Iran. J. Biotechnol. April 2026;24(2): e4148 DOI: 10.30498/ijb.2025.524798.4148

Research Article

Single-Cell and Mendelian Randomization Analyses Identify Macrophage Polarization and Efferocytosis Biomarkers in Osteoarthritis and Reveal Their Regulatory Mechanisms

Hongtao Zhang †, Zhen Zhao†, Kui Xu, Zhenwei Ji, Chuan Dong, Hua Long, Shun Niu*

Department of Orthopedic, Tangdu Hospital of the Forth Military Medical University, Xi'an, 710038, China

**Corresponding author*: Shun Niu, Department of Orthopedic, Tangdu Hospital of the Forth Military Medical University, Xi'an, 710038, China. E-mail: 21379067@qq.com

† First and second authors contributed equally to this work.

Received: 2025/05/20; **Accepted:** 2025/10/19

Background: Osteoarthritis (OA) is a progressive joint disease influenced by macrophage-associated efferocytosis.

Objective: This study aimed to identify key OA biomarkers by integrating transcriptomic and single-cell RNA sequencing (scRNA-seq) data with Mendelian randomization (MR).

Materials and Methods: We integrated three OA-related transcriptomic datasets (GSE89408, GSE55457, and GSE152805) from synovial tissue with efferocytosis- and macrophage polarization-related genes (ERGs and MPRGs). Differentially expressed genes (DEGs) were identified and analyzed using gene set enrichment (ssGSEA), weighted gene co-expression network analysis (WGCNA), and functional enrichment (GO/KEGG). MR was performed with GWAS data to evaluate causal links between candidate genes and OA. Machine learning methods (LASSO and SVM-RFE), supported by ROC analysis and artificial neural network (ANN) modeling, and were used to screen potential biomarkers. Single-cell RNA-seq analysis was conducted with Seurat, incorporating clustering, functional enrichment, cell–cell communication (CellChat), and pseudotime trajectory analysis (Monocle) to characterize cellular heterogeneity and differentiation pathways in OA.

Results: MR and machine learning highlighted FMO4 and GPR65 as risk biomarkers, validated in independent datasets (AUC > 0.7). GPR65 showed strong associations with neutrophils, regulatory T cells, and antigen-presenting cell (APC) co-inhibition, while both genes were enriched in immune and metabolic pathways. Single-cell analysis identified synovial subintimal fibroblasts and HLA-DRA+ cells as key contributors with distinct differentiation patterns. An ANN model further improved OA classification.

Conclusion: These findings provide new insights into OA pathogenesis and support FMO4 and GPR65 as promising diagnostic and therapeutic targets.

Keywords: Efferocytosis; Macrophage polarization; Mendelian randomization; Osteoarthritis

1. Background

Osteoarthritis (OA) is a widespread degenerative joint disorder characterized by progressive cartilage breakdown, synovial inflammation, subchondral bone sclerosis, and osteophyte (bone spur) formation, ultimately leading to pain, stiffness, and impaired mobility (1). This condition affects approximately 500 million people worldwide, with nearly one-third of adults over the age of 65 exhibiting clinical manifestations (2).

The burden of OA continues to increase due to

aging populations and rising obesity rates, posing substantial challenges to healthcare systems (3). Current treatment strategies — including non-steroidal anti-inflammatory drugs (NSAIDs), physical therapy, lifestyle modifications, and joint replacement surgery — primarily aim to relieve symptoms but do not halt disease progression or reverse cartilage damage (4). Moreover, these treatments involve considerable financial costs and carry risks of adverse effects, underscoring the urgent need for improved biomarkers (5). A particular diagnostic challenge lies in OA's frequently asymptomatic early phase, during which irreversible cartilage degeneration and inflammatory processes progress unnoticed (6). Therefore, elucidating the pathogenic mechanisms of OA and identifying reliable biomarkers are essential for enabling timely interventions and the development of targeted therapies.

Macrophages are primary phagocytic cells with diverse biological functions, including pathogen clearance, antigen presentation, cytokine secretion, and tissue remodeling (7). They can polarize into either pro-inflammatory M1 or anti-inflammatory M2 phenotypes in response to various environmental stimuli, regulated by cytokines, microbial components, and growth factors (8). In osteoarthritic joints, synovial macrophages produce inflammatory mediators such as TNF-α, IL-1β, and IFN-γ, which promote M1 polarization and accelerate cartilage degradation (9). In contrast, IL-4- and IL-13-induced M2 macrophages mediate anti-inflammatory effects and facilitate tissue repair (10). Modulating macrophage polarization to favor the M2 phenotype has emerged as a promising therapeutic strategy in OA management. However, the precise mechanisms by which macrophage polarization contributes to OA pathogenesis remain poorly understood, necessitating further investigation.

Efferocytosis, the process by which macrophages engulf and remove apoptotic cells, plays a critical role in maintaining tissue homeostasis and resolving inflammation (11). Impaired clearance of apoptotic cells leads to the accumulation of cellular debris and the release of inflammatory mediators, thereby exacerbating synovial inflammation (12). The GAS6-TAM receptor signaling axis, a key pathway in efferocytosis, has been shown to play an important role in regulating inflammation associated with OA (13). However, the mechanistic interplay between macrophage polarization dynamics, efferocytosis efficiency, and OA progression remains insufficiently understood and warrants comprehensive investigation.

In addition to transcriptomic and single-cell analyses, we used Mendelian randomization (MR) analysis. MR is an epidemiological approach that uses genetic variants as instrumental variables to estimate the causal effects of molecular traits on clinical outcomes. As genotypes are determined at conception and not affected by environmental or behavioral confounders, MR provides an unbiased means to determine if candidate biomarkers are causally associated with OA (14). MR typically utilizes data from genome-wide association studies (GWAS) to associate genetic variants with specific exposures and outcomes (15). Recent MR studies have yielded valuable insights into the genetic determinants and potential therapeutic targets for OA. Alhassan et al. demonstrated that metabolic and inflammatory pathways — particularly body mass index (BMI) and lipid metabolism — play a causal role in OA development. Zou et al. identified plasma proteins such as MAPK3, GZMK, and ITIH1 as novel therapeutic targets for knee and hip OA through proteome-wide MR and colocalization analysis (16,17). Recent bioinformatics studies have reported novel OA biomarkers, including aging- and circadian rhythm-related genes, ferroptosis-associated immune regulators (18), and immune-infiltration-associated markers (HTRA1, DPT, MXRA5) (19). Few studies have specifically examined macrophage polarization- or efferocytosis-related biomarkers in OA (20,21). Therefore, further research is needed to clarify their molecular roles and identify potential diagnostic and therapeutic targets.

2. Objectives

This study aimed to elucidate the molecular mechanisms linking macrophage polarization and efferocytosis to OA pathogenesis and to identify potential causal biomarkers associated with these processes. Specifically, we integrated bulk transcriptomic datasets, single-cell RNA sequencing (scRNA-seq), and MR analyses to identify and validate key genes, including FMO4 and GPR65, as potential biomarkers of OA. Differential expression analysis, LASSO regression modeling, and MR were employed to provide functional and causal evidence, while machine learning algorithms evaluated diagnostic performance. Furthermore, scRNA-seq analysis was conducted to characterize relevant cellular

subtypes, explore differentiation trajectories, and delineate intercellular communication networks within osteoarthritic tissues.

3. Materials and Methods

3.1. Data Source

Three OA -related transcriptome datasets were utilized in this investigation; all obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The GSE89408 (platform: GPL11154) dataset included 22 OA and 28 control samples, while the GSE55457 (platform: GPL96) dataset comprised 10 OA and 10 control samples. Additionally, GSE152805 (scRNA-seq, platform: GPL20301) contained three OA samples. All three datasets were based on synovial tissue samples.

A total of 50 efferocytosis-related genes (ERGs) and 35 macrophage polarization-related genes (MPRGs) were collected from previously published literature (22, 23). Genome-wide association study (GWAS) datasets for exposure factors and OA (ebi-a-GCST005814) were retrieved from the IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). The OA dataset included 15,845,511 single nucleotide polymorphisms (SNPs) derived from 50,508 European participants.

Batch effects in bulk RNA-seq datasets were identified by PCA and corrected using the ComBat function in the sva R package. For scRNA-seq data, integration was performed with Seurat's canonical correlation analysis (CCA) and mutual nearest neighbor (MNN) alignment to reduce technical variability.

3.2. Differential Expression Analysis and Scoring Calculation

The GSE89408 dataset was analyzed using DESeq2 (v1.38.3) (24) to identify differentially expressed genes (DEGs) between OA and control samples, with thresholds set at $|\log_2 FC| > 0.5$ and adjusted *P*-value (P.adj) < 0.05. A volcano plot and heatmap were generated using ggplot2 (v3.4.2) (25) and circlize (v0.4.15) (26), respectively. The cutoff of |log2FC| > 0.5 was selected to capture biologically relevant but moderate changes in expression while maintaining sensitivity, whereas the adjusted $P < 0.05$ threshold ensured statistical robustness after multiple testing correction. Subsequently, the identified DEGs were intersected with efferocytosis-related genes (ERGs) and macrophage polarization-related genes (MPRGs), yielding differentially expressed ERGs (DE-ERGs) and differentially expressed MPRGs (DE-MPRGs). Next, single-sample gene set enrichment analysis (ssGSEA) was performed using the GSVA package (v1.48.3) (26), with DE-ERGs and DE-MPRGs as background gene sets. Differential expression analysis was performed using DESeq2, with multiple testing correction applied via the Benjamini–Hochberg false discovery rate (FDR) method. Genes with an adjusted *P*-value < 0.05 and |log2FC| > 0.5 were considered significantly differentially expressed. The resulting enrichment scores were then compared between OA and control samples from the GSE89408 dataset.

