Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Glucose Metabolism-Related Crosstalk Genes Between Osteoporosis and Frozen Shoulder Identified by Integrated Bioinformatics and Machine Learning Reveal Shared Diagnostic Biomarkers and Immune Microenvironment Characteristics

Wei Hao
Affiliated Hospital of Shandong University of Traditional Chinese Medicine
Long Fang
Shandong Provincial Third Hospital Affiliated with Shandong University
Baolong Wang
Shandong Provincial Third Hospital Affiliated with Shandong University
Ziwei Hou
Affiliated Hospital of Shandong University of Traditional Chinese Medicine
Lizhong Jing
Affiliated Hospital of Shandong University of Traditional Chinese Medicine
Jiushan Yang
yjs_ortho@163.com

Affiliated Hospital of Shandong University of Traditional Chinese Medicine

Article

Keywords: Osteoporosis, Frozen Shoulder, Glucose Metabolism, Immune Microenvironment, LTF

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10384204/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 30th, 2026. See the published version at https://doi.org/10.1038/s41598-026-78295-7.

Glucose Metabolism-Related Crosstalk Genes Between Osteoporosis and Frozen Shoulder Identified by Integrated Bioinformatics and Machine Learning Reveal Shared Diagnostic Biomarkers and Immune Microenvironment Characteristics

Wei Hao1[#], Long Fang2[#], Baolong Wang2, Ziwei Hou1, Lizhong Jing1, Jiushan Yang1*

[#]the authors contribute to this article equally.

*corresponding author, e-mail: yjs_ortho@163.com

[1]Department of Orthopedics, The Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong, 250011, China.

[2]Department of Spine surgery, Shandong Provincial Third Hospital Affiliated with Shandong University, Jinan, Shandong, 250031, China.

Abstract

Osteoporosis (OP) and frozen shoulder (FS) are debilitating musculoskeletal diseases frequently complicated by metabolic disorders, while their overlapping molecular mechanisms are largely unclear. This study explored their shared pathogenesis based on glucose metabolism-related genes (GMRGs) and constructed dual-disease diagnostic models. We integrated transcriptomic data from the Gene Expression Omnibus database, screened core crosstalk genes using differential expression analysis, weighted gene co-expression network analysis and three machine learning algorithms[least absolute shrinkage and selection operator , logistic regression, and support vector machine recursive feature elimination]. Four shared key GMRGs[MARCKS, MAP3K1, LTF and ZMIZ1] were identified. The established models showed excellent diagnostic performance for OP in training set GSE56815, lower performance in validation set GSE7429, moderate-to-high efficacy for FS in training set GSE238053, and high accuracy in validation set GSE140731, with LTF showing the greatest contribution in the nomogram analysis. Protein–protein interaction analysis identified NFKBIA, LTF and HP as hub genes. Immune infiltration analysis showed that HP was positively correlated with neutrophils in OP, NFKBIA was negatively correlated with memory B cells in OP, and LTF was positively correlated with M1 macrophages in FS, suggesting a potential glucose metabolism-immune regulatory axis. Consensus clustering classified OP into two subtypes with distinct immune

microenvironments. Virtual screening using DrugCLIP identified potential LTF-targeting candidate compounds, including P207-9104 and Z169905102, with molecular docking showing strong predicted binding affinities for LTF-P207-9104 and LTF-Z169905102. This study provide novel insights into the comorbidity mechanism of OP and FS, provides promising diagnostic biomarkers and therapeutic candidates, warranting further experimental validation.

Keywords: Osteoporosis, Frozen Shoulder, Glucose Metabolism, Immune Microenvironment, LTF

Introduction

Osteoporosis (OP) and frozen shoulder (FS) represent two prevalent yet distinct pathological conditions within the musculoskeletal system, each imposing a substantial burden on global healthcare systems. Osteoporosis is a systemic skeletal disorder characterized by compromised bone strength, predisposing individuals to an increased risk of fragility fractures. This condition arises from a progressive loss of bone mass and deterioration of bone microarchitecture, often remaining clinically silent until fracture occurs. In contrast, frozen shoulder, also known as adhesive capsulitis, is a chronic inflammatory condition manifesting as progressive pain and severe restriction of shoulder motion. Despite their distinct clinical manifestations, both diseases share several epidemiological and pathophysiological features, including a higher prevalence in postmenopausal women and a significant association with aging and metabolic disorders[1–3]. The World Health Organization has identified OP as a major public health concern, affecting an estimated 200 million women worldwide[4] Similarly, FS has a global incidence ranging from 2% to 5%, with a substantial proportion of patients experiencing persistent functional limitations[5]. The conventional diagnostic and therapeutic strategies for these conditions remain suboptimal: OP diagnosis mainly relies on bone mineral density (BMD) measurement, which often identifies the disease after substantial bone loss, whereas FS diagnosis is primarily clinical and lacks early molecular biomarkers. Although anti-osteoporotic drugs can reduce fracture risk[6], and conservative or surgical treatments for FS can relieve pain and improve mobility[7], these strategies are limited by delayed diagnosis, variable responses, slow recovery, and persistent stiffness in some patients[7]. Therefore, there is an urgent need to elucidate the shared molecular mechanisms underlying these two conditions and to identify reliable diagnostic biomarkers that may facilitate early intervention and future personalized management.

Compelling evidence has emerged over the past decade linking metabolic disturbances, particularly glucose metabolism dysregulation, to the pathogenesis of both OP and FS. Diabetes mellitus, characterized by hyperglycemia and insulin resistance, is a well-established metabolic risk

factor for accelerated bone loss and increased fracture risk[8]. Hyperglycemia exerts detrimental effects on bone marrow mesenchymal stem cells (BMSCs), shifting their differentiation potential from osteogenesis toward adipogenesis, thereby impairing bone formation[9]. Furthermore, glucose metabolic pathways, including glycolysis and oxidative phosphorylation, provide the energetic foundations for both osteoblast and osteoclast activity, and disturbances in these pathways have been implicated in the pathogenesis of type 2 diabetic osteoporosis[10]. The Warburg effect, a phenomenon of enhanced glycolysis even in the presence of oxygen, has been proposed as a key energy metabolic event in OP, influencing osteoblast maturation through the modulation of osteocalcin and WNT signaling[11]. In parallel, hyperglycemia-induced inflammatory pathways such as NF-κB signaling and oxidative stress exacerbate inflammatory and fibrotic changes in the shoulder capsule, thereby promoting the development of frozen shoulder (FS)[12,13]. Despite these parallel lines of evidence, the specific molecular mechanisms through which glucose metabolism-related genes (GMRGs) may contribute to the shared pathogenesis of OP and FS remain largely unexplored, representing a significant gap in our understanding of their comorbidity.

The interrelationship between FS and OP has been increasingly recognized in clinical practice, yet the underlying molecular crosstalk remains poorly defined. A recent study has identified the *p38 MAPK* signaling pathway as a potential shared mechanism, with the *p38α* inhibitor *TAK715* demonstrating efficacy in reversing both fibrosis and osteoporosis in a rat model[5]. This finding underscores the potential for a common therapeutic strategy targeting shared molecular pathways. Beyond signaling pathways, the immune microenvironment plays a crucial and distinct role in each disease. In OP, chronic low-grade inflammation drives osteoclast activation and bone resorption, with aberrant immune cell infiltration, such as neutrophils and macrophages, contributing to bone loss[14]. In FS, a pronounced inflammatory and fibrotic response, involving macrophages and T cells, leads to capsular thickening and contracture[15]. Crucially, immune cells are heavily dependent on glucose metabolism for their activation and effector functions; for instance, M1 macrophages rely on aerobic glycolysis (Warburg effect) to fuel their pro-inflammatory cytokine production, while regulatory T cells preferentially utilize oxidative phosphorylation[16]. This suggests that GMRGs may act as a critical bridge connecting metabolic dysregulation with immune dysfunction, thereby orchestrating the progression of both diseases. However, bioinformatics studies systematically screening shared GMRGs between OP and FS and exploring their associations with immune infiltration and disease-related processes remain limited.

To address this knowledge gap, the present study was designed to harness the power of integrative bioinformatics approaches, leveraging publicly available transcriptomic datasets from the

Gene Expression Omnibus (GEO) database. We collected gene expression profiles for both OP (GSE56815) and FS (GSE140731), along with independent validation datasets (GSE7429 for OP and GSE238053 for FS). Our analytical framework was methodologically robust, employing a multi-step strategy to identify disease-specific and disease-shared GMRGs. First, we performed differential expression analysis between disease and control samples for each condition. Second, we employed weighted gene co-expression network analysis (WGCNA) to identify modules of co-expressed genes significantly correlated with each disease phenotype, thereby capturing broader biological networks. Third, OP-related WGCNA module genes were intersected with OP glucose metabolism-related differentially expressed genes (GMRDEGs), and the resulting module genes were further overlapped with FS differentially expressed genes to identify crosstalk genes (CGs). Finally, and most critically, we applied a suite of three distinct machine learning algorithms-Logistic regression, support vector machine recursive feature elimination (SVM-RFE), and the least absolute shrinkage and selection operator (LASSO)-to systematically and conservatively screen for the most robust and reproducible cross-talk genes (CGs) between the two diseases. This multi-algorithm, multi-dataset screening and validation strategy was used to improve the reliability of candidate gene selection and reduce false-positive findings associated with single-method analyses.

The primary objective of this study was three-fold. Firstly, we aimed to systematically identify the key GMRGs that constitute the molecular crosstalk between OP and FS, thereby elucidating shared pathogenic mechanisms linking metabolic dysregulation to these two musculoskeletal disorders. Secondly, we sought to construct and validate diagnostic prediction models based on these identified crosstalk genes and to evaluate their diagnostic performance and potential clinical utility for both conditions. Finally, we endeavored to characterize the immune microenvironment associated with these key crosstalk genes through comprehensive immune infiltration analysis, to investigate the hypothesis that GMRGs influence disease pathogenesis via modulation of the immune landscape. By achieving these aims, this study may provide a molecular perspective on the comorbidity of OP and FS and offer candidate biomarkers and potential therapeutic targets for future experimental validation and translational research.

1. Materials and Methods

1.1 Data Acquisition and Preprocessing

The Osteoporosis (OP) datasets GSE56815[17] and GSE7429[18] were downloaded from the GEO database[19] using the R package GEOquery (Version 2.70.0). GSE56815 contained 20 OP samples and 20 control samples, whereas GSE7429 contained 10 OP samples and 10 control

samples. Both datasets contained samples from *Homo sapiens*. The tissue source for GSE56815 was monocytes, while that for GSE7429 was B cells in blood. Both datasets utilized the GPL96 microarray platform. All OP and Control samples were included in this study. Detailed information is provided in Table 1.

Table 1 GEO Microarray Chip Information

	GSE56815	GSE7429	GSE140731	GSE238053
Platform	GPL96	GPL96	GPL24676	GPL24676
Species	Homo sapiens	Homo sapiens	Homo sapiens	Homo sapiens
Tissue	monocytes	B cells in blood	Shoulder Capsule	synovium
Samples in OP group	20	10	/	/
Samples in FS group	/	/	22	4
Samples in Control group	20	10	26	4

GEO, Gene Expression Omnibus;

Glucose metabolism-related genes (GMRGs) were collected from the GeneCards database[20]. Using “Glucose metabolism” as the search keyword, we retained only protein-coding genes with a relevance score > 0, yielding 3,263 GMRGs. Additionally, seven GMRGs were retrieved from published literature via PubMed using the same keyword. After merging and removing duplicates, a total of 3,266 GMRGs were obtained (Table S1).

Table S1: Full list of glucose metabolism-related genes (GMRGs) collected from GeneCards and PubMed literature(as shown in supplementary material)

The Frozen Shoulder (FS) datasets GSE140731[21] and GSE238053[22] were also downloaded from the GEO database using the GEOquery package. GSE140731 contained 22 FS samples and 26 control samples, whereas GSE238053 contained 4 FS samples and 4 control samples. Samples were derived from *Homo sapiens*, with tissue sources being the shoulder capsule for GSE140731 and synovium for GSE238053. Both datasets used the GPL24676 microarray platform (see Table 1). All FS and Control samples were included in the analysis.

Normalization was performed using the normalizeBetweenArrays function from the R package limma (Version 3.62.1) for both the OP datasets (GSE56815, GSE7429) and the FS datasets (GSE140731, GSE238053).

1.2 Screening of Differentially Expressed Genes Related to Osteoporosis and Glucose Metabolism

For dataset GSE56815, samples were divided into OP (20 cases) and Control (20 cases) groups. Differential expression analysis was performed using the limma package. Differentially expressed genes (DEGs) were defined by thresholds of $|logFC| > 0$ and p-value < 0.05, with Benjamini-Hochberg (BH) correction applied. Genes with $logFC > 0$ and p-value < 0.05 were considered up-regulated, while those with $logFC < 0$ and p-value < 0.05 were considered down-regulated. Volcano plots were generated using the ggplot2 package.

To identify glucose metabolism-related differentially expressed Genes (GMRDEGs), the intersection of DEGs ($|logFC| > 0$, p-value < 0.05) from GSE56815 and the GMRG list was calculated and visualized using a Venn diagram. Heatmaps displaying the top 10 up-regulated and down-regulated GMRDEGs were generated using the pheatmap package (Version 1.0.12).

1.3 Screening of Differentially Expressed Genes Related to Frozen Shoulder and Glucose Metabolism

For dataset GSE140731, samples were divided into FS (n = 22) and Control (n = 26) groups. Differential expression analysis was performed using the DESeq2 package. DEGs were defined by thresholds of $|logFC| > 0$ and p-value < 0.05, with BH correction. Up-regulated and down-regulated genes were identified based on logFC direction. Volcano plots were generated using ggplot2. Heatmaps showing the top 10 up-regulated and down-regulated DEGs for GSE140731 were generated using pheatmap.

1.4 Weighted Gene Co-expression Network Analysis (WGCNA)

WGCNA was performed using the R package WGCNA[23] to identify gene modules correlated with clinical traits. First, the variance of all genes in GSE56815 was calculated, and the top 50% most variable genes were selected. Pairwise correlations were computed, and a weighted adjacency matrix was constructed to approximate a scale-free topology. A hierarchical clustering tree was built, and dynamic tree cutting was used to define modules. Parameters included: minimum module size = 25, soft threshold power = 6 (chosen to achieve a scale-free fit index ≥ 0.8), no sample trimming, and a module merging cut height of 0.3. Module-trait relationships were calculated for OP and Control groups. Modules with $|correlation\ coefficient| > 0.3$ and p-value < 0.05 were selected. Genes within these significant modules were intersected with GMRDEGs to identify Module Genes. Finally, Module Genes were intersected with FS DEGs to identify Crosstalk Genes (CGs).

1.5 Gene Ontology (GO) and KEGG Pathway Enrichment Analysis

GO[24] and KEGG[25] enrichment analyses were conducted on CGs using the R package clusterProfiler (Version 4.10.0)[26]. GO terms were categorized into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Enrichment significance was determined by *p*-value < 0.05 and False Discovery Rate (FDR/q-value) < 0.05, with the *p*-value adjustment method set to none. The top 5 enriched terms for BP, CC, MF, and KEGG pathways were selected based on ascending *p*-value order.

1.6 Gene Set Variation Analysis (GSVA)

GSVA[27] was performed to evaluate pathway enrichment differences between groups. Gene sets were obtained from the Molecular Signatures Database (MSigDB; h.all.v2023.1.Hs.symbols.gmt)[28]. Using the R package GSVA (Version 1.50.0), GSVA was applied to all genes in GSE56815 (OP vs. Control) and GSE140731 (FS vs. Control). Significant pathways were identified by *p*-value < 0.05. The top 10 positively enriched and top 10 negatively enriched pathways, ranked by logFC, were selected for further analysis.

1.7 Construction of Diagnostic Models for Osteoporosis and Frozen Shoulder

Three machine learning algorithms were employed to identify model genes: logistic regression, SVM-RFE, and LASSO regression.

Logistic Regression: CGs with *p*-value < 0.05 were selected to construct the initial model.

SVM-RFE[29**]:** Based on the genes selected by Logistic Regression (seed set to 2024), SVM-RFE was performed to select features with minimal classification error and maximal accuracy.

LASSO Regression: Using the R package glmnet[30](seed set to 2024, family="binomial"), LASSO regression was applied to the genes selected by SVM-RFE. This method adds a penalty term (lambda × absolute slope) to reduce overfitting and improve generalization. Results were visualized using diagnostic model plots and variable trajectory plots. The intersection of model genes from OP and FS analyses yielded key genes. Subsequently, LASSO regression was re-applied to these key genes to finalize the diagnostic models for both conditions.

Validation of Diagnostic Models

Risk scores (RiskScore) were calculated based on the final Key Genes using the formula: $RiskScore = \sum_i Coefficient(gene_i) * mRNA\ Expression(gene_i)$.

Receiver Operating Characteristic (ROC) curves were plotted using the pROC package (Version 1.18.5)[31] to assess diagnostic performance, where an Area Under the Curve (AUC) closer to 1 indicates better accuracy. Nomograms[32] were constructed using the R package rms[33](Version 6.7-1) to visualize the contribution of each key gene to diagnostic prediction. Decision Curve Analysis (DCA)[34] was performed using the R package ggDCA (Version 1.1) to evaluate clinical utility. Functional similarity among key genes was assessed using the R package GOSemSim[35]. Internal validation was performed on independent datasets (GSE7429 for OP; GSE140731 for FS) using similar metrics.

1.8 Expression Validation and Correlation Analysis

Differences in key gene expression between disease and control groups were visualized using grouped comparison plots. Spearman correlation analysis was performed to assess relationships among key genes, with results displayed as heatmaps using pheatmap. Correlation strength was interpreted as follows: $|r| < 0.3$ (weak/no correlation), $0.3 \leq |r| < 0.5$ (weak), $0.5 \leq |r| < 0.8$ (moderate), and $|r| \geq 0.8$ (strong).

1.9 Protein-Protein Interaction (PPI) Network and Hub Gene Identification

A PPI network for CGs was constructed using the STRING database[36] with a minimum required interaction score > 0.15. Hub genes were identified using four centrality algorithms [Degree, Maximum Neighborhood Component (MNC), Edge Percolated Component(EPC), Degree Maximum Neighborhood Component(DMNC)] from the CentiServer package[37]. The top 5 genes from each algorithm were selected, and their intersection was defined as the hub genes.

1.10 Immune Infiltration Analysis

Immune cell infiltration was estimated using the CIBERSORT algorithm[38]. For GSE56815 (OP) and GSE140731 (FS), immune cell fractions were calculated, and bar plots were generated to show proportions. Differences in immune cell abundance between groups were compared using ggplot2. Spearman correlation was used to analyze relationships between immune cells and between hub genes and immune cells, visualized via heatmaps and bubble plots, respectively.

1.11 Regulatory Network Construction

Transcription factors (TFs) regulating hub genes were identified using the ChIPBase database[39], and mRNA-TF regulatory networks were visualized using Cytoscape[40]. miRNAs

targeting hub genes were retrieved from StarBase v3.0[41], and mRNA-miRNA regulatory networks were constructed and visualized.

1.12 Subtype Construction and Immune Analysis in Osteoporosis

Consensus clustering[42] (using R package ConsensusClusterPlus[43]) was performed on GSE56815 based on model genes to identify disease subtypes. Parameters included k = 2–9, 50 iterations, sampling 80% of samples, clusterAlg = 'km', and distance = 'euclidean'. The optimal number of clusters was determined according to the consensus clustering results, and subsequent analyses were performed to compare gene expression and immune infiltration patterns among the identified OP subtypes. Expression differences of model genes across subtypes were analyzed. Immune infiltration in subtypes was assessed using both CIBERSORT and Single-Sample Gene Set Enrichment Analysis (ssGSEA)[44] for 28 immune cell types. For ssGSEA, the relative enrichment scores of 28 immune cell types were calculated based on a previously published immune cell marker gene set. Correlations between CGs and immune cells in each subtype were analyzed using Spearman correlation analysis.

