al. [53] noted that these cells are crucial for IL-17 production, which is linked to the senescence-related response to injury in OA. Further research has confirmed diminished levels of γδ T cells in individuals with OA [54]. We speculate that the role of T cells varies at different stages of OA. The complicated mechanism of γδ T cells in OA progression requires further exploration. We attempted to identify evidence from the PPI protein network and found that the four genes interacted with 188 miRNAs and 156 TFs. All four genes are important genes that affect OA. Among these, MAP3K8 and PER1 were the most strongly associated with other small molecules.

We further analyzed cell typing using a scRNA-seq dataset and found that OA cells could be clustered into eight cell types, with the highest proportion being SIF and SSF. In Wang X et al.'s study [55], single-cell analysis was also used to analyze the cell distribution of cartilage tissue in healthy individuals. It was found that chondrocytes such as fibrochondrocytes and regulatory chondrocytes accounted for the vast proportion of cells. In our study, we did not find any clustering of chondrocytes, which might be related to our study of synovial tissue, where there were fewer chondrocytes but a higher proportion of SIF and SSF. Synovial fibroblasts, essential for normal joint function, are key in the initiation and progression of OA. When activated, these cells secrete cytokines, chemokines, and enzymes that degrade the matrix, thus promoting inflammatory cell recruitment, expediting the breakdown of the cartilage matrix, and ultimately leading to the destruction of joint cartilage and bone [56]. The cell evolution and pseudo-temporal analyses in this research found a marked enhancement in cartilage development processes in both SIF and SSF, consistent with earlier studies.

Enrichment analyses of cellular pathways revealed the activation of "Detoxification of copper ion" and "Stress response to copper ion" in OA patients, indicating a potential role for copper ion metabolism in the progression of OA. Research examining levels of five metal ions in the blood of OA patients showed that high copper ion concentrations were associated with a greater likelihood of developing OA, aligning with our observations [57]. Yet, further research is needed to elucidate the specific pathways involved.

We also found that the four Hub-OADEGs were abnormally expressed in SIF, which implied that the four genes might be involved in the differentiation of cells. Our pseudo-temporal series analysis further demonstrated that the four Hub-OADEGs are intimately linked with the progression of OA. As the disease progressed, their expression levels exhibited either an increase or a decrease. In addition, we confirmed the crucial involvement of these genes in OA formation by finding a negative association between MAP3K8 expression and the "cartilage development" pathway.

We also validated the expression of the four genes in cell experiment, which showed the same expression trends as the bioinformatics except *XIST*. That could be a performance bias due to the small sample of our cell experiment, or differences in experimental conditions. Nevertheless, the validation of *AGMAT, MAP3K8, PER* demonstrated that these genes are involved in the progress of OA and might become treatment targets.

The novelty of this study is reflected in the following aspects. To our knowledge, this study marks the inaugural application of the CIBERSORT and ssGSEA algorithms to systematically examine differences in immune cell infiltration among various OA subtypes, thereby shedding light on the immunological mechanisms underlying OA. We identified four key hub genes that were significantly associated with OA (Hub-OADEGs: AGMAT, MAP3K8, PER1, and XIST) and investigated their expression patterns and potential functions in various ICs. By integrating scRNA-seq data, we provided detailed annotation and analysis of different cell types in patients with OA, revealing the high heterogeneity and complexity of OA cells. Furthermore, using CellChat, we analyzed intercellular communication and explored the crucial role of synovial fibroblasts in OA progression. These findings offer new perspectives for the targeted therapy of OA. The study faces a number of limitations. The modest sample size could undermine the statistical robustness of the results. Additionally, reliance on publicly available data sources introduces risks related to sample heterogeneity and potential biases in the data. The analysis of immune cell infiltration using the CIBERSORT and ssGSEA algorithms also carries inherent assumptions and constraints, necessitating further validation with larger, independent datasets. Moreover, the bioinformatics outcomes of this research require additional experimental work to confirm the roles of the identified hub genes and their impact on immune cell infiltration within OA contexts.

