



OPEN Comprehensive bioinformatics analysis reveals novel potential biomarkers associated with aging and mitochondria in osteoporosis

Ke Bi, Yuxi Chen, Yuhang Hu, Song Li, Weiming Li, Zhange Yu & Lei Yu

Osteoporosis (OP) is a prevalent age-related bone metabolic disease. Aging and mitochondrial dysfunction are involved in the onset and progression of OP, but the specific mechanisms have not been elucidated. The aim of this study was to identify novel potential biomarkers associated with aging and mitochondria in OP. In this study, based on GEO database, aging-related and mitochondria-related differentially expressed genes (AR&MRDEGs) were screened. The AR&MRDEGs were enriched in mitochondrial structure and function. Then, 6 key genes were identified by WGCNA and multiple machine learning, and a novel diagnostic model was constructed. The efficacy of diagnostic model was validated using external datasets. The results showed that diagnostic model had favorable diagnostic prediction ability. Next, key gene regulatory networks were constructed and single-gene GSEA analysis was performed. In addition, based on a single-cell dataset from OP, single-cell differentially expressed genes (scDEGs) were identified. The results revealed that aging-related and mitochondria-related genes (AR&MRGs) were enriched in the ERK pathway in tissue stem cells (TSCs), and in mitochondrial membrane potential depolarization in monocytes. Cellular communication analysis showed that TSCs were active, with numerous signaling interactions with monocytes, macrophages and immune cells. Finally, the expression of key gene was verified by quantitative real-time PCR (qRT-PCR). This study is expected to provide strategies for the diagnosis and treatment of OP targeting aging and mitochondria.

Keywords Osteoporosis, Aging, Mitochondria, WGCNA, Diagnostic model, Single-cell bioinformatics analysis

Osteoporosis (OP) is a bone metabolic disease characterized by bone microstructural destruction, leading to increased bone fragility and susceptibility to fracture¹. Osteoporotic fracture is extremely harmful and is one of the major causes of disability and death in elderly patients². With the aging population, the prevalence of OP is rapidly escalating and has become an essential public health challenge³. Dualenergy X-ray absorptiometry (DXA) is a well-recognized standard for OP diagnosis⁴. However, DXA is insensitive to early bone loss, fails to accurately assess the severity of OP, and fails to predict fracture risk. Based on gene microarrays, identification of key genes by transcriptome analysis is a valuable approach. Multi-level, multi-dimensional bioinformatics analysis and validation of multiple datasets contribute to the identification of more reliable genetic biomarkers.

Studies have shown that bone homeostasis, maintained by a complex balance between bone formation and bone resorption, suffers disruption with aging⁵. Genetic mouse models have demonstrated that elimination of aging cells in vivo significantly attenuates the aging-related OP process⁶. Meanwhile, mitochondrial dysfunction is one of critical pathogenic mechanisms of OP⁷. Mitochondrial DNA alterations, oxidative phosphorylation damage, phagocytosis dysfunction, and defects in mitochondrial biogenesis and dynamics have all been associated with OP⁸. Moreover, mitochondrial dysfunction contribute in cellular aging and aging-related stem cell viability decline⁹. Aging and mitochondria are potential targets for OP. However, genetic biomarkers of aging and mitochondria-related genes for OP are still lacked.

In this study, aging-related and mitochondria-related differentially expressed genes (AR&MRDEGs) were captured by differential expression analysis. Potential biological functions and associated pathways of AR&MRDEGs were defined by GO, KEGG enrichment analysis, GSEA and GSVA. Through WGCNA and

Department of Orthopedic Surgery at the First Affiliated Hospital, Harbin Medical University, Harbin, China. email: hmuyulei@hrbmu.edu.cn

multiple machine learning algorithms, key genes were screened and a novel diagnostic model was constructed. Single-cell bioinformatics analysis was performed to explore the biological processes involved in aging-related and mitochondria-related genes (AR&MRGs) in specific cell type and intercellular communication in OP.

Materials and methods

Data collection

The GSE35959¹⁰, GSE56815¹¹ and GSE7158¹² datasets were downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>)¹³ by “GEOquery” package (Version 2.70.0)¹⁴ (Table S1). GSE35959 acts as training dataset, and GSE56815 and GSE7158 serve as validation dataset in this study. The GSE35959, GSE56815, GSE7158 datasets were normalized and annotated by the “limma” package (Version 3.58.1). Aging-related genes (ARGs) and mitochondria-related genes (MRGs) were collected through GeneCards (<https://www.genecards.org/>)¹⁵ and MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/>)¹⁶. The 485 AR&MRGs were acquired by intersecting ARGs with MRGs.

Identification of aging-related and mitochondria-related differentially expressed genes

The DEGs between OP and control were identified by “limma” package (Version 3.58.1)¹⁷. The threshold of DEGs was $|\log FC| > 1$ and adjusted $P < 0.05$. The AR&MRDEGs were obtained by venne diagram of DEGs and AR&MRGs. Heatmaps of AR&MRDEGs were demonstrated by “pheatmap” package (Version 1.0.12) and chromosomal localization maps were created using “RCircos” package (Version 1.2.2)¹⁸.

GO and KEGG enrichment analysis

Enrichment analysis of AR&MRDEGs for GO¹⁹ and KEGG²⁰ was performed using the “clusterProfiler” package (version 4.10.0)²¹. Entries were selected by adj. $P < 0.05$ and FDR (q value) < 0.25 . The value of P correction method was Benjamini-Hochberg.

Gene set enrichment analysis and gene set variation analysis

Genes were ranked according to logFC, and then subjected to GSEA by “clusterProfiler” package²². Meanwhile, GSE35959 were analyzed for GSVA using “GSVA” package (Version 1.50.0)²³. The filtering criterion of entries was $P < 0.05$ and FDR < 0.25 .

Expression difference analysis and correlation analysis of AR&MRDEGs

The group comparison plots were drawn based on AR&MRDEGs expression. Meanwhile, the correlation of AR&MRDEGs was analyzed by spearman algorithm. Correlation heatmaps were plotted via “pheatmap” package (Version 1.0.12) and correlation scatterplots were drawn by “ggplot2” package (Version 3.4.4).

Weighted gene co-expression network analysis

WGCNA²⁴ was run by the “WGCNA” package²⁵. The variance of genes was calculated and the top 8000 genes were screened. The minimum genes number of module was 100; the ideal soft threshold was 16; the module shear height was 0.4. The correlation of modules with OP was measured. Modules with $|r \text{ value}| > 0.5$ and $P < 0.05$ were picked to subsequent analysis.

Machine learning and construction of diagnostic model

Firstly, candidate genes were screened using “randomForest” (RF) package²⁶. After cross-validation, the number of variables with smaller errors was selected, and then vital variables were selected based on MeanDecreaseGini for subsequent analysis.

Then, least absolute shrinkage and selection operator (LASSO) regression²⁷ was run through “glmnet” package²⁸ to calculate risk score and construct risk model. The LASSO riskscore was calculated as follows: $RiskScore = \sum_i Coefficient (gene_i) * mRNA \text{ Expression } (gene_i)$

The LASSO regression analysis was visualized by LASSO coefficient path and cross verification curve.

Validation of diagnostic model

The receiver operating characteristic curves (ROC) of risk score and key genes were plotted by “pROC” package (Version 1.18.5) and area under the curve (AUC) was calculated to evaluate diagnosis efficiency. The interrelationships of key genes were demonstrated by nomogram based on logistic regression analysis through “rms” package (Version 6.7-1)²⁹. Calibration curve was plotted to assess effectiveness and discrimination of diagnostic model. The net benefits of genetic information decisions were assessed by Decision Curve Analysis (DCA) with “ggDCA” package³⁰.

Protein-protein interaction networks

The GeneMANIA database (<https://genemania.org/>)³¹ allows discovering genes with similar functions through massive genomics and proteomics data, as well as locating genes sharing functions, predicting gene functions, and prioritizing genes. Protein-protein interaction (PPI) network of key genes was mapped based on GeneMINIA.

Regulatory network of key genes

The miRNAs related to key genes were available from StarBase v3.0 database (<https://starbase.sysu.edu.cn/>)³². In addition, drugs targeting key genes were predicted by comparative toxicogenomics database (CTD) (<https://ctdbase.org/>)³³. Finally, transcription factors (TFs) regulating key genes were analyzed by ChIPBase database (<http://rna.sysu.edu.cn/chipbase/>)³⁴. Based on StarBase v3.0 database, RNA-binding proteins (RBPs) targeting key genes were selected³⁵. The regulatory networks of key genes were visualized by Cytoscape software³⁶.

GSEA for correlated genes of key genes

First, the correlation coefficients between the key genes and the remaining genes were calculated and ranked. This process covered all relevant genes in OP and Control groups in dataset GSE35959. Subsequently, we selected the top five genes with the highest correlation with each key gene and performed GSEA on these genes to demonstrate the normalized enrichment scores corresponding to the biological functions or pathways in which they are involved.

protein structure domain prediction

The protein structures of key genes were predicted and visualized by AlphaFoldDB database (<https://alphafold.com>)³⁷. AlphaFoldDB calculated predicted local distance difference test (pLDDT) for each residue of key proteins. The pLDDT < 50, the predicted structure with low confidence; 50 < pLDDT < 70, the predicted structure with medium confidence; 70 < pLDDT < 90, the predicted structure with high confidence; pLDDT > 90, the predicted structure with extremely high confidence.

Immune infiltration analysis

Firstly, immune cells were labeled and infiltration abundance of immune cells was calculated using ssGSEA to derive the immune cell infiltration matrix³⁸. Group comparison plots were drawn to demonstrate differences between OP and control groups in immune cell infiltration. Subsequently, the correlation among immune cells was calculated based on Spearman algorithm. Finally, correlation bubble plots were drawn to display correlation between immune cells and key genes.

Quality control of single-cell dataset

The GSM4423510 from GSE147287 (scRNA-seq)³⁹ was acquired by “Seurat v4.0” package. The samples source derived from homo sapiens bone marrow. The GSE147287 dataset was normalized by the “NormalizeData” function. The top2000 highly variable genes were truncated using “vst” method via “FindVariableFeatures” function. Cell subtypes were captured by “FindNeighbors” and “FindClusters” functions, and cell clusters were identified by “clustree” function at 0.6 resolution. Finally, dimension reduction was performed by “RunUMAP” function to visualization.