1.13 Virtual Screening (DrugCLIP)

Potential small molecule compounds targeting the hub gene *LTF* were screened using DrugCLIP[45]. Protein structures were obtained from the PDB database. Binding pockets were predicted using CB-Dock2, and the largest pocket (C1 site) was selected as the target region, with its center coordinates and pocket radius recorded. Compounds were ranked in descending order according to DrugCLIP scores, and top-ranked candidates were selected for molecular docking and visualization. Binding affinity was classified as: weak/non-binding (Vina score > -5 kcal/mol), moderate (-7 < Vina score < -5 kcal/mol), or strong binding (Vina score < -7 kcal/mol).

2. Results

2.1 Flow chart(Figure. 1)

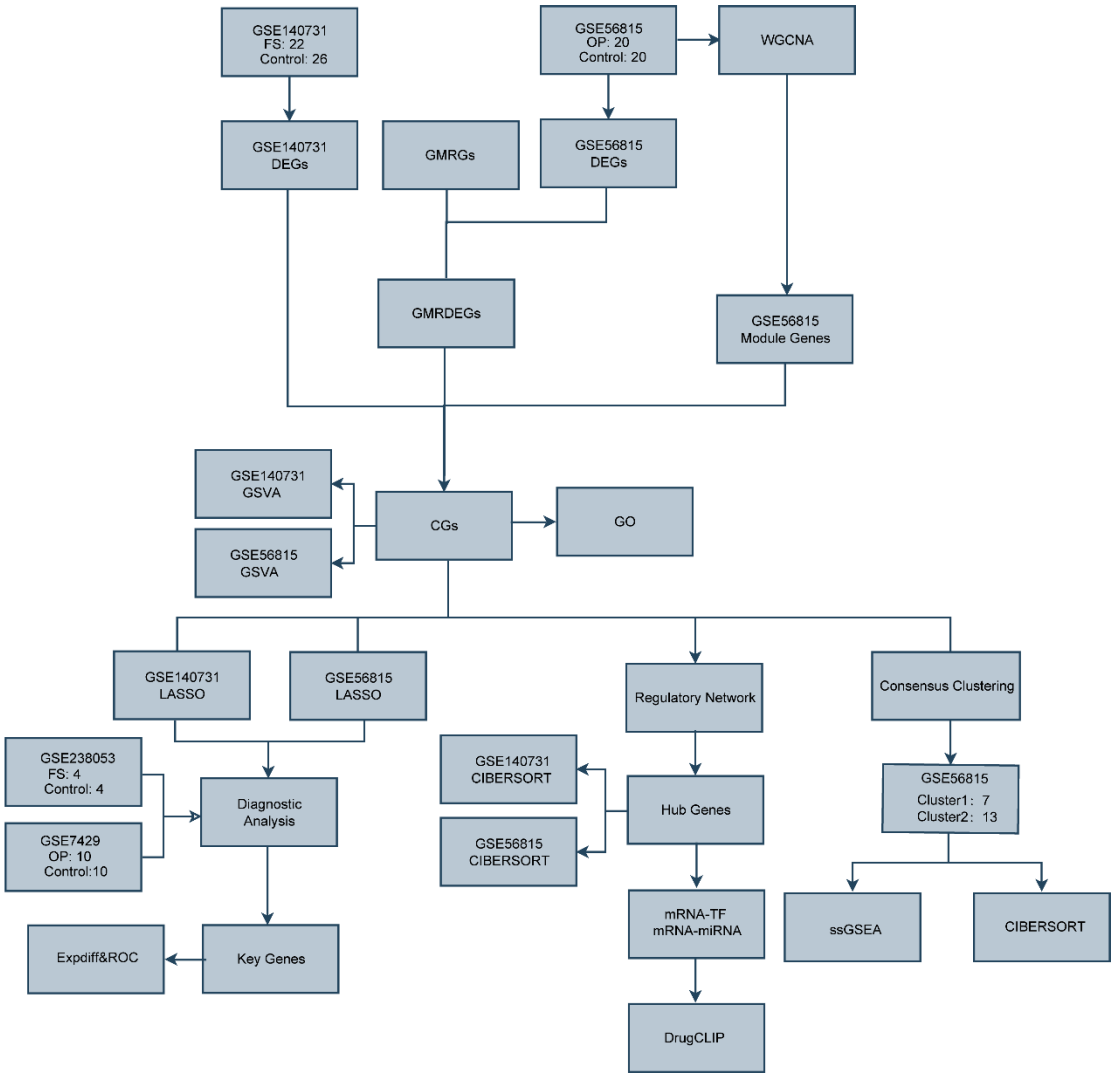


Fig.1 Flow Chart for the Comprehensive Analysis of CGs

OP, Osteoporosis; FS, Frozen shoulder; WGCNA, Weighted Gene Co-expression Network Analysis; DEGs, Differentially Expressed Genes; CGs, Crosstalk Genes; GSVA, Gene Set Variation Analysis; GO, Gene Ontology; LASSO, Least Absolute Shrinkage and Selection Operator; ROC, Receiver Operating Characteristic.

2.2 Data Preprocessing and Normalization

Normalization using limma resulted in consistent expression distributions for both OP (GSE56815, GSE7429) and FS (GSE140731, GSE238053) datasets, as confirmed by boxplots (Figures 2A-D and 3A-D), indicating successful normalization and improved comparability among samples.

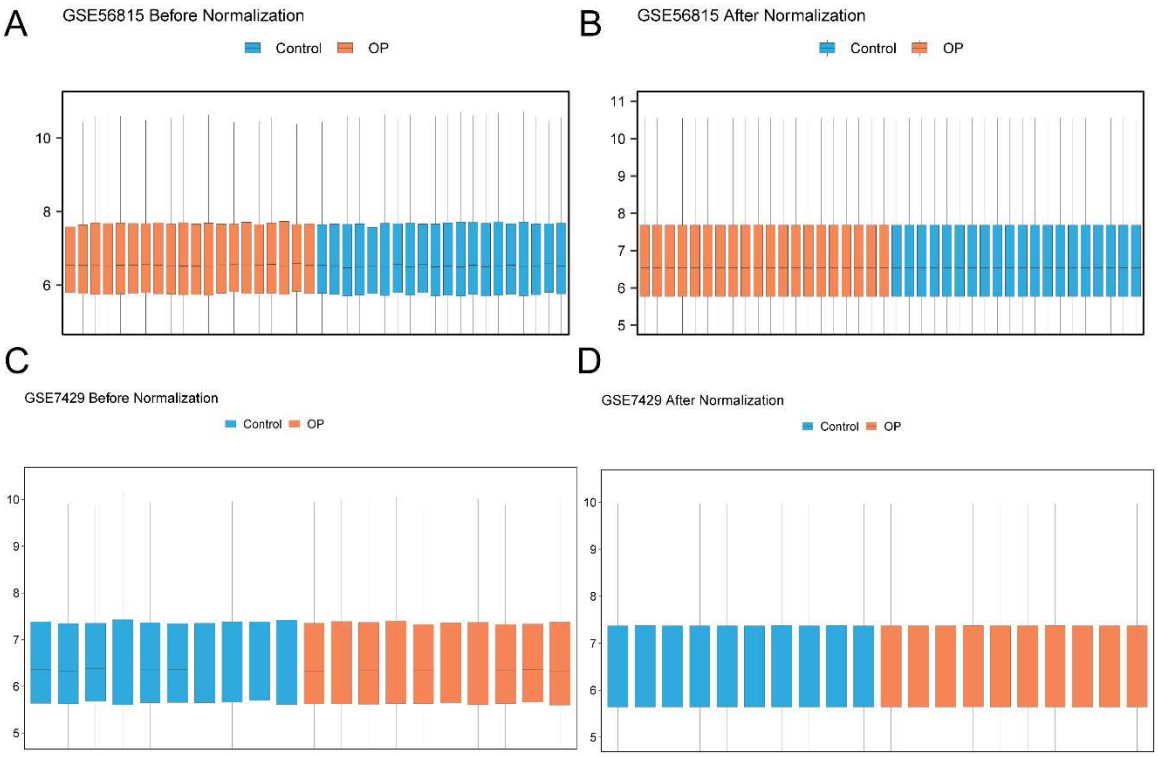


Fig.2 Data cleaning of GSE56815& GSE7429

A. Boxplot of GSE56815 distribution before standardized processing. B. Boxplot of GSE56815 distribution after standardized treatment. C. Boxplot of GSE7429 distribution before standardized treatment. D. Boxplot of GSE7429 distribution after standardized treatment.

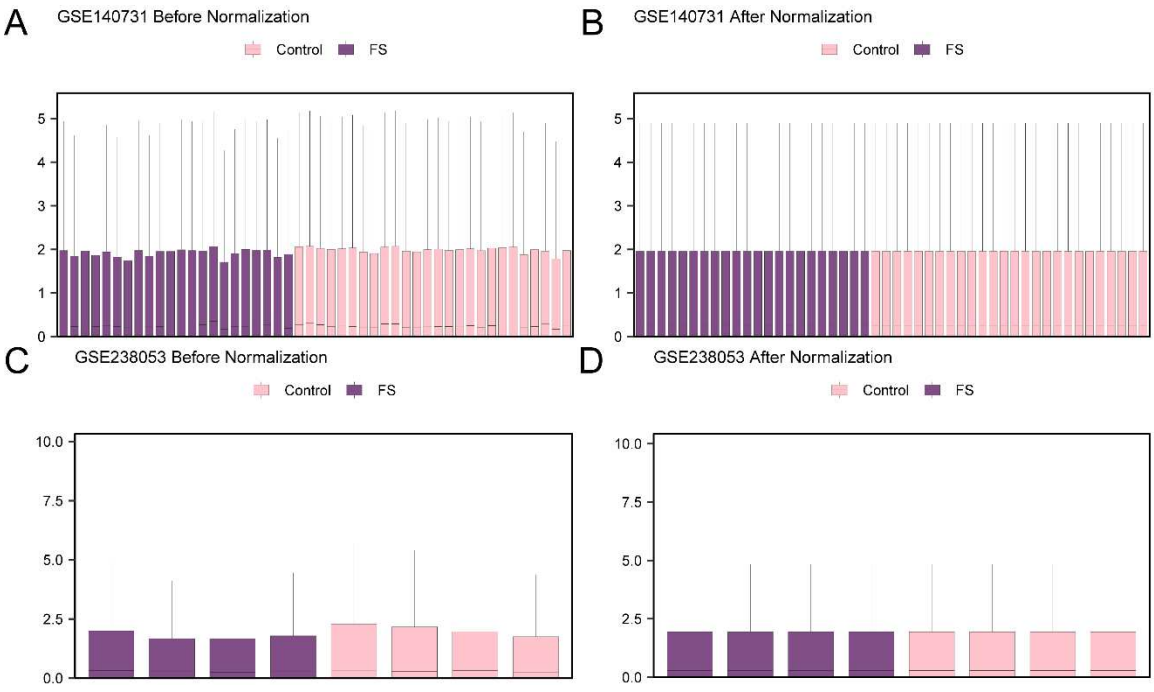


Fig.3 Data cleaning of GSE140731& GSE238053

A. Boxplot of distribution before standardized processing GSE140731. B. Distribution boxplot GSE140731 after standardized treatment. C. Distribution boxplot GSE238053 before standardized treatment. D. Distribution boxplot GSE238053 after standardized treatment.

2.3 Identification of DEGs and GMRDEGs in Osteoporosis

Differential analysis of GSE56815 identified 2,240 DEGs (|logFC| > 0, $p < 0.05$), including 1,226 up-regulated and 1,014 down-regulated genes (Figure 4A). Intersection with GMRGs yielded 522 GMRDEGs (Figure 4B). Heatmaps of the top 10 up- and down-regulated GMRDEGs are shown in Figure 4C.

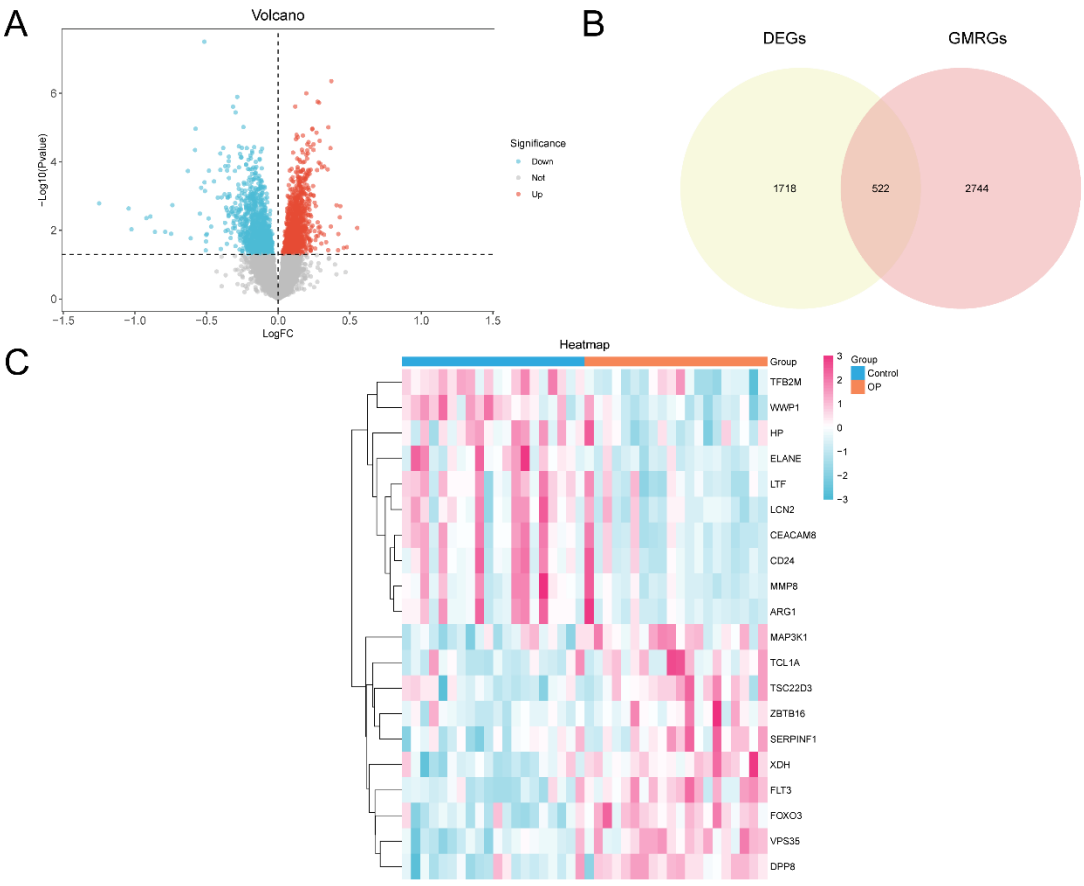


Fig.4 Differential Gene Expression Analysis

A. Volcano plot of differentially expressed genes analysis of OP and Control in GSE56815. B. Differentially expressed genes (DEGs) and glucose metabolism-related genes (GMRGs) Venn diagram in GSE56815. C. Heat map of differentially expressed genes (GMRDEGs) related to glucose metabolism in GSE56815. OP, Osteoporosis; DEGs, Differentially Expressed Genes; GMRGs, Glucose metabolism-Related Genes; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes. Orange is OP, blue is Control. Red represents high expression and blue represents low expression in the heat map.

2.4 Identification of DEGs in Frozen Shoulder

Differential analysis of GSE140731 using DESeq2 identified 10,238 DEGs, including 5,372 up-regulated and 4,866 down-regulated genes (Figure 5A). Heatmaps of the top 10 DEGs are presented (Figure 5B).

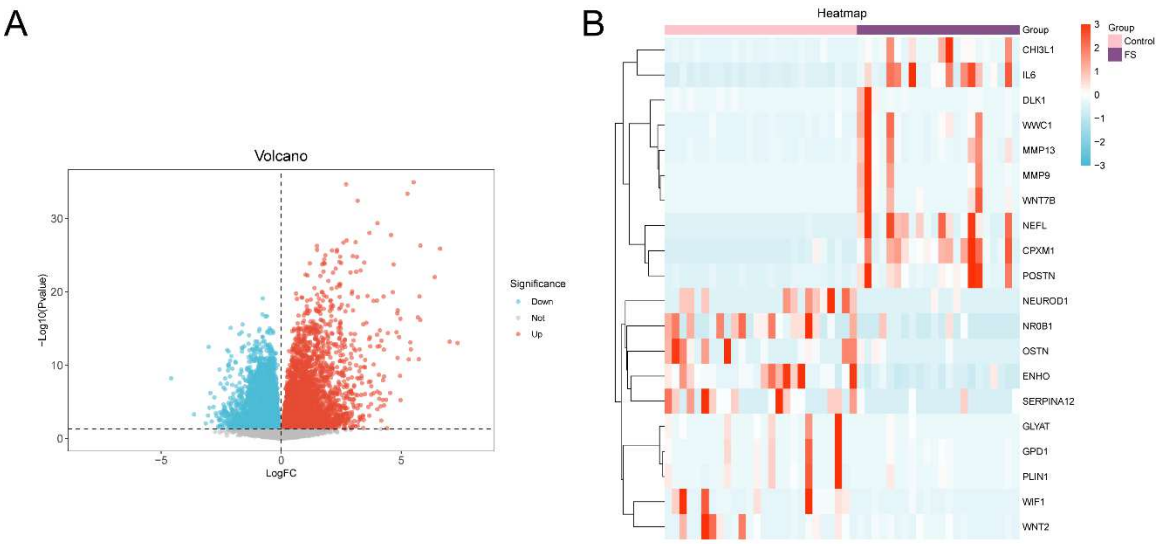


Fig.5 Differential Gene Expression Analysis

A. Volcano plot of differentially expressed genes analysis between FS and Control in GSE140731. B. Heat map of differentially expressed genes (DEGs) in GSE140731. FS, Frozen shoulder; DEGs, Differentially Expressed Genes; GMRGs, Glucose metabolism-Related Genes; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes. FS in purple and Control in pink. Red represents high expression and blue represents low expression in the heat map.

2.5 WGCNA Analysis

In GSE56815, the optimal soft threshold power was determined to be 6 (scale-free fit index = 0.8) (Figure 6A). Hierarchical clustering identified 12 modules (Figure 6B-C). Three modules (MEmagenta, MEgreen, MEsalmon) showed significant correlation with the OP trait ($|r| > 0.3$, $p < 0.05$) (Figure 6D). Intersection of genes in these modules with GMRDEGs resulted in 37 module genes (Figure 6E-G), including representative genes such as *MAPK8, TM6SF2, SLC37A1, LTF, MPO, HP, NFKBIA, MARCKS, MAP3K1, and SFXN3*.

