■ BONE BIOLOGY

Development and validation of a gene signature predicting the risk of postmenopausal osteoporosis

W. Yuan,
M. Yang,
Y. Zhu

From The First Hospital of China Medical University, Shenyang, China

Aims

We aimed to develop a gene signature that predicts the occurrence of postmenopausal osteoporosis (PMOP) by studying its genetic mechanism.

Methods

Five datasets were obtained from the Gene Expression Omnibus database. Unsupervised consensus cluster analysis was used to determine new PMOP subtypes. To determine the central genes and the core modules related to PMOP, the weighted gene co-expression network analysis (WCGNA) was applied. Gene Ontology enrichment analysis was used to explore the biological processes underlying key genes. Logistic regression univariate analysis was used to screen for statistically significant variables. Two algorithms were used to select important PMOP-related genes. A logistic regression model was used to construct the PMOP-related gene profile. The receiver operating characteristic area under the curve, Harrell's concordance index, a calibration chart, and decision curve analysis were used to characterize PMOP-related genes. Then, quantitative real-time polymerase chain reaction (qRT-PCR) was used to verify the expression of the PMOP-related genes in the gene signature.

Results

We identified three PMOP-related subtypes and four core modules. The muscle system process, muscle contraction, and actin filament-based movement were more active in the hub genes. We obtained five feature genes related to PMOP. Our analysis verified that the gene signature had good predictive power and applicability. The outcomes of the GSE56815 cohort were found to be consistent with the results of the earlier studies. qRT-PCR results showed that *RAB2A* and *FYCO1* were amplified in clinical samples.

Conclusion

The PMOP-related gene signature we developed and verified can accurately predict the risk of PMOP in patients. These results can elucidate the molecular mechanism of *RAB2A* and *FYCO1* underlying PMOP, and yield new and improved treatment strategies, ultimately helping PMOP monitoring.

Article focus

- A new subtype of postmenopausal osteoporosis (PMOP) was identified.
- The biological pathways related to hub genes were explored.
- A new PMOP-related gene signature was established.

Key messages

- We identified three subtypes of PMOP and obtained three related modules through weighted gene co-expression network analysis (WCGNA). The brown module is most related to PMOP and contains 137 genes.
- Muscle system processes, muscle contraction, and actin filament-based

Correspondence should be sent to Yue Zhu; email: zhuyuedr@163.com

movement may participate in PMOP through these hub genes.

- We constructed a PMOP-related gene signature based on five characteristic genes. This gene signature has good prediction ability and clinical applicability.

Strengths and limitations

- The PMOP-related gene markers we identified can accurately predict the risk of PMOP in patients.
- These results help to clarify the molecular mechanism of PMOP, generate new and improved treatment strategies, and help monitor PMOP.
- However, this study could only use quantitative real-time polymerase chain reaction (qRT-PCR) for simple verification of the results. Therefore, a larger sample size and more experiments are needed to further explore the function of PMOP-related gene signature.

Introduction

Osteoporosis (OP) is a common systemic bone disease characterized by decreased bone mineral density (BMD) and bone structure destruction, as well as substantial increases in bone fragility and fracture susceptibility, whose incidences are high.[1-5] Primary and secondary are the two categories into which OP is divided. Further, idiopathic, senile, and postmenopausal osteoporosis (PMOP) are the three types of primary OP. PMOP, a common type of OP, is caused by a lack of oestrogen and continuous loss of calcium with ageing.[6,7] Postmenopausal women are the main targets of the disease. With an ageing population, the number of patients with this disease is expected to increase steadily in the near future.[8] Nearly 40% of women suffer from OP and fragility fractures during their lifetime.[8] Therefore, PMOP is an important factor threatening the health and life quality of women, as well as a critical burden on the global public health system.[9] The basic pathogenesis of PMOP is related to an imbalance between bone resorption by osteoclasts and bone deposition by osteoblasts.[9] Although there has been notable progress in the early diagnosis, treatment, and management of PMOP,[10,11] there is no gold standard treatment method. Therefore, new and improved PMOP treatment strategies need to be developed. Previous studies have proved that the genetic variation of *ARHGEF3* plays a role in BMD measurement and proposed that the Rho guanosine triphosphatase (GTPase)-Rho family guanine nucleotide exchange factor (RhoGEF) pathway is related to PMOP.[12] Jemtland et al[13] identified new candidate genes for OP and showed that the expression of the transcription factor (TF) SRY-box transcription factor 4 (*SOX4,*) and the bone matrix proteins matrix metallopeptidase 13 (*MMP13*) and matrix extracellular phosphoglycoprotein (*MEPE*) are downregulated in OP. This indicates that the occurrence and development of PMOP is a complex biological process related to a series of genomic changes and different molecular pathogenetic

mechanisms. Therefore, we thrive on developing a PMOP-related gene signature by studying its genetic mechanism and providing a new and improved PMOP treatment strategy.