ScDEGs and single-cell enrichment analysis

Firstly, cell annotation was performed based on ImmGenData database and “Singer” package (Version 2.4.1)⁴⁰. Subsequently, the “DotPlot” and “FeaturePlot” functions were utilized to exhibit key genes expression in different cell. Finally, DEGs were identified by “FindAllMarkers” function. The top10 DEGs of each cell were selected as single cell differentially expressed genes (scDEGs). Common genes of scDEGs and AR&MRGs were subjected to GO and KEGG enrichment analyses.

Cell communication analysis

The intercellular communication was inferred and quantified by “CellChat” package (Version 2.1.1)⁴¹ and the CellPhoneDB.human database. Significantly ligand-receptor pairs were identified by ligand-receptor interaction probability and perturbation tests. Then, cell communication network was integrated by summing the number or intensity of ligand-receptor pairs. The essential receptor-ligand pairs during immune cells signaling and receive signals were shown through bubble plots.

pseudotime analysis

Differentiation start and end were identified based on trace of cell types and expression of characterized genes. Pseudotime analysis was run by “monocle” package (Version 2.30.1)^{42–44} to predict differentiation and developmental trace and to analyze key genes variations during pseudotime process.

Bone loss model mice

The research was approved by the Ethics Committee of the First Affiliated Hospital of Harbin Medical University. All experimental methods followed the relevant guidelines and regulations, as well as the ARRIVE guidelines. Female 8-week-old C57BL/6J mice were bought from Liaoning Changsheng Biotechnology Co. Following 1 week acclimatization, mice were anesthetized by intraperitoneal injection of chloral hydrate (10%), and then mice were subjected to OVX or sham surgery. Mice were housed under identical pathogen-free conditions and provided with sufficient food and water. After 8 weeks of rearing, mice were euthanized by cervical dislocation and femurs were collected.

RT-qPCR

Total RNA was extracted from mouse bone tissue using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). NANODrop2000 (Thermo Scientific, Carlsbad, USA) was utilized to verify the quality of RNA. Total RNA was reverse transcribed into cDNA in a total reaction volume of 10 μ L by cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, Massachusetts, United States) following manufacturer’s instructions. The ABI 7500 fast Real-Time PCR system (Applied Biosystems, Foster City, USA) was used to measure the expression levels of mRNAs by using SYBR Green (Applied Biosystems, USA). 18s was served as control. The primer sequences are as follows:

Genes	Forward Primer (5'–3')	Reverse Primer (5'–3')
ABAC4	CCAGCCTTTCCGAGAG	CTGAGCGCCACCTACAA
MAOA	GCTTTATCTCCCGTCCATT	CGAATCACCTTCCATACA
KIF5A	AAAGAAGCCGTGCGGTA	GGTTGGTGGGTGAGGAG
NFKB2	ATCTGGGTGTCCTGCATGT	CCTTGGCCTCCTGCTCT
BAX	GCCTCGCTCACCATCTG	CCCACCCCTCCAATAA
YWHAE	GCATTGAAGGTGGTATGGA	AACAAAAGAGGTTGAGCGA

statistical analysis

All data processing and analysis were based on R software (Version 4.3.3). For comparisons of continuous variables in two groups, Student's independent t-test was applied to statistical significance for normally distributed variables, and Wilcoxon rank sum test was applied to non-normally distributed variables. Kruskal-Wallis test was utilized for comparison of three and more groups. Correlation coefficients between variables were calculated by Spearman correlation analysis. The $P < 0.05$ was identified as statistically significant.

Results

The flowchart of this study is shown in Fig. 1.

Identification of AR&MRDEGs in OP and enrichment analysis

Firstly, the GSE35959, GSE56815 and GSE7158 datasets were normalized. The box plots demonstrated differences before and after normalization (Fig. S1). Based on GSE35959 dataset, DEGs between OP and control groups were identified. According to $|\log_{2}FC| > 1$ and adjusted $P < 0.05$, 2908 DEGs were displayed in volcano plot, including 2523 up-regulated genes and 385 down-regulated genes (Fig. 2A). And then 47 AR&MRDEGs were available by venn diagram (Fig. 2B and Table S2) of DEGs and AR&MRGs. Heatmap presented the expression

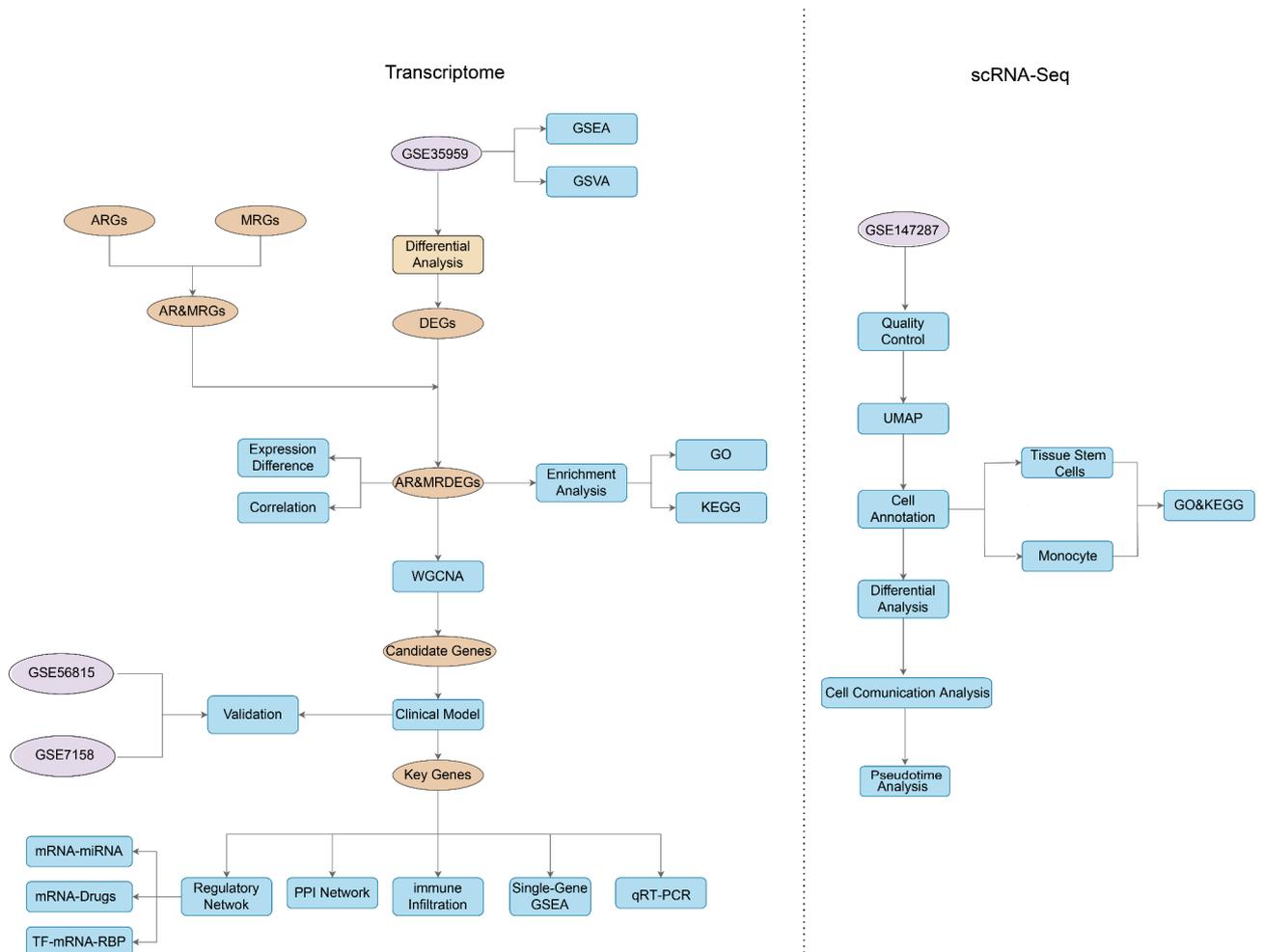


Fig. 1. Flow chart for the comprehensive analysis of AR&MRDEGs.