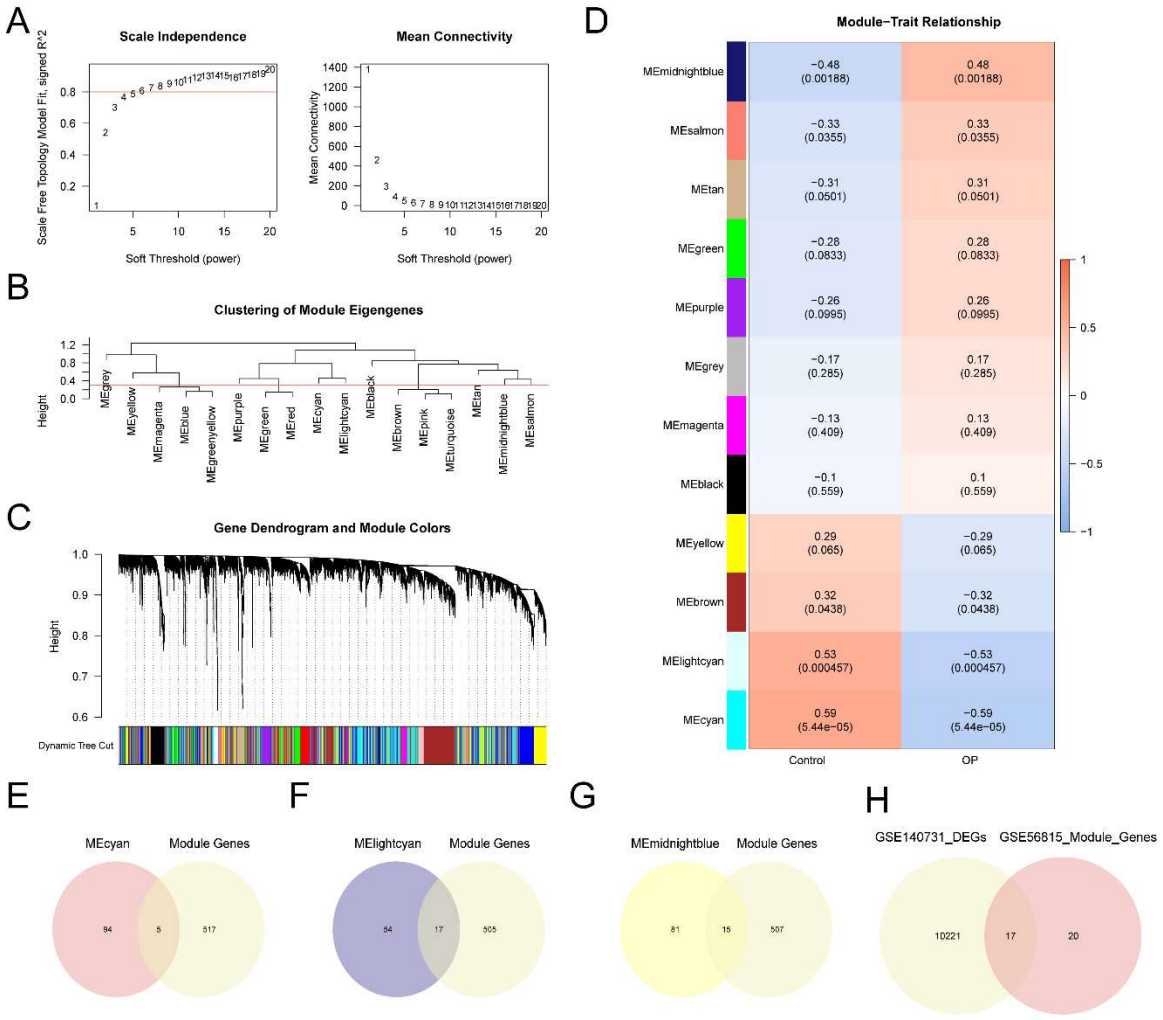


Fig.6 WGCNA Analysis

A. Scale-free network display of the best soft threshold from weighted gene co-expression network analysis (WGCNA) , the left panel shows the best soft threshold and the right panel shows the network connectivity under different soft threshold conditions. B. Display of module clustering results of genes with top 50% variance. C. Presentation of cluster results for genes with top 50% variance. The upper part is divided into hierarchical clustering dendrogram, and the lo wer part is divided into gene modules. D. The results of correlation analysis between the gene cluster modules with th e top 50% variance and the OP and Control groups are shown. E-G. Venn diagram of 585 differentially expressed gene s related to glucose metabolism and genes contained in MEmagenta, MEgreen, and MEsalmon modules. H. Venn diag ram display of module genes and differential genes containing genes of periarthritis of shoulder. OP, Osteoporosis; W GCNA, Weighted Gene Co-Expression Network Analysis; GMRDEGs, Glucose metabolism-Related Differentially Expres sed Genes. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, between 0.3 a nd 0.5 was weak correlation, between 0.5 and 0.8 was moderate correlation, and above 0.8 was strong correlation. R ed represents positive correlation, blue represents negative correlation.

2.6 GO and KEGG Enrichment Analysis of CGs

Enrichment analysis of the 17 CGs revealed significant involvement in biological processes such as defense response to fungus, response to fungus, hydrogen peroxide catabolic process, regulation of hemopoiesis, and regulation of myeloid cell differentiation. Molecular functions included antioxidant activity and heparin binding. Cellular components were enriched in endocytic vesicle lumen and secretory granule lumen. KEGG analysis highlighted the Phagosome pathway(permission granted by Kanehisa Laboratories) (Table 2, Figures 7A-E).

Table 2 Results of GO and KEGG Enrichment Analysis for CGs

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
BP	GO:0050832	defense response to fungus	3/16	67/18986	2.28E-05	2.28E-05	5.49E-03
BP	GO:0009620	response to fungus	3/16	78/18986	3.59E-05	3.59E-05	5.49E-03
BP	GO:0042744	hydrogen peroxide catabolic process	2/16	31/18986	3.05E-04	3.05E-04	2.98E-02
BP	GO:1903706	regulation of hemopoiesis	4/16	433/18986	3.91E-04	3.91E-04	2.98E-02
BP	GO:0045637	regulation of myeloid cell differentiation	3/16	219/18986	7.59E-04	7.59E-04	4.02E-02
CC	GO:0071682	endocytic vesicle lumen	3/17	23/19960	8.99E-07	8.99E-07	4.73E-05
CC	GO:0034774	secretory granule lumen	4/17	324/19960	1.37E-04	1.37E-04	1.94E-03
CC	GO:0060205	cytoplasmic vesicle lumen	4/17	329/19960	1.46E-04	1.46E-04	1.94E-03
CC	GO:0031983	vesicle lumen	4/17	330/19960	1.47E-04	1.47E-04	1.94E-03
CC	GO:0042581	specific granule	3/17	160/19960	3.16E-04	3.16E-04	2.98E-03
MF	GO:0016209	antioxidant activity	2/16	86/18737	2.40E-03	2.40E-03	8.79E-02

MF	GO:0008201	heparin binding	2/16	176/18737	9.66E-03	9.66E-03	8.79E-02
MF	GO:0004252	serine-type endopeptidase activity	2/16	178/18737	9.87E-03	9.87E-03	8.79E-02
MF	GO:0008236	serine-type peptidase activity	2/16	195/18737	1.17E-02	1.17E-02	8.79E-02
MF	GO:0061665	SUMO ligase activity	1/16	14/18737	1.19E-02	1.19E-02	8.79E-02
KEGG	hsa04145	Phagosome	3/11	155/9364	6.66E-04	4.93E-02	4.42E-02

GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes;

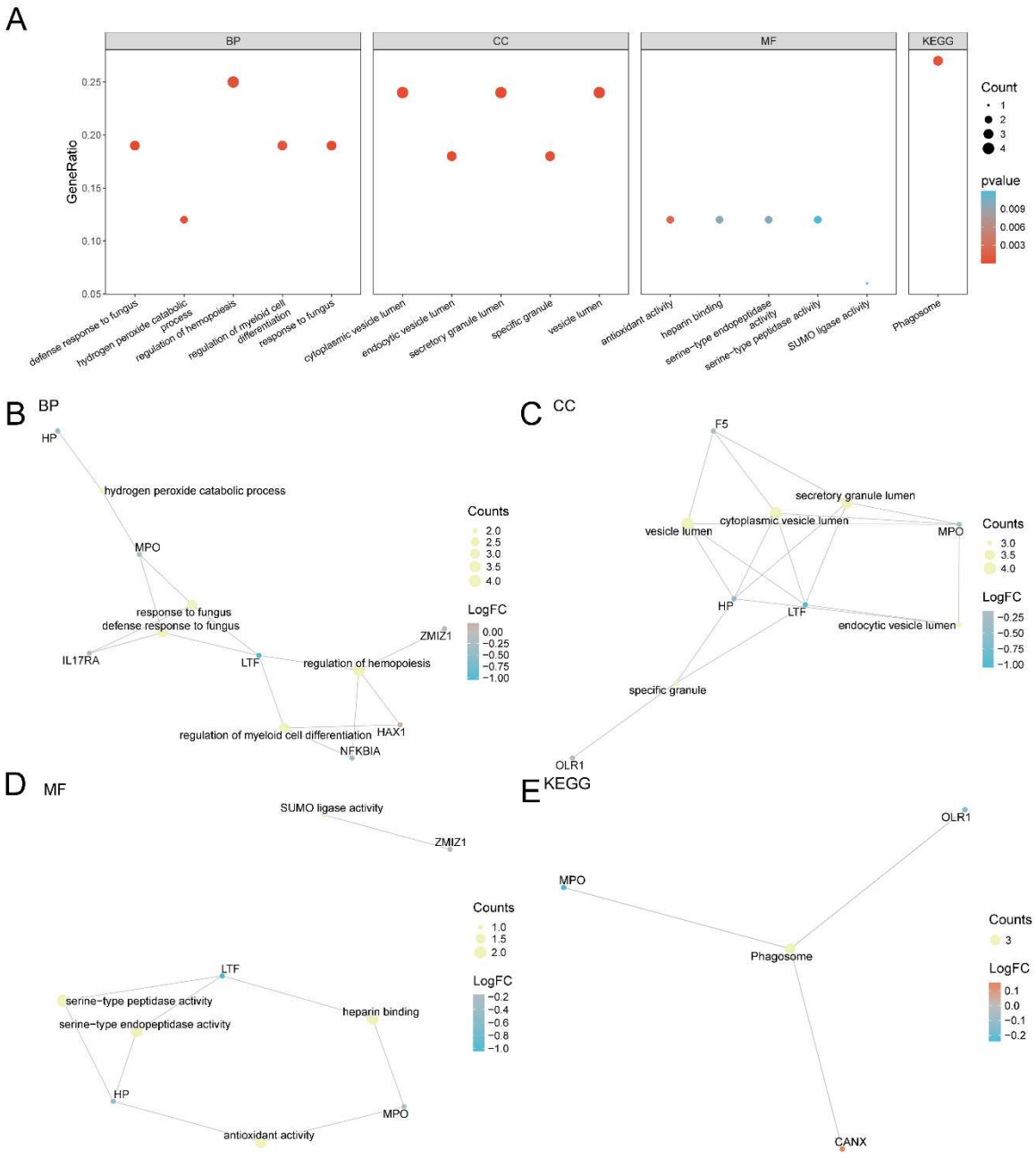


Fig.7 GO and KEGG Enrichment Analysis for CGs

Bubble diagram of gene ontology (GO) and pathway (KEGG) enrichment analysis results of A. CGS showed: biological p rocess (BP), cellular component (CC), molecular function (MF) and biological pathway (KEGG). GO terms and KEGG ter ms are shown on the abscissa. B-e. Network diagram of gene ontology (GO) enrichment analysis results of CGs: BP (B), CC (C), MF (D) and KEGG (E). The yellow node represents the entry, and the line represents the relationship between the entry and the molecule. The more red the color in the molecular node, the larger the LogFC, represents up-regulat ion, the more blue the color, the smaller the LogFC, represents down-regulation. ; GO, Gene Ontology; KEGG, Kyoto E

ncyclopedia of Genes and Genomes; BP, Biological Process; CC, Cellular Component; MF, Molecular Function. The bub ble size in the bubble plot represents the number of genes, and the color of the bubble represents the size of the p-va lue value. The reder the color, the smaller the p-value value, and the bluer the larger the p-value value. The screening criteria for gene ontology (GO) enrichment analysis were p value < 0.05 and FDR value (q value) < 0.05.

2.7 GSVA Results

GSVA identified distinct pathway enrichments between OP and Control groups in GSE56815 (Table 3, Figure 8A-B) and between FS and Control groups in GSE140731 (Table 4, Figure 9A-B). Representative positively and negatively enriched pathways for each comparison are listed in Tables 3 and 4 and visualized in Figures 8B and 9B. The top 10 positively and negatively enriched pathways were visualized.

Table 3 Results of GSVA for GSE56815

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
REACTOME_SIGNALING_BY_LRP5_MUTANTS	5.80E-01	-1.46E-02	4.83E+00	1.37E-05	1.05E-02	3.04E+00
REACTOME_TRYPTOPHAN_CATABOLISM	5.80E-01	-1.37E-02	5.21E+00	3.70E-06	8.50E-03	4.21E+00
KEGG_MEDICUS_REFERENCE_DNA_DEGRADATION_BY_EXTRACELLULAR_ENDOLYSOSOMAL_DNASE	5.20E-01	1.35E-02	4.72E+00	1.99E-05	1.05E-02	2.70E+00
REACTOME_SPERM_MOTILITY_AND_TAXES	4.90E-01	-1.14E-01	2.46E+00	1.76E-02	1.90E-01	-3.28E+00
KEGG_MEDICUS_REFERENCE_NOTCH_MESP2_SIGNALING	4.86E-01	-4.29E-02	3.43E+00	1.22E-03	6.80E-02	-9.71E-01
KEGG_MEDICUS_REFERENCE_STEROID_HORMONE_BIOSYNTHESIS_PROGESTERONE_TO_CORTISOL_CORTISONE	4.51E-01	2.95E-02	3.22E+00	2.28E-03	8.93E-02	-1.52E+00
REACTOME_REACTIONS_SPECIFIC_TO_THE_COMPLEX_N_GLYCAN_SYNTHESIS_PATHWAY	4.28E-01	2.70E-02	3.86E+00	3.33E-04	3.43E-02	1.81E-01
BIOCARTA_GHRELIN_PATHWAY	4.25E-01	1.94E-02	3.94E+00	2.61E-04	3.30E-02	3.99E-01
REACTOME_RUNX1_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_INTERLEUKIN_SIGNALING	4.21E-01	-7.93E-02	2.73E+00	8.73E-03	1.41E-01	-2.68E+00
REACTOME_FORMATION_OF_INTERMEDIATE_MESODERM	4.19E-01	-1.82E-02	3.43E+00	1.24E-03	6.80E-02	-9.85E-01
WP_MIRNA_BIOGENESIS	-4.67E-01	1.38E-02	-2.98E+00	4.52E-03	1.17E-01	-2.11E+00
KEGG_MEDICUS_VARIANT_MUTATION_CAUSED_ABERRANT_HTT_TO_REST_MEDIATED_TRANSCRIPTIONAL_REPRESSION	-4.67E-01	3.04E-02	-4.04E+00	1.91E-04	2.89E-02	6.79E-01

REACTOME_MET_ACTIVATES_RAS_SIGNALING	-4.84E-01	2.79E-03	-4.66E+00	2.48E-05	1.05E-02	2.50E+00
KEGG_MEDICUS_REFERENCE_NOTCH_LIGAND_UBIQUITYLATION	-4.85E-01	-9.80E-02	-3.33E+00	1.65E-03	7.49E-02	-1.23E+00
BIOCARTA_HBX_PATHWAY	-4.89E-01	3.20E-02	-4.13E+00	1.39E-04	2.50E-02	9.62E-01
REACTOME_SUMO_IS_CONJUGATED_TO_E1_UBA2_SAE1	-4.92E-01	5.81E-02	-3.39E+00	1.38E-03	7.06E-02	-1.07E+00
REACTOME_CREB_PHOSPHORYLATION	-5.00E-01	-2.13E-02	-4.31E+00	7.72E-05	1.85E-02	1.48E+00
REACTOME_REGULATION_OF_HMOX1_EXPRESSION_AND_ACTIVITY	-5.15E-01	-2.26E-02	-3.70E+00	5.43E-04	4.45E-02	-2.53E-01
REACTOME_SHC_RELATED_EVENTS_TRIGGERED_BY_IGF1R	-5.22E-01	-3.03E-02	-4.78E+00	1.62E-05	1.05E-02	2.88E+00
WP_EXRNA_MECHANISM_OF_ACTION_AND_BIOGENESIS	-5.78E-01	6.96E-02	-3.02E+00	3.98E-03	1.15E-01	-2.00E+00

GSVA, Gene Set Variation Analysis.

Table 4 Results of GSVA for GSE140731

	logFC	AveExpr	t	P.Value	adj.P.Val	B
WP_MIR_509_3P_ALTERATION_OF_YAP1_ECM_AXIS	8.56E-01	-1.86E-02	9.22E+00	5.20E-13	2.47E-10	1.94E+01
KEGG_MEDICUS_REFERENCE_BLOOD_GROUP_H_O_ANTIGEN_TYPE_2_BIOSYNTHESIS	8.20E-01	-5.51E-02	9.37E+00	2.98E-13	2.10E-10	1.99E+01
REACTOME_RELEASE_OF_APOPTOTIC_FACTORS_FROM_THE_MITOCHONDRIA	7.76E-01	-4.16E-02	7.69E+00	1.93E-10	1.62E-08	1.36E+01
REACTOME_RUNX3_REGULATES_YAP1_MEDIATED_TRANSCRIPTION	7.61E-01	-9.71E-03	7.96E+00	6.81E-11	6.80E-09	1.46E+01
REACTOME_CROSSLINKING_OF_COLLAGEN_FIBRILS	7.49E-01	-8.69E-03	8.47E+00	9.43E-12	1.63E-09	1.66E+01

KEGG_MEDICUS_PATHOGEN_EBV_EBNA3C_TO_P53_MEDIATED_TRANSCRIPTION	7.34E-01	-6.14E-02	8.50E+00	8.25E-12	1.54E-09	1.67E+01
REACTOME_PLATELET_ADHESION_TO_EXPOSED_COLLAGEN	7.22E-01	-3.06E-02	1.13E+01	2.68E-16	1.02E-12	2.67E+01
KEGG_MEDICUS_REFERENCE_NOTCH_LIGAND_UBIQUITYLATION	7.22E-01	-2.13E-02	6.64E+00	1.12E-08	4.66E-07	9.69E+00
KEGG_MEDICUS_VARIANT_TMPRSS2_ERG_FUSION_TO_TRANSCRIPTIONAL_ACTIVATION	7.17E-01	-1.02E-02	6.91E+00	3.90E-09	2.06E-07	1.07E+01
KEGG_MEDICUS_VARIANT_MUTATION_INACTIVATED_TP53_TO_TRANSCRIPTION	7.11E-01	-7.64E-02	8.03E+00	5.15E-11	5.58E-09	1.49E+01
REACTOME_UPTAKE_AND_FUNCTION_OF_DIPHTHERIA_TOXIN	-7.73E-01	1.15E-02	-7.41E+00	5.78E-10	4.00E-08	1.26E+01
REACTOME_COHESIN_LOADING_ONTO_CHROMATIN	-7.79E-01	3.88E-03	-6.95E+00	3.36E-09	1.82E-07	1.09E+01
WP_MAPK_AND_NFKB_SIGNALING_PATHWAYS_INHIBITED_BY_YERSINIA_YOPJ	-7.92E-01	-1.21E-02	-8.72E+00	3.59E-12	1.01E-09	1.75E+01
WP_COVID_19_AND_ENDOTHELIAL_CELL_SENESCENCE	-7.93E-01	4.29E-03	-8.30E+00	1.79E-11	2.61E-09	1.59E+01
KEGG_MEDICUS_VARIANT_SCRAPIE_CONFORMATION_PRPSC_TO_PRNP_PI3K_NOX2_SIGNALING_PATHWAY	-8.02E-01	-5.96E-03	-8.01E+00	5.55E-11	5.85E-09	1.48E+01
REACTOME_ARACHIDONATE_PRODUCTION_FROM_DAG	-8.17E-01	-1.77E-02	-7.01E+00	2.73E-09	1.53E-07	1.11E+01
KEGG_MEDICUS_PATHOGEN_SALMONELLA_SOPB_TO_ANXA2_S100A10_REGULATED_ACTIN_CYTOSKELETON	-8.24E-01	-2.87E-02	-6.57E+00	1.51E-08	5.84E-07	9.40E+00
REACTOME_NR1H2_NR1H3_REGULATE_GENE_EXPRESSION_TO_LIMIT_CHOLESTEROL_UPTAKE	-8.41E-01	-2.52E-02	-7.54E+00	3.39E-10	2.58E-08	1.31E+01
REACTOME_RSK_ACTIVATION	-8.96E-01	1.09E-02	-9.14E+00	7.22E-13	2.86E-10	1.91E+01

KEGG_MEDICUS_REFERENCE_ERK_RSK_SIGNALING	-9.02E-01	9.32E-03	-9.35E+00	3.20E-13	2.10E-10	1.99E+01

GSVA, Gene Set Variation Analysis.