With the development of bioinformatics technology, an increasing number of tools can be used to analyze gene expression profiles, study the molecular mechanisms of diseases, and identify disease biomarkers.[14] Unsupervised cluster analysis is one of the most useful data mining techniques in cancer research. Consensus-ClusterPlus extends and visualizes unsupervised cluster analysis. In this study, we used cluster analysis to identify clusters and their number in the dataset. We combined it with a consistency matrix heat map, consistency cumulative distribution function (CDF) map, and δ area map to determine the number of clusters.[15] As a newly devised system biology method, the analysis of the relationship between the various clinical features and the modules, the gene clusters through identical patterns of gene expressions and form modules, and the connectivity of the gene clusters in a comprehensive network could be described by using weighted gene co-expression network analysis (WGCNA).[16,17] For screening the characteristic genes, the Least Absolute Contraction and Selection Operator (LASSO), a regression analysis algorithm could be used. In classification and regression, there is widescale use of the Support Vector Machine (SVM), a supervised machine-learning technology. A recursive feature elimination (RFE) algorithm can select the best gene from the metadata queue to avoid overfitting. In order to identify the gene set with the highest discrimination ability, support vector machine-recursive feature elimination (SVM-RFE) was used to select the appropriate features.[18] In addition, previous studies have shown that a large number of biomarkers, such as messenger RNA (mRNA), microRNA, and circular RNA (circRNAs), can be used as potential targets for the diagnosis and treatment of cancer and other diseases.[19-22] Biomarkers can reflect the molecular mechanisms of the occurrence and development of diseases. Therefore, identifying valuable biomarkers can contribute to improving the current status of disease treatments.[23,24] A previous study has shown that OP is a polygenic disease affected by the combined action of multiple genes.[25] Therefore, it is necessary to identify reliable PMOP-related biomarkers and develop a PMOP-related gene signature.

In this study, we downloaded and integrated a PMOP dataset from the Gene Expression Omnibus (GEO) and ArrayExpress databases. Three subtypes related to PMOP were identified using unsupervised cluster analysis. Based on the differentially expressed genes (DEGs) of the three subtypes, the core modules related to PMOP were constructed using the WGCNA algorithm, and the hub genes of PMOP were obtained. We also used the LASSO and SVM-RFE algorithm to screen PMOP-related feature genes based on these hub genes. Finally, logistic regression was used to construct PMOP-related gene characteristics according
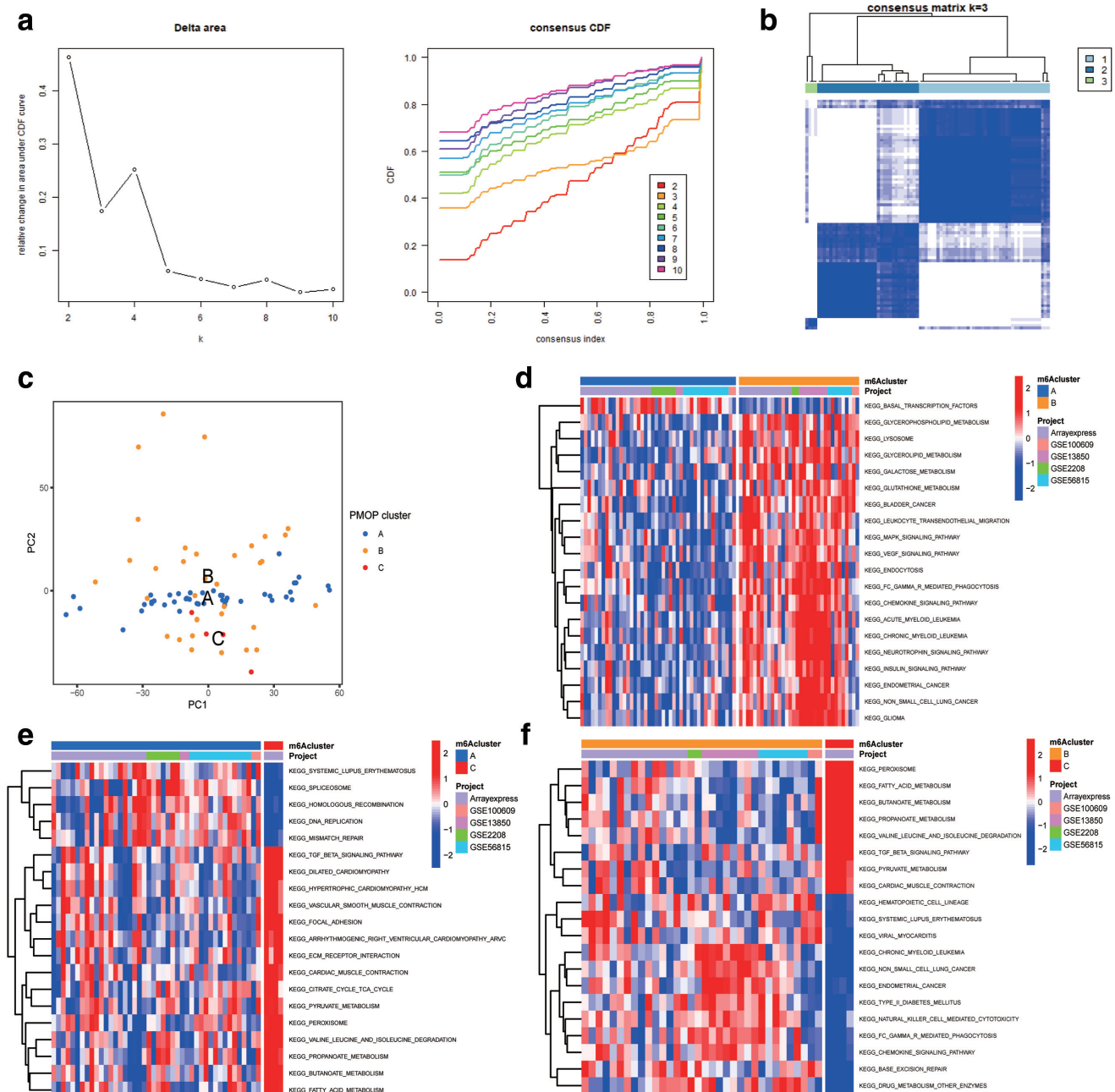
Fig. 1