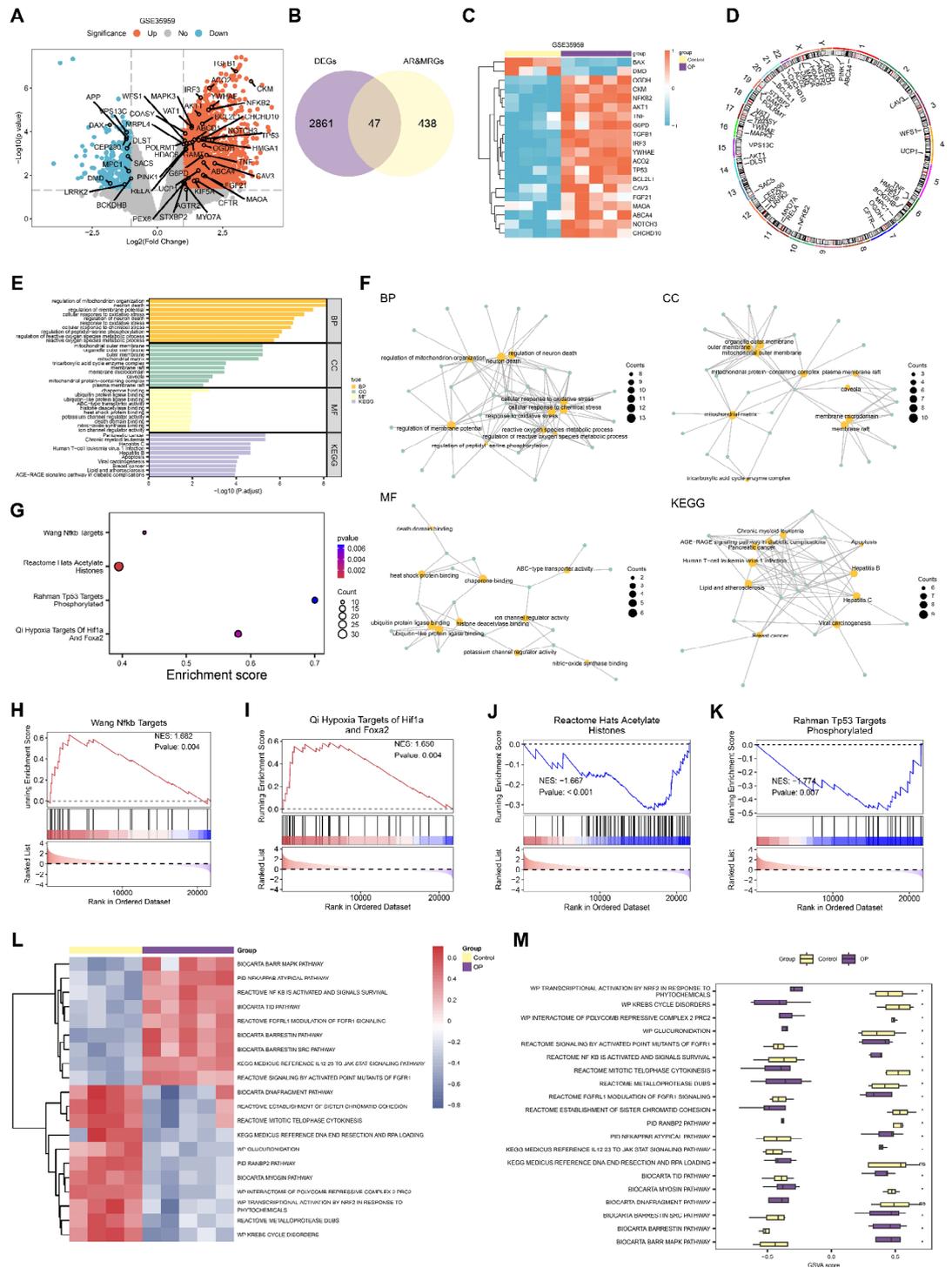


Fig. 2. Identification of AR&MRDEGs in OP and enrichment analysis. **(A)** Volcano plot of DEGs between OP and control in GSE35959. **(B)** Venn diagram of DEGs and AR&MRGs. **(C)** Heatmap of top20 of AR&MRDEGs. Purple represents OP and yellow represents control. Red means high expression and blue means low expression. **(D)** Chromosome localization map of AR&MRDEGs. **(E)** Histogram of GO and KEGG enrichment analysis of AR&MRDEGs. BP, Biological Process; CC, Cellular Component; MF, Molecular Function. **(F)** Network map of GO and KEGG enrichment analysis of AR&MRDEGs. Orange nodes represent entries, green nodes represent molecules, and lines represent the relationship between entries and molecules. **(G)** Bubble diagram of GSEA. The bubble indicates gene set size, and color represents *P* value. **(H-K)** Details of GSEA in Nfkb Targets (H), Targets of Hif1a and Foxa2 (I), Acetylate Histones (J) and Tp53 Targets Phosphorylated (K). **(L)** Heatmap of GSVA. Blue represents low enrichment and red represents high enrichment. **(M)** Group comparison plots of GSVA. The ns represents $P \geq 0.05$ and no statistical significance; * represents $P < 0.05$ and statistically significant.

of top20 of AR&MRDEGs (Fig. 2C). Finally, the chromosome localization of AR&MRDEGs was mapped. Several AR&MRDEGs including *DMD*, *MAOA*, *HDAC6*, *AGTR2*, *ABCD1* and *G6PD* were located on the X chromosome (Fig. 2D).

To explore action of AR&MRDEGs in OP, GO and KEGG enrichment analysis were performed. AR&MRDEGs were primarily enriched in biological processes (BP) such as mitochondrion organization regulation, response to oxidative stress and reactive oxygen species metabolic process; cellular components (CC) such as mitochondrial outer membrane, mitochondrial matrix and tricarboxylic acid cycle enzyme complex; molecular functions (MF) such as chaperone binding, ubiquitin protein ligase binding and death domain binding. Simultaneously, AR&MRDEGs enriched in pathways such as apoptosis, lipid and atherosclerosis and age-race signaling pathway in diabetic complications (Fig. 2E). Meanwhile, the number of AR&MRDEGs contained in entries and links among entries were visualized by network diagram (Fig. 2F). The results showed that AR&MRDEGs were closely related to mitochondrial structure and function.

To research total genes variation in OP, GSEA was run (Fig. 2G). The all genes primarily enriched in Nfkb targets (Fig. 2H), targets of Hif1a and Foxa2 (Fig. 2I), acetylate histones (Fig. 2J), and tp53 targets phosphorylated (Fig. 2K). To explore differences of gene set, GSEA was performed. The top20 entries were displayed in heatmap (Fig. 2L). Subsequently, difference validation was performed by Mann-Whitney U test. The group comparison graph indicated that krebs cycle disorders, glucuronidation and nfkb pathway were significant between OP and control (Fig. 2M).

Differential expression analysis and correlation analysis of AR&MRDEGs

To explore differential expression of AR&MRGs, group comparison plots demonstrated 41 AR&MRGs were statistically expressed between OP and control group (Fig. S2A). Next, correlations among 47 AR&MRGs were calculated and correlation heatmaps were drawn (Fig. S2B). The correlation heatmap showed that positive correlations among most AR&MRDEGs, with significant positive correlations between *HDAC6* and *CAV3*, *VAT1* and *OGDH* ($r=0.983$, $P<0.001$). Finally, 4 pairs genes with strongest correlation were demonstrated by correlation scatterplots (Fig. S2C-F).

Weighted gene co-expression network analysis

To filter key module genes, WGCNA was performed. When the scale-free fitting index is 0.85, the optimal soft threshold is 16 (Fig. 3A). When the screening criterion was 0.4, the genes were clustered in 6 modules, including: darkgreen, orange, grey, brown, darkorange and green (Fig. 3B). Then, the relationships between genes and merged modules were visualized (Fig. 3C). The correlations of 6 modules with OP were acquired (Fig. 3D). Finally, darkgreen module served as the key module for subsequent analysis. The 22 candidate genes were obtained by venn diagram of darkgreen module genes and AR&MRDEG (Fig. 3E).

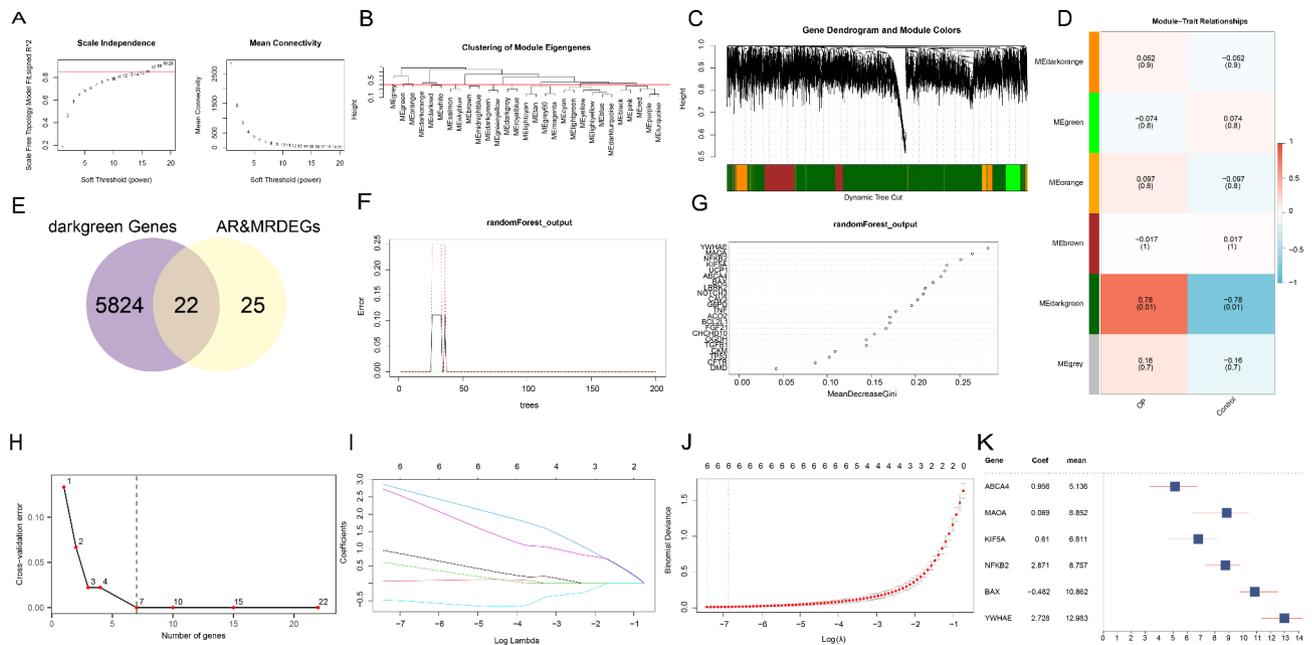


Fig. 3. WGCNA and machine learning. (A) Scale-free networks of soft threshold power and mean connectivity. (B) Module aggregation of top 8000 genes in variance. (C) Gene clustering dendrogram by hierarchical clustering. (D) Correlation analysis between modules and OP. (E) Venn diagram of AR&MRDEGs and darkgreen module genes. (F) Training error plot of RF. (G) MeanDecreaseGini scatterplot of candidate genes. (H) Error plots of cross-validation. (I) LASSO coefficient path. (J) Cross verification curve in LASSO. (K) Forest map of key genes in LASSO.

Machine learning and construction of diagnostic model

To investigate clinical value of 22 candidate genes, the RF was applied. The error plot of decision trees displayed that the error leveled off when the number of decision trees was 50 (Fig. 3F). Meanwhile, meandecreasegini scatter plots of 22 candidate genes were plotted (Fig. 3G). The cross-validation error plot demonstrated that the model error is smaller when the number of genes is 7 (Fig. 3H).