Future studies should focus on several key areas. It is imperative to increase the sample size to strengthen the statistical validity and applicability of the results. Experimental studies are crucial to ascertain the precise functions of the hub genes and ICs identified as significant in OA. Moreover, in-depth exploration of the regulatory mechanisms of these hub genes and their dynamic changes during OA progression is warranted. Utilizing more advanced single-cell sequencing technologies to investigate cell communication and signaling pathway changes at different stages of OA, reveal the underlying pathophysiological mechanisms. In conclusion, this study introduces a new perspective on the immunological aspects of OA, and it is expected that subsequent detailed investigations will enhance the guidance available for clinical diagnosis and treatment.

5. Conclusions

AGMAT, MAP3K8, PER1, and *XIST* could serve as early molecular diagnostic markers for OA, and these four genes could independently and accurately predict the occurrence of OA. Based on these four genes, which play regulatory roles in the differentiation of the subtypes, OA samples can be clustered into different subtypes, Moreover, these four genes are closely associated with the progression of OA.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This work was supported by the National Natural Science Foundation of China [grant number: 82072472].

Data availability statement

The datasets used during the present study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Enming Yu: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Mingshu Zhang:** Visualization, Validation, Software, Resources, Methodology, Investigation. **Chunyang Xi:** Visualization, Validation, Methodology, Formal analysis, Data curation. **Jinglong Yan:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37047.

References

[1] B. Abramoff, F.E. Caldera, Osteoarthritis: pathology, diagnosis, and treatment options, Med. Clin. 104 (2) (2020) 293–311, https://doi.org/10.1016/j.mcna.2019.10.007.

[2] A. Colletti, A.F.G. Cicero, Nutraceutical approach to chronic osteoarthritis: from molecular research to clinical evidence, Int. J. Mol. Sci. 22 (23) (2021) 12920, https://doi.org/10.3390/ijms222312920. Published 2021 Nov 29.

[3] H. Long, Q. Liu, H. Yin, et al., Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the global burden of disease study 2019, Arthritis Rheumatol. 74 (7) (2022) 1172–1183, https://doi.org/10.1002/art.42089.

[4] S.L. Kolasinski, T. Neogi, M.C. Hochberg, et al., 2019 American College of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee [published correction appears in arthritis care res (hoboken). 2021;73(5):764], Arthritis Care Res. 72 (2) (2020) 149–162, https://doi.org/10.1002/acr.24131.

[5] Z.W. Sun, M.Y. Yan, J.J. Wang, et al., Single-cell RNA sequencing reveals different chondrocyte states in femoral cartilage between osteoarthritis and healthy individuals, Front. Immunol. 15 (2024) 1407679, https://doi.org/10.3389/fimmu.2024.1407679.

[6] K.B. Hu, Y.H. Ou, L.Y. Xiao, et al., Identification and construction of a disulfidptosis-mediated diagnostic model and associated immune microenvironment of osteoarthritis from the perspective of PPPM, J. Inflamm. Res. 17 (2024) 3753–3770, https://doi.org/10.2147/JIR.S462179.

[7] M. Geyer, C. Schönfeld, Novel insights into the pathogenesis of osteoarthritis, Curr. Rheumatol. Rev. 14 (2) (2018) 98–107, https://doi.org/10.2174/1573397113666170807122312.

[8] V. Molnar, V. Matišić, I. Kodvanj, et al., Cytokines and chemokines involved in osteoarthritis pathogenesis, Int. J. Mol. Sci. 22 (17) (2021) 9208, https://doi.org/10.3390/ijms22179208.

[9] X. Hu, S. Ni, K. Zhao, J. Qian, Y. Duan, Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates, Front. Immunol. 13 (2022) 871008, https://doi.org/10.3389/fimmu.2022.871008.

[10] C.H. Chou, V. Jain, J. Gibson, et al., Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis, Sci. Rep. 10 (1) (2020) 10868, https://doi.org/10.1038/s41598-020-67730-y. Published 2020 Jul 2.

[11] R. Huber, C. Hummert, U. Gausmann, et al., Identification of intra-group, inter-individual, and gene-specific variances in mRNA expression profiles in the rheumatoid arthritis synovial membrane, Arthritis Res. Ther. 10 (4) (2008) R98, https://doi.org/10.1186/ar2485.

[12] D. Woetzel, R. Huber, P. Kupfer, et al., Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation, Arthritis Res. Ther. 16 (2) (2014) R84, https://doi.org/10.1186/ar4526.

[13] M.E. Ritchie, B. Phipson, D. Wu, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007.

[14] A. Robert, J. Michael, et al., The string-to-string correction problem, J. ACM 21 (1) (1974) 168–173.