Identification of postmenopausal osteoporosis (PMOP)-related subtypes. a) A consistent cumulative distribution function (CDF) graph and a δ area graph with different cluster numbers (k = 2 to 9). b) The consensus matrix heat map yielded the identification of three clusters. c) Principal component analysis (PC1 and PC2) shows evident clusters among the three PMOP subtypes. d) to f) Gene set variation analysis (GSVA) enrichment analysis shows the biological pathways between different PMOP subtypes. The red colour represents the activated pathway, and the blue colour represents the inhibited pathway. d) PMOP cluster A and PMOP cluster B; e) PMOP cluster A and PMOP cluster C; f) PMOP cluster B and PMOP cluster C.

to the characteristic genes of PMOP. Gene enrichment analysis was used to investigate the potential biological pathways of central genes. Quantitative real-time polymerase chain reaction (qRT-PCR) was used to verify the genes used to build the gene signature. The results of this study will elucidate the molecular mechanism of PMOP, providing new insights for improved PMOP diagnosis and treatment. We hypothesized that we

could develop genetic characteristics that predict the occurrence of PMOP.

Methods

Data download and processing. We downloaded five datasets (GSE2208, GSE13580, GSE56815, GSE100609, and E-MEXP-1618) from the GEO[26] and ArrayExpress[27] databases, which included 231 sample information and
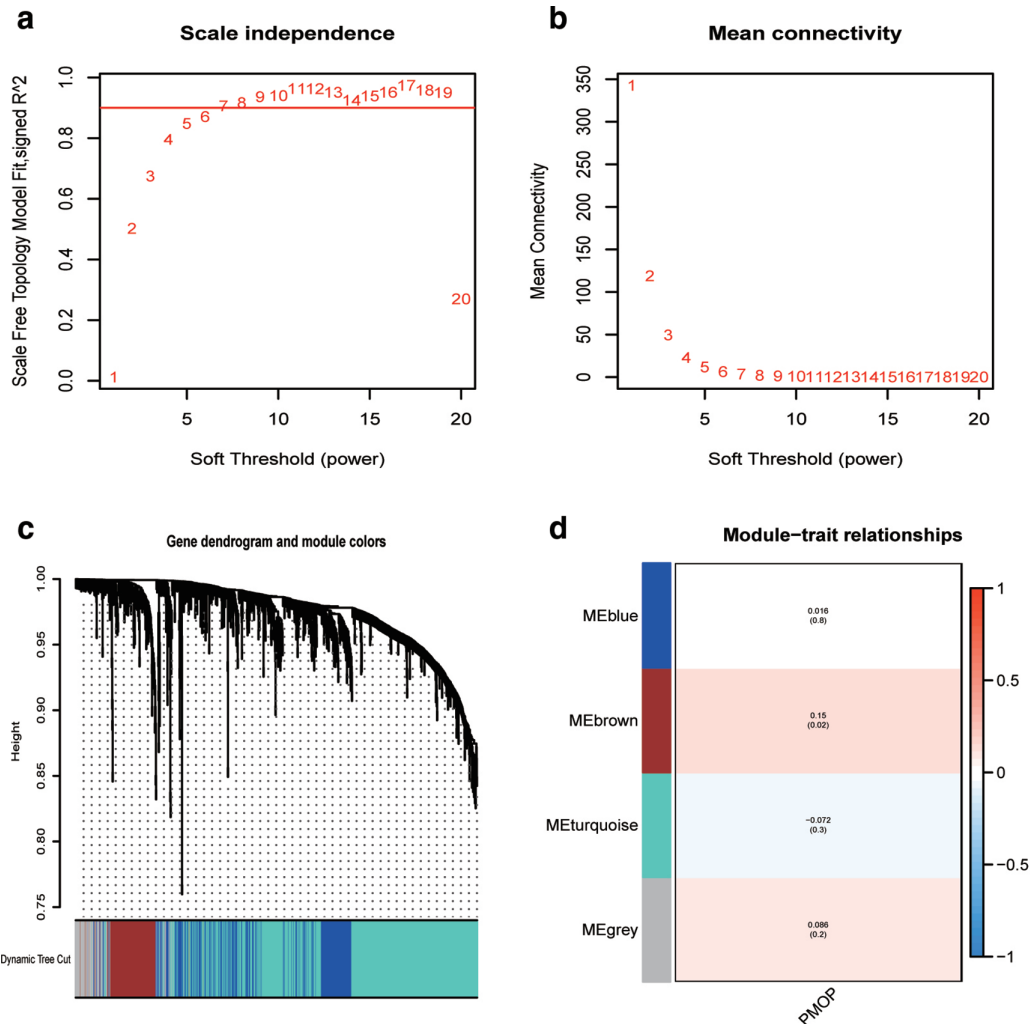
Fig. 2