Then, based on RF, LASSO regression analysis was constructed. LASSO coefficient path (Fig. 3I) and cross verification curve (Fig. 3J) were plotted. The LASSO derived 6 key genes, including *ABCA4*, *MAOA*, *KIF5A*, *NFKB2*, *BAX* and *YWHAE*. The forest maps of key genes were drawn (Fig. 3K). The diagnostic model was constructed according to key genes and their risk coefficients. The LASSO riskscore was calculated as follows:

$$\text{RiskScore} = \sum_i \text{Coefficient}(\text{gene}_i) * \text{mRNA Expression}(\text{gene}_i)$$

Validation of diagnostic model

To validate the diagnostic model, the ROC curve of risk score was plotted. The risk score presented high accuracy ($\text{AUC} > 0.9$) in GSE35959 (Fig. S3A). The ROC curves of key genes indicated that *ABCA4*, *MAOA*, *KIF5A*, *NFKB2*, *BAX*, and *YWHAE* all provided high diagnosis accuracy ($\text{AUC} > 0.9$) (Fig. S3B–D). Then, the nomogram plot assessed potency of key genes for diagnostic model, with higher potency of *NFKB2* expression and lower potency of *BAX* expression (Fig. S3E). In addition, calibration curve was plotted to determine precision and discrimination of diagnostic model. The calibration curve plot demonstrated that dashed line of calibration line substantially coincided with diagonal line of ideal model (Fig. S3F). Finally, the clinical utility of diagnostic model was assessed by DCA. The net benefit of diagnostic model was consistently higher than all positive and all negative within most risk thresholds (Fig. S3G).

Meanwhile, external validation of diagnostic model was conducted based on GSE56815 and GSE7158 datasets. The ROC curves displayed both risk scores and key genes with high accuracy ($0.5 < \text{AUC} < 0.7$) (Fig. 4A, D, H and K). The nomogram indicated that the expression of key genes had high utility for diagnostic model (Fig. 4E and L). The calibration curve demonstrated that calibration line is close to diagonal line, and the model has better discrimination ability (Fig. 4F and M). The DCA showed that strategies based on genetic information bring higher net benefits over a wide range of risk thresholds (Fig. 4G and N).

Construction of PPI networks and regulatory networks

The PPI networks were constructed through GeneMANIA database, containing 6 key genes and 20 functionally similar genes (Fig. S4). Firstly, miRNAs related to key genes were extracted from StarBase database, and the mRNA-miRNA regulatory network was constructed, containing *BAX*, *KIF5A*, *YWHAE3* and 46 miRNAs (Fig. 5A). Potential drugs for key genes were then identified through CTD database, and mRNA-drug network was constructed, containing *MAOA*, *BAX2* and 38 drugs or molecular compounds (Fig. 5B). Finally, TFs combining with key genes were available through ChIPBase database and RBPs associated with key genes were predicted through StarBase database. The TF-mRNA-RBP network was constructed, containing 6 key genes, 66 TFs and 67 RBPs (Fig. 5C).

GSEA for correlated genes of key genes

To explore pathways associated with key genes, GSEA was run. GSEA was performed on remaining genes and the biological functions or pathways with top5 normalized enrichment score (NES) were demonstrated (Fig. S5). The *smarca2*-targets showed association with *ABCA4*, *BAX*, *KIF5A*, *MAOA*, *NFKB2* and *YWHAE*; *MIR21*-targets was association with *BAX*, *NFKB2* and *YWHAE*; *Myc*-targets by serum had negative association with *KIF5A* and *MAOA*.

Prediction of protein structural domains

The protein structures of 6 key genes were available by AlphaFoldDB (Fig. S6). The major structural domains of 3 key genes had extremely high confidence ($\text{pLDDT} > 90$), including: *BAX*, *MAOA*, *YWHAE*. The major structural domains of 3 key genes had high confidence ($70 < \text{pLDDT} < 90$), including: *ABCA4*, *KIF5A*, *NFKB2*.

Immune infiltration analysis

Studies have confirmed that immune disorders were involved in the development of OP⁴⁵. Hence, immune infiltration analysis was carried out in this study. Firstly, group comparison plots demonstrated significant differences in immune cells infiltration abundance, including: activated dendritic cell, natural killer T cell and CD56dim natural killer cell (Fig. S7A). Next, correlation heatmaps revealed strong correlations among most immune cells, with central memory CD8 T cell and activated CD4 T cell exhibiting the strongest negative correlation ($r = -0.95$, $P < 0.05$) (Fig. S7B). Finally, correlation bubble plots indicated that strong correlations were presented between key genes and immune cells, with *BAX* and activated dendritic cell having strongest negative correlation ($r = -0.9$, $P < 0.05$) (Fig. S7C).

cell annotation and key genes expression

Single-cell counts matrix of GSM4423510 in GSE147287 from OP was investigated. The violin plot presented the number of gene features, the count of genes, and the percent of mitochondria genes per cell (Fig. 6A). UMAP was applied for dimension reduction and visualization. When the resolution was 0.6, 7335 cells were categorized into 17 separate clusters (Fig. 6B). The cell clusters were then defined into 11 cell types, including tissue stem cells (TSCs), myelocytes, hematopoietic stem cells (HSC), neutrophils, CD8⁺ T cells, macrophages, monocytes, adipocytes, erythrocytes, erythrocytes, B cells and chondrocytes (Fig. 6C). The bar graph showed the percentage

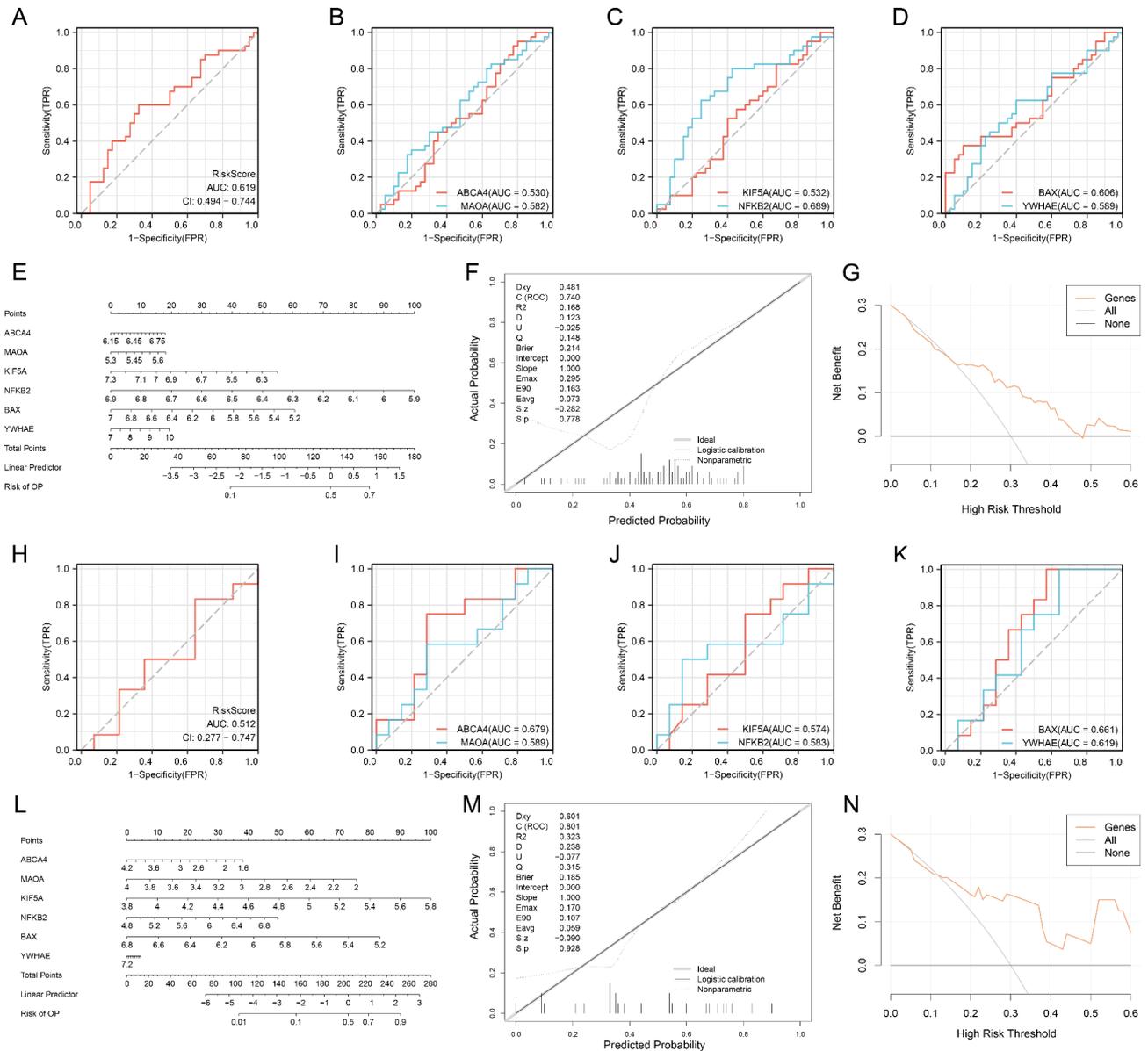


Fig. 4. The validation of diagnostic model in GSE56815 and GSE7158. **(A)** ROC curves of risk scores in GSE56815. **(B)** ROC curves of *ABCA4* and *MAOA*. **(C)** ROC curves *KIF5A* and *NFKB2*. **(D)** ROC curves of *BAX* and *YWHAE*. **(E)** Nomogram of key genes in diagnostic model. **(F)** Calibration curve of diagnostic model. **(G)** Decision curve analysis of diagnostic model. **(H)** ROC curves of risk scores in GSE7158. **(I)** ROC curves of *ABCA4* and *MAOA*. **(J)** ROC curves *KIF5A* and *NFKB2*. **(K)** ROC curves of *BAX* and *YWHAE*. **(L)** Nomogram of key genes in diagnostic model. **(M)** Calibration curve of diagnostic model. **(N)** Decision curve analysis of diagnostic model.

of 11 cell types, with TSCs occupying highest percentage (Fig. 6D). Bubble plots demonstrated the expression levels and percentages of 4 key genes (*BAX*, *MAOA*, *NFKB2* and *YWHAE*) in different cell (Fig. 6E). UMAP plots separately presented the expression of 4 key genes in GSM4423510 (Fig. 6F).

Single-cell DEGs and enrichment analysis

The scDEGs among cells were identified by $|\log FC| > 2$ and $adj. P < 0.05$ (Fig. 6G). The expression of top10 scDEGs was visualized (Fig. 6H). The heatmap displayed that *COL14A1*, *CHRDD1*, etc. were primarily expressed in TSCs; *FABP4*, *CIQB*, etc. were mainly expressed in macrophages; *KLF4*, *ASGR1*, etc. were mainly expressed in monocytes.