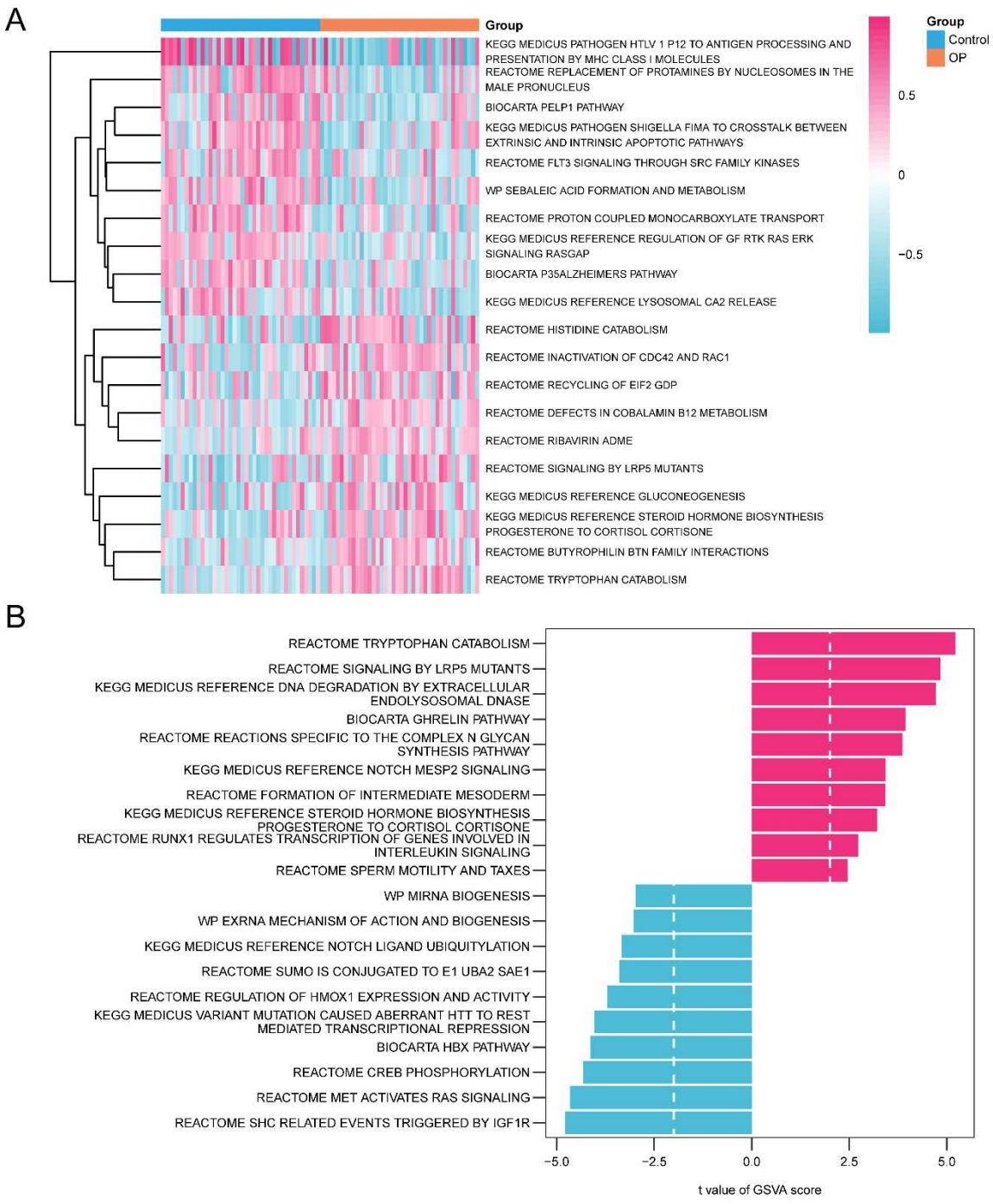


Fig.8 GSVA Analysis

Heat map (A) and bidirectional bar graph (B) of gene set variation analysis (GSVA) results between OP and Control of GSE56815. OP, Osteoporosis; GSVA, Gene Set Variation Analysis. Orange for OP, blue for Control. The screening criteria for gene set variation analysis (GSVA) were p value<0.05 and logFC ranked Top10 positive enriched pathways and Top10 negative enriched pathways. In the heat map, blue represents low enrichment and red represents high enrichment.

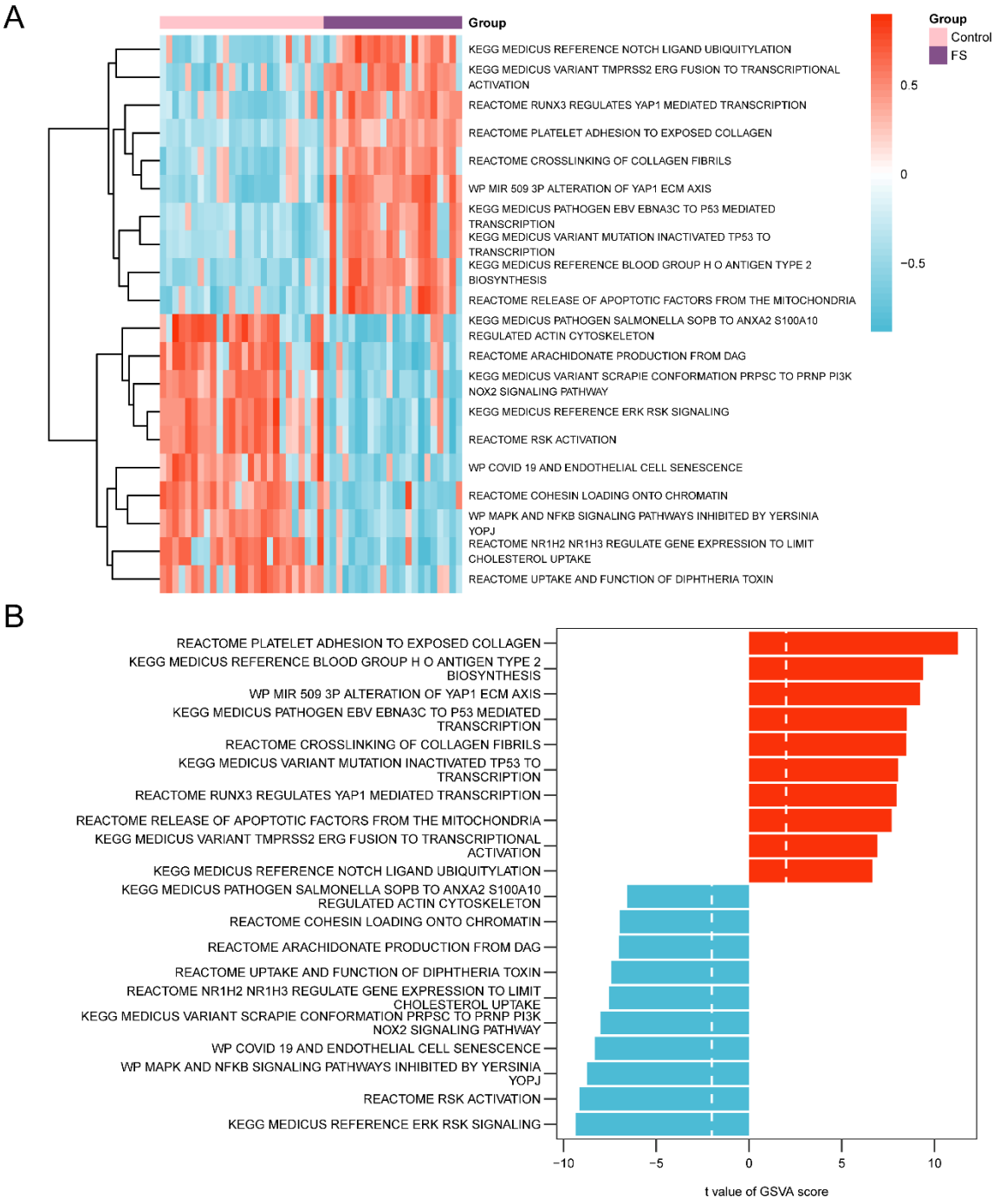


Fig.9 GSVA Analysis

Heat map (A) and bidirectional bar graph (B) of gene set variation analysis (GSVA) results between FS and Control of GSE140731. FS, Frozen shoulder; GSVA, Gene Set Variation Analysis. Purple for FS, pink for Control. The screening criteria for gene set variation analysis (GSVA) were p value<0.05 and logFC ranked Top10 positive enriched pathways and Top10 negative enriched pathways. In the heat map, blue represents low enrichment and red represents high enrichment.

2.8 Construction of Diagnostic Models

Osteoporosis Model: Logistic regression identified 13 significant genes (Figure 10A). SVM-RFE narrowed this to 12 genes (*MAP3K1, ZMIZ1, HP, MPO, SSR4, TTC27, HAX1, IL17RA, F5, NFKBIA, MARCKS, LTF*) with optimal accuracy (Figure 10B-C). LASSO regression further refined the model to 8 key genes: *MAP3K1, ZMIZ1, HP, MPO, HAX1, F5, MARCKS, LTF* (Figure 10D-E).

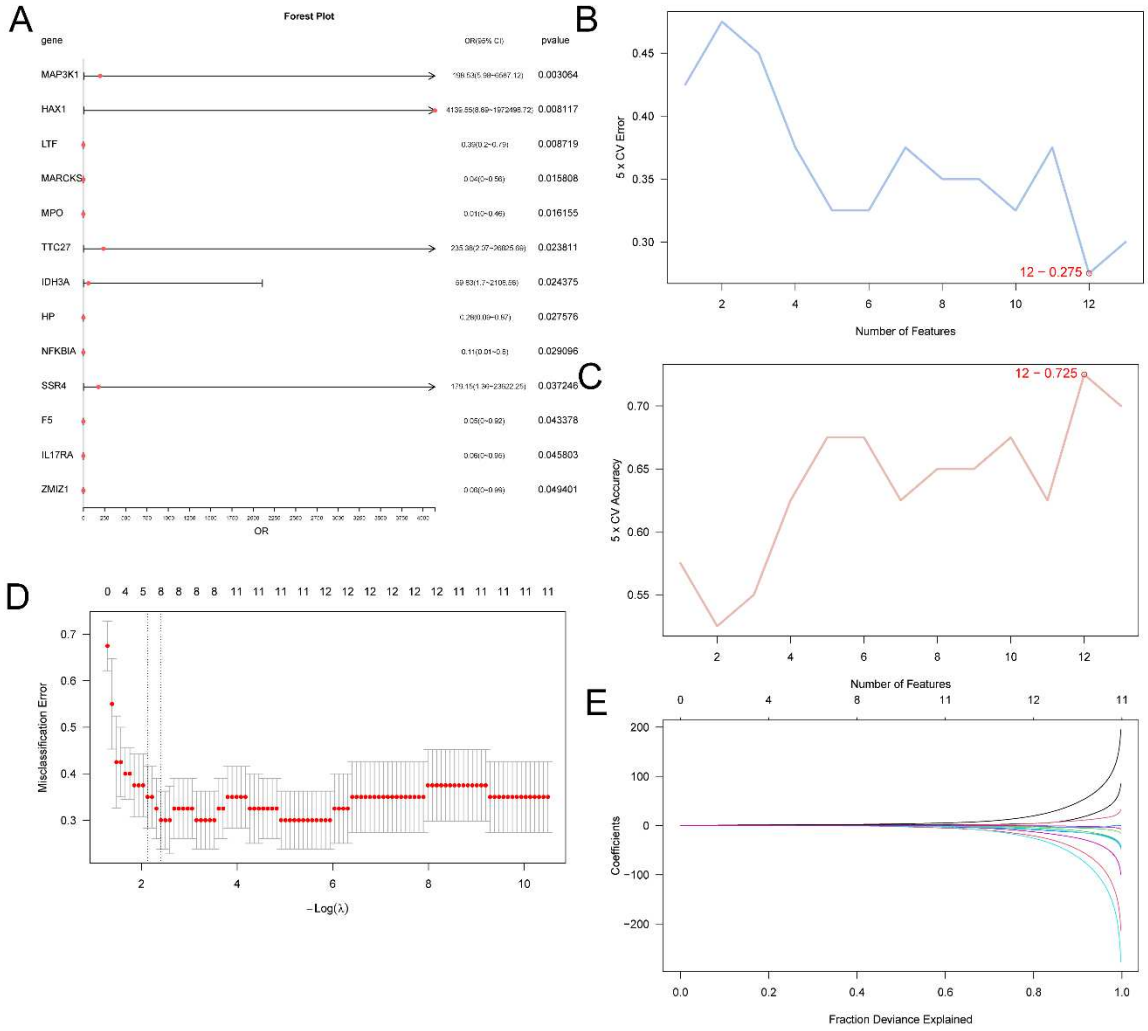
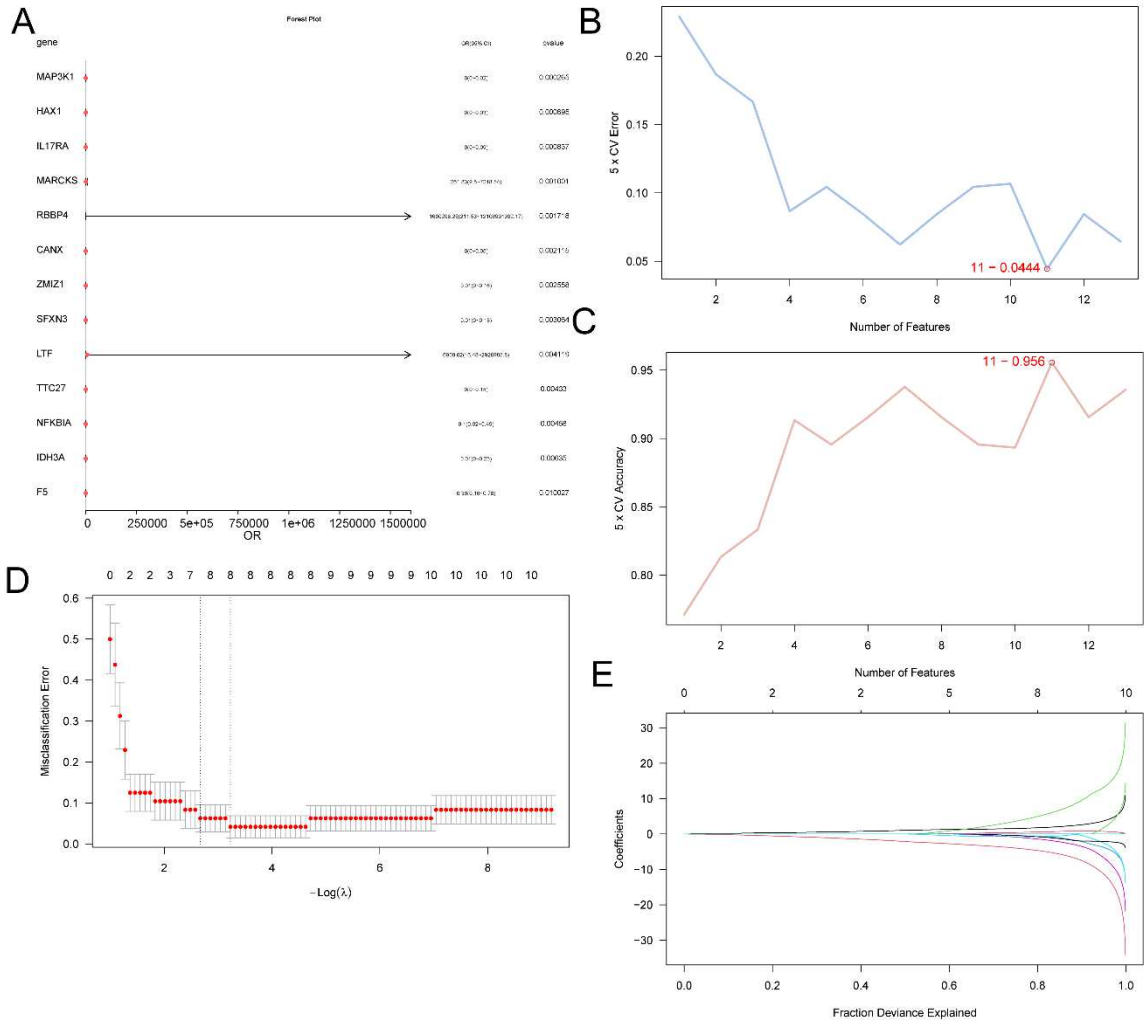


Fig.10 Machine Learning

A. Forest Plot of Logistic regression model. B.The number of genes with the lowest error rate. C. the number of genes with the highest accuracy obtained by the SVM algorithm. D-E. Diagnostic model plot (D) and variable trajectory plot (E) of LASSO regression model. OP, Osteoporosis; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes.

Frozen Shoulder Model: Logistic regression identified 13 significant genes(Figure 11A). SVM-RFE selected 11 genes (*MARCKS, MAP3K1, RBBP4, SFXN3, TTC27, CANX, IL17RA, LTF, IDH3A, F5, ZMIZ1*) (Figure 11B-C). LASSO regression retained 8 FS model genes: *MARCKS, MAP3K1, RBBP4, SFXN3, CANX, IL17RA, LTF, and ZMIZ1* (Figure 11D-E).



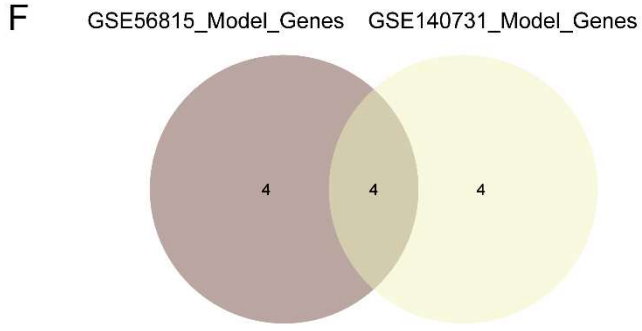


Fig.11 Machine Learning

A. Forest Plot of Logistic regression model. B-C. The number of genes with the lowest error rate (B) and the number of genes with the highest accuracy (C) obtained by the SVM algorithm are visualized. D-E. Diagnostic model plot (D) and variable trajectory plot (E) of LASSO regression model. F. Venn diagram of Model Genes intersection of osteoporosis and scapulohumeral periarthritis. FS, Frozen shoulder; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes.

Final Key Genes: The intersection of OP and FS Model Genes yielded 4 key genes: *MARCKS, MAP3K1, LTF, and ZMIZ1* (Figure 11F). Final LASSO models for both conditions confirmed these 4 key genes, namely *MARCKS, MAP3K1, LTF, and ZMIZ1* (Figures 12A-B, 13A-B).