Identification of the core modules related to postmenopausal osteoporosis (PMOP). a) Analysis of the scale-free index of various soft threshold powers. b) Analysis of the mean connectivity of various soft threshold powers. c) Dendrogram of all differentially expressed genes clustered based on dissimilarity measurements (1-TOM). d) Heat map of the characteristic genes of the module. We chose the blue module for subsequent analysis. ME, module eigengene; TOM, topological overlap matrix.

5,242 gene expression levels. We integrated the five datasets. For chip data, probes that match genes or multiple genes were removed. If different probes mapped to the same gene, the mean expression value was taken as the expression value of one gene. The 'sva' R package (The Comprehensive R Archive Network (CRAN), R Foundation for Statistical Computing, Austria) was used to remove the batch effect in different datasets, and the data were standardized.

Determination of different PMOP subtypes. Based on the gene expression profile of PMOP, we used unsupervised cluster analysis to identify different PMOP subtypes. The parameter was set as a Partitioning Around Medoids (PAM) algorithm based on Euclidean distance. In total 80% of samples were randomly selected in each iteration, and the number of iterations was set to 1,000, ensuring the stability of the classification. To determine the optimal value of K, the tracking graph, the δ area, the consensus CDF, and the consensus matrix were applied with

the value of K between 2 and 9. The R software package ConsensusClusterPlus (Bioconductor, R Foundation for Statistical Computing) was used for cluster analysis.

Gene set variation analysis and functional annotation between PMOP subtypes. We downloaded the gene set 'c2.cp.kegg.v7.2.symbols.gmt' from the Molecular Signatures Database (MSigDB)[28] and used the R software package 'GSVA' (Bioconductor) for gene set variation analysis (GSVA), to determine the biological process differences between PMOP subtypes. The R package 'clusterProfiler' (Bioconductor) was used to perform functional annotation. The cut-off value was a false discovery rate of < 0.05.

Correlation between different PMOP subtypes and immune cell infiltration. To examine the correlation between the immune cells and the different PMOP subtypes, a single-sample gene set enrichment analysis (GSEA) was conducted. We evaluated immune cells such as CD8 T cells, macrophages, natural killer T (NKT)
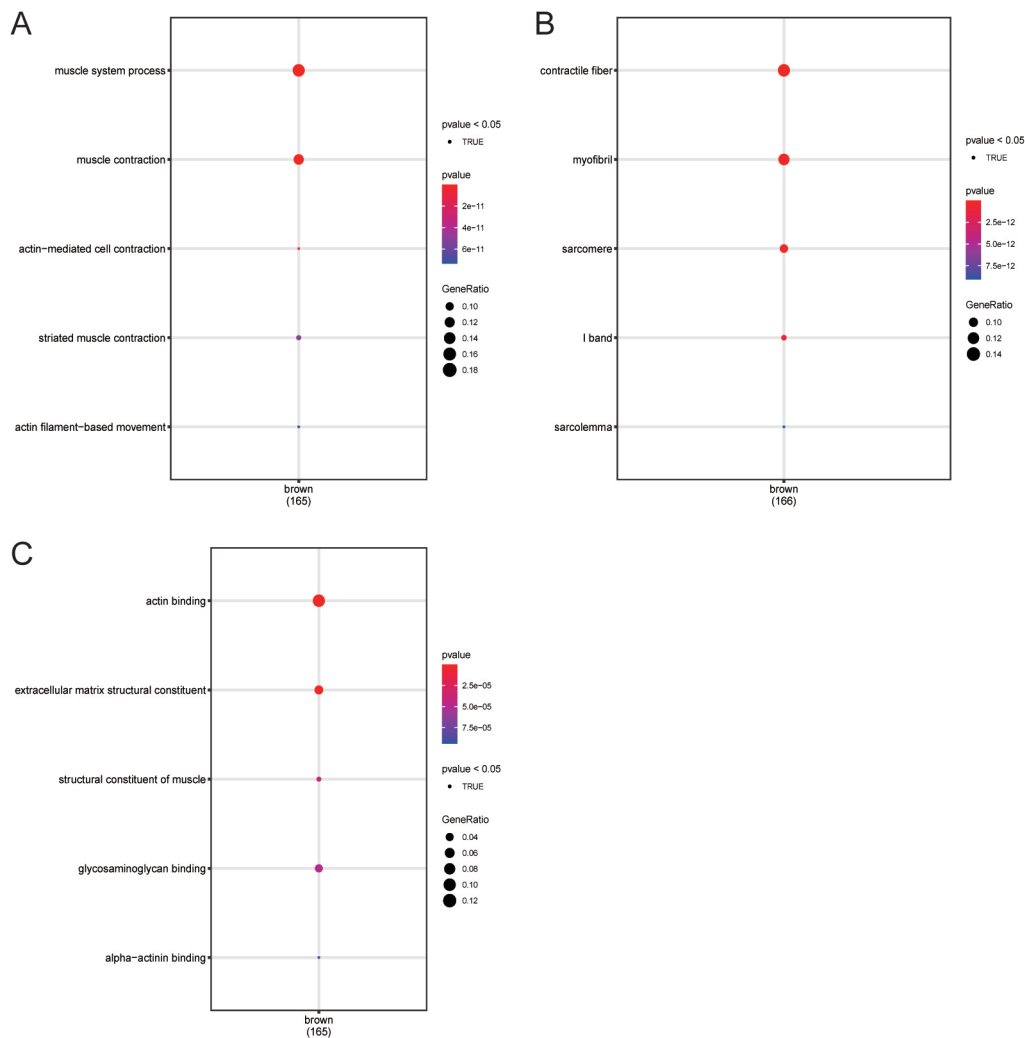
Fig. 3