Next, GO and KEGG enrichment was carried out for common genes of scDEGs with AR&MRGs. The common genes of scDEGs and AR&MRGs in TSCs were primarily enriched in myeloid leukocyte differentiation, regulation of ERK1 and ERK2 cascade and other BP; vesicle lumen, endocytic vesicle and other CC; ephrin receptor binding, lamin binding and other MF; ErbB signaling pathway, cholesterol metabolism and other pathways (Fig. 6I). The common genes of scDEGs and AR&MRGs in monocyte were primarily enriched in

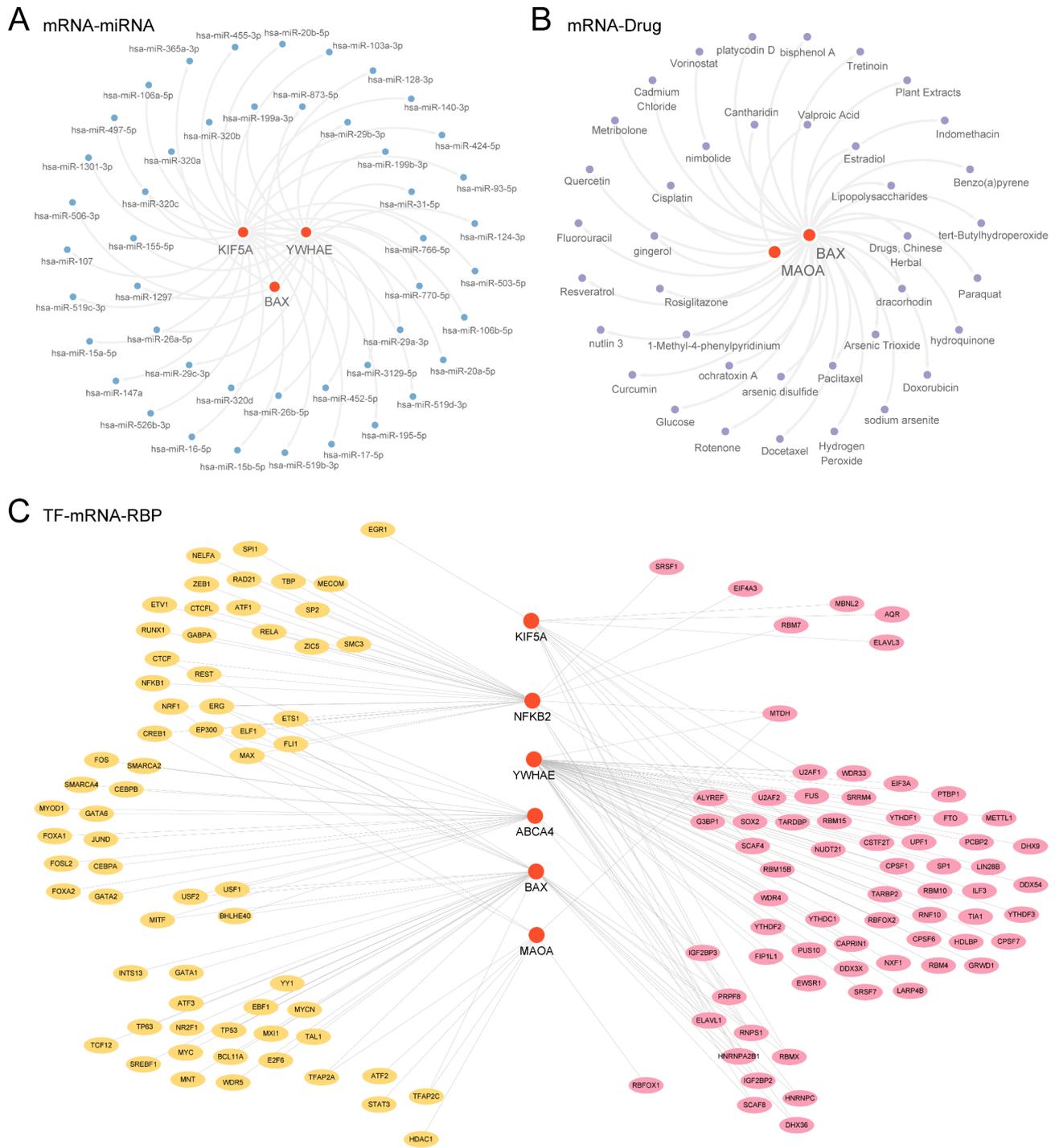


Fig. 5. Regulatory network of key genes. **(A)** The mRNA-miRNA network. **(B)** The mRNA-Drug network. **(C)** TF-mRNA-RBP network. Red represents mRNA, blue represents miRNA, purple represents drug, yellow represents TF, and pink represents RBP.

regulation of mitochondrial depolarization, regulation of membrane depolarization and other BP; endocytic vesicle lumen, lysosomal lumen and other CC; proteoglycan binding, glycosaminoglycan binding and other MF (Fig. 6).

Sell communication analysis

Communication among 11 cell types was inferred and quantified. Heatmaps and circle plots demonstrated the number (Fig. 7A) and intensity (Fig. 7B) of cell communication. TSCs was active in cell communication networks. There was plentiful signal transduction among TSCs, macrophages and monocytes in OP (Fig. 7C-E). Cell communication among TSCs, macrophages, monocytes and immune cells revealed the presence of

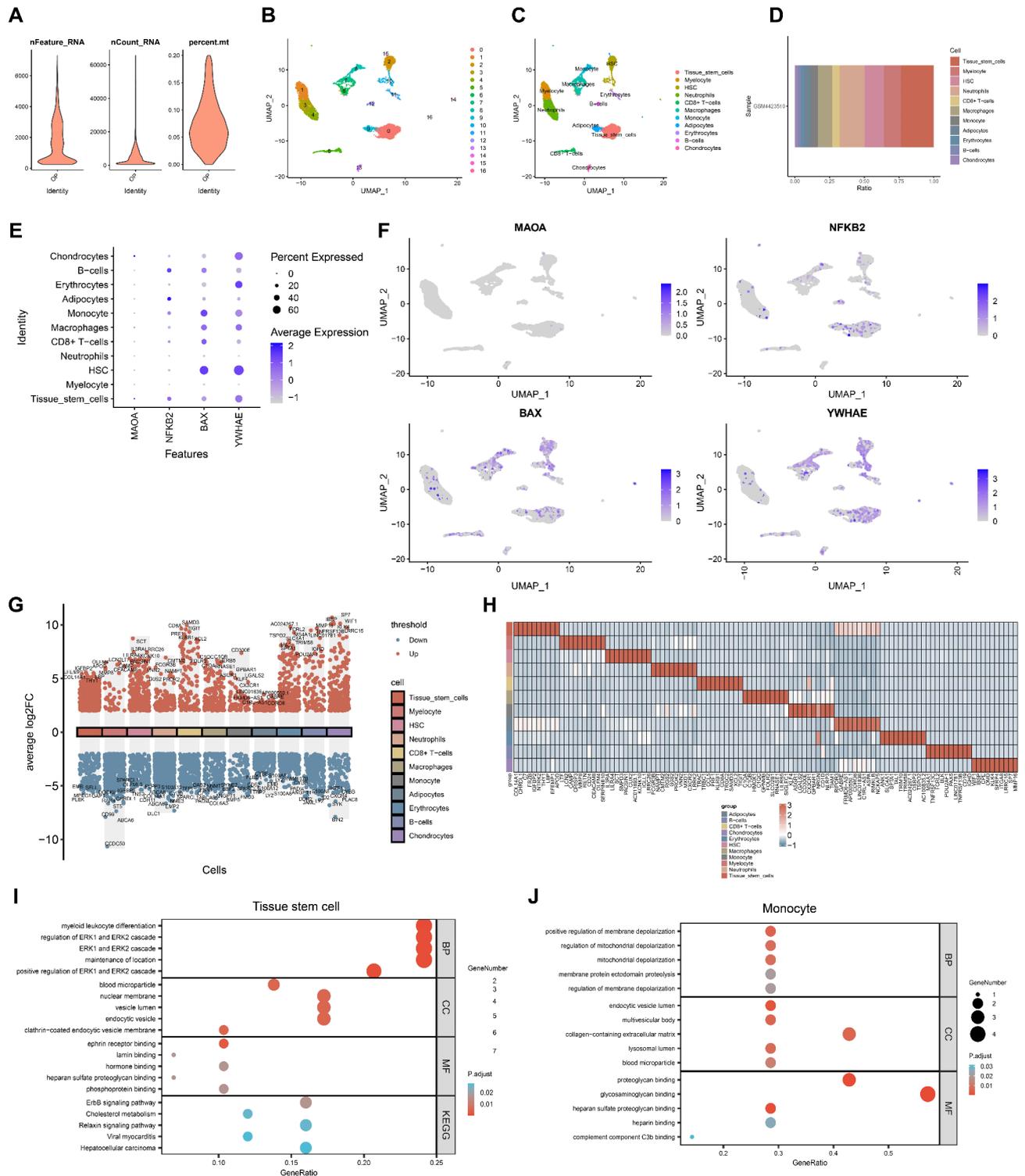


Fig. 6. Key genes expression, single-cell DEGs and enrichment analysis. **(A)** Violin map of gene expression. **(B)** The 7335 cells were clustered into 17 cell clusters by UMAP. **(C)** Cells were annotated into 11 cell types. **(D)** Histogram of cell proportions. **(E)** Bubble diagram of expression levels of 4 key genes. **(F)** The expression levels of key genes *MAOA*, *NFKB2*, *BAX*, *YWHAE* in GSE147287. **(G)** Volcano plot of scDEGs. **(H)** Heatmap of scDEGs expression. **(I)** Enrichment analyses of common genes of AR&MRGs and scDEGs in TSCs. **(J)** Enrichment analyses of common genes of AR&MRGs and scDEGs in monocyte.