3.3. Weighted Gene Co-Expression Network Analysis (WGCNA)

The primary objective of performing WGCNA was to identify co-expression modules that are significantly associated with ssGSEA-derived scores of ERGs and MPRGs. This step was designed to reduce noise, highlight biologically relevant modules, and prioritize candidate genes for downstream causal inference, biomarker selection, and functional validation. WGCNA (v1.72-1) (27) was performed to identify genes associated with DE-MPRGs and DE-ERGs. First, all samples from the GSE89408 dataset were clustered to detect and exclude outliers. To ensure that gene interactions conformed to a scale-free topology, the R^2 value was set to 0.85, and the mean connectivity was assessed to approach zero in order to determine an appropriate soft-thresholding power. Subsequently, the dynamic tree cutting algorithm was applied to identify gene modules, with a minimum module size of 200 genes. The correlation between module eigengenes and ssGSEA scores was calculated across all samples in the GSE89408 dataset. Intra-module genes were defined as those moderately correlated with ssGSEA scores ($|r| > 0.3$, $P < 0.05$). The soft-thresholding power was set to 17 based on the scale-free topology criterion ($R^2 > 0.85$), in line with standard WGCNA practice to generate biologically meaningful co-expression modules.

3.4. Functional Exploration of MP-ERGs

To identify genes that were differentially expressed and associated with both ERGs and MPRGs, overlapping genes from the DEGs and module genes were selected

using eulerr (v7.0.0) and designated as MP-ERGs. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of MP-ERGs were performed using clusterProfiler (v4.8.2) (28) to identify functional pathways associated with these genes (P.adj < 0.05).

3.5. Analysis of MR

The three core assumptions of MR are: (1) instrumental variables (IVs) must be strongly associated with the exposure; (2) IVs should not be related to confounding factors; and (3) IVs should influence the outcome only through the exposure.

To assess the causal relationship between MP-ERGs and OA, MR was performed with MP-ERGs as the exposure and OA as the outcome. TwoSampleMR (v0.5.8) (29) was first used to select IVs strongly associated with MP-ERGs but not with OA ($P < 5 \times 10^{-8}$, clump = TRUE, $r^2 = 0.001$, kb = 10,000). Effect alleles and effect sizes were harmonized using the mv_harmonise_data function from the TwoSampleMR package. MR analyses were then conducted with the mv_multiple function across five approaches: MR-Egger (30), Weighted Median (31), Inverse Variance Weighted (IVW), Simple Mode (32), and Weighted Mode (33). A *P*-value < 0.05 in the IVW analysis indicated a causal association between MP-ERGs and OA. Otherwise, an odds ratio (OR) of 1 was used as the cut-off, with OR > 1 suggesting MP-ERGs as a risk factor for OA and OR < 1 suggesting a protective effect. Results were visualized using scatter, forest, and funnel plots.

3.6. Sensitivity and Steiger Test

To ensure the reliability of the MR results, several sensitivity analyses were performed, including heterogeneity (34), horizontal pleiotropy (35), and leave-one-out (LOO) tests (36). Cochran's Q test, implemented with the mr_heterogeneity function, was used to evaluate heterogeneity between exposure factors and OA ($P > 0.05$). Horizontal pleiotropy was assessed to detect potential confounding effects of IVs on OA through exposure variables ($P > 0.05$). The LOO test was then applied to validate the robustness of the findings by sequentially removing individual SNPs. Finally, the Steiger orientation test was conducted to confirm the correct causal direction of the MR analysis ($P < 0.05$).

3.7. Selection of Biomarkers

To identify feature genes, Least Absolute Shrinkage and Selection Operator (LASSO) and Support Vector Machine–Recursive Feature Elimination (SVM-RFE) were applied to MP-ERGs. For LASSO, triple cross-validation with glmnet (v4.1-7) (37) was used to determine the optimal penalty parameter (Lambda.min), yielding the genes with the lowest model error rate. For SVM-RFE, e1071 (v1.7-13) (38) was used to select the subset of genes corresponding to the minimum classification error. The minimum error criterion, rather than the one-standard-error rule, was adopted because it produced more stable results across validation sets. Both methods employed 10-fold cross-validation to reduce overfitting. Model performance was assessed using AUC, sensitivity, specificity, precision, and recall.

Candidate biomarkers were defined as genes identified by both machine learning algorithms. Their expression levels ($P < 0.05$) and diagnostic performance were evaluated in the GSE89408 and GSE55457 datasets. Diagnostic ability was assessed using receiver operating characteristic (ROC) curves generated with the pROC package (v1.18.0) (39), with an area under the curve (AUC) > 0.7 indicating good performance. Genes showing consistent differential expression in both datasets and an AUC > 0.7 were confirmed as biomarkers.

3.8. Construction of Artificial Neural Networks (ANN)

The association between biomarkers and OA was first assessed using univariate logistic regression, with OR calculated ($P < 0.05$). Biomarker-based ANNs were then constructed in the GSE89408 and GSE55457 datasets using neuralnet (v1.44.2) (40), which normalized the data and specified five hidden layers. Several ANN architectures with 2–5 hidden layers were evaluated, and performance was compared using cross-validation. ROC curves for each dataset were generated to assess discrimination between OA and control samples. The five-layer model demonstrated the most stable performance with minimal variance across folds and consistent validation in independent datasets, indicating that the results were not attributable to overfitting.

3.9. Functional Enrichment and Immune Infiltration

To investigate the functions and pathways of the

biomarkers, single-gene gene set enrichment analysis (GSEA) was performed individually. First, Spearman correlations between each biomarker and all other genes were calculated in the GSE89408 dataset using the psych package (v2.3.6) (41), and the results were ranked in descending order. GSEA was then conducted with cluster Profiler, using KEGG pathways from the Molecular Signatures Database (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb) as the reference gene set (|NES| > 1, P.adj < 0.05). Similarly, gene set variation analysis (GSVA) was carried out with GSVA software, using the c2.cp.kegg.v2023.1.Hs.symbols.gmt file from MSigDB as the background gene set. Intergroup differences in enriched pathways were further analyzed using limma (v3.56.2) (42) (P.adj < 0.05).

To explore the involvement of immune-related mechanisms in OA, the ssGSEA algorithm from the GSVA package was applied to estimate the relative abundance of 16 immune cell types and 13 immune-related functions in the GSE89408 dataset. Intergroup comparisons were performed, and Spearman correlations were further computed between biomarkers and immune cells or immunological processes ($P < 0.05$).

3.10. Molecular Regulation and Drug Prediction

To construct a biomarker-based molecular regulatory network, miRNAs targeting the biomarkers were predicted using the HMDD (http://www.cuilab.cn/hmdd) and miRWalk (http://mirwalk.umm.uni-heidelberg.de/) databases. Only miRNAs predicted by both databases were retained as candidate target miRNAs, and their corresponding lncRNAs were further predicted using the starBase database (http://starbase.sysu.edu.cn/) with the criterion *clipExpNum > 15*. In parallel, transcription factors (TFs) regulating the biomarkers were identified using the NetworkAnalyst platform in conjunction with the JASPAR database (http://jaspar.genereg.net/). To explore potential therapeutic relevance, the Comparative Toxicogenomics Database (CTD) (http://ctdbase.org/) was applied to predict drugs targeting the biomarkers (Reference Count > 2). Finally, mRNA–miRNA–lncRNA, mRNA–TF, and biomarker–drug regulatory networks were constructed.

3.11. The Scrna-Seq Data Filtering, Downgrading and Clustering

To explore the roles of different cell types in OA, scRNA-seq data were analyzed using Seurat (v4.3.0) (43). Genes detected in fewer than 200 cells, cells with nFeature_RNA ≥ 6000, cells with nCount_RNA ≥ 40,000, and cells with percent.mt ≥ 10% were excluded. The filtered data were log-normalized, and highly variable genes were selected before performing principal component analysis (PCA) using the vst method. Cells were then clustered via Uniform Manifold Approximation and Projection (UMAP) using the FindNeighbors and FindClusters functions (resolution = 0.3). Clusters were annotated based on known marker genes reported in the literature (44).

3.12. Analysis of Cell Function and Differentiation Status

To investigate functional differences among cell populations in OA, ReactomeGSA (v1.12.0) (45) was used to examine pathway enrichment across clusters. CellChat (v1.6.1) (46) was applied to explore intercellular interactions. Cell populations exhibiting high expression levels in OA and strong biomarker expression were selected as key cells. These key cells were reclustered twice (resolution = 0.1) and analyzed with GSVA using the Hallmark gene set from MSigDB as the reference. Finally, pseudotime trajectories of the key cells were inferred using Monocle (v2.26.0) (47) to investigate their differentiation states.

3.13. Statistical Analysis

All bioinformatics analyses were performed in R (v4.2.2). Differences between OA and control groups were assessed using the Wilcoxon test, with $P < 0.05$ considered statistically significant.

4. Results

4.1. Macrophage Polarization and Efferocytosis were Remarkably Associated with OA

Following differential expression analysis of the GSE89408 dataset, 2,181 DEGs were identified between OA and control groups, including 1,347 upregulated and 834 downregulated genes (**Fig. 1A–1B**). Intersecting the DEGs with MPRGs and ERGs yielded 13 DE-ERGs and 7 DE-MPRGs (**Fig. 1C–1D**). Score calculations showed significant differences in DE-ERG and DE-MPRG scores between OA and control samples (**Fig. 1E–1F**).