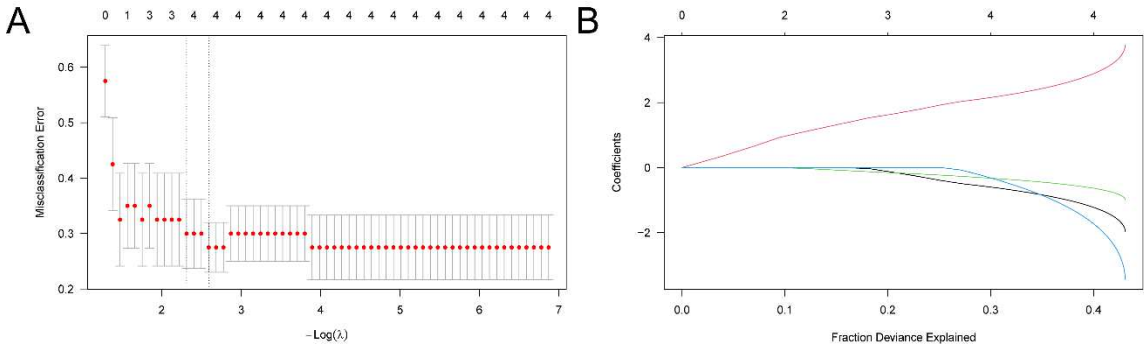


Fig.12 Machine Learning

A-B. Diagnostic model plot (A) and variable trajectory plot (B) of LASSO regression model. OP, Osteoporosis;

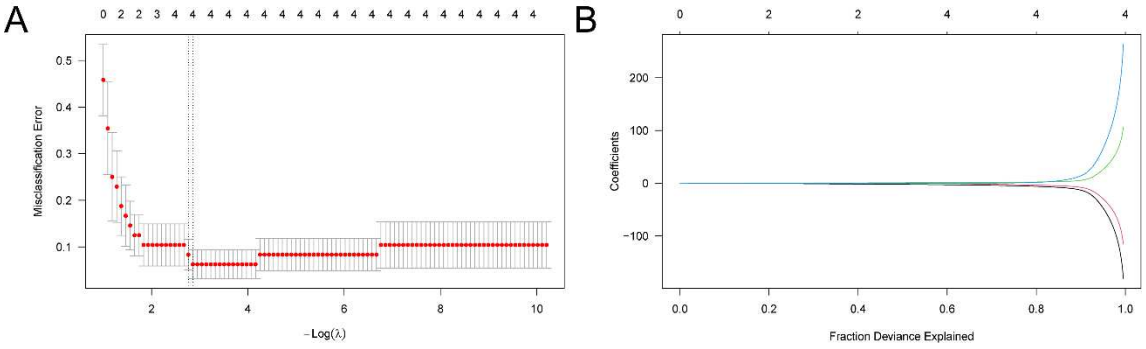


Fig.13 Machine Learning

A-B. Diagnostic model plot (A) and variable trajectory plot (B) of LASSO regression model. FS, Frozen shoulder; GMRD EGs, Glucose metabolism-Related Differentially Expressed Genes;

2.9 Model Validation

Osteoporosis: In the training set (GSE56815), the LASSO model demonstrated high diagnostic accuracy (AUC > 0.9) (Figure 14A). DCA indicated good clinical net benefit (Figure 14B). The nomogram showed *LTF* had the highest predictive weight, while *MARCKS* had the lowest (Figure 14C). Functional similarity analysis among the OP model genes again highlighted *TTC27* (Figure 14D). In the validation set (GSE7429), the Logistic model showed lower accuracy (0.5 < AUC < 0.7) (Figure 14E), while DCA suggested potential clinical net benefit (Figure 14F).

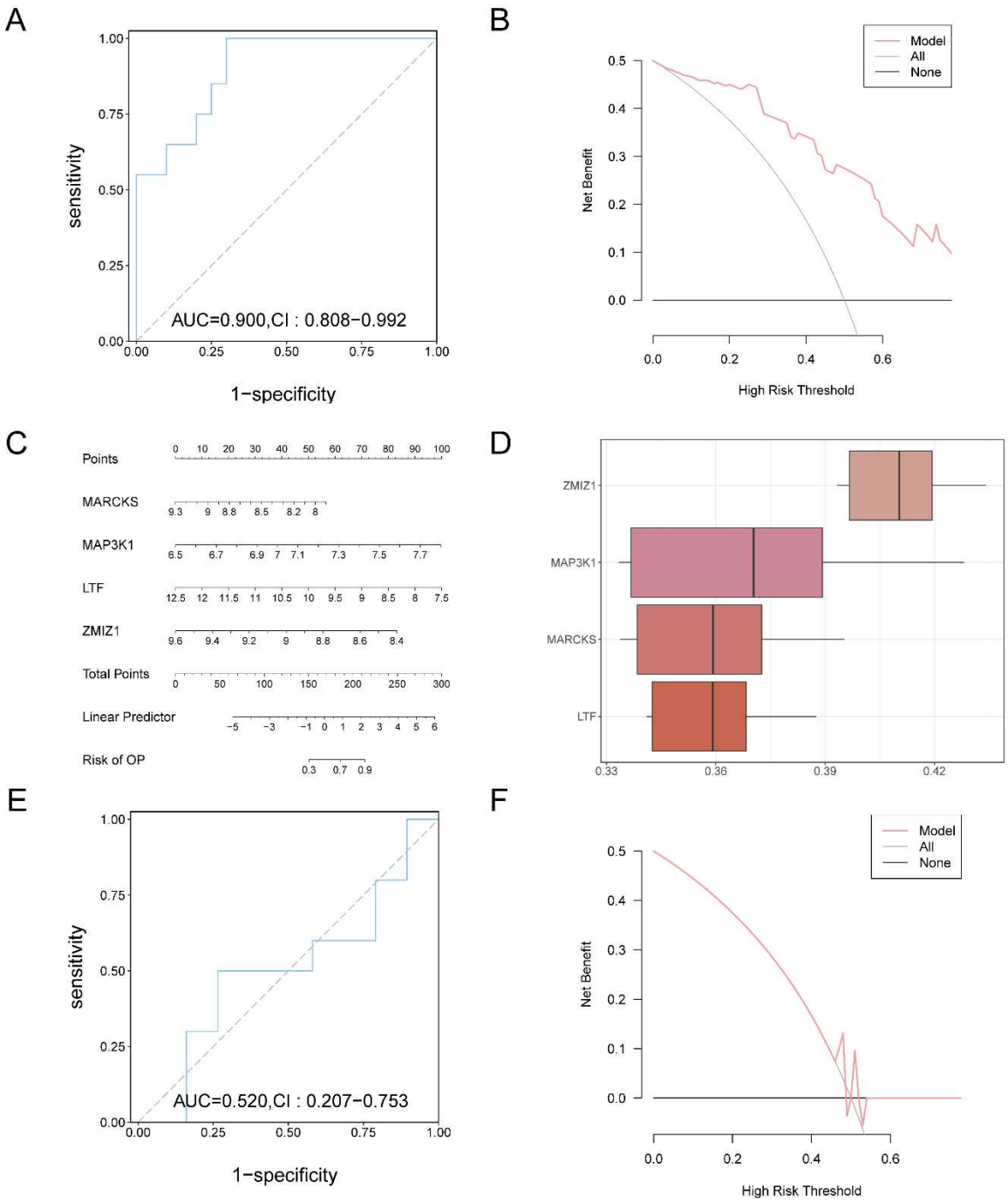


Fig.14 Diagnostic and Validation Analysis of OP and Friends Analysis

A. ROC curves of Lasso model in osteoporosis (OP) samples and Control (Control) samples. AUC is close to 1, diagnose the better the results. When AUC was between 0.5 and 0.7, the accuracy was low, when AUC was between 0.7 and 0. 9, the accuracy was moderate, and when AUC was above 0.9, the accuracy was high. B. Decision curve analysis (DCA) plot of predicted probabilities based on Lasso model. C. Nomogram of GSE56815 Key Genes in osteoporosis (OP) diag nostic model. D. Box plots of functional similarity (Friends) analysis of Key Genes. E. ROC curves of logistic regression

model in Osteoporosis samples and Control samples. The closer the AUC is to 1, the better the diagnostic performanc e. When AUC was between 0.5 and 0.7, the accuracy was low, when AUC was between 0.7 and 0.9, the accuracy was moderate, and when AUC was above 0.9, the accuracy was high. F. Decision curve analysis (DCA) diagram of osteoporo sis (OP) diagnostic model based on model prediction probability in GSE7429. The ordinate of decision curve analysis (DCA) chart is the net benefit, and the abscissa is the Probability Threshold or Threshold Probability. OP, Osteoporosis; DCA, Decision Curve Analysis; ROC, Receiver Operating Characteristic; AUC, Area Under Curve.

Frozen Shoulder: In the training set (GSE238053), the Lasso model showed moderate-to-high accuracy (0.7 < AUC < 0.9) (Figure 15A). DCA confirmed clinical benefit (Figure 15B). The nomogram indicated *LTF* as the most significant predictor (Figure 15C). Functional analysis again highlighted *TTC27* (Figure 15D). In the validation set (GSE140731), the Logistic model showed high accuracy (AUC > 0.9) (Figure 15E) with positive DCA results (Figure 15F).

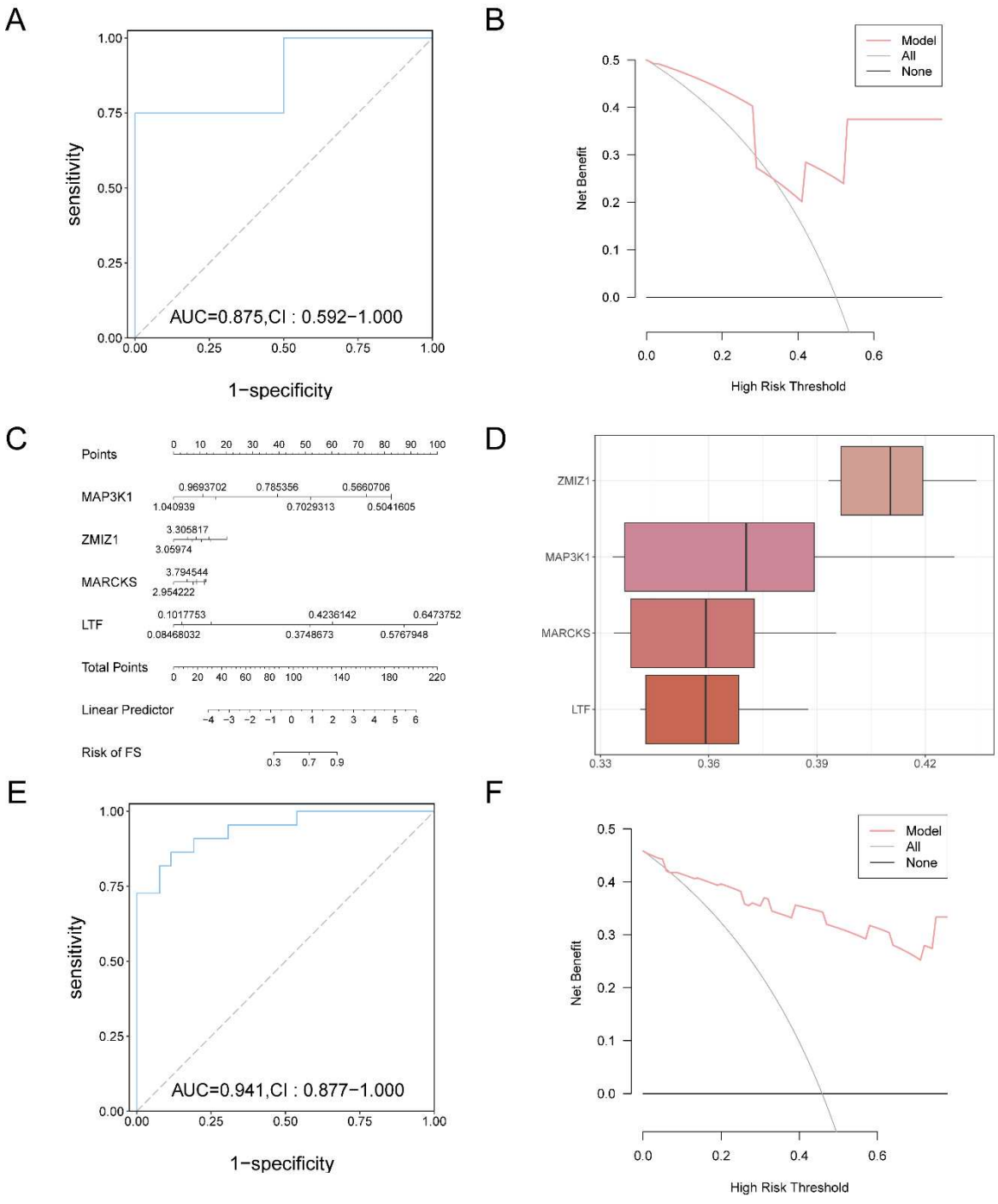


Fig.15 Diagnostic and Validation Analysis of FS and Friends Analysis

A. ROC curves of Lasso model in frozen shoulder (FS) samples and Control (Control) samples. The closer the AUC is to 1, the better the diagnostic effect is. When AUC was between 0.5 and 0.7, the accuracy was low, when AUC was between 0.7 and 0.9, the accuracy was moderate, and when AUC was above 0.9, the accuracy was high. B. Decision curve analysis (DCA) plot of predicted probabilities based on Lasso model. C. Nomogram of GSE238053 Key Genes in the diagnosis model of scapulohumeral periarthritis (FS). D. Box plots of functional similarity (Friends) analysis of Key

Genes. E. ROC curves of logistic regression model in frozen shoulder (FS) samples and Control samples. The closer the AUC is to 1, the better the diagnostic effect is. When AUC was between 0.5 and 0.7, the accuracy was low, when AUC was between 0.7 and 0.9, the accuracy was moderate, and when AUC was above 0.9, the accuracy was high. F. Decision curve analysis (DCA) plot of periarthritis of shoulder (FS) diagnostic model based on the prediction probability of the model in GSE140731. The ordinate of decision curve analysis (DCA) chart is the net benefit, and the abscissa is the Probability Threshold or Threshold Probability. FS, Frozen shoulder; DCA, Decision Curve Analysis; ROC, Receiver Operating Characteristic; AUC, Area Under Curve.

2.10 Expression and Correlation of Key Genes

Osteoporosis (GSE56815): All 4 key genes showed significant differential expression (*MARCKS, ZMIZ1*: $p < 0.05$; *LTF*: $p < 0.01$; *MAP3K1*: $p < 0.001$) (Figure 16A). Most key genes were positively correlated, except for a significant negative correlation between *MARCKS* and *MAP3K1* (r = -0.381) (Figure 16B).

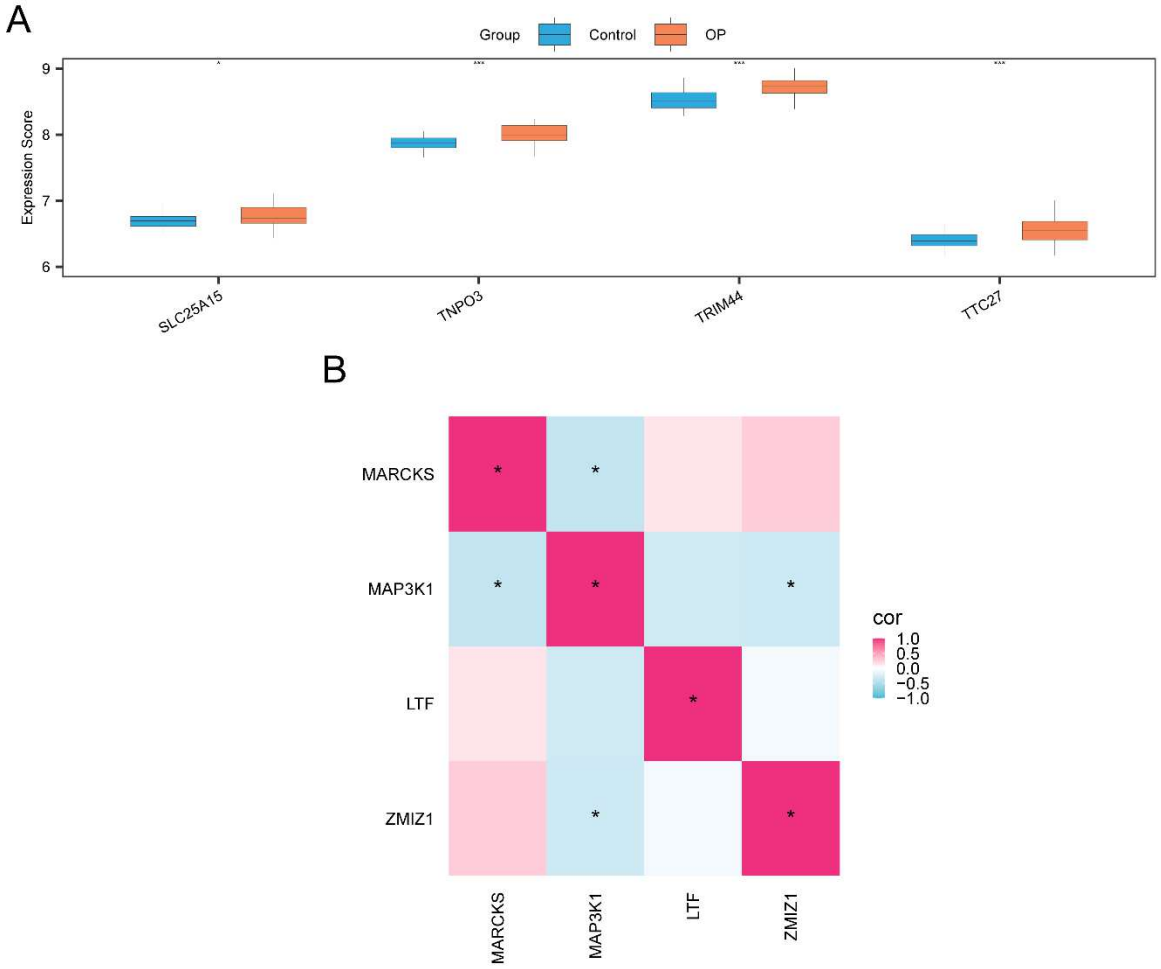


Fig.16 Differential Expression, Correlation Analysis

A. Group comparison of Key Genes in OP and Control of GSE56815. B. Heat map of the correlation of Key Genes between OP and Control in GSE56815. OP, Osteoporosis; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes; ns represents p value ≥ 0.05, not statistically significant; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significant. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, 0.3 to 0.5 was weak correlation, 0.5 to 0.8 was moderate correlation, and above 0.8 was strong correlation. In the group comparison graph, orange is OP and blue is Control. In the correlation heat map, red is positive correlation, blue is negative correlation, and the depth of color represents the strength of correlation.

Frozen Shoulder (GSE140731): All 4 key genes were significantly differentially expressed (*ZMIZ1*: $p < 0.01$; *LTF*, *MAP3K1*, *MARCKS*: $p < 0.001$) (Figure 17A). Most key genes showed negative correlations, with a significant negative correlation between *MAP3K1* and *MARCKS* (r = -0.619) (Figure 17B).

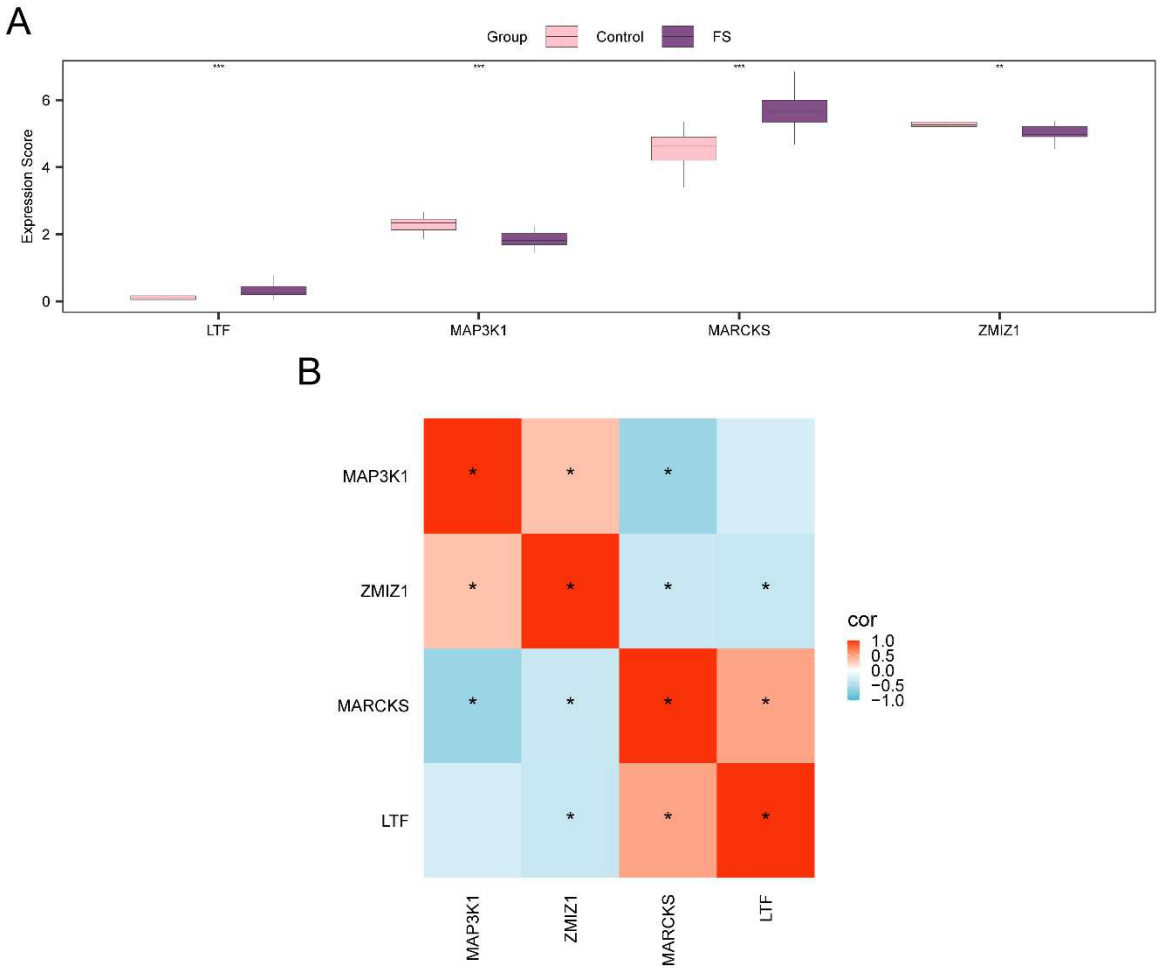


Fig.17 Differential Expression, Correlation Analysis

A. Group comparison of Key Genes in FS and Control of GSE140731 B. Heat map of the correlation of Key Genes betwe en FS and Control in GSE140731. FS, Frozen shoulder; GMRDEGs, Glucose metabolism-Related Differentially Expressed Genes; ns represents p value ≥ 0.05, not statistically significant; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significa nt. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, 0.3 to 0.5 was weak co rrelation, 0.5 to 0.8 was moderate correlation, and above 0.8 was strong correlation. In the group comparison plot, pu rple denotes FS and pink denotes Control. In the correlation heat map, red is positive correlation, blue is negative corr elation, and the depth of color represents the strength of correlation.