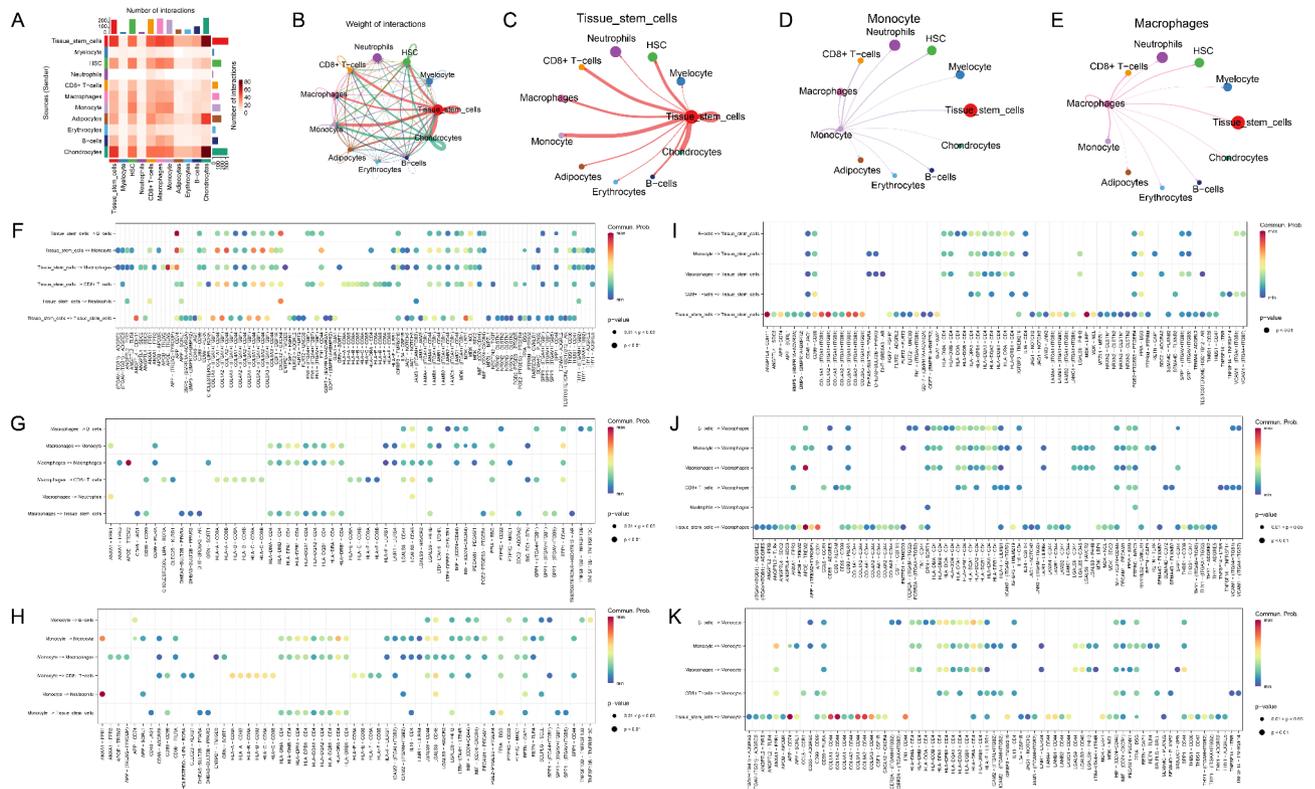


Fig. 7. Intercellular communication analysis. **(A)** Heat map of interactions number. among 11 cells types. **(B)** Network diagram of interactions strength among 11 cell types. The nodes represent different cell types, the arrows indicate signaling direction, and thicker lines mean higher strength of interactions. **(C-E)** Intercellular communication. between TSCs **(C)**, macrophages **(D)**, monocytes **(E)** and other cells, respectively. **(F-H)** Cell communication ligand-receptor pairs for signals sent to immune cells by TSCs, macrophages, and monocytes. **(I-K)** Cell communication ligand-receptor pairs in TSCs, macrophages, and monocytes receiving immune cell signals.

significant receptor-ligand pairs (Fig. 7F and K). In TSCs-B cells, the APP-CD74 receptor-ligand pair had a high intensity of action.

pseudotime analysis

The development trace of 11 cell was demonstrated by differentiation developmental chronograms (Fig. 8A) and trace diagrams (Fig. 8B). The development starting point was state 1, and the developmental end points were state 2 and state 3. The distribution of 11 cell in cell track skeleton map was plotted (Fig. 8C). The bar graphs demonstrated 4 key genes expression at different development nodes (Fig. 8D). The *BAX* was expressed highest in state 2 and lowest in state 1; *MAOA* was expressed low in all developmental stages; *NFKB2* was expressed high in state 3 and low in state 1; and *YWHAE* was expressed highest in state 2 and low in state 1. Finally, heatmap exhibited expression of key genes in pseudotime (Fig. 8E).

qRT-PCR

The expression of key genes was verified by qRT-PCR (Fig. 9). The expression of *ABCA4*, *MAOA*, *KIF5A* and *NFKB2* in OP were higher than in control, and the expression of *BAX* showed the opposite trend. The expression of *YWHAE* showed no difference between OP and control groups.

Discussion

OP is a bone disease characteristically associated with reduced bone strength and elevated risk of fracture, and occurs primarily in postmenopausal women and older men⁴⁶. OP is prevalent and extremely dangerous, but diagnosis and treatment rates remain low⁴⁷. Currently, the OP diagnosis relies on imaging and BMD testing by DXA, but all have limitations⁴⁸. OP imaging is highly subjective and difficult to detect early bone loss. DXA is insensitive to early bone loss, fails to assess the severity of OP and predict fracture risk⁴⁹. Early diagnosis of bone loss is critical for the management of OP patients.

Aging has recognized as an essential mechanism independently of estrogen deficiency leading to OP⁵⁰. Meanwhile, mitochondrial dysfunction is one of key hallmarks of cell aging⁵¹. Studies have demonstrated that bone marrow mesenchymal stem cells (BMSCs) aging affects viability, differentiation, and thus decrease osteogenesis, leading to OP⁵². Aging and mitochondria are potential targets for OP. However, identified genetic

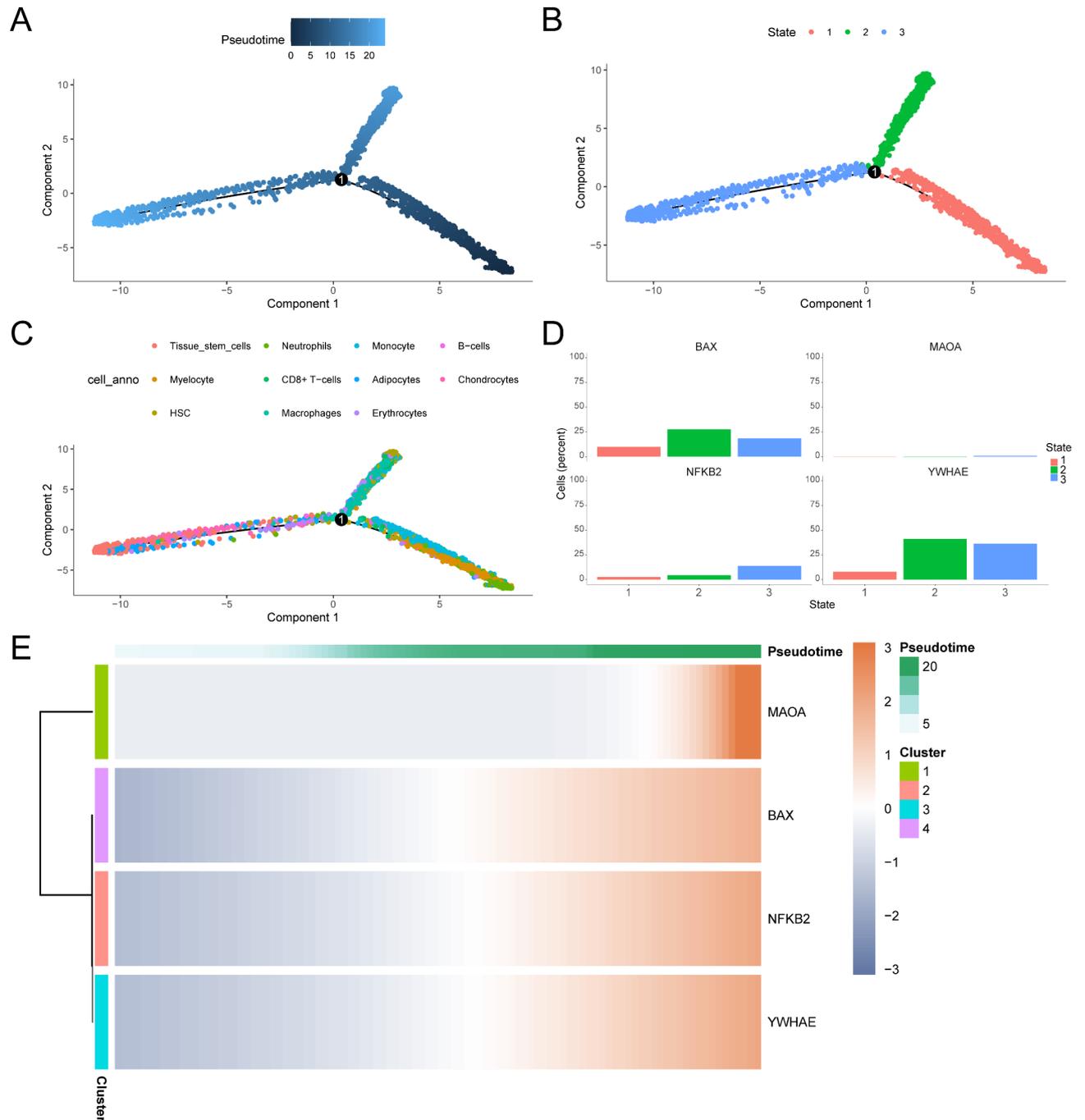


Fig. 8. Pseudotime analysis. **(A)** Cell track skeleton in pseudotime. The color from dark to light represents pseudotime order. **(B)** Stage of cell track skeleton. **(C)** Distribution of 11 cell types in cell track skeleton. **(D)** Histogram of key genes expression at different developmental stages. **(E)** Heatmap of pseudotime expression of key genes.

biomarkers of AR&MRGs in OP have not been reported. Multi-level, multi-dimensional bioinformatics analysis might contribute to identify more accurate genetic biomarkers. Therefore, this study utilized public data to screen key AR&MRGs and construct a diagnostic model that might contribute to the diagnosis of OP in future.