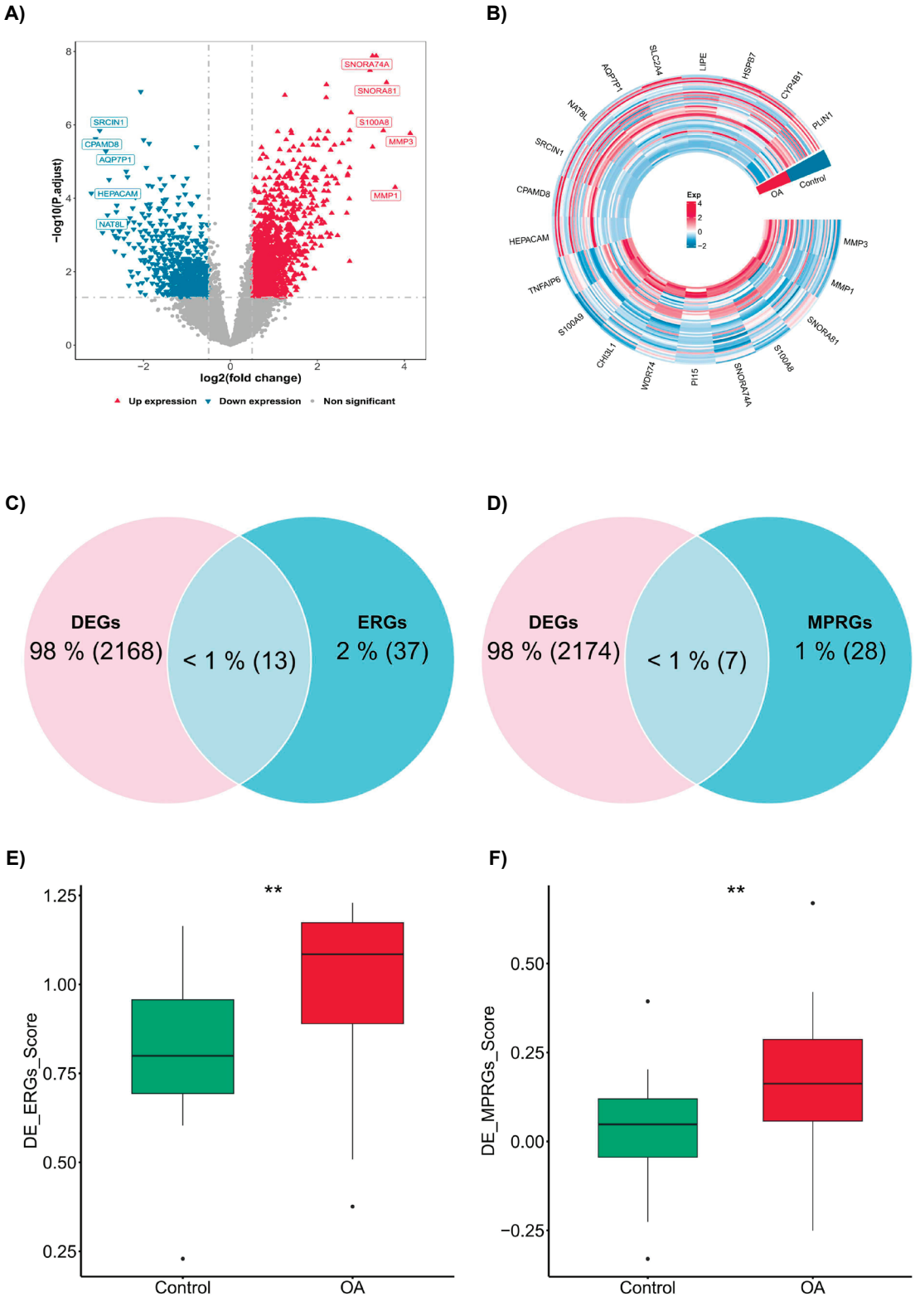

Figure 1. Differential expression and enrichment analysis in bulk synovial tissue samples (GSE89408). (**A, B**) Differential gene expression analysis in the GSE89408 dataset identified 2,181 DEGs (1,347 upregulated, 834 downregulated) . (**C, D**) Venn diagrams showed 13 overlapping DE-ERGs and 7 DE-MPRGs. **(E, F)** DE-ERG and DE-MPRG scores differed significantly between OA and controls.

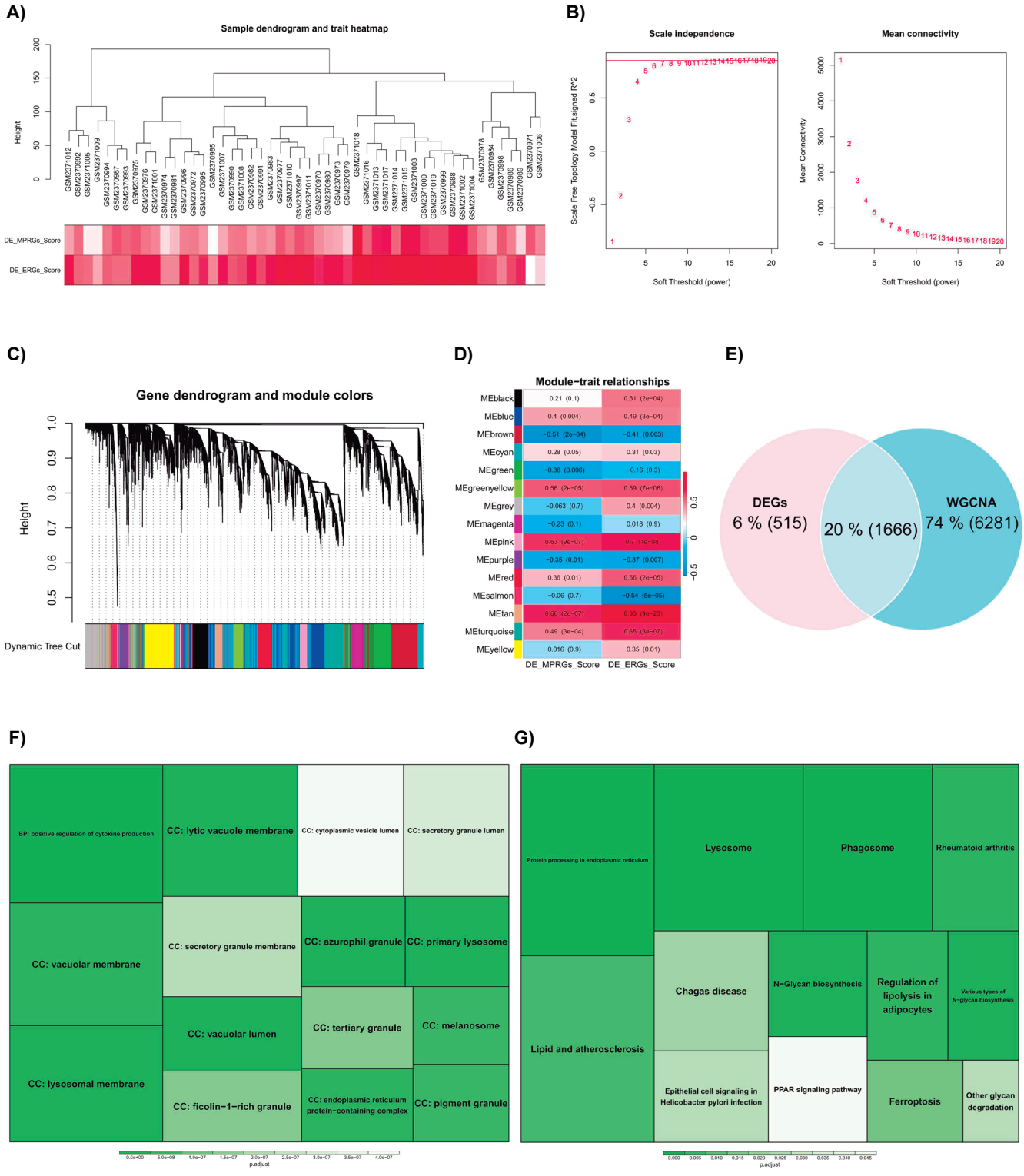

Figure 2. WGCNA and functional enrichment of MP-ERGs in bulk synovial tissue samples (GSE89408). **A**) WGCNA was performed using DE-ERG and DE-MPRG scores as traits. **B**) No outliers were detected in the GSE89408 dataset. A soft-threshold power of 17 was selected based on scale-free topology and connectivity. (**C, D**) Fifteen gene modules were identified, with 8 modules significantly associated with both scores. **E**) These 7,947 genes were selected as module genes, and their intersection with DEGs identified 1,666 MP-ERGs. (**F, G**) Functional enrichment analysis showed MP-ERGs were involved in 482 GO terms (e.g., cytokine production, vacuolar lumen) and 13 pathways (e.g., protein processing in the endoplasmic reticulum, phagosome).

To further explore gene co-expression patterns, WGCNA was performed using DE-ERG and DE-MPRG scores as traits. No outlier samples were detected in the GSE89408 dataset following clustering (**Fig. 2A**). Based on scale-free topology criteria (R^2 and mean connectivity), a soft-thresholding power of 17 was selected (**Fig. 2B**). Hierarchical clustering identified 15 gene modules, eight of which (blue, brown, green-yellow, pink, purple, red, tan, and turquoise) were significantly associated with both ssGSEA scores (**Fig. 2C–2D**). The 7,947 genes in these eight modules were designated as module genes for further analysis.

Intersecting the 7,947 module genes with the 2,181 DEGs identified 1,666 MP-ERGs (**Fig. 2E**). Functional enrichment analysis indicated that MP-ERGs were primarily associated with 482 GO terms (e.g., positive regulation of cytokine production, vacuolar lumen) and 13 KEGG pathways (e.g., protein processing in the endoplasmic reticulum, phagosome) (**Fig. 2F–2G**).

4.2. FMO4 and GPR65 were Considered as Biomarkers

To identify genes causally related to OA, MR analyses were performed using MP-ERGs as the exposure and OA as the outcome. Fourteen MP-ERGs were found to have a causal association with OA. LASSO and SVM-RFE were then applied to these 14 MP-ERGs. The LASSO model showed the lowest error rate at Lambda.min = 0.0498, identifying five feature genes: ATOH8, COQ2, DHRS12, FMO4, and GPR65 (**Fig. 3A–3B**). Concurrently, SVM-RFE achieved the highest accuracy with eight feature genes: PRDX1, ATOH8, GPR65, CD59, CKAP2, FMO4, MDFIC, and RGS18 (**Fig. 3C**).