2.11 PPI Network and Hub Genes

The PPI network constructed from 17 CGs included 16 connected nodes, indicating that one CG was not connected in the STRING network under the selected threshold (Figure 18A). Intersection of top 5 genes from four centrality algorithms identified 3 Hub Genes: *NFKBIA*, *LTF*, and *HP* (Figure 18B).

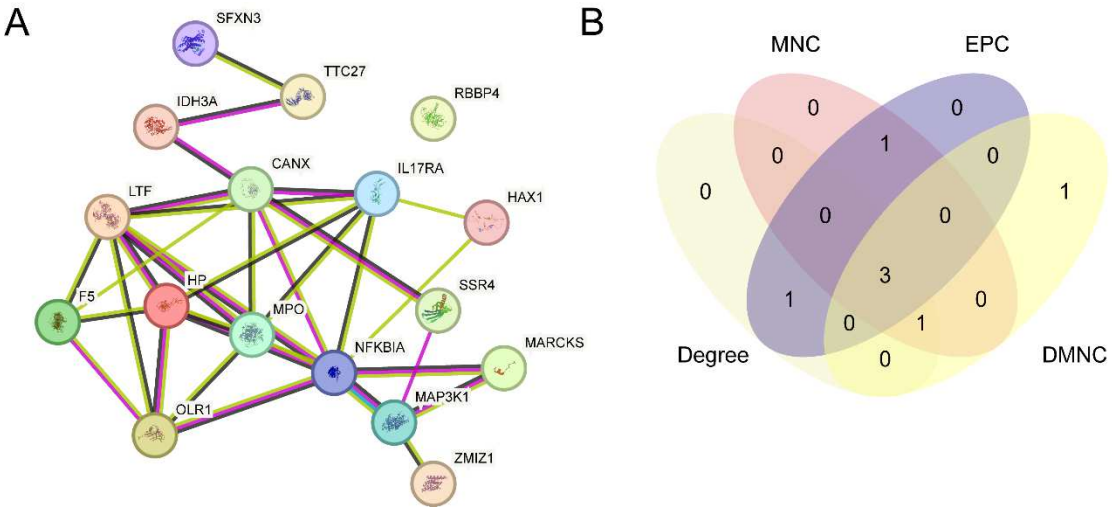


Fig18 PPI Network and hub genes Analysis

A. Protein-protein interaction Network (PPI Network) of CGs calculated from STRING database. B. Venn diagram of Top 5CGs calculated by four algorithms in CentiServer package.

2.12 Immune Infiltration Analysis

Osteoporosis (GSE56815): CIBERSORT detected 19 immune cell types (Figure 19A). Strong correlations were observed among immune cells, with Monocytes and Eosinophils showing the strongest negative correlation (r = -0.877) (Figure 19B). Hub gene analysis showed that *HP* was positively correlated with Neutrophils (r = 0.489), wheraas *NFKBIA* was negatively correlated with Memory B cells (r = -0.462) (Figure 19C).

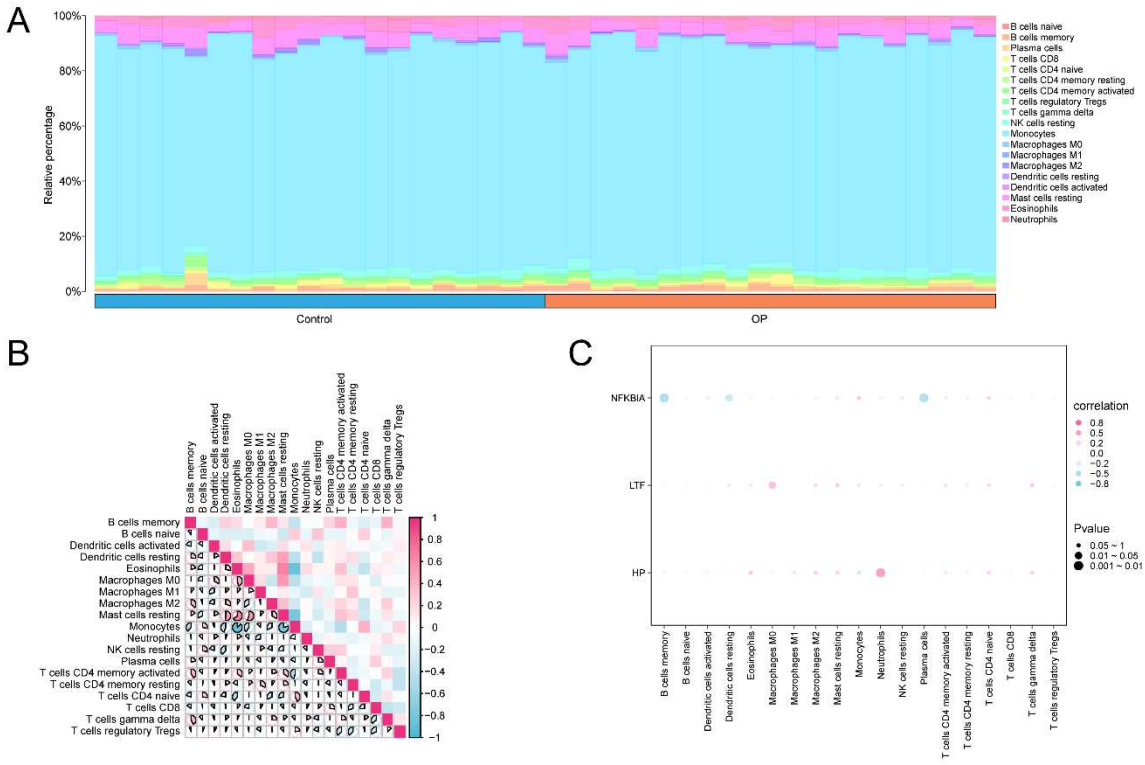


Fig.19 GSE56815 Immune Infiltration Analysis by CIBERSORT Algorithm

A. Bar chart of the proportion of immune cells in GSE56815 (A). B. Correlation heatmap of immune cells in GSE56815. C. Bubble plot of correlation between immune cell infiltration abundance and hub genes in GSE56815. ns represents p value ≥ 0.05, no statistical significance; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significant. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, between 0.3 and 0.5 was weak correlation, between 0.5 and 0.8 was moderate correlation, and above 0.8 was strong correlation. Red is positive correlation, blue is negative correlation. The depth of the color represents the strength of the correlation.

Frozen Shoulder (GSE140731): CIBERSORT detected 19 immune cell types(Figure 20A). Eight immune cell types including memory B cells , T cells CD4 memory resting, T cells follicular helper, Monocytes, Macrophages M0, Macrophages M1, Macrophages M2, Mast cells activated differed significantly between groups (Figure 20B). Monocytes and M0 Macrophages showed the strongest negative correlation (r = -0.580) (Figure 20C). *LTF* positively correlated with M1 Macrophages (r = 0.556), while *IL17RA* negatively correlated with M1 Macrophages (r = -0.449) (Figure 20D).

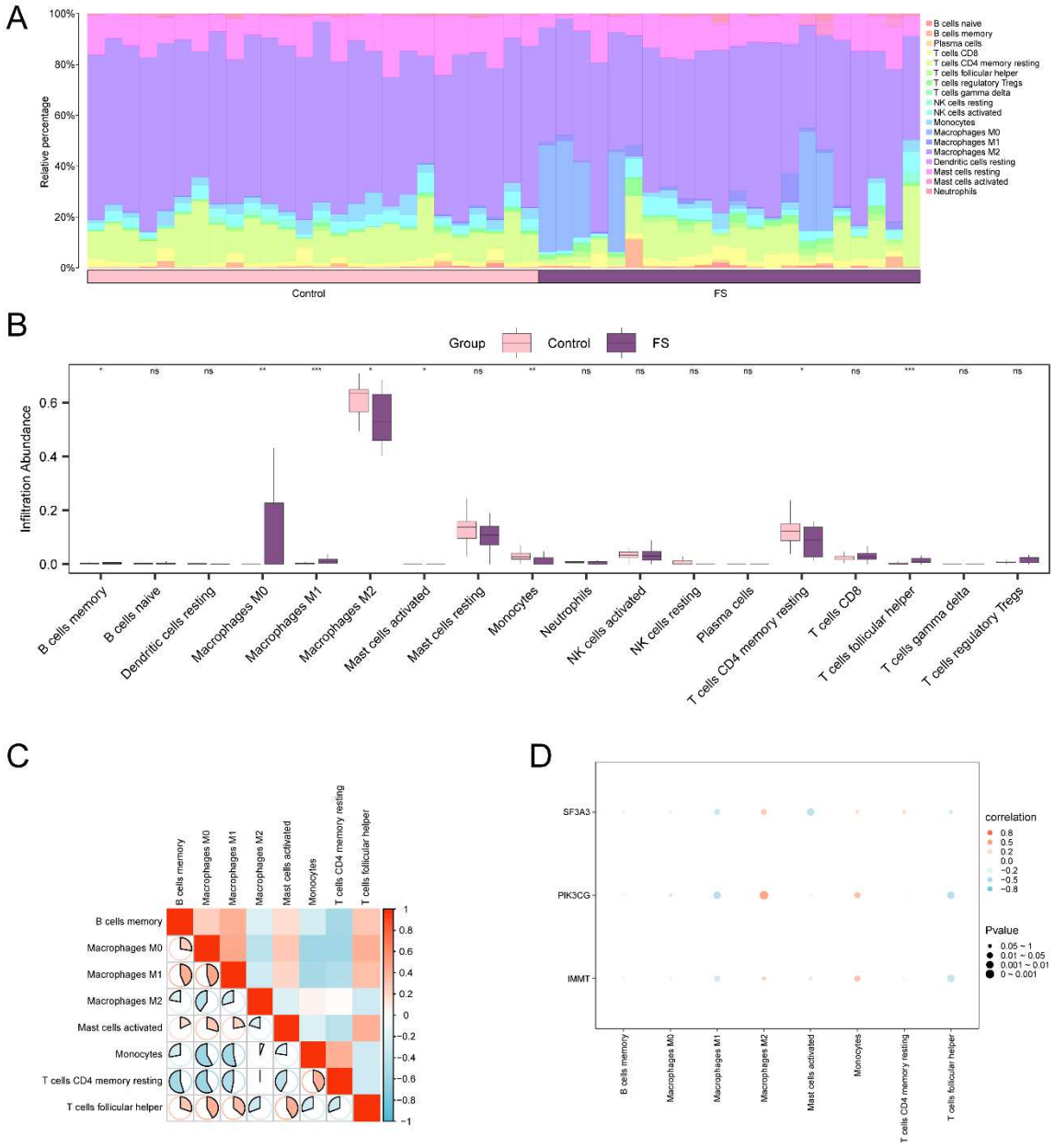


Fig.20 GSE140731 Immune Infiltration Analysis by CIBERSORT Algorithm

A-B. The proportion of immune cells in GSE140731 bar chart (A) and group comparison chart (B). C. Correlation heat map of immune cells in GSE140731. D. Bubble plot of correlation between immune cell infiltration abundance and hub genes in GSE140731. ns represents p value ≥ 0.05, no statistical significance; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significant. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, between 0.3 and 0.5 was weak correlation, between 0.5 and 0.8 was moderate correlation, and above 0.8 was strong correlation. Red is positive correlation, blue is negative correlation. The depth of the color represents the strength of the correlation.

2.13 Regulatory Networks

The mRNA-TF regulatory network included 1 mRNA and 5 TFs (Figure 21A, Table S2). The mRNA-miRNA regulatory network included 1 mRNA and 12 miRNAs (Figure 21B, Table S3).

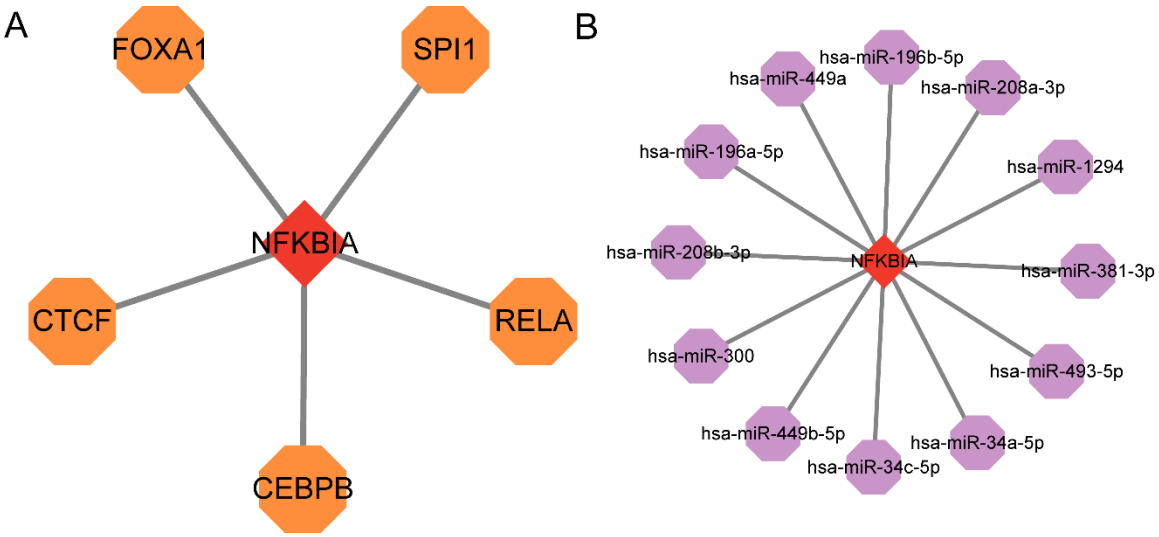


Fig.21 Regulatory Network of hub genes

A. mRNA-TF Regulatory Network of hub genes. B. mRNA-miRNA Regulatory Network of hub genes. Red is mRNA, orange is TF, and purple is miRNA.

Table S2: Transcription factors (TFs) predicted to regulate hub genes based on ChIPBase database(as shown in supplementary material)**.**

Table S3: miRNAs targeting hub genes retrieved from StarBase v3.0 database(as shown in supplementary material)

2.14 Osteoporosis Subtyping

Consensus clustering identified two distinct OP subtypes (Cluster1, n = 7; Cluster2, n = 13) in GSE56815 (Figure 22A-C). t-SNE analysis confirmed clear separation (Figure 22D). Heatmaps showed differential expression of model genes between subtypes (Figure 22E).

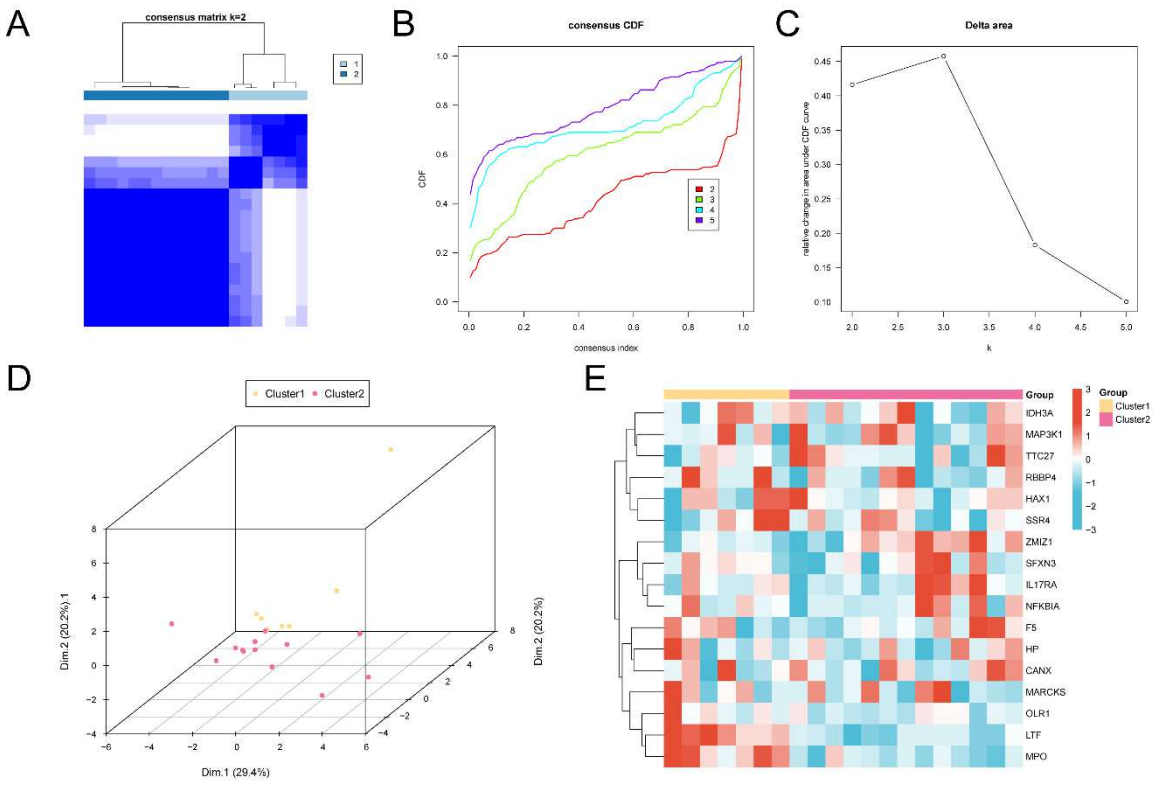


Fig.22 Consensus Clustering Analysis for OP

A. Plot of consensus clustering results for GSE56815. B-c. Consistency cumulative distribution function (CDF) plot (B) a nd Delta plot (C) of consistency clustering analysis. D. 3D t-SNE cluster plot of 2 disease subtypes of osteoporosis (OP). E. Heat map of expression values of model genes in osteoporosis (OP) subtypes. OP, Osteoporosis; CDF, Empirical Cum ulative Distribution Function; PCA, Principal Component Analysis. ns stands for p value ≥ 0.05, not statistically significa nt; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significant. Yellow is subtype 1 (Cluster1), pink is subtype 2 (Cluster 2).

2.15 Subtype-Specific Immune Analysis

2.15.1 CIBERSORT Results:

Cluster1: Strong positive correlation between Naive B cells and Naive CD4+ T cells (r = 1.000). *IL17RA* positively correlated with Monocytes (r = 0.929); *SFXN3* negatively correlated with Tregs (r = -0.964) (Figure 23B, D).