[15] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7.

[16] A. Ameri, E.J. Scheme, K.B. Englehart, P.A. Parker, Bagged regression trees for simultaneous myoelectric force estimation, in: In2014 22nd Iranian Conference on Electrical Engineering (ICEE), IEEE, 2014 May 20, pp. 2000–2003.

[17] R. van de Schoot, S. Depaoli, R. King, et al., Bayesian statistics and modelling, Nature Reviews Methods Primers 1 (1) (2021) 1–14.

[18] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogrammetry Remote Sens. 114 (2016) 24–31.

[19] M.A. Hall, L.A. Smith, Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper, InFLAIRS conference (1999: 1999) 235–239.

[20] Machado MD, Schirru R. A New Evolutionary Algorithm with LQV Learning for Combinatorial Problems Optimization.

[21] H. Wang, Q. Xu, L. Zhou, Large unbalanced credit scoring using lasso-logistic regression ensemble, PLoS One 10 (2) (2015) e0117844.

[22] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, H. Pfister, UpSet: visualization of intersecting sets, IEEE Trans. Vis. Comput. Graph. 20 (12) (2014) 1983–1992, https://doi.org/10.1109/TVCG.2014.2346248.

[23] H. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots, BMC Bioinf. 14 (2013) 244, https://doi.org/10.1186/1471-2105-14-244.

[24] A. Iasonos, D. Schrag, G.V. Raj, K.S. Panageas, How to build and interpret a nomogram for cancer prognosis, J. Clin. Oncol. 26 (8) (2008) 1364–1370, https://doi.org/10.1200/JCO.2007.12.9791.

[25] C. Dawkins, T.N. Srinivasan, J. Whalley, InHandbook of Econometrics, vol. 5, Elsevier, Calibration, 2001, pp. 3653–3703.

[26] X. Robin, N. Turck, A. Hainard, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77. Published 2011 Mar 17.

[27] M. Fitzgerald, B.R. Saville, R.J. Lewis, Decision curve analysis, JAMA 313 (4) (2015) 409–410.

[28] E.F. Lock, D.B. Dunson, Bayesian consensus clustering, Bioinformatics 29 (20) (2013) 2610–2616, https://doi.org/10.1093/bioinformatics/btt425.

[29] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (12) (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.

[30] J.C. Gower, Principal Coordinates Analysis, Wiley StatsRef: Statistics Reference Online, 2014, pp. 1–7.

[31] M.A. Harris, J. Clark, A. Ireland, et al., The Gene Ontology (GO) database and informatics resource, Nucleic Acids Res. 32 (2004) D258–D261, https://doi.org/10.1093/nar/gkh036.

[32] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.

[33] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[34] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259, https://doi.org/10.1007/978-1-4939-7493-1_12.

[35] H. Wickham, Data Analysis. Ggplot2: Elegant Graphics for Data Analysis, 2016, pp. 189–201.

[36] R. Kolde, M.R. Kolde, Package 'pheatmap', R package 1 (10) (2018).

[37] C.S. McGinnis, L.M. Murrow, Z.J. Gartner, DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors, Cell Syst 8 (4) (2019) 329–337.e4, https://doi.org/10.1016/j.cels.2019.03.003.

[38] I. Korsunsky, N. Millard, J. Fan, et al., Fast, sensitive and accurate integration of single-cell data with Harmony, Nat. Methods 16 (12) (2019) 1289–1296, https://doi.org/10.1038/s41592-019-0619-0.

[39] H. Abdi, L.J. Williams, Principal component analysis, Wiley interdisciplinary reviews: Comput. Stat. 2 (4) (2010) 433–459.

[40] S. Aibar, C.B. González-Blas, T. Moerman, et al., SCENIC: single-cell regulatory network inference and clustering, Nat. Methods 14 (11) (2017) 1083–1086, https://doi.org/10.1038/nmeth.4463.

[41] S. Jin, C.F. Guerrero-Juarez, L. Zhang, et al., Inference and analysis of cell-cell communication using CellChat, Nat. Commun. 12 (1) (2021) 1088, https://doi.org/10.1038/s41467-021-21246-9. Published 2021 Feb 17.

[42] P. Perešíni, M. Kuźniar, D. Kostić, Monocle: dynamic, fine-grained data plane monitoring. InProceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies, 2015, pp. 1–13.