By overlapping the results of both machine learning approaches, ATOH8, FMO4, and GPR65 were selected as candidate biomarkers (**Fig. 3D**). However, ATOH8 displayed inconsistent diagnostic performance in independent validation datasets (AUC < 0.7) and weaker functional associations with OA-related pathways. Consequently, only FMO4 and GPR65 were retained as robust biomarkers.

Both FMO4 and GPR65 were significantly upregulated in the OA group in the GSE89408 and GSE55457 datasets, with AUC values exceeding 0.7, confirming their biomarker status (**Fig. 3E–3H**). Additional performance metrics demonstrated high sensitivity (>0.80), specificity (>0.75), precision (>0.78), and recall

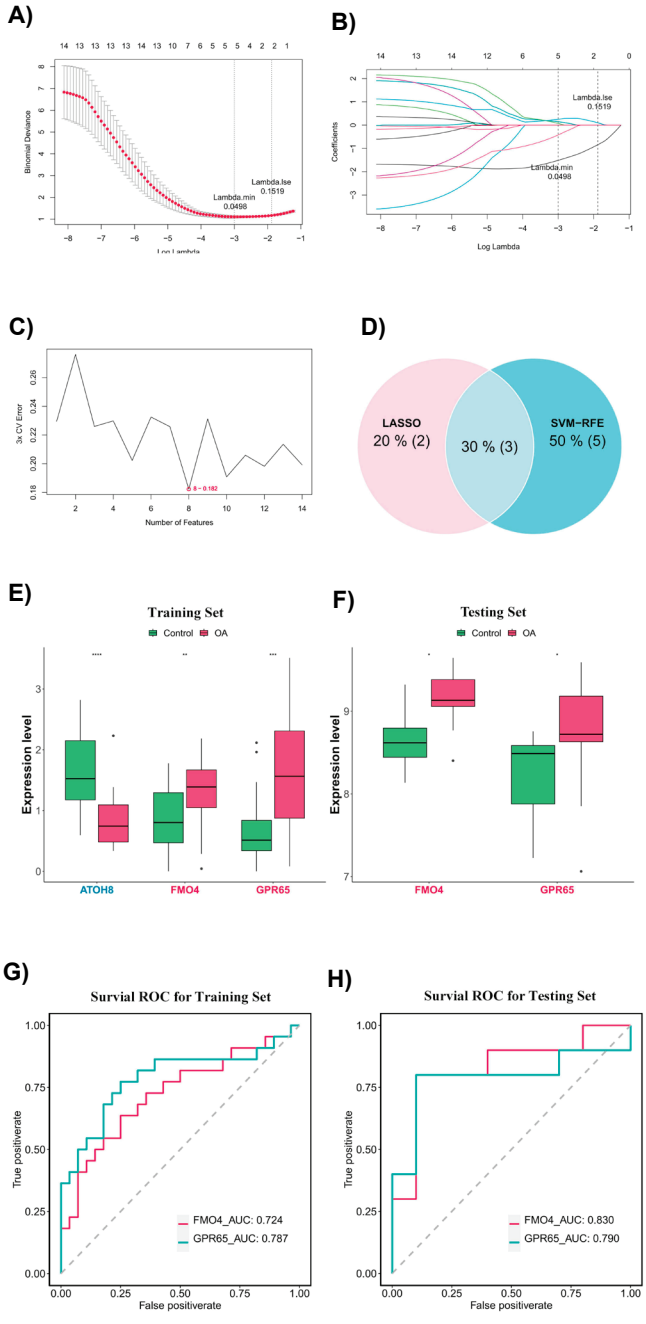

Figure 3. Biomarker selection via machine learning in bulk synovial tissue samples (GSE89408 and GSE55457). (**A, B**) Mendelian randomization (MR) identified 14 MP-ERGs causally associated with OA. LASSO regression identified 5 feature genes (ATOH8, COQ2, DHRS12, FMO4, GPR65) at optimal Lambda.min = 0.0498. **C**) SVM-RFE selected 8 feature genes (PRDX1, ATOH8, GPR65, CD59, CKAP2, FMO4, MDFIC, RGS18) with the highest accuracy. **D**) ATOH8, FMO4, and GPR65 were chosen as candidate biomarkers from both methods. (**E–H**) Expression analysis in GSE89408 and GSE55457 datasets confirmed significantly higher FMO4 and GPR65 levels in OA, with AUC values > 0.7, supporting their diagnostic potential.

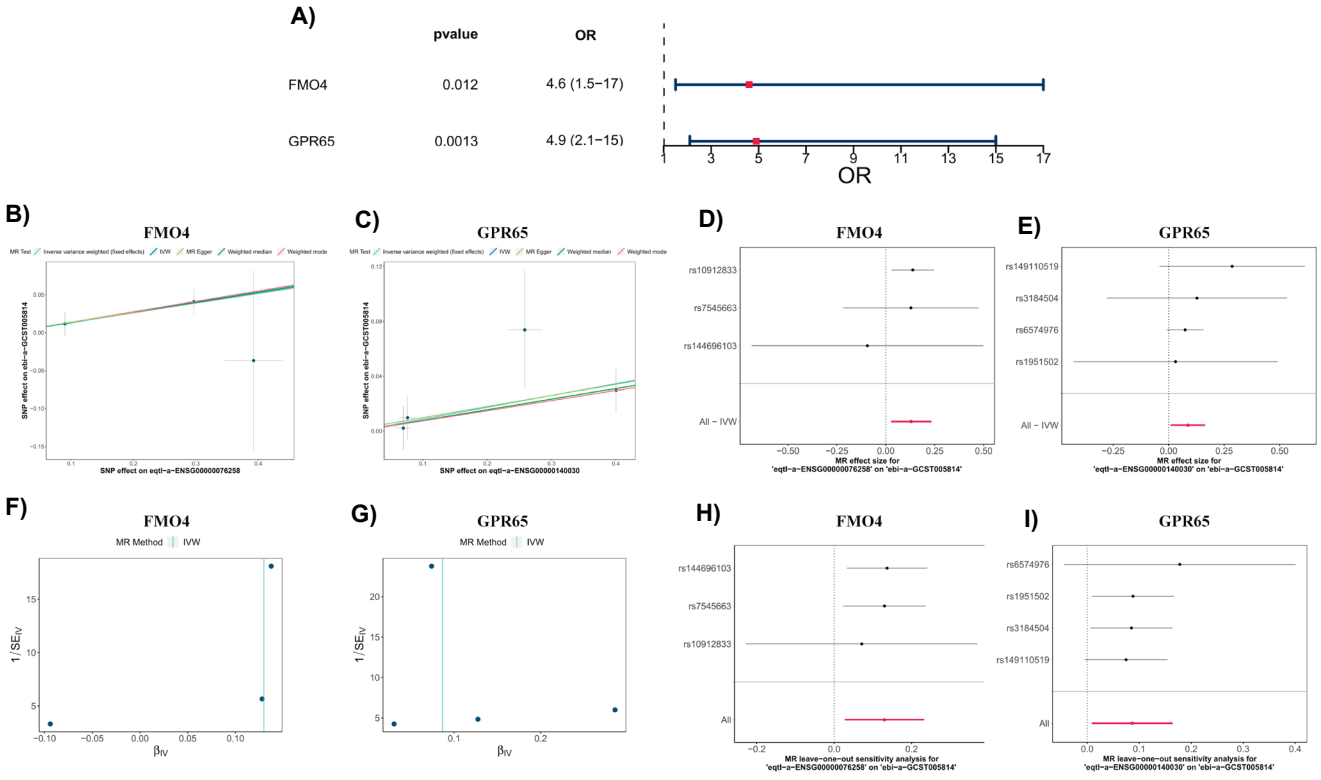

Figure 4. Mendelian randomization and sensitivity analyses for FMO4 and GPR65 (GWAS Dataset: ebi-a-GCST005814). A) Univariate logistic regression showed that FMO4 and GPR65 were risk factors for OA, with ORs > 1. **(B–E)** MR analysis confirmed their genetic association with OA. (**F, G**) Scatter and forest plots indicated a positive effect, while funnel plots showed symmetric SNP distribution, supporting Mendel's second law. (**H, I**) Sensitivity analyses revealed no significant heterogeneity, pleiotropy, or influential SNPs. Steiger direction tests confirmed no reverse causality.

(>0.80) across independent datasets, further supporting their diagnostic robustness beyond AUC alone.

4.3. Both FMO4 and GPR65 were Risk Factors for OA

Univariate logistic regression demonstrated that both biomarkers had ORs greater than 1, indicating they are risk factors for OA (**Fig. 4A**). MR analyses further confirmed this, showing FMO4 (P = 0.0125, OR = 1.1386; 95% CI: 1.0284–1.2606) and GPR65 (P = 0.0283, OR = 1.0904; 95% CI: 1.0092–1.178) as risk factors. The positive slope in scatter plots and MR effect values greater than 0 in forest plots supported these findings (**Fig. 4B-4E**). Funnel plots showed a balanced distribution of SNPs around the IVW estimate, suggesting adherence to Mendel's second law (**Fig. 4F-4G**).

Sensitivity analyses confirmed the reliability of the MR results: no heterogeneity was observed (FMO4: P = 0.7519; GPR65: P = 0.6569), no horizontal pleiotropy was detected (FMO4: P = 0.9763; GPR65: P = 0.9193), and no single SNP significantly influenced the outcomes (**Fig. 4H-4I**). Finally, the Steiger direction test returned TRUE, indicating no evidence of reverse causation between the biomarkers and OA.

4.4. Biomarker-Based ANN Demonstrate Robust Discrimination between OA and Control Samples

To evaluate the predictive value of the biomarkers for OA, ANNs were constructed using expression data from the GSE89408 and GSE55457 datasets (**Fig. 5A-5B**). The resulting ROC curves showed AUC values > 0.7 in both datasets, indicating strong discriminatory power of the ANN between OA and control samples (**Fig. 5C-5D**).