Cluster2: Strong positive correlation between Naive B cells and Naive CD4+ T cells (r = 1.00). *IDH3A* positively correlated with Activated Dendritic Cells (r = 0.720) and negatively correlated with Resting NK cells (r = -0.615) (Figure 23C, E).

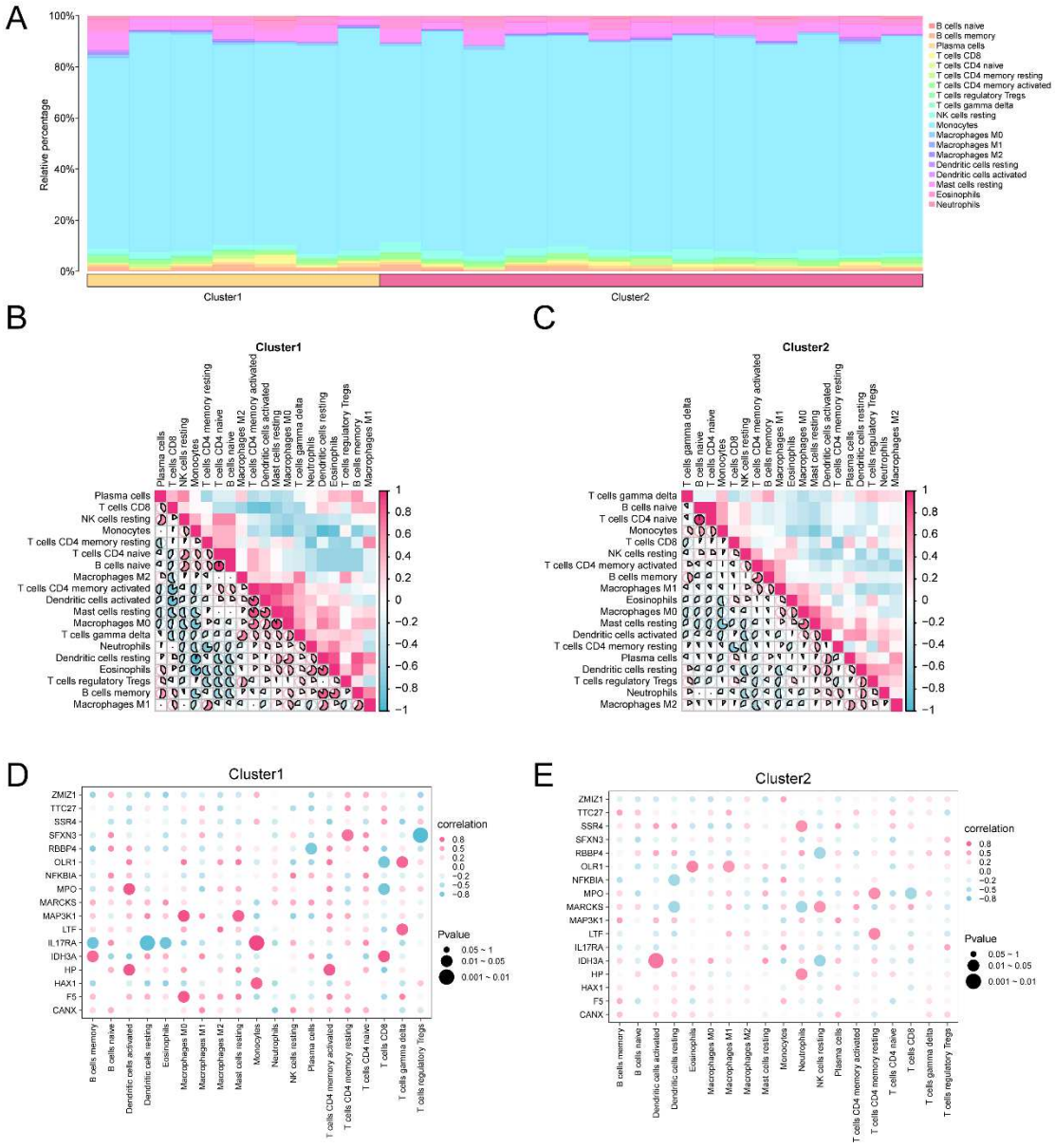


Fig.23 Risk Group Immune Infiltration Analysis by CIBERSORT Algorithm

A. Bar chart of the proportion of immune cells in Cluster1 and Cluster2 of osteoporosis (OP) disease group samples. B-c. Results of correlation analysis of immune cell infiltration abundance in Cluster1 (B) and Cluster2 (C) of osteoporosis (OP) disease group samples are shown. Bubble plot of correlation between immune cell infiltration abundance and C Gs in Cluster1 (D) and Cluster2 (E) of osteoporosis (OP). The bubble sizes represent different P value intervals, which a re 0.05-1, 0.01-0.05, 0.001-0.01, and 0-0.001, respectively. The larger the bubble, the more significant the correlation. OP, Osteoporosis. ns stands for p value ≥ 0.05, statistically significant; * represents p value < 0.05, indicating statistica l significance; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly st atistically significant. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, bet ween 0.3 and 0.5 was weak correlation, between 0.5 and 0.8 was moderate correlation, and above 0.8 was strong cor

relation. Pink is Cluster2, and yellow is Cluster1. Red is a positive correlation, blue is a negative correlation, and the de pth of the color represents the strength of the correlation.

2.15.2 ssGSEA Results:

Cluster1: Four immune cells including central memory CD8+ T cell, effector memory CD8 T cell, activated dendritic cell, mast cell showed significant differences(Figure 24A). Activated DCs and Mast cells were negatively correlated (r=-0.643). *IDH3A* positively correlated with Effector Memory CD8+ T cells (r=0.893); *MPO* negatively correlated with them (r=-0.893) (Figure 24B, D).

Cluster2: Central Memory CD8+ T cells and Mast cells were negatively correlated (r=-0.335). *LTF* positively correlated with Activated DCs (r=0.802) (Figure 24C, E).

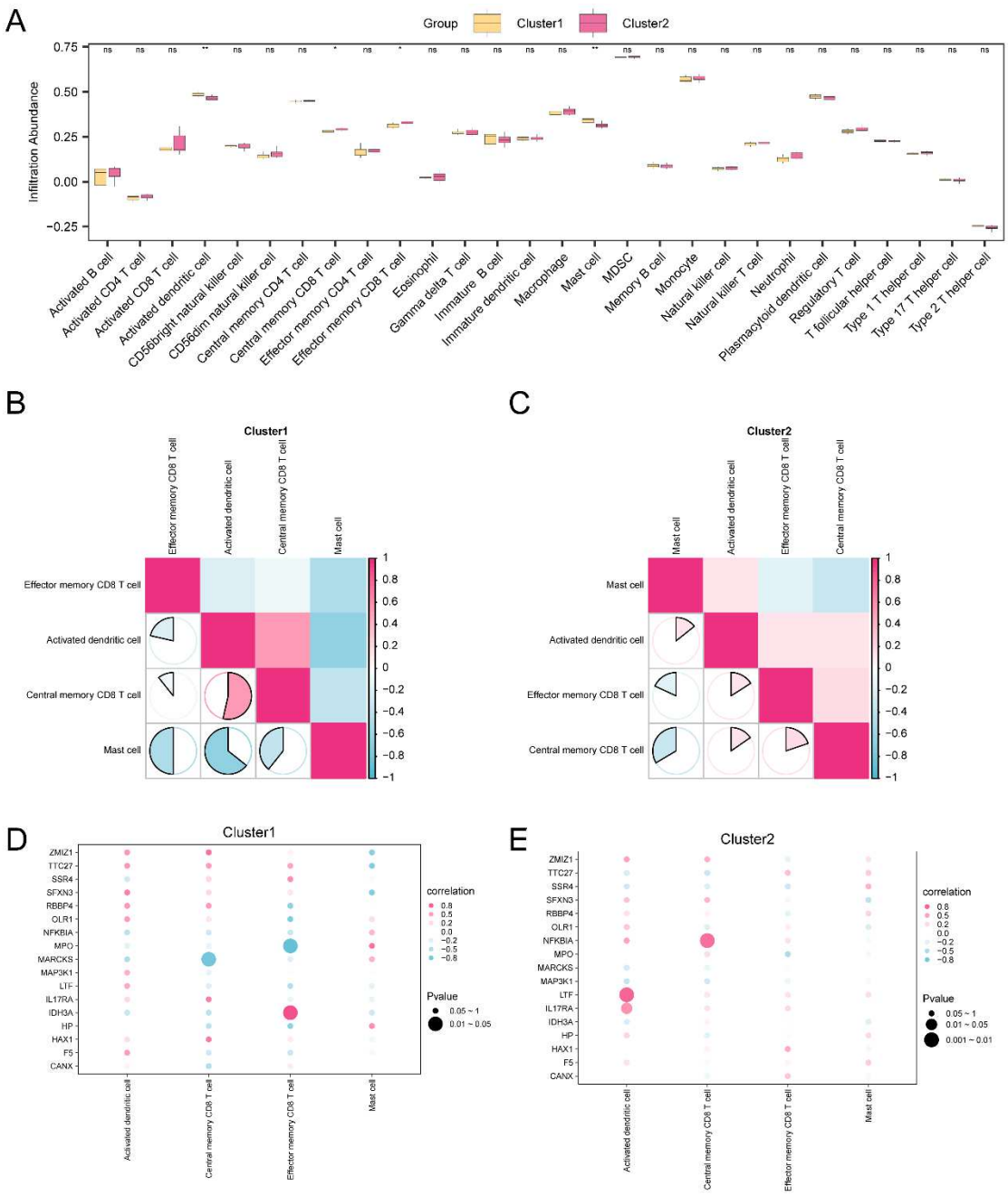


Fig.24 Immune Infiltration Analysis by ssGSEA Algorithm

A. Comparison of grouping of immune cells in Cluster1 and Cluster2 of osteoporosis (OP) disease group samples. B-c. Results of correlation analysis of immune cell infiltration abundance in Cluster1 (B) and Cluster2 (C) of osteoporosis (O P) disease group samples are shown. Bubble plot of correlation between immune cell infiltration abundance and CGs i n Cluster1 (D) and Cluster2 (E) of osteoporosis (OP). The bubble sizes represent different P value intervals, which are 0. 05-1, 0.01-0.05, 0.001-0.01, and 0-0.001, respectively. The larger the bubble, the more significant the correlation. ssGS EA, single-sample Gene-Set Enrichment Analysis; OP, Osteoporosis. ns stands for p value ≥ 0.05, not statistically signific

ant; * represents p value < 0.05, statistically significant; ** represents p value < 0.01, highly statistically significant; *** represents p value < 0.001 and highly statistically significant. The absolute value of correlation coefficient (r value) below 0.3 was weak or no correlation, between 0.3 and 0.5 was weak correlation, between 0.5 and 0.8 was moderate correlation, and above 0.8 was strong correlation. Pink is Cluster2, and yellow is Cluster1. Pink is a positive correlation, blue is a negative correlation, and the depth of the color represents the strength of the correlation.

2.16 Virtual Screening

DrugCLIP screening identified potential compounds targeting *LTF*. Top candidates included P207-9104 and Z169905102. Molecular docking suggested strong predicted binding affinities for *LTF*-P207-9104 (Vina score = -8.45 kcal/mol), *LTF*-Z169905102 (Vina score = -8.19 kcal/mol), and a second docking result for *LTF*-P207-9104 (Vina score = -7.58 kcal/mol) (Figure 25A-C).

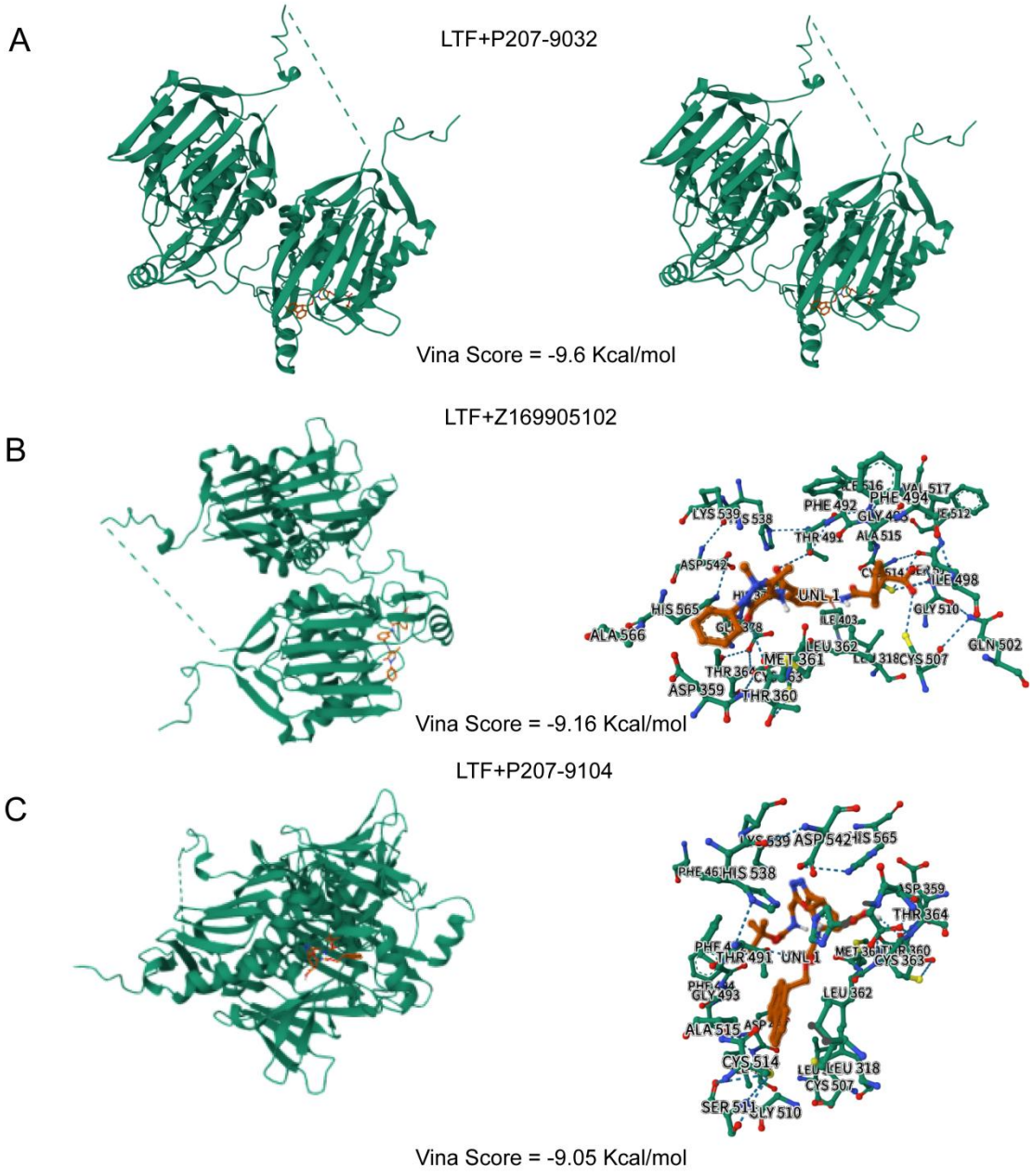


Fig.25 Molecular docking

A-C. Visualization of docking results between LTF and P207-9104 (A), LTF and Z169905102 (B), and LTF and P207-9104 (C), from left to right are docking global maps and interaction force maps, respectively. Blue dashed lines -- hydrogen bonds, light blue dashed lines -- weak hydrogen bonds, gray dashed lines -- hydrophobic forces, cyan dashed lines -- halogen bonds, yellow dashed lines -- ionic bonds, orange dashed lines -- cation-π interactions, and green dashed lines -- π-π conjugation.

3. Discussion

Osteoporosis (OP) and frozen shoulder (FS) are two prevalent musculoskeletal disorders that impose substantial burdens on aging populations and may be clinically associated through overlapping metabolic and inflammatory features. OP is characterized by systemic bone loss and microarchitectural deterioration, leading to increased fracture risk, while FS manifests as chronic shoulder pain and progressive restriction of joint movement due to capsular inflammation and fibrosis. Despite their distinct clinical presentations, both conditions share fundamental pathophysiological hallmarks, including metabolic dysregulation and aberrant immune-inflammatory responses[46]. However, current diagnostic approaches for OP rely primarily on bone mineral density measurements, which lack sensitivity for early-stage disease, and FS diagnosis remains largely clinical and non-specific. Although existing treatments can reduce fracture risk or relieve shoulder pain and stiffness, their effectiveness may be limited by delayed diagnosis, variable responses, and incomplete recovery in some patients. These limitations highlight the urgent need for shared molecular biomarkers that can improve early detection and mechanistic understanding of these comorbidities.

In this study, we performed an integrative bioinformatics analysis to systematically investigate the crosstalk between OP and FS from the perspective of glucose metabolism-related genes (GMRGs). By combining transcriptomic data from the GEO database with WGCNA and multiple machine learning algorithms, we identified four key genes-*MARCKS*, *MAP3K1*, *LTF*, and *ZMIZ1*-as candidate shared molecular signatures. A LASSO regression model based on these genes showed diagnostic potential, particularly in the FS validation analysis, whereas its limited performance in the OP validation dataset indicates the need for further validation. PPI network analysis further identified *NFKBIA*, *LTF*, and *HP* as hub genes associated with immune microenvironment alterations. Notably, virtual screening via DrugCLIP identified potential candidate compounds targeting *LTF*. These findings provide novel insights into the potential shared molecular mechanisms underlying OP and FS and provide candidate biomarkers and therapeutic targets for future validation. The following discussion elaborates on the biological significance of these results, their clinical implications, and the limitations that warrant future investigation.

Building upon the identification of four key genes through the integration of WGCNA and multiple machine learning algorithms, the specific biological functions of these molecules provide a compelling narrative for their dual involvement in osteoporosis (OP) and frozen shoulder (FS). Among these, *MARCKS*, *MAP3K1*, and *LTF* are particularly noteworthy. *MARCKS* is a critical substrate

for protein kinase C and is involved in cell adhesion, motility, and inflammatory responses; therefore, its dysregulation may be related to cytoskeletal and inflammatory alterations relevant to bone remodeling and capsular fibrosis, although this hypothesis requires experimental validation[47,48]. *MAP3K1* is a central node in multiple stress-activated protein kinase cascades, most notably the *p38* and *JNK* pathways, which are master regulators of inflammatory cytokine production[49]. Lactoferrin (*LTF*), an iron-binding glycoprotein, exerts pleiotropic effects by modulating immune cell activity and participating in bone metabolism-related processes, including the promotion of osteoblast proliferation and inhibition of osteoclastogenesis[50]. This gene's dual role in immune modulation and bone homeostasis positions it as a plausible molecular hub for the co-occurrence of an inflammatory condition like FS and a metabolic bone disease like OP. The strong contribution of *LTF* to the diagnostic model in the nomogram analysis, coupled with the well-established link between hyperglycemia and impaired *LTF* function, suggests that a dysregulation in iron homeostasis and innate immunity may be a shared pathogenic driver[51].

Further dissection of the molecular framework reveals that the three hub genes identified from the PPI network-*NFKBIA*, *LTF*, and *HP*-offer a more refined perspective on the specific "glucose metabolism-immune" axis. *NFKBIA* encodes *IκBα*, the primary inhibitor of the *NF-κB* transcription factor, a master switch for pro-inflammatory cytokine production[52]. The negative correlation between *NFKBIA* and memory B cells in OP may reflect an association between *NF-κB*-related inflammatory regulation and B-cell-related immune alterations, but this relationship requires further functional validation. In contrast, the positive correlation between *LTF* and M1 macrophages in FS is consistent with previous reports describing the involvement of *LTF* in immune regulation[53]. These findings suggest that although glucose metabolism-related dysregulation may represent a shared background, OP and FS may involve different immune-associated patterns: in OP, *NFKBIA* and *HP* were correlated with memory B cells and neutrophils, respectively, whereas in FS, *LTF* and *IL17RA* were correlated with M1 macrophages. Further experimental studies are needed to clarify whether these associations reflect disease-specific immune mechanisms. This is consistent with known mechanisms in rheumatoid arthritis, where defective metabolic repair in T cells shifts glucose toward biosynthetic pathways, ultimately leading to a metabolically stressed microenvironment that drives effector cell function and tissue destruction[54]. The hub gene HP further supports this link, as it is a hemoglobin-binding acute-phase protein whose expression is tightly linked to oxidative stress and metabolic inflammation[55].