[43] A. Kumar, P. Palit, S. Thomas, et al., Osteoarthritis: prognosis and emerging therapeutic approach for disease management, Drug Dev. Res. 82 (1) (2021) 49–58, https://doi.org/10.1002/ddr.21741.

[44] M. Khazzam, A.O. Gee, M. Pearl, Management of glenohumeral joint osteoarthritis, J. Am. Acad. Orthop. Surg. 28 (19) (2020) 781–789, https://doi.org/10.5435/JAAOS-D-20-00404.

[45] A. Mathiessen, P.G. Conaghan, Synovitis in osteoarthritis: current understanding with therapeutic implications, Arthritis Res. Ther. 19 (1) (2017) 18, https://doi.org/10.1186/s13075-017-1229-9.

[46] Y. Yang, P. Shen, T. Yao, et al., Novel role of circRSU1 in the progression of osteoarthritis by adjusting oxidative stress, Theranostics 11 (4) (2021) 1877–1900.

[47] K. Kanbe, K. Inoue, C. Xiang, Q. Chen, Identification of clock as a mechanosensitive gene by large-scale DNA microarray analysis: downregulation in osteoarthritic cartilage, Mod. Rheumatol. 16 (3) (2006) 131–136.

[48] J.E. Woodell-May, S.D. Sommerfeld, Role of inflammation and the immune system in the progression of osteoarthritis, J. Orthop. Res. 38 (2) (2020) 253–257, https://doi.org/10.1002/jor.24457.

[49] E.B.P. Lopes, A. Filiberti, S.A. Husain, M.B. Humphrey, Immune contributions to osteoarthritis, Curr. Osteoporos. Rep. 15 (6) (2017) 593–600, https://doi.org/10.1007/s11914-017-0411-y.

[50] Y. Han, J. Wu, Z. Gong, et al., Identification and development of a novel 5-gene diagnostic model based on immune infiltration analysis of osteoarthritis, J. Transl. Med. 19 (1) (2021) 522, https://doi.org/10.1186/s12967-021-03183-9.

[51] B. Wen, M.N. Liu, X.Y. Qin, Z.Y. Mao, X.W. Chen, Identifying immune cell infiltration and diagnostic biomarkers in heart failure and osteoarthritis by bioinformatics analysis, Medicine (Baltim.) 102 (26) (2023) e34166, https://doi.org/10.1097/MD.0000000000034166.

[52] B. Fan, B. Fan, N. Sun, H. Zou, X. Gu, A radiomics model to predict gammadelta T-cell abundance and overall survival in head and neck squamous cell carcinoma, Faseb. J. 38 (5) (2024) e23529, https://doi.org/10.1096/fj.202301353RR.

[53] H.J. Faust, H. Zhang, J. Han, et al., IL-17 and immunologically induced senescence regulate response to injury in osteoarthritis, J. Clin. Invest. 130 (10) (2020) 5493–5507, https://doi.org/10.1172/JCI134091.

[54] B. Wen, M.N. Liu, X.Y. Qin, Z.Y. Mao, X.W. Chen, Identifying immune cell infiltration and diagnostic biomarkers in heart failure and osteoarthritis by bioinformatics analysis, Medicine (Baltim.) 102 (26) (2023) e34166, https://doi.org/10.1097/MD.0000000000034166.

[55] X. Wang, Y. Ning, P. Zhang, et al., Comparison of the major cell populations among osteoarthritis, Kashin-Beck disease and healthy chondrocytes by single-cell RNA-seq analysis, Cell Death Dis. 12 (6) (2021) 551, https://doi.org/10.1038/s41419-021-03832-3.

[56] B. Tu, R. Fang, Z. Zhu, et al., Comprehensive analysis of arachidonic acid metabolism-related genes in diagnosis and synovial immune in osteoarthritis: based on bulk and single-cell RNA sequencing data, Inflamm. Res. 72 (5) (2023) 955–970, https://doi.org/10.1007/s00011-023-01720-4.

[57] J. Zhou, C. Liu, Y. Sun, et al., Genetically predicted circulating levels of copper and zinc are associated with osteoarthritis but not with rheumatoid arthritis, Osteoarthritis Cartilage 29 (7) (2021) 1029–1035, https://doi.org/10.1016/j.joca.2021.02.564.