To explore the functional processes associated with the biomarkers, GSEA and GSVA were performed. GSEA

revealed enrichment of both biomarkers in ribosome and proteasome pathways (**Fig. 5E-5F**), whereas GSVA showed significant upregulation in allograft rejection and riboflavin metabolism, with downregulation in basal cell carcinoma and circadian rhythm–related pathways (**Fig. 5G-5H**).

4.5. GPR65 May Play a Greater Role than FMO4 in Immunoregulation in OA

Immune infiltration analysis revealed significant enrichment of five immune cell types—macrophages, neutrophils, Th2 cells, tumor-infiltrating lymphocytes (TILs), and Treg cells—as well as four immune-related functions (APC co-inhibition, APC co-stimulation, checkpoint, and T cell co-inhibition) between OA and control groups (**Fig. 6A-6B**). Both biomarkers exhibited positive correlations with these immune cells and functions, with GPR65 showing stronger associations than FMO4 (**Fig. 6C-6D**).

In the regulatory networks, diverse interactions

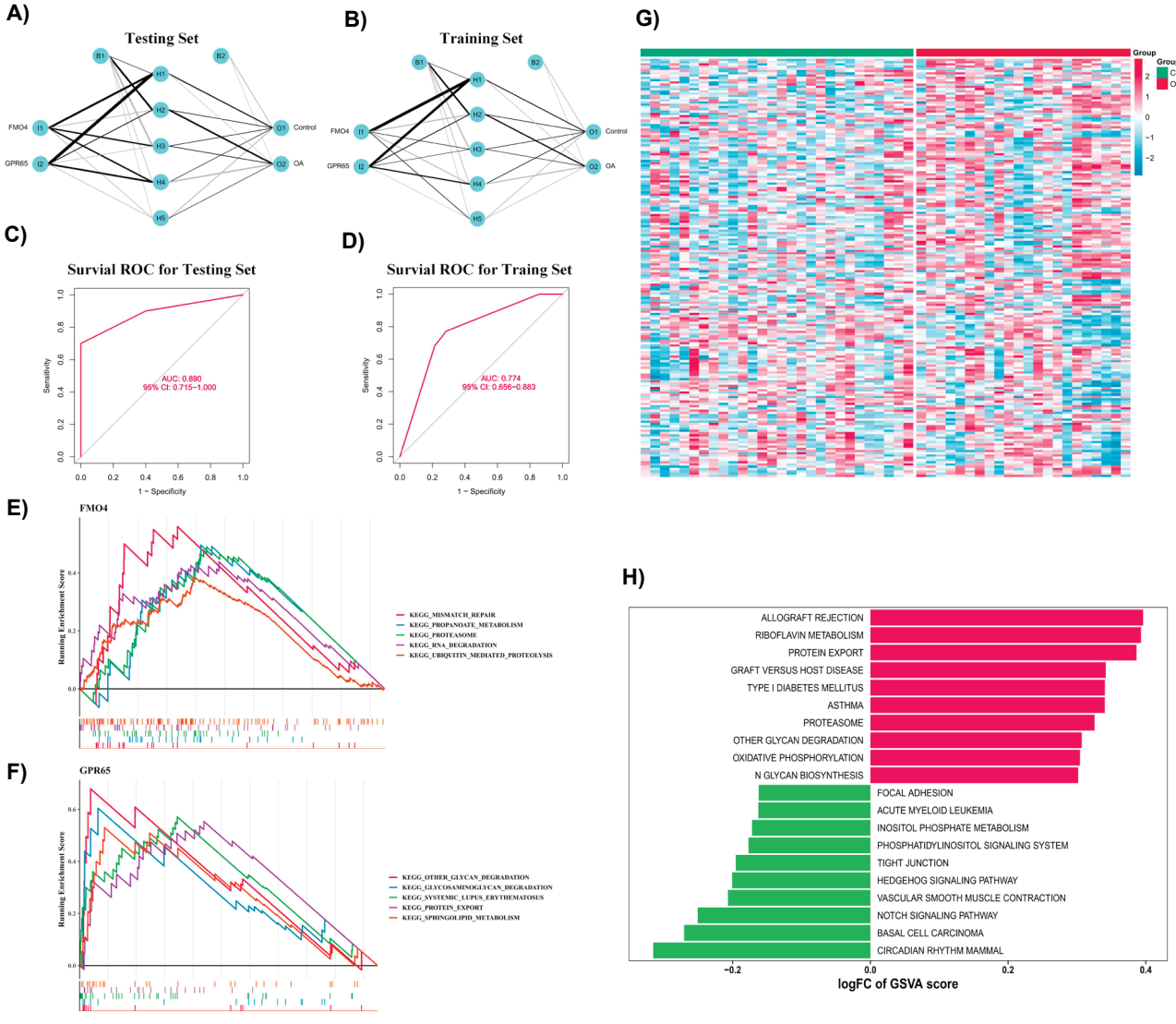

Figure 5. Artificial neural network and functional enrichment analyses in bulk synovial tissue samples (GSE89408 and GSE55457). (**A, B**) ANN models were constructed using GSE89408 and GSE55457 expression data. (**C, D**) ROC curve analysis indicated AUC values above 0.7 for both networks. (**E, F**) GSEA and GSVA showed enrichment in ribosome and proteasome-related pathways, (**G, H**) with significant up-regulation in pathways like allograft rejection and riboflavin metabolism, and down-regulation in basal cell carcinoma and circadian rhythm.

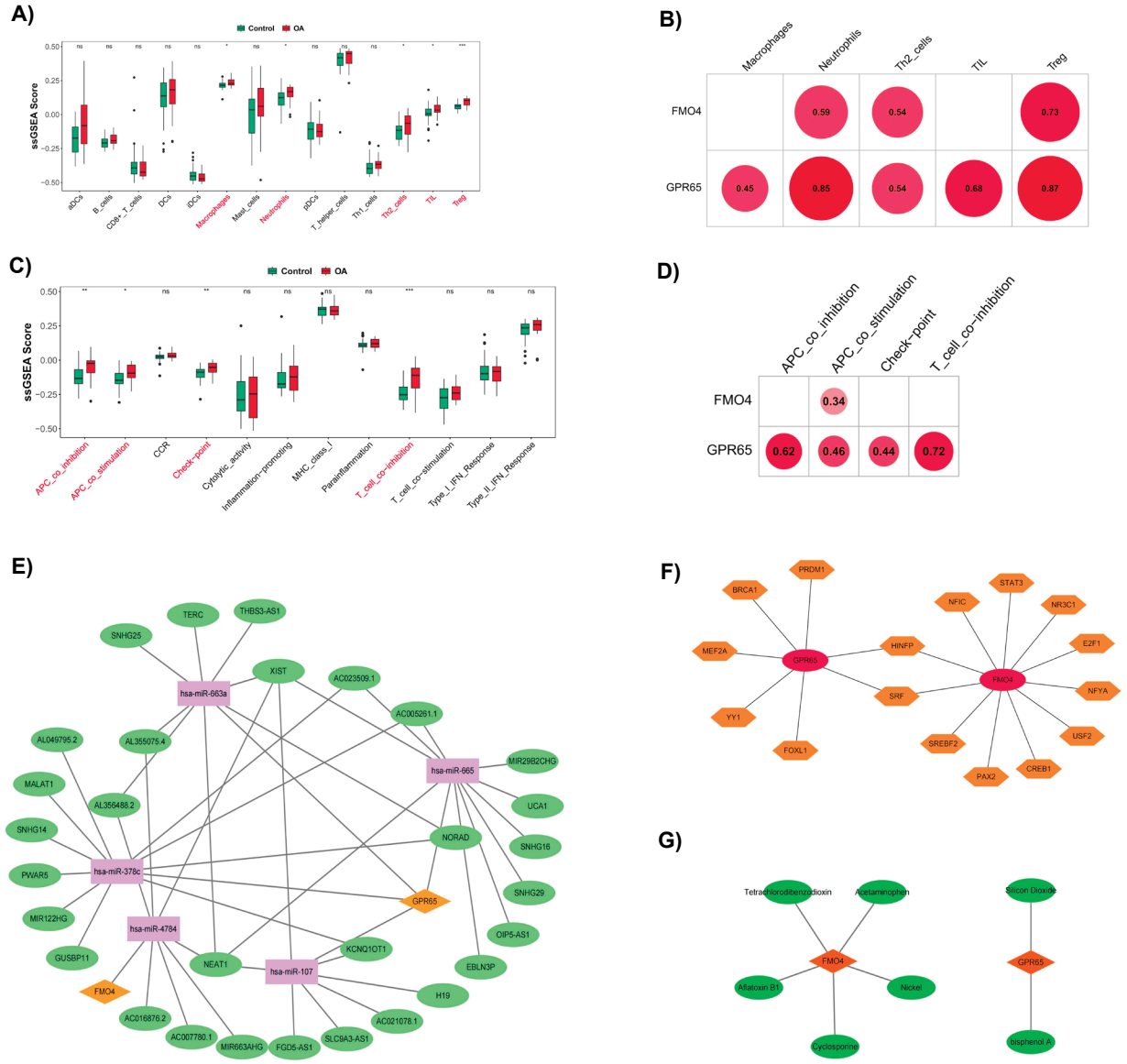

Figure 6. Immune infiltration and regulatory network analyses in bulk synovial tissue samples (GSE89408). (**A, B**) Immune infiltration analysis showed significant enrichment of five immune cell types and four immune-related functions in OA compared to controls. (**C, D**) Both biomarkers correlated with immune cells and functions, with GPR65 showing stronger associations. (**E, F**) Regulatory roles were identified in mRNA-miRNA-lncRNA and mRNA-TF networks. **G**) Drug prediction linked FMO4 to five chemicals, and GPR65 to two.

involving the biomarkers were observed, including FMO4–hsa-miR-4784–NEAT1, GPR65–hsa-miR-663a–XIST, FMO4–SRF, and GPR65–YY1 in the mRNA–miRNA–lncRNA and mRNA–TF networks (**Fig. 6E-6F**). Drug prediction analysis indicated that FMO4 was associated with five compounds (tetrachlorodibenzodioxin, aflatoxin B1, cyclosporine, nickel, and acetaminophen), whereas GPR65 was linked to two compounds (bisphenol A and silicon dioxide) (**Fig. 6G**). While these findings highlight potential avenues for drug repurposing, *in vitro* and *in vivo* validation is needed to confirm these interactions.