Due to difficulty in obtaining human BMSCs, GSE35959 is available as the largest sample size dataset. In this study, 47 AR&MRDEGs were available from differential expression analysis and WGCNA. AR&MRDEGs were enriched in mitochondrial structure and function such as mitochondrial tissue, mitochondrial outer membrane, mitochondrial matrix, regulation of mitochondrial membrane potential, response to oxidative stress, metabolism of reactive oxygen species and tricarboxylic acid cycle enzyme complex. When organisms exposed to oxidative stress stimuli, excessive reactive oxygen species were released, mitochondrial membrane potential depolarized, mitochondrial membrane permeability increased, and mitochondrial dysfunction

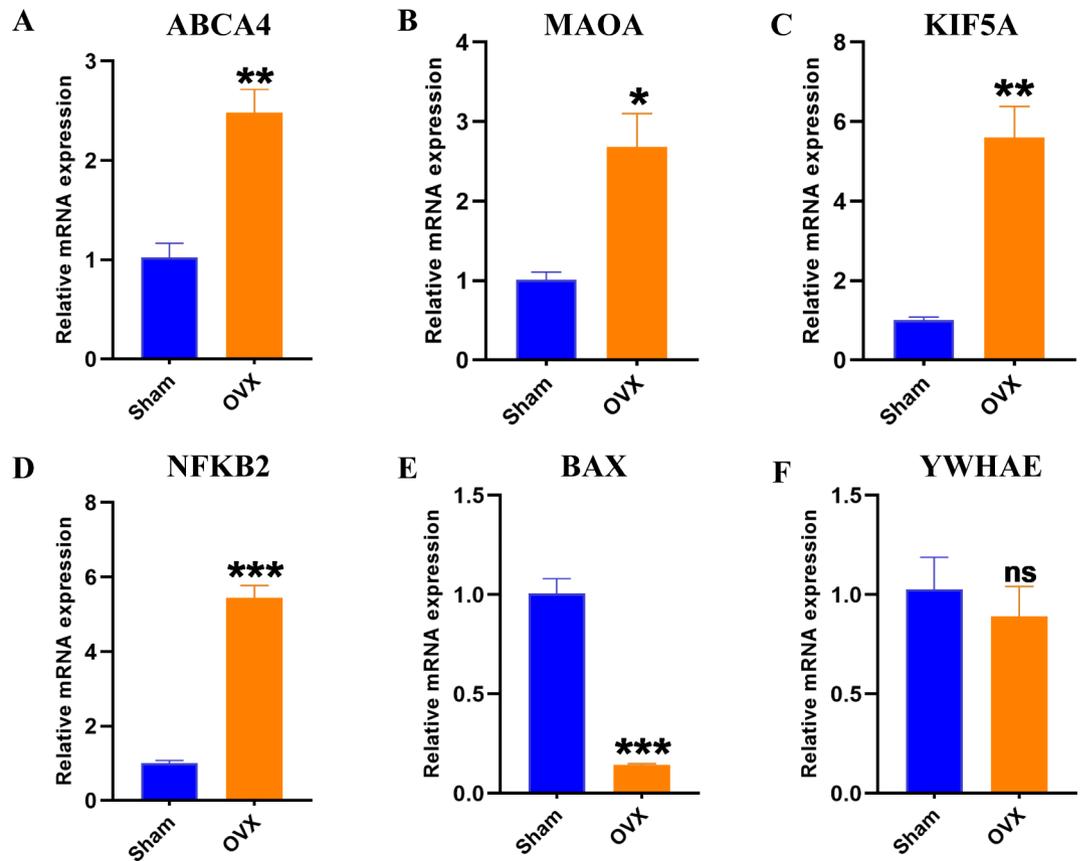


Fig. 9. The qRT-PCR validation of key genes. (A)*ABCA4*(B)*MAOA*(C)*KIF5A*(D)*NFKB2*(E)*BAX*(F)*YWHAE*. The ns represents $P \geq 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

including oxidative phosphorylation, tricarboxylic acid cycle and ATP generation even mitochondria-derived apoptosis and mitochondrial cleavage were observed⁵³. The present study revealed key biological processes and signaling pathways associated with aging and mitochondria in OP, providing new insights for understanding OP. Modulation of mitochondria to reverse aging to ameliorate bone loss needs more experiments and more animal models to explore. Meanwhile, mitochondria supplements might be one of potential strategies for OP patients⁵⁴.

In this study, the diagnostic model was constructed based on 6 key genes (*YWHAE*, *MAOA*, *NFKB2*, *KIF5A*, *ABCA4* and *BAX*). The calibration curves revealed that model had certain discrimination and no systematic bias. Decision curves indicated that net benefits from genetic information strategy were favorable in most risk threshold ranges. The diagnostic model had moderate external prediction ability, but it also laid foundation for more accurate diagnosis tools in future. The present diagnostic model focused on aging and mitochondria in BMSCs. However, OP was a systemic, multi-organ, multi-phenotype disease. The pathogenesis of OP was complex and involved genetic and environmental factors. Adding key genes of osteoclasts, key genes of estrogen pathway, and even clinical features such as age and basal disease to diagnostic model might greatly improve generalizability and comprehensive prediction ability.

The PPI network was built by GeneMANIA database, including 6 key genes and 20 functionally similar proteins. *YWHAE*, *BAX* and *MAOA* were located in core nodes of the network, interacting with functionally similar proteins, and could serve as targets to regulate aging and mitochondrial of BMSCs. *BAX*, a proapoptotic protein, initiates osteoblasts apoptosis by increasing mitochondrial membrane permeability and releasing cytochrome *c*⁵⁵. *MAOA* is associated with BMD, especially in postmenopausal women, and is critical locus for bone mass in Japanese women⁵⁶. It is reported that *NFKB2* may be one of potential pathogenic genes and has diagnosis value for OP, which is consistent with this study conclusion⁵⁷. Studies have identified *YWHAE*, a hub gene for differentially expressed iron metabolism-related genes in OP⁵⁸. There are few reports about *KIF5A*, *ABCA4* associated with OP.

The mRNA-miRNA network showed that *KIF5A*, *YWHAE*, *BAX* were regulated by multiple miRNAs. The post-transcriptional regulatory mechanisms were complicated. MiRNAs could regulate BMSCs osteogenesis differentiation through key proteins. MiR-124-3p inhibits osteogenic differentiation of BMSCs by suppressing PI3K/Akt/mTOR⁵⁹. MiR-128-3p targets directly nuclear factor 5 of activated T cells (NFAT5), a protein that binds to osteoprotegerin and modulates osteoclastogenesis in the presence of nuclear factor κ B receptor activator ligand⁶⁰. The TF-mRNA-RBP network indicated that *NFKB2* shares TF with *ABCA4*, *BAX*, while *YWHAE* shares RBP with *KIF5A*, *BAX*. Collaborative mechanisms exist among key genes at transcriptional level. Both EP300 and ERG were potential TFs for *ABCA4*, *BAX* and *NFKB2*, but no studies have been reported in OP, and more basic

experiments were needed to confirm their roles. The mRNA-Drug network revealed that *MAOA*, *BAX* interacted with various drugs. *MAOA* and *BAX* were potential targets for drug screening of OP. Studies have found that targeted delivery of curcumin to bone marrow of diabetic OP patients promoted osteogenic differentiation of BMSCs and increased bone mass by enhancing mitochondrial function and inhibiting ferroptosis⁶¹. Dasatinib and quercetin, which improved aging BMSCs function and promoted bone regeneration by targeting aging cells and senescence-associated secretory phenotypes (SASPs), was recently explored as a novel therapy for a variety of age-related diseases⁶².

Heatmap of scDEGs showed that *APOD* and *CHRD1* were highly expressed in TSCs. Studies confirmed that *APOD* influences OP progression by regulating BMSCs osteogenic differentiation through the *PI3K/Akt* pathway⁶³. Meanwhile, *CHRD1* promotes BMSCs osteogenic differentiation by activating *BMP4-SMAD1/5/9* pathway⁶⁴. TSCs enrichment showed AR&MRGs enriched in *ERK* pathway. Studies have demonstrated that *ERK* pathway activation promotes BMSCs differentiation to osteoblasts, facilitates osteogenesis, and prevents bone loss⁶⁵. Monocyte enrichment revealed AR&MRGs enrichment in mitochondrial membrane potential depolarization. Studies have demonstrated that estrogen promotes mitochondrial apoptotic death of early osteoclast progenitors via *Bak/Bax*, reducing osteoclast numbers and reducing bone loss⁶⁶. Meanwhile, cell communication analysis revealed substantial cell communication among TSCs, monocytes, and macrophages. In OP, M1-like macrophages have increased mitochondrial transfer to BMSCs, and abnormal metabolism in BMSCs regulates bone homeostasis⁵⁴. This study provided insights for aging and mitochondria at the single-cell level in OP.

There are some limitations to this study. First, we performed the batch effect removal independently for the three datasets (GSE35959, GSE56815, and GSE7158).

This choice is intended to minimize the interference of potential batch effects in the results, but it may also lead to the loss of some information, which may weaken the comprehensiveness of the results. Differential analysis was based only on the test set GSE35959 and further screened for 47 AR&MRDEGs. Nevertheless, the presence of dataset bias needs to be interpreted with caution to avoid overstating the generalizability of the findings. However, we recognize that computational analyses have limitations of their own. For example, we used t-tests and Wilcoxon rank sum tests in differential expression analysis. These methods may not fully capture all aspects of biological complexity when dealing with small sample sizes, especially if the data are unevenly distributed or highly noisy.

Although our study has made some progress in identifying key genes associated with OP and constructing a diagnostic model, there are still some limitations. First, the construction of the model is mainly based on the training set GSE35959, which has a relatively small number of samples, which may lead to overfitting of the model during the training process, thus affecting its generalization ability on external validation sets (GSE56815 and GSE7158). Meanwhile, despite our efforts to improve the robustness of the model through cross-validation of multiple algorithms on the test set, it is still difficult to rule out potential biases and the influence of external variables on the results. More diverse datasets and more sophisticated algorithms are needed in the future to further validate and enhance the reliability and applicability of the models.

Conclusion

In this study, 6 key genes (*ABCA4*, *MAOA*, *KIF5A*, *NFKB2*, *BAX* and *YWHAE*) were identified. A novel diagnostic model was constructed and preliminarily validated. Meanwhile, this study enriched insights of aging and mitochondria in OP, and laid foundation for the subsequent deep mechanism research to drive OP research advances.

Data availability

The datasets analyzed during the current study are available in public databases such as GEO (<https://www.ncbi.nlm.nih.gov/geo/>), MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb>) and GeneCards.