Functional enrichment analysis of postmenopausal osteoporosis (PMOP) hub genes. a) Biological process analysis. b) Cell component analysis. c) Molecular function analysis.

cells, dendritic cells, and regulatory T cells. The relative abundance of each immune cell type was determined using the enrichment score determined by the single-sample GSEA.

Identification of DEGs in different PMOP subtypes. To identify the DEGs under different PMOP subtypes, the empirical Bayesian method of the 'Limma' R package (Bioconductor) was used.[29] A p-value < 0.05 was the threshold for determining DEGs.

Identification of key co-expression modules of PMOP using WGCNA. Based on the identified DEGs, we used the 'WGCNA' R package (CRAN) to construct the core module of PMOP. The reliability of the results was ensured by eliminating abnormal samples. Applying the power function, the degree of adjacency between all the differential genes was determined, while the selection of the most suitable soft threshold power was done based on the standard scale-free network. Subsequently, the

corresponding degree of dissimilarity (1-TOM) was determined after transforming the adjacency into a topological overlap matrix (TOM). For module recognition, we adopted the dynamic tree-cutting method, which uses 1-TOM as the distance metric to hierarchically cluster genes; the minimum size cut-off value of the generated dendrogram was 40. Maintaining a height cut-off value of 0.25, the highly identical modules were merged after identification by clustering. The brown module was most related to PMOP and contained 173 PMOP-related hub genes.

Gene Ontology enrichment analysis. To explore the potential biological pathways of the PMOP hub genes, we used the 'clusterProfiler' package to explore the functions of the central genes. The adjusted p < 0.05 was used as the selection criterion.