4.6. SSF and HLA-DRA Were Defined as Key Cells in OA

To explore the regulatory roles of different cells in OA, scRNA-seq data were analyzed. After data filtering, 10,248 cells and 21,417 genes were retained.

Screening of 2,000 highly variable genes followed by PCA yielded 30 principal components (**Fig. S2-S3**). UMAP clustering identified 11 cell clusters in OA **(Fig. 7A**), which were further annotated into eight major cell types: synovial subintimal fibroblasts (SSF), synovial intimal fibroblasts (SIF), smooth muscle cells (SMC), endothelial cells (EC), mast cells, HLA-DRA+ cells, T cells, and B cells (**Fig. 7B-7C**).

Functional enrichment analysis revealed that SMC differed in ATP-sensitive potassium channel activity, while HLA-DRA+ cells varied in hydroxycarboxylic acid–binding receptor function (**Fig. 7D**). Cell–cell communication analysis showed extensive interactions among all eight cell types, with SSF exhibiting the highest number and intensity of interactions (**Fig. 7E**). Examination of cell abundance indicated that SSF, SIF, and HLA-DRA+ cells were dominant in OA (**Fig. 7F**), and biomarkers were highly expressed in SSF and HLA-DRA+ cells, designating them as key cells (**Fig. 7G-7H**).

Independent reclustering of these two key cell types revealed that SSF comprised two subpopulations, whereas HLA-DRA+ cells contained three subpopulations (**Fig. 8A-8B**). Functional enrichment indicated

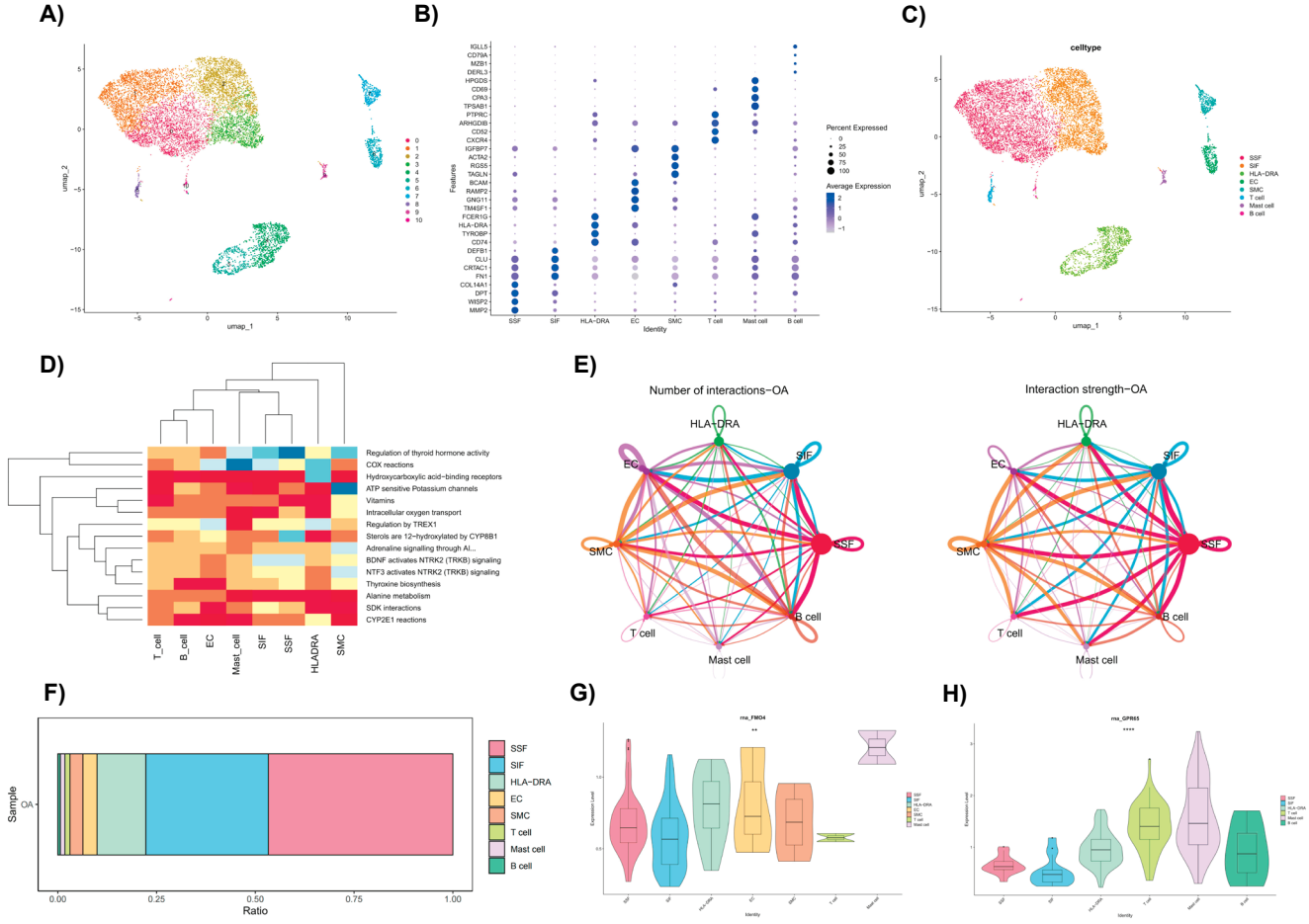

Figure 7. Cell clustering and interaction analysis in OA synovial tissue (scRNA-seq Dataset: GSE152805). **A**) UMAP clustering identified 11 cell clusters, (**B, C**) annotated into 8 cell types: SSF, SIF, SMC, EC, mast cells, HLA-DRA+ immune cells, T cells, and B cells. **D**) Functional enrichment linked SMCs to ATP-sensitive potassium channels and HLA-DRA+ cells to hydroxycarboxylic acid-binding receptors. **E**) Cellular communication analysis showed extensive interactions, with SSFs interacting most with other cells. **F**) SSF, SIF, and HLA-DRA were dominant in OA, (**G, H**) with higher biomarker expression in SSF and HLA-DRA, marking them as key cells.

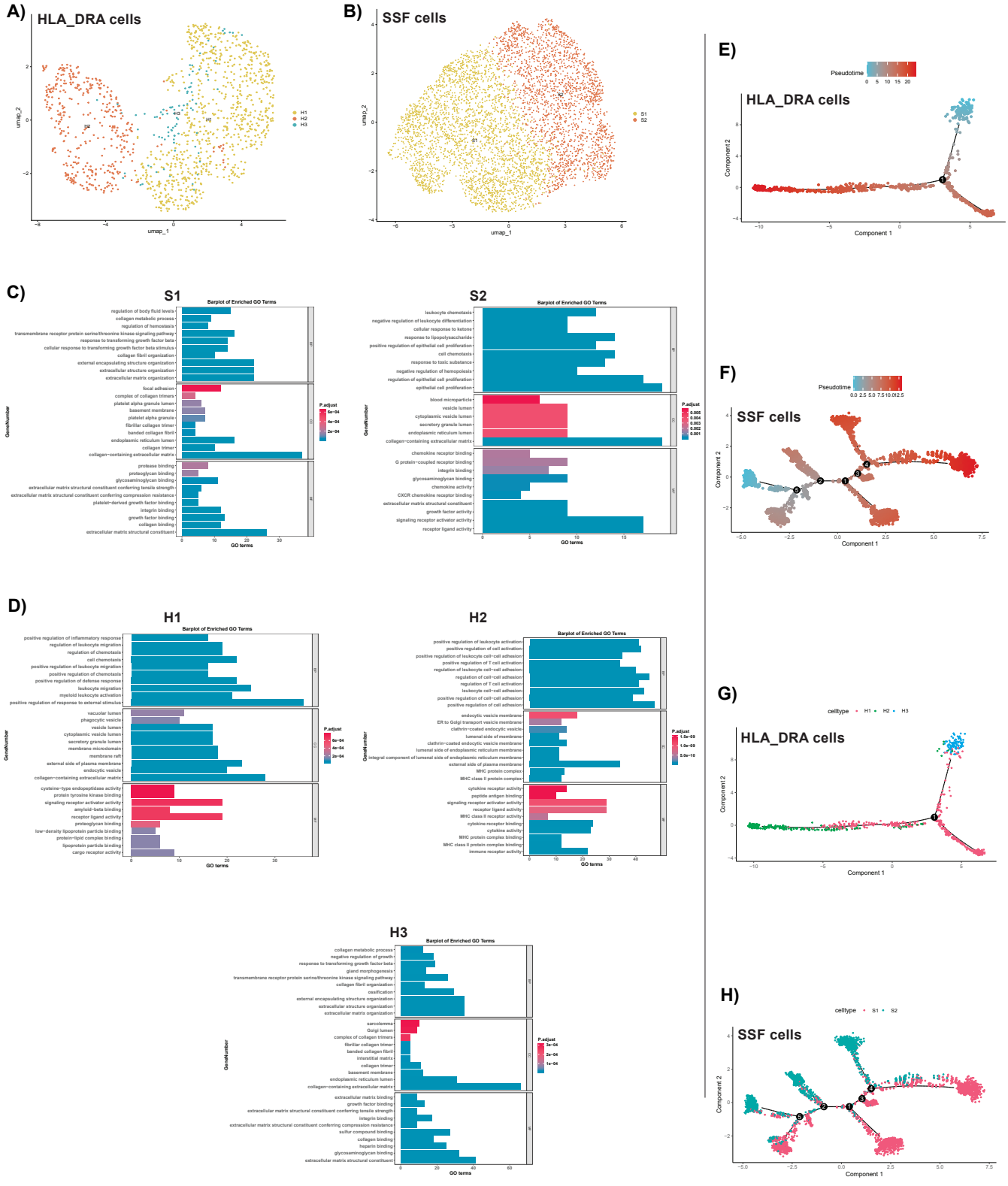

Figure 8. Reclassification of SSF and HLA-DRA+ immune cells revealed distinct subpopulations. **(A, B)** SSF was split into two subpopulations, and HLA-DRA+ into three. **C)** Functional enrichment showed oppositely enriched functions in SSF subpopulations, including protein secretion and mitotic spindle organization. **D)** H1, H2, and H3 SSF subpopulations were linked to response to external stimuli, cell adhesion, and collagen matrix. **(E-H)** Pseudotime analysis revealed different differentiation stages for the SSF subpopulations, with H2 and H3 HLA-DRA subpopulations at the start and end of differentiation, respectively.