The significant correlation patterns between these hub genes and specific immune cells allow for the construction of a more detailed mechanistic pathway. Furthermore, the enrichment of CGs in terms such as “defense response” and “phagosome” suggests that these genes may be associated with immune defense and phagocyte-related processes involving macrophages and neutrophils. This convergence of metabolic stress (high glucose) and immune dysfunction (altered phagocytosis and cytokine release) provides a robust mechanistic framework for the co-morbidity of OP and FS.

The identification of distinct molecular subtypes of OP via consensus clustering of the model genes adds another layer of complexity to the disease pathology, moving beyond a monolithic view of bone loss. The existence of two subtypes (C1 and C2) with profoundly different immune microenvironments suggests that the metabolic dysregulation driving the disease is not uniform. In C1, *IL17RA* showed the strongest positive correlation with monocytes ($r = 0.929$, $p < 0.05$), suggesting a possible association with *IL-17/Th17*-related inflammatory signaling[56]. Meanwhile, Warburg metabolism augments glycolytic ATP production to support *TH-17* responses[57]. This pathway is central to the pathogenesis of rheumatoid arthritis, where *IL-17* produced by T cells acts on monocytes and other myeloid cells to fuel chronic inflammation and bone erosion[54]. In C2, *IDH3A*, a tricarboxylic acid (TCA) cycle-related enzyme, showed the strongest positive correlation with activated dendritic cells ($r = 0.720$, $p < 0.05$), suggesting a possible association between mitochondrial metabolism-related genes and dendritic-cell-related immune regulation. This pattern mirrors findings in metabolic disorders like diabetes, where altered oxidative metabolism primes dendritic cells for pro-inflammatory roles[58]. This subtype-specific immune signature is a critical finding, as it implies that the mechanistically-diverse metabolic defects in OP, such as altered glycolytic flux versus compromised mitochondrial function, can lead to distinct immunological outcomes.

Consequently, the biological implications of these two subtypes are profound for treatment personalization. C1, with its *IL17RA*-high/Monocyte signature, could be particularly susceptible to therapies targeting the *IL-17* pathway, which have shown efficacy in autoimmune arthritis[59]. In contrast, C2, with its *IDH3A*-high/Dendritic cell signature, might benefit from interventions that modulate dendritic cell maturation, such as metformin which can impact mitochondrial complex I activity[60]. This bipartite model aligns with the concept of “immunometabolic” disease endotypes seen in other chronic conditions, where the underlying metabolic lesion determines the immune effector mechanism. For example, a T cell with a mitochondrial defect may be more prone to differentiate into a *Th17* cell, while a cell with a defective pentose phosphate pathway may favor a cytotoxic profile[61,62].

The identification of these two subtypes suggests that OP may involve different molecular and immune states rather than a single uniform pathway. This finding encourages future studies to correlate these subtypes with clinical parameters like bone mineral density, fracture risk, and response to conventional OP therapies, thereby enabling a shift towards precision medicine in the treatment of osteoporosis. Besides, this heterogeneity may affect patients' susceptibility to FS or their patterns of disease progression, providing a molecular stratification basis for further investigating which osteoporosis (OP) patients are more susceptible to developing comorbid FS.

The virtual screening for potential therapeutics targeting *LTF* yields a high-value, testable hypothesis for a shared treatment strategy, complementing the diagnostic marker discovery. The DrugCLIP predictions identified two candidate compounds, P207-9104 and Z169905102, with strong predicted binding affinity to *LTF*, as indicated by Vina scores of -8.45 and -8.19 kcal/mol, respectively. This finding is mechanistically significant as it provides an entry point for pharmacological intervention. However, whether these compounds can modulate *LTF*-related immune or bone-metabolic processes remains unknown and requires biochemical and functional validation. The chemical structures of these compounds and their potential off-target effects are critical unknowns, but the methodology itself is valid. Similar computational screening strategies have been used to prioritize candidate compounds in other disease contexts, but experimental validation remains essential[63].

These findings suggest that *LTF* may represent a candidate target worthy of further pharmacological and functional validation. By focusing on a central hub protein like *LTF*, which sits at the intersection of the shared “glucose metabolism – immune response – bone homeostasis” axis, the therapeutic potential extends beyond single-disease management. Future studies could test whether *LTF*-targeting candidate compounds influence M1 macrophage-related inflammation in FS models or osteoblast-osteoclast balance in OP-related models. However, the study’s reliance on purely computational docking requires rigorous follow-up. The next steps should include biophysical confirmation of binding (e.g., Surface Plasmon Resonance), followed by functional assays in relevant cell types, such as assessing the effect on macrophage polarization in a high-glucose environment or osteogenic differentiation in bone marrow stromal cells.

4. Limitations

Despite the comprehensive bioinformatics approach employed, this study has several limitations. First, the reliance solely on publicly available transcriptomic datasets (e.g., GSE56815, GSE7429) with relatively small sample sizes-ranging from as few as four cases per group-may

compromise statistical robustness and generalizability. The heterogeneity in tissue origins, such as monocytes, B cells, shoulder capsule, and synovium, as well as platform differences between OP and FS datasets, may introduce technical heterogeneity and influence the reproducibility of the identified genes and diagnostic models. Second, the absence of wet-laboratory validation, such as in vitro cell experiments, in vivo animal models, flow cytometry, or immunohistochemistry, means that the functional roles of the proposed hub genes (*NFKBIA*, *LTF*, *HP*), the immune infiltration patterns, and the efficacy of the virtually screened compounds remain predictive. These findings, while insightful, lack direct causal evidence, and the diagnostic model's limited external performance in osteoporosis (AUC 0.5–0.7) underscores the need for independent, large-scale clinical validation. Future studies should prioritize multi-omics integration and prospective cohort designs to confirm these molecular targets.

5. Conclusion

In summary, this study provides insights into potential shared molecular mechanisms between osteoporosis and frozen shoulder from the perspective of glucose metabolism-related genes. By integrating transcriptomic data with machine learning and network analyses, we identified four shared key genes (*MARCKS*, *MAP3K1*, *LTF*, and *ZMIZ1*) with diagnostic potential and three hub genes (*NFKBIA*, *LTF*, and *HP*) associated with glucose metabolism-related immune alterations. The identification of distinct molecular subtypes in osteoporosis and *LTF*-targeting candidate compounds with predicted high binding affinity further supports the potential value of these findings for future validation. These results provide a foundational framework for developing shared diagnostic biomarkers and therapeutic targets for both diseases. Future research should focus on functional validation of these genes in cellular and animal models, independent validation of the diagnostic models in prospective clinical cohorts, and experimental assessment of the binding and biological effects of the *LTF*-targeting candidate compounds.

References

1. Eastell R, O'Neill TW, Hofbauer LC, et al. Postmenopausal osteoporosis. *Nat Rev, Dis Primers*. 2016;2:16069. doi:10.1038/nrdp.2016.69

2. Polyzos SA, Anastasilakis AD, Efstathiadou ZA, Yavropoulou MP, Makras P. Postmenopausal osteoporosis coexisting with other metabolic diseases: Treatment considerations. *Maturitas*. 2021;147:19-25. doi:10.1016/j.maturitas.2021.02.007

3. Kingston K, Curry EJ, Galvin JW, Li X. Shoulder adhesive capsulitis: Epidemiology and predictors of surgery. *J Shoulder Elbow Surg*. 2018;27(8):1437-1443. doi:10.1016/j.jse.2018.04.004

4. Alabadi B, Civera M, Moreno-Errasquin B, Cruz-Jentoft AJ. Nutrition-based support for osteoporosis in postmenopausal women: a review of recent evidence. *Int J Women's Health*. 2024;16:693-705. doi:10.2147/IJWH.S409897

5. Li X, Yan Y, Wang Z, et al. The link between osteoporosis and frozen shoulder: exploring the therapeutic effect of TAK715 on reversing fibrosis and protecting against osteoporosis via the p38 MAPK signaling pathway. *BMC musculoskelet disord*. 2024;25(1):942. doi:10.1186/s12891-024-08068-8

6. Eastell R, Rosen CJ, Black DM, Cheung AM, Murad MH, Shoback D. Pharmacological management of osteoporosis in postmenopausal women: an endocrine society* clinical practice guideline. *J Clin Endocrinol Metab*. 2019;104(5):1595-1622. doi:10.1210/jc.2019-00221

7. Pandey V, Madi S. Clinical guidelines in the management of frozen shoulder: an update! *Indian J Orthop*. 2021;55(2):299-309. doi:10.1007/s43465-021-00351-3

8. Li M, Li W, Mai G, et al. Targeting dysregulated glycolysis in type 2 diabetic osteoporosis: identification of a diagnostic gene signature and therapeutic validation of calcifediol. *Life Sci*. 2026;400:124473. doi:10.1016/j.lfs.2026.124473

9. Shirazi S, Esawi E, Nassar ZD, Gronthos S, Cakouros D. Hyperglycemia Leads to BMSC Impaired Osteogenesis, Enhanced Adipogenesis, and Altered Metabolism. *J Cell Biochem*. 2026;127(4):e70090. doi:10.1002/jcb.70090

10. Song F, Lee WD, Marmo T, et al. Osteoblast-intrinsic defect in glucose metabolism impairs bone formation in type II diabetic male mice. *eLife*. 12:e85714. doi:10.7554/eLife.85714

11. Vadivalagan C, Krishnan A, Chen SJ, et al. The warburg effect in osteoporosis: cellular signaling and epigenetic regulation of energy metabolic events to targeting the osteocalcin for phenotypic alteration. *Cell Signalling*. 2022;100:110488. doi:10.1016/j.cellsig.2022.110488

12. Navarro-Ledesma S, Hamed-Hamed D, Pruimboom L. A new perspective of frozen shoulder pathology; the interplay between the brain and the immune system. *Front Physiol*. 2024;15:1248612. doi:10.3389/fphys.2024.1248612

13. Navarro-Ledesma S. Frozen shoulder as a systemic immunometabolic disorder: the roles of estrogen, thyroid dysfunction, endothelial health, lifestyle, and clinical implications. *J Clin Med*. 2025;14(20):7315. doi:10.3390/jcm14207315

14. Kushioka J, Chow SKH, Toya M, et al. Bone regeneration in inflammation with aging and cell-based immunomodulatory therapy. *Inflamm Regen*. 2023;43(1):29. doi:10.1186/s41232-023-00279-1

15. Akbar M, Crowe LAN, McLean M, et al. Translational targeting of inflammation and fibrosis in frozen shoulder: Molecular dissection of the T cell/IL-17A axis. *Proc Natl Acad Sci U S A*. 2021;118(39):e2102715118. doi:10.1073/pnas.2102715118

16. Loftus RM, Finlay DK. Immunometabolism: Cellular Metabolism Turns Immune Regulator. *J Biol Chem*. 2016;291(1):1-10. doi:10.1074/jbc.R115.693903

17. Zhou Y, Gao Y, Xu C, Shen H, Tian Q, Deng HW. A novel approach for correction of crosstalk effects in pathway analysis and its application in osteoporosis research. *Sci Rep*. 2018;8(1):668. doi:10.1038/s41598-018-19196-2

18. Xiao P, Chen Y, Jiang H, et al. In vivo genome-wide expression study on human circulating B cells suggests a novel ESR1 and MAPK3 network for postmenopausal osteoporosis. *J Bone Miner Res: Off J Am Soc Bone Miner Res*. 2008;23(5):644-654. doi:10.1359/jbmr.080105

19. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue):D991-995. doi:10.1093/nar/gks1193

20. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.1-1.30.33. doi:10.1002/cpbi.5

21. Kamal N, McGee SL, Eng K, et al. Transcriptomic analysis of adhesive capsulitis of the shoulder. *J Orthop Res: Off Publ Orthop Res Soc*. 2020;38(10):2281-2289. doi:10.1002/jor.24686

22. Yan Y, Li X, Wang Z, Yang R. Salvianolic acid B attenuates inflammation and prevent pathologic fibrosis by inhibiting CD36-mediated activation of the PI3K-Akt signaling pathway in frozen shoulder [dataset 2]. Published online July 24, 2023.

23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. 2008;9:559. doi:10.1186/1471-2105-9-559

24. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47(D1):D419-D426. doi:10.1093/nar/gky1038

25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

26. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J Integr Biol*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118

27. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. 2013;14:7. doi:10.1186/1471-2105-14-7

28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260

29. 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 Bioinf*. 2018;19(1):432. doi:10.1186/s12859-018-2451-4

30. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenet*. 2019;11(1):123. doi:10.1186/s13148-019-0730-1

31. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77. doi:10.1186/1471-2105-12-77

32. Wu J, Zhang H, Li L, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. *Cancer Commun*. 2020;40(7):301-312. doi:10.1002/cac2.12067

33. Harrell FE Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd ed. Cham: Springer; 2015.

34. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol*. 2018;74(6):796-804. doi:10.1016/j.eururo.2018.08.038

35. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 2010;26(7):976-978. doi:10.1093/bioinformatics/btq064

36. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

37. Jalili M, Salehzadeh-Yazdi A, Asgari Y, et al. CentiServer: a comprehensive resource, web-based application and R package for centrality analysis. *PLOS One*. 2015;10(11):e0143111. doi:10.1371/journal.pone.0143111

38. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. doi:10.1038/nmeth.3337

39. Zhou KR, Liu S, Sun WJ, et al. ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res*. 2017;45(D1):D43-D50. doi:10.1093/nar/gkw965

40. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303

41. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92-97. doi:10.1093/nar/gkt1248

42. Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics*. 2013;29(20):2611-2616. doi:10.1093/bioinformatics/btt425

43. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572-1573. doi:10.1093/bioinformatics/btq170

44. Gong W, Kuang M, Chen H, et al. Single-sample gene set enrichment analysis reveals the clinical implications of immune-related genes in ovarian cancer. *Front Mol Biosci*. 2024;11:1426274. doi:10.3389/fmolb.2024.1426274

45. Jia Y, Gao B, Tan J, et al. Deep contrastive learning enables genome-wide virtual screening. *Science*. 2026;391(6781):eads9530. doi:10.1126/science.ads9530

46. Yu X, Ma Z, Jin L. Identification of common genetic features and pathways for osteoporosis with frozen shoulder by integrated bioinformatics analysis and machine learning. *Korean J Physiol Pharmacol: Off J Korean Physiol Soc Korean Soc Pharmacol*. 2026;30(2):125-138. doi:10.4196/kjpp.25.259

47. Pang C, Wen L, Qin H, Zhu B, Lu X, Luo S. Sotrastaurin, a PKC inhibitor, attenuates RANKL-induced bone resorption and attenuates osteochondral pathologies associated with the development of OA. *J Cell Mol Med*. 2020;24(15):8452-8465. doi:10.1111/jcmm.15404

48. Lee SM, Suk K, Lee WH. Myristoylated alanine-rich C kinase substrate (MARCKS) regulates the expression of proinflammatory cytokines in macrophages through activation of p38/JNK MAPK and NF-κB. *Cell Immunol*. 2015;296(2):115-121. doi:10.1016/j.cellimm.2015.04.004

49. Mordente K, Ryder L, Bekker-Jensen S. Mechanisms underlying sensing of cellular stress signals by mammalian MAP3 kinases. *Mol Cell*. 2024;84(1):142-155. doi:10.1016/j.molcel.2023.11.028

50. Wu Y, Zhang C, Lv C. Direct and indirect regulation of bone metabolism by lactoferrin. *Front Endocrinol*. 2025;16:1660312. doi:10.3389/fendo.2025.1660312

51. Xiao Z, Shen D, Lan T, et al. Reduction of lactoferrin aggravates neuronal ferroptosis after intracerebral hemorrhagic stroke in hyperglycemic mice. *Redox Biol*. 2022;50:102256. doi:10.1016/j.redox.2022.102256

52. Liu T, Zhang L, Joo D, Sun SC. NF-κB signaling in inflammation. *Signal Transduction Targeted Ther*. 2017;2:17023-. doi:10.1038/sigtrans.2017.23

53. de la Rosa G, Yang D, Tewary P, Varadhachary A, Oppenheim JJ. Lactoferrin acts as an alarmin to promote the recruitment and activation of APCs and antigen-specific immune responses. *J Immunol*. 2008;180(10):6868-6876. doi:10.4049/jimmunol.180.10.6868

54. Qiu J, Wu B, Goodman SB, Berry GJ, Goronzy JJ, Weyand CM. Metabolic control of autoimmunity and tissue inflammation in rheumatoid arthritis. *Front Immunol*. 2021;12:652771. doi:10.3389/fimmu.2021.652771

55. Purushothaman M, Krishnan P, Purushothaman KR, et al. Genotype-dependent impairment of hemoglobin clearance increases oxidative and inflammatory response in human diabetic atherosclerosis. *Arterioscler, Thromb, Vasc Biol*. 2012;32(11):2769-2775. doi:10.1161/ATVBAHA.112.252122

56. Shahrara S, Pickens SR, Dorfleutner A, Pope RM. IL-17 induces monocyte migration in rheumatoid arthritis. *J Immunol*. 2009;182(6):3884-3891. doi:10.4049/jimmunol.0802246

57. Xu K, Yin N, Peng M, et al. Glycolytic ATP fuels phosphoinositide 3-kinase signaling to support effector T helper 17 cell responses. *Immunity*. 2021;54(5):976-987.e7. doi:10.1016/j.immuni.2021.04.008

58. Brombacher EC, Everts B. Shaping of dendritic cell function by the metabolic micro-environment. *Front Endocrinol*. 2020;11:555. doi:10.3389/fendo.2020.00555

59. Mease PJ, McInnes IB, Kirkham B, et al. Secukinumab inhibition of interleukin-17A in patients with psoriatic arthritis. *N Engl J Med*. 2015;373(14):1329-1339. doi:10.1056/NEJMoa1412679

60. Reczek CR, Chakrabarty RP, D'Alessandro KB, et al. Metformin targets mitochondrial complex I to lower blood glucose levels. *Sci Adv*. 2024;10(51):eads5466. doi:10.1126/sciadv.ads5466

61. Daneshmandi S, Cassel T, Lin P, et al. Blockade of 6-phosphogluconate dehydrogenase generates CD8+ effector T cells with enhanced anti-tumor function. *Cell Rep*. 2021;34(10):108831. doi:10.1016/j.celrep.2021.108831

62. Baixauli F, Piletic K, Puleston DJ, et al. An LKB1-mitochondria axis controls TH17 effector function. *Nature*. 2022;610(7932):555-561. doi:10.1038/s41586-022-05264-1

63. Li W, Yuan B, Zhao Y, et al. Transcriptome profiling reveals target in primary myelofibrosis together with structural biology study on novel natural inhibitors regarding JAK2. *Aging (Milano)*. 2021;13(6):8248-8275. doi:10.18632/aging.202635

Statements and Declarations

Funding：

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declaration of Interests：

The authors declare no competing interests.

Author Contributions：

All authors contributed to the conception and design of this study. Long Fang completed the screening of core crosstalk genes and the construction of diagnostic models. Baolong Wang performed protein-protein interaction analysis. Ziwei Hou conducted immune infiltration analysis. Lizhong Jing finished consensus clustering analysis and molecular virtual screening. Wei Hao drafted the original manuscript, and Jiushan Yang revised and edited the manuscript. All authors provided critical comments on the earlier versions of the manuscript. All authors have read and approved the final version of the manuscript.

Clinical trial number: not applicable.

Ethics approval:

This research mainly focused on integrated bioinformatics mining and machine learning analysis. All datasets used in this work are publicly accessible, and thus exemption from ethical approval was granted by the Research Ethics Committee of our hospital.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- TableS2mRNATF.tsv
- TableS3mRNAmiRNA.tsv
- TableS4.csv