Logistic regression univariate analysis of PMOP hub genes. We used logistic regression univariate analysis
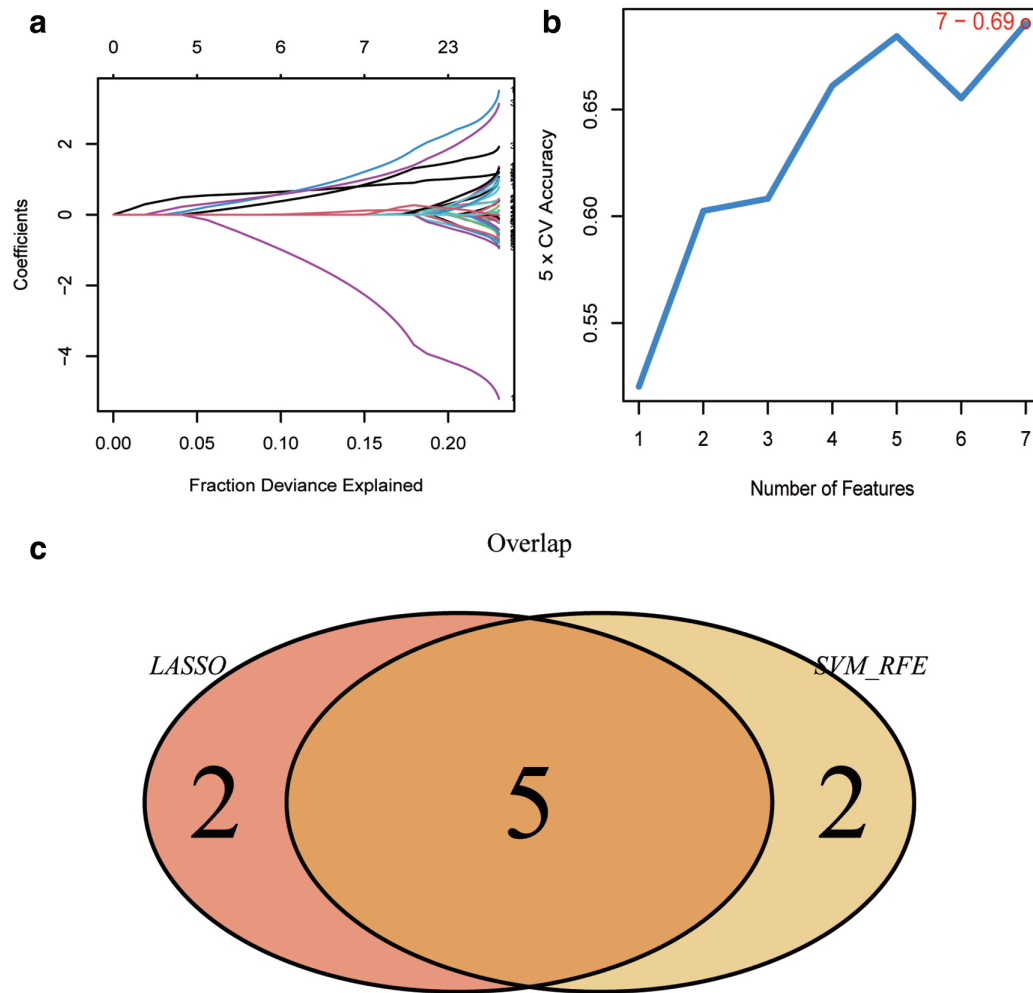
Fig. 4

The selection of postmenopausal osteoporosis (PMOP) feature genes by two algorithms: a) Least Absolute Contraction and Selection Operator (LASSO) algorithm; b) support vector machine-recursive feature elimination (SVM-RFE) algorithm; c) the intersection of characteristic genes screened by the two algorithms. CV accuracy, cross-validation accuracy.

for risk factor analysis. We used variables that showed statistical significance for the subsequent screening of PMOP characteristic genes. The threshold standard was set at p < 0.05.

Screening the characteristic genes of PMOP. We used the LASSO algorithm based on the PMOP-related genes obtained in the previous step, and adjusted the penalty parameters through ten-fold cross-validation to screen PMOP-related feature genes.[30] Simultaneously, for screening the PMOP-related feature genes, the SVM-RFE algorithms were applied.[31,32] Finally, the PMOP-related feature genes screened by the LASSO and SVM-RFE algorithms were intersected to obtain the final feature genes.

Construction and verification of the prognostic gene signature related to PMOP. The LASSO regression model does not have coefficients of zero.[33] We used the LASSO regression model to determine the variables used to construct the gene signature based on the obtained PMOP-related feature genes. To develop a prognostic gene signature

represented by a nomogram, the logistic regression analysis was applied. A calibration curve was used to evaluate the performance of the PMOP-related gene signature. The receiver operating characteristic (ROC) area under the curve (AUC) and Harrell's concordance index (C-index) were used to assess the accuracy of the PMOP-related gene signature, namely, to analyze the predictive power of gene signature for patients with different risk outcomes. The larger the values of the AUC and C-index, the higher the accuracy of the gene signature. Decision curve analysis (DCA) was drawn through the 'rmda' R package (CRAN) to evaluate the clinical utility of gene signature.[34] Additionally, the GSE56815 cohort was used to verify the gene signature.