that some SSF subpopulations had contrasting roles, such as protein secretion versus mitotic spindle (**Fig. 8C**). In HLA-DRA+ subpopulations, positive regulation of response to external stimulus, cell adhesion, and collagen-containing extracellular matrix were markedly enriched in H1, H2, and H3, respectively (**Fig. 8D**). Pseudotime analysis showed that the two SSF subpopulations were distributed differently along differentiation trajectories, while HLA-DRA+ subpopulations H2 and H3 appeared at the end and beginning of differentiation, respectively (**Fig. 8E-8H**).

5. Discussion

OA is a progressive joint disease characterized by chronic inflammation, cartilage degradation, and synovial fibrosis. Although key inflammatory pathways in OA have been identified, the precise molecular mechanisms linking immune regulation, cartilage breakdown, and disease progression remain incompletely understood. In this study, we integrated bulk transcriptomics, scRNA-seq, and MR to investigate the interplay between macrophage polarization, efferocytosis, and OA pathology. Differential gene expression profiling highlighted FMO4 and GPR65 as potential regulators, both of which were validated as genetic risk factors for OA, demonstrating strong causal associations in MR analyses.

Functional enrichment analysis indicated that FMO4 and GPR65 are involved in immune modulation, lysosomal function, and protein homeostasis, suggesting they may influence synovial inflammation and joint remodeling. Single-cell analysis further highlighted SSF and HLA-DRA+ immune cells as key cellular mediators, reinforcing the relevance of these biomarkers in OA pathogenesis.

As a proton-sensing G-protein–coupled receptor (GPCR), GPR65 responds to the acidic microenvironment's characteristic of inflamed osteoarthritic synovium. In fibroblast-like synoviocytes (FLS), GPR65 activation promotes secretion of inflammatory cytokines, including IL-6, IL-8, and MCP-1, which contribute to synovial inflammation and pain sensitization (48). Unlike prior studies focused solely on FLS, our investigation provides a broader perspective on GPR65's role in OA by integrating bioinformatics analyses, including functional MR and immune cell infiltration profiling. Functional enrichment implicates GPR65 in glycosaminoglycan degradation, sphingolipid metabolism, and immune dysregulation, pathways critical to cartilage breakdown and synovial remodeling. MR analysis identifies GPR65 as a genetic risk factor for OA, while logistic regression and ROC analysis (AUC > 0.7) support its diagnostic potential. Moreover, ANN modeling validates its ability to distinguish OA patients from healthy controls.

At the cellular level, scRNA-seq analysis revealed GPR65 enrichment in SSF and HLA-DRA+ antigen-presenting cells, both of which are key regulators of joint inflammation and extracellular matrix remodeling. CellChat-based intercellular communication analysis identified SSF as the primary hub of fibro-inflammatory crosstalk, suggesting that GPR65 may drive fibroblast activation and immune interactions within the synovial microenvironment. Immune infiltration analysis showed a strong positive correlation between GPR65 expression and regulatory T cells (Tregs, $r = 0.87$, $p < 0.05$) as well as M2 macrophage polarization. Although this correlation is unusually high for biological data, it was consistently observed across analyses, indicating a robust association and a potential role for GPR65 in modulating immune homeostasis and resolving chronic inflammation in OA.

From a therapeutic perspective, drug repurposing analysis highlighted bisphenol A and silicon dioxide as potential GPR65 inhibitors, suggesting molecular interactions that could guide targeted interventions. However, some predicted compounds, such as bisphenol A and aflatoxin B1, are primarily environmental toxicants rather than therapeutic agents; thus, these findings should be interpreted as indicators of molecular interactions rather than direct treatment recommendations.

Notably, these results align with previous research on proton-sensing GPCRs, such as GPR68, which is upregulated in OA cartilage and associated with disease progression (49). Collectively, our findings suggest that GPR65 is a key regulator of synovial inflammation, immune dysregulation, and fibroblast activation in OA, with significant diagnostic and therapeutic implications.

FMO4 (Flavin-Containing Monooxygenase 4) is well known for its role in detoxification and oxidative stress regulation, primarily in hepatic metabolism (50). Our study is the first to implicate FMO4 in OA. Functional analyses indicate that FMO4 is

associated with lysosomal degradation, endoplasmic reticulum protein processing, and phagosome function, processes that may help maintain cellular homeostasis and extracellular matrix remodeling in OA joints. Immune infiltration analysis revealed a strong positive correlation between FMO4 expression and regulatory T cells (Tregs), suggesting a role in immune tolerance and inflammation control. GSEA further implicated FMO4 in RNA degradation, ubiquitin-mediated proteolysis, and mismatch repair, essential processes for synovial cell turnover and inflammation regulation. MR analysis provided genetic evidence supporting FMO4 as a risk factor for OA. Single-cell transcriptomics showed high FMO4 expression in SSF and HLA-DRA+ antigen-presenting cells, key contributors to OA progression through inflammation regulation and tissue remodeling. Cell–cell communication analysis confirmed SSF as a primary cellular hub, suggesting that FMO4 may modulate fibroblast activation and immune cell recruitment. Collectively, these findings indicate that FMO4 contributes to OA pathogenesis by regulating oxidative stress, immune cell dynamics, and protein homeostasis, positioning it as a promising biomarker and therapeutic target.

In contrast, GPR65 plays a distinct immunomodulatory role in OA. As a proton-sensing G-protein–coupled receptor (GPCR), GPR65 responds to acidic microenvironments typical of inflamed synovium. In synovial fibroblasts, GPR65 activation promotes secretion of pro-inflammatory cytokines, including IL-6 and MCP-1. Similar proton-sensing GPCRs, such as GPR68, are upregulated in OA cartilage and implicated in extracellular matrix degradation. Together, FMO4 and GPR65 highlight complementary pathways in OA pathogenesis, linking oxidative stress, immune regulation, and fibroblast-mediated tissue remodeling.

In this study, we identified GPR65 and FMO4 as key biomarkers involved in macrophage polarization and efferocytosis in OA. Comprehensive bioinformatics analyses indicated that these genes play crucial roles in immune regulation, metabolic reprogramming, and extracellular matrix remodeling. MR analysis further validated GPR65 and FMO4 as genetic risk factors for OA, reinforcing their pathological significance. scRNA-seq and immune infiltration profiling revealed that GPR65 is highly expressed in SSF and HLA-DRA+ antigen-presenting cells, implicating its role in synovial inflammation and fibroblast–immune interactions. These findings provide novel insights into OA pathogenesis by integrating genetic risk assessment, functional characterization, and immune regulation. However, further experimental validation — including quantitative PCR, gene knockout models, and in vitro cellular studies — is necessary to fully elucidate the underlying mechanisms.

Our study has several limitations. First, it was primarily computational, and experimental validation of the proposed biomarkers was not performed. Future work should include qPCR verification of FMO4 and GPR65, functional analyses using gene knockdown/knockout models, and validation in larger clinical cohorts to determine their diagnostic and therapeutic potential. Second, the transcriptomic and GWAS datasets used in this study were largely derived from individuals of European ancestry, so the applicability of our findings to other populations remains uncertain and requires further testing. Third, our biomarker discovery was based on bulk synovial tissue RNA-seq, where differential expression represents an average across all cell types and may mask opposing changes in specific populations. While scRNA-seq partially addressed this limitation by localizing biomarker expression, further cell-type–specific validation is needed to clarify their roles in distinct synovial cell populations.

6. Conclusion

FMO4 and GPR65 were identified as novel OA biomarkers through integrated transcriptomic, single-cell, and MR analyses. Validation across multiple datasets and machine learning models demonstrated their strong diagnostic potential and highlighted their involvement in immune regulation, fibroblast activation, and extracellular matrix remodeling. These findings provide new insights into the pathogenesis of OA, although further experimental studies are required to confirm their clinical utility.

Acknowledgements

We would like to express our gratitude to Junjie Yuan for their support with data analysis and statistical guidance.

Funding

This study was supported by the Fourth Militarily Medical University, Tangdu Hospital, Academic Acceleration Plan, Interdisciplinary Convergence

Project (Grant Number 2024JCRH008).