Received: 5 October 2024; Accepted: 30 December 2024

Published online: 06 January 2025

References

- Snyder, S. Postmenopausal osteoporosis. *N. Engl. J. Med.* **390** (7), 675–676 (2024).
- Ensrud, K. E. & Crandall, C. J. Osteoporosis. *Annals Intern. Med.*, **177**(1). (2024).
- Huang, J. et al. β -Receptor blocker enhances the anabolic effect of PTH after osteoporotic fracture. *Bone Res.* **12** (1), 18 (2024).
- Zhang, K. et al. DeepmdQCT: a multitask network with domain invariant features and comprehensive attention mechanism for quantitative computer tomography diagnosis of osteoporosis. *Comput. Biol. Med.* **170**, 107916 (2024).
- Zheng, J., He, J. & Li, H. *FAM19A5 in vascular aging and osteoporosis: Mechanisms and the calcification paradox*. Ageing Research Reviews, 99: p. 102361. (2024).
- Yilmaz, D. et al. Mouse models of accelerated aging in musculoskeletal research for assessing frailty, Sarcopenia, and osteoporosis - A review. *Ageing Res. Rev.* **93**, 102118 (2024).
- Zheng, C. X. et al. Mitochondrial regulation of stem cells in bone homeostasis. *Trends Mol. Med.*, **26**(1). (2020).
- Wang, F. S. et al. *Biophysical Modulation of the Mitochondrial Metabolism and Redox in Bone Homeostasis and Osteoporosis: How Biophysics Converts into Bioenergetics*. Antioxidants (Basel, Switzerland), **10**(9). (2021).
- Amorim, J. A. et al. Mitochondrial and metabolic dysfunction in ageing and age-related diseases. *Nat. Rev. Endocrinol.* **18** (4), 243–258 (2022).
- Benisch, P. et al. The transcriptional profile of mesenchymal stem cell populations in primary osteoporosis is distinct and shows overexpression of osteogenic inhibitors. *PLoS One.* **7** (9), e45142 (2012).
- Zhou, Y. et al. A novel approach for correction of crosstalk effects in pathway analysis and its application in osteoporosis research. *Sci. Rep.* **8** (1), 668 (2018).
- Lei, S. F. et al. An in vivo genome wide gene expression study of circulating monocytes suggested GBP1, STAT1 and CXCL10 as novel risk genes for the differentiation of peak bone mass. *Bone* **44** (5), 1010–1014 (2009).

13. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41** (Database issue), D991–D995 (2013).
14. Davis, S. & Meltzer, P. S. *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor*. Bioinformatics (Oxford, England), **23**(14): pp. 1846–1847. (2007).
15. Stelzer, G. et al. *The GeneCards suite: from Gene Data Mining to Disease Genome sequence analyses*. *Curr. Protocols Bioinf.*, **54**. (2016).
16. Liberson, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinf. (Oxford England)*. **27** (12), 1739–1740 (2011).
17. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** (7), e47 (2015).
18. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for Circos 2D track plots. *BMC Bioinform.* **14**, 244 (2013).
19. Mi, H. et al. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47** (D1), D419–D426 (2019).
20. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000).
21. Yu, G. et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: J. Integr. Biology*. **16** (5), 284–287 (2012).
22. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102** (43), 15545–15550 (2005).
23. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
24. Liu, W. et al. [Weighted gene co-expression network analysis in biomedicine research]. *Sheng Wu Gong. Cheng Xue Bao = Chin. J. Biotechnol.* **33** (11), 1791–1801 (2017).
25. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
26. Liu, Y. & Zhao, H. *Variable importance-weighted Random Forests*. Quantitative Biology (Beijing, China), **5**(4): pp. 338–351. (2017).
27. Cai, W. & van der Laan, M. *Nonparametric Bootstrap Inference for the Targeted Highly Adaptive Least Absolute Shrinkage and Selection Operator (LASSO) Estimator* (The International Journal of Biostatistics, 2020).
28. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenetics*. **11** (1), 123 (2019).
29. Wu, J. et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. *Cancer Commun. (London England)*. **40** (7), 301–312 (2020).
30. Van Calster, B. et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* **74** (6), 796–804 (2018).
31. Franz, M. et al. GeneMANIA update 2018. *Nucleic Acids Res.* **46** (W1), W60–W64 (2018).
32. Li, J. H. et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42** (Database issue), D92–D97 (2014).
33. Grondin, C. J. et al. Predicting molecular mechanisms, pathways, and health outcomes induced by Juul e-cigarette aerosol chemicals using the comparative Toxicogenomics database. *Curr. Res. Toxicol.* **2**, 272–281 (2021).
34. Zhou, K. R. et al. CHIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* **45** (D1), D43–D50 (2017).
35. Singh, A. RNA-binding protein kinetics. *Nat. Methods*. **18** (4), 335 (2021).
36. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (11), 2498–2504 (2003).
37. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596** (7873), 583–589 (2021).
38. Xiao, B. et al. Identification and Verification of Immune-related gene prognostic signature based on ssGSEA for Osteosarcoma. *Front. Oncol.* **10**, 607622 (2020).
39. Wang, Z. et al. Single-cell RNA sequencing deconvolutes the in vivo heterogeneity of human bone marrow-derived mesenchymal stem cells. *Int. J. Biol. Sci.* **17** (15), 4192–4206 (2021).
40. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20** (2), 163–172 (2019).
41. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12** (1), 1088 (2021).
42. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods*. **14** (3), 309–315 (2017).
43. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. **14** (10), 979–982 (2017).
44. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32** (4), 381–386 (2014).
45. Fischer, V. & Haffner-Luntzer, M. Interaction between bone and immune cells: implications for postmenopausal osteoporosis. *Semin. Cell Dev. Biol.* **123**, 14–21 (2022).
46. Walker, M. D. & Shane, E. Postmenopausal osteoporosis. *N. Engl. J. Med.* **389** (21), 1979–1991 (2023).
47. Bellido, T. *Bisphosphonates for osteoporosis: from bench to clinic*. *J. Clin. Investig.*, **134**(6). (2024).
48. Reid, I. R. & Billington, E. O. Drug therapy for osteoporosis in older adults. *Lancet (London England)*. **399** (10329), 1080–1092 (2022).
49. Zeng, R. et al. Identification of a potential diagnostic signature for postmenopausal osteoporosis via transcriptome analysis. *Front. Pharmacol.* **13**, 944735 (2022).
50. Li, J. et al. TGFβ1 + CCR5+ neutrophil subset increases in bone marrow and causes age-related osteoporosis in male mice. *Nat. Commun.* **14** (1), 159 (2023).
51. López-Otín, C. et al. Hallmarks of aging: an expanding universe. *Cell* **186** (2), 243–278 (2023).
52. Zhang, H. et al. LncRNA NEAT1 controls the lineage fates of BMSCs during skeletal aging by impairing mitochondrial function and pluripotency maintenance. *Cell Death Differ.* **29** (2), 351–365 (2022).
53. Guo, Y. et al. Mitochondrial dysfunction in aging. *Ageing Res. Rev.* **88**, 101955 (2023).
54. Cai, W. et al. Mitochondrial transfer regulates cell fate through metabolic remodeling in osteoporosis. *Adv. Sci. (Weinheim Baden-Württemberg Germany)*. **10** (4), e2204871 (2023).
55. Jia, B. et al. Baicalin attenuates dexamethasone-induced apoptosis of bone marrow mesenchymal stem cells by activating the hedgehog signaling pathway. *Chin. Med. J.* **136** (15), 1839–1847 (2023).
56. Yamada, Y., Ando, F. & Shimokata, H. Association of genetic variants of MAOA and SH2B1 with bone mineral density in community-dwelling Japanese women. *Mol. Med. Rep.* **1** (2), 269–274 (2008).
57. Xia, B. et al. Identification of potential pathogenic genes associated with osteoporosis. *Bone Joint Res.* **6** (12), 640–648 (2017).
58. Li, Z. et al. Identification and validation of iron metabolism genes in osteoporosis. *BMC Med. Genom.* **17** (1), 5 (2024).
59. Gan, L. et al. Kaempferol promotes the osteogenesis in rBMSCs via mediation of SOX2/miR-124-3p/PI3K/Akt/mTOR axis. *Eur. J. Pharmacol.* **927**, p174954 (2022).
60. Zhang, H. et al. Long noncoding RNA KCNQ1OT1 inhibits osteoclast differentiation by regulating the miR-128-3p/NFAT5 axis. *Ageing* **14** (10), 4486–4499 (2022).
61. Li, Y. et al. A DNA tetrahedron-based ferroptosis-suppressing nanoparticle: superior delivery of curcumin and alleviation of diabetic osteoporosis. *Bone Res.* **12** (1), 14 (2024).
62. Wang, Y. et al. Repurpose dasatinib and quercetin: targeting senescent cells ameliorates postmenopausal osteoporosis and rejuvenates bone regeneration. *Bioactive Mater.* **25**, 13–28 (2023).

63. Yu, R. H. et al. Apolipoprotein D alleviates glucocorticoid-induced osteogenesis suppression in bone marrow mesenchymal stem cells via the PI3K/Akt pathway. *J. Orthop. Surg. Res.* **15** (1), 307 (2020).
64. Liu, T. et al. Chordin-Like 1 improves osteogenesis of bone marrow mesenchymal stem cells through enhancing BMP4-SMAD pathway. *Front. Endocrinol.* **10**, 360 (2019).
65. Chen, M. et al. Fgf9 regulates bone marrow mesenchymal stem cell fate and bone-fat balance in osteoporosis by PI3K/AKT/Hippo and MEK/ERK signaling. *Int. J. Biol. Sci.* **20** (9), 3461–3479 (2024).
66. Kim, H. N. et al. Estrogens decrease osteoclast number by attenuating mitochondria oxidative phosphorylation and ATP production in early osteoclast precursors. *Sci. Rep.* **10** (1), 11933 (2020).

Acknowledgements

no.

Author contributions

L.Y. conducted the bioinformatic analyses and wrote the manuscript. L.Y., K.B. and Y.C. designed and revised the manuscript. Y.H. and S.L. did statistical analysis. W.L. and Z.Y. gave useful suggestions and polished the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the postdoctoral program of heilongjiang province [No. LBH-Z22230] and the doctoral foundation program of the first affiliated hospital of harbin medical university [No. 2023B14].

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-84926-8>.

Correspondence and requests for materials should be addressed to L.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025