Quantitative real-time PCR analysis. Using the total RNA extraction reagent RNAiso plus reagent (Takara, Japan), the total RNA from the tissue samples was extracted. Further, using a complementary DNA (cDNA) reverse transcription kit, the cDNA was synthesized from the RNA that was extracted. With the use of the SLAN-96S
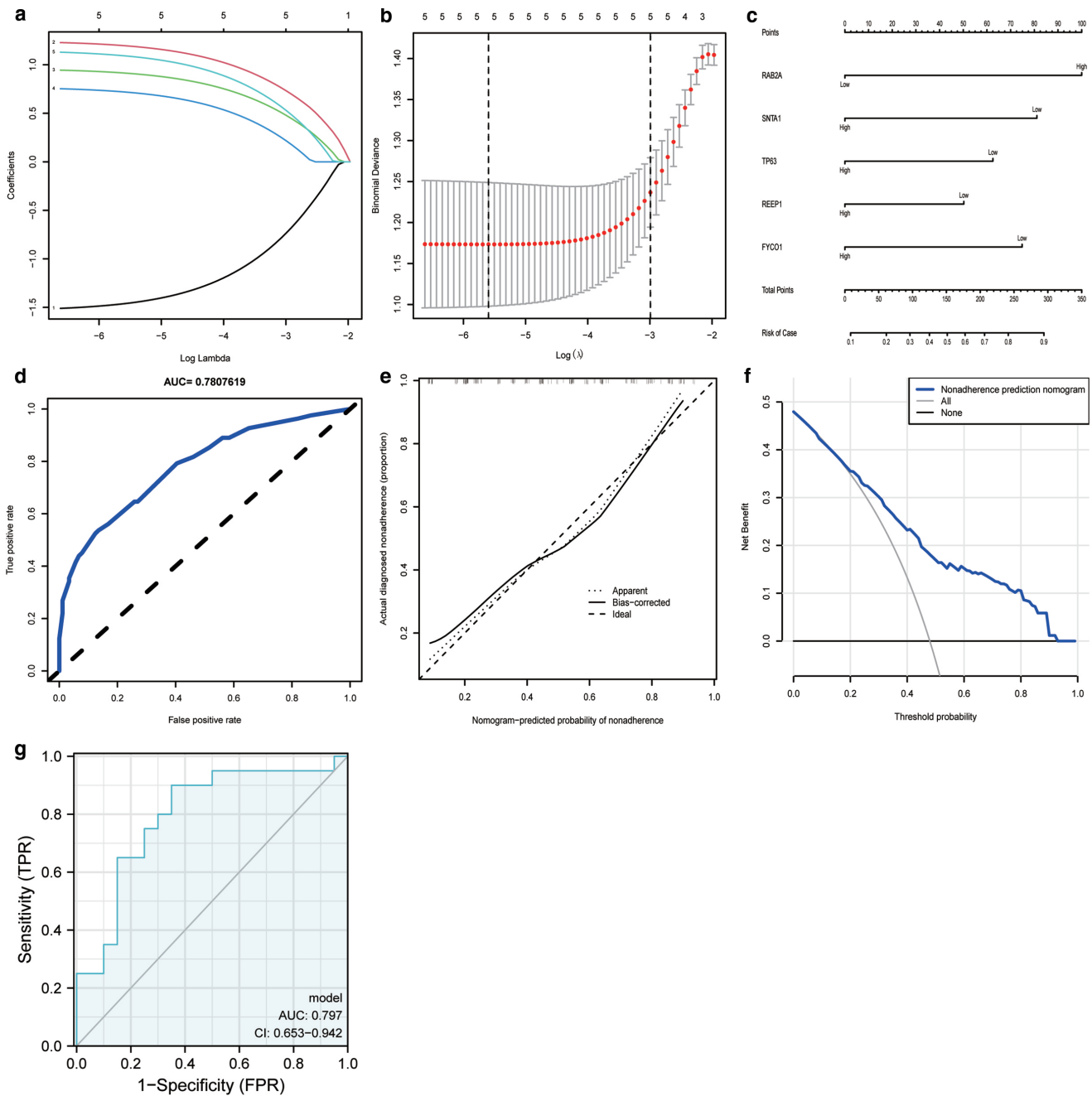
Fig. 5

Construction and verification of a postmenopausal osteoporosis (PMOP)-related gene signature. a) The best parameter (lambda) in the least absolute contraction and selection operator (LASSO) model is selected to pass the minimum criterion using five-fold cross-validation. b) Distribution map of LASSO coefficients of five features. c) The nomogram of PMOP gene characteristics. d) The receiver operating characteristic (ROC) curve is used to evaluate the prognostic value of the gene signature. e) The calibration curve of the nomogram. f) Decision curve analysis of the nomogram. g) The ROC curve of the validation group verifies that the gene signature have good prognostic value. AUC, area under the curve; CI, confidence interval; FPR, false positive rate – the proportion of the number of negative samples misjudged as positive in the actual negative samples; TPR, true positive rate – the proportion of predicted correct positive samples to all actual positive samples.

fluorescence quantizer (Shanghai Hongshi, China), the mRNA expression levels were quantified by qRT-PCR. The relative expression of the target genes was determined using the $2^{-\Delta\Delta CT}$ method. Primer sequences are shown in Supplementary Table i.

Gene set enrichment analysis in high- and low-risk groups. GSEA was used to analyze biological processes

between high- and low-risk groups. Marker gene set 'c2. cp.kegg.v7.4.symbols.gmt' was obtained from MSigDB database. Phenotypic label selection: tumour vs. PMOP. P.adjust < 0.05 was set as the threshold for statistical significance.

Statistical analysis. Spearman's rank correlation coefficient was used to evaluate the correlation coefficients
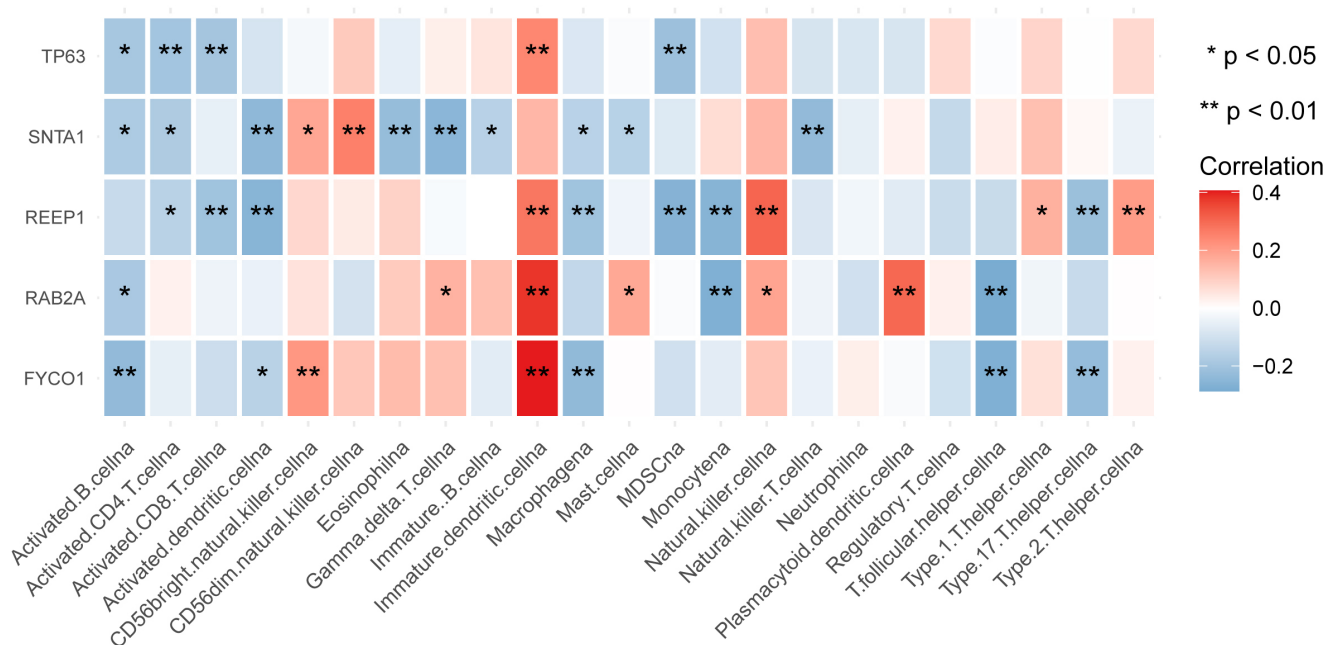
Fig. 6

Correlation analysis between genes for constructing gene signatures and immune cells. Red represents positive correlation; blue represents negative correlation. *p < 0.05; **p < 0.01.

between genes that constructed gene signatures and immune cells. The Wilcoxon signed-rank test was used to analyze and evaluate the correlation between the expression of characteristic genes in the normal group and the case group. The ROC curve was used to evaluate the predictive value of the PMOP-related feature gene, and the timeROC package was used to calculate the AUC. An AUC > 0.60 was considered a general predictive value, and an AUC > 0.70 was considered a good predictive value. The statistical analysis in this study was performed using the R-4.0.3 software (R Foundation for Statistical Computing). Statistical significance was set at p < 0.05.

Results

Determination of PMOP-related molecular subtypes. To explore new molecular subtypes of PMOP, we performed an unsupervised cluster analysis of 43 PMOP patients based on the expression profile of PMOP-related genes. The relative change in the area under the CDF curve and the consensus heat map showed that the optimal value of K was 3 (Figures 1a and 1b). Therefore, we identified three PMOP-related analysis subtypes (PMOP cluster A, PMOP cluster B, and PMOP cluster C). The principal component analysis showed a clear separation between the molecular subtypes of PMOP (Figure 1c). Moreover, we performed GSVA enrichment analysis to study the biological processes between different PMOP molecular subtypes. As shown in Figures 1d to 1f, PMOP cluster A positively correlated with basic TFs and DNA replication. PMOP cluster B positively correlated with the vascular endothelial growth factor (VEGF), mitogen-activated protein kinase (MAPK), and chemokine signalling pathways. Fatty acid metabolism, transforming growth factor-β

(TGF-β) signalling pathway, and extracellular matrix (ECM) receptor interaction were more active in PMOP cluster C.

Construction of a PMOP-weighted gene co-expression module. In this study, we identified 1,434 PMOP-related and specifically upregulated genes using the 'limma' package to study the potential genetic changes within PMOP subtypes. We constructed a co-expression network of PMOP genes based on 1,434 PMOP-specific upregulated genes using the 'WGCNA' package, with a soft threshold of 7 (Figures 2a and 2b). The mean hierarchical clustering and dynamic tree pruning identified a total of four modules (Figure 2c). The correlation between PMOP and each module was evaluated using the heat map of the module-feature relationship. The brown module had the highest correlation with PMOP (r = 0.15, p = 0.020, Spearman's rank correlation coefficient), including 