Conflict of Interests

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

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Reference

1. Van den Bosch MHJ. Osteoarthritis year in review 2020: biology. *Osteoarthritis Cartilage*. 2021;**29**(2):143-150. doi: 10.1016/j.joca.2020.10.006.
2. Driban JB, Harkey MS, Barbe MF, Ward RJ, MacKay JW, Davis JE, Lu B, *et al*. Risk factors and the natural history of accelerated knee osteoarthritis: a narrative review. *BMC Musculoskelet Disord*. 2020;**21**(1):332. doi: 10.1186/s12891-020-03367-2.
3. Sokolove JD, Lepus CM. The emergence of osteoarthritis as a public health concern. *Nat Rev Rheumatol*. 2016;**12**(3):203-212. doi: 10.1038/nrrheum.2016.1.
4. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with a focus on management. *Nat Rev Rheumatol*. 2019;**15**(2):82-96. doi: 10.1038/s41584-018-0114-3.
5. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, Hawker GA, *et al*. Osteoarthritis: a comprehensive review. *JAMA*. 2019;**321**(14):1395-1406. doi: 10.1001/jama.2019.0943.
6. Liang Y, Lin F, Huang Y. Identification of Biomarkers Associated with Diagnosis of Osteoarthritis Patients Based on Bioinformatics and Machine Learning. *J Immunol Res*. 2022;**2022**:5600190. doi: 10.1155/2022/5600190.
7. Gordon S, Martinez FO. Alternative activation of macrophages: mechanism and functions. *Immunity*. 2010;**32**(5):593-604. doi: 10.1016/j.immuni.2010.05.007.
8. Estrada McDermott J, Pezzanite L, Goodrich L, Santangelo K, Chow L, Dow S, Wheat W. Role of Innate Immunity in Initiation and Progression of Osteoarthritis, with Emphasis on Horses. *Animals (Basel)*. 2021;**11**(11):3247. doi: 10.3390/ani11113247.
9. Chen Y, Jiang W, Yong H, He M, Yang Y, Deng Z, Li Y. Macrophages in osteoarthritis: pathophysiology and therapeutics. *Am J Transl Res*. 2020;**12**(1):261-268.
10. Ko CY, Lin YY, Achudhan D, Chang JW, Liu SC, Lai CY, Huang YL, Tsai CH *et al*. Omentin-1 ameliorates the progress of osteoarthritis by promoting IL-4-dependent anti-inflammatory responses and M2 macrophage polarization. *Int J Biol Sci*. 2023;**19**(16):5275-5289. doi: 10.7150/ijbs.86701.
11. Nagata S. Apoptosis and clearance of apoptotic cells. *Annu Rev Immunol*. 2018;**36**:489-517. doi: 10.1146/annurev-immunol-042617-053105.
12. Yao Z, Qi W, Zhang H, Zhang Z, Liu L, Shao Y, Zeng H, *et al*. Down-regulated GAS6 impairs synovial macrophage efferocytosis and promotes obesity-associated osteoarthritis. *Elife*. 2023;**12**:e83069. doi: 10.7554/eLife.83069.
13. Lemke G, Burstyn-Cohen T. TAM receptors and the clearance of apoptotic cells. *Ann N Y Acad Sci*. 2010;**1209**:23-29. doi: 10.1111/j.1749-6632.2010.05752.x.
14. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;**23**(R1):R89-R98. doi: 10.1093/hmg/ddu395.
15. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;**181**(4):251-260. doi: 10.1093/aje/kwu283.
16. Alhassan E, Nguyen K, Hochberg MC, Mitchell BD. Causal Factors for Osteoarthritis: A Scoping Review of Mendelian Randomization Studies. *Arthritis Care Res* (Hoboken). 2024; **76**(3):366-375. doi: 10.1002/acr.25252.
17. Zou M, Shao Z. Proteome-Wide Mendelian Randomization and Colocalization Analysis Identify Therapeutic Targets for Knee and Hip Osteoarthritis. *Biomolecules*. 2024;**14**(3):355. doi: 10.3390/biom14030355.
18. Wu Y, Liu J, Yu W, Wang X, Li J, Zeng W. Single-cell transcriptome and multi-omics integration reveal ferroptosis-driven immune microenvironment remodeling in knee osteoarthritis. *Front Immunol*. 2025;**16**:1608378. doi: 10.3389/fimmu.2025.1608378. PMID: 40636124.
19. Sun Y, Yang H, Guo J, Du J, Han S, Yang X. Identification of HTRA1, DPT and MXRA5 as potential biomarkers associated with osteoarthritis progression and immune infiltration. *BMC Musculoskelet Disord*. 2024;**25**(1):647. doi: 10.1186/s12891-024-07758-7.
20. Ko CY, Lin YY, Achudhan D, Chang JW, Liu SC, Lai CY, *et al*. Omentin-1 ameliorates the progress of osteoarthritis by promoting IL-4-dependent anti-inflammatory responses and M2 macrophage polarization. *Int J Biol Sci*. 2023;**19**(16):5275-5289. doi: 10.7150/ijbs.86701. PMID: 37928270.
21. Yao Z, Qi W, Zhang H, Zhang Z, Liu L, Shao Y, Zeng H, *et al*. Down-regulated GAS6 impairs synovial macrophage efferocytosis and promotes obesity-associated osteoarthritis. *Elife*. 2023;**12**:e83069. doi: 10.7554/eLife.83069.
22. Zhou C, Gan X, Sun S, Wang L, Zhang Y, Zhang J. Construction of an efferocytosis-related long non-coding ribonucleic acid scoring system to predict clinical outcome and immunotherapy response in pancreatic adenocarcinoma. *Biochem Biophys Rep*. 2023;**35**:101540. doi: 10.1016/j.bbrep.2023.101540.
23. Zhao Y, Li M, Yang Y, Wu T, Huang Q, Wu Q, Ren C. Identification of Macrophage Polarization-Related Genes as Biomarkers of Chronic Obstructive Pulmonary Disease Based on Bioinformatics Analyses. *Biomed Res Int*. 2021; **2021**:9921012. doi: 10.1155/2021/9921012.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;**15**(12):550. doi: 10.1186/s13059-014-0550-8.
25. Wang B, Feng Y, Li Z, Zhou F, Luo J, Yang B, Long S, *et al*. Identification and validation of chromatin regulator-related signatures as a novel prognostic model for low-grade gliomas using translational bioinformatics. *Life Sci*. 2024;**336**:122312. doi: 10.1016/j.lfs.2023.122312.
26. Hänzelmann, S., Castelo, R. Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*.2013;**14**:7. doi:10.1186/1471-2105-14-7.
27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*.

2008;**9**:559. doi: 10.1186/1471-2105-9-559.
28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;**16**(5):284-287. doi: 10.1089/omi.2011.0118.
29. Huang P, Zhang PF, Li Q. Causal relationship between cannabis use and cancer: a genetically informed perspective. *J Cancer Res Clin Oncol*. 2023;**149**(11):8631-8638. doi: 10.1007/s00432-023-04807-x.
30. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;**44**(2):512-525. doi: 10.1093/ije/dyv080.
31. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;**40**(4):304-314. doi: 10.1002/gepi.21965.
32. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG; EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;**30**(7):543-552. doi: 10.1007/s10654-015-0011-z.
33. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;**46**(6):1985-1998. doi: 10.1093/ije/dyx102.
34. Nichter M. Cultural dimensions of hot, cold and sema in Sinhalese health culture. *Soc Sci Med*. 1987;**25**(4):377-387. doi: 10.1016/0277-9536(87)90276-0.
35. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;**23**(R1):R89-R98. doi: 10.1093/hmg/ddu328.
36. Cui Z, Feng H, He B, He J, Tian Y. Relationship Between Serum Amino Acid Levels and Bone Mineral Density: A Mendelian Randomization Study. *Front Endocrinol (Lausanne)*. 2021;**12**:763538. doi: 10.3389/fendo.2021.763538.
37. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;**33**(1):1-22. doi: 10.1016/j.jspi.2009.07.020.
38. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics*. 2018;**19**(1):432. doi: 10.1186/s12859-018-2451-4.
39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;**12**:77. doi: 10.1186/1471-2105-12-77.
40. Ngoc Van TT, Kim C, Lee H, Kim J, Shong B. Machine learning-based exploration of molecular design descriptors for area-selective atomic layer deposition (AS-ALD) precursors. *J Mol Model*. 2023;**30**(1):10. doi: 10.1007/s00894-023-05806-y.
41. Penney SR, Simpson AI. Correction to Lancet Psychiatry *Lancet*. 2023;**10**:197-208.
42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;**43**(7):e47. doi: 10.1093/nar/gkv007.
43. Gribov A, Sill M, Lück S, Rücker F, Döhner K, Bullinger L, Benner A, Unwin A. SEURAT: visual analytics for the integrated analysis of microarray data. *BMC Med Genomics*. 2010;**3**:21. doi: 10.1186/1755-8794-3-21.
44. Yang K, Xie Q, Tang T, Zhao N, Liang J, Shen Y, Li Z, *et al*. Astragaloside IV as a novel CXCR4 antagonist alleviates osteoarthritis in the knee of monosodium iodoacetate-induced rats. *Phytomedicine*. 2023;**108**:154506. doi: 10.1016/j.phymed.2022.154506.
45. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol Cell Proteomics*. 2020;**19**(12):2115-2125. doi: 10.1074/mcp.TIR120.002155.
46. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, *et al*. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;**12**(1):1088. doi: 10.1038/s41467-021-21246-9.
47. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, *et al*. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;**32**(4):381-386. doi: 10.1038/nbt.2859.
48. Pattison LA, Rickman RH, Hilton H, Dannawi M, Wijesinghe SN, Ladds G, *et al*. Activation of the proton-sensing GPCR, GPR65 on fibroblast-like synoviocytes contributes to inflammatory joint pain. *Proc Natl Acad Sci U S A*. 2024;**121**(51):e2410653121. doi: 10.1073/pnas.2410653121.
49. Khan NM, Diaz-Hernandez ME, Martin WN, Patel B, Chihab S, Drissi H. pH-sensing G protein-coupled orphan receptor GPR68 is expressed in human cartilage and correlates with degradation of extracellular matrix during OA progression. *PeerJ*. 2023;**11**:e16553. doi: 10.7717/peerj.16553.
50. Cashman JR, Zhang J. Human flavin-containing monooxygenases. *Annu Rev Pharmacol Toxicol*. 2006;**46**:65-100. doi: 10.1146/annurev.pharmtox.46.120604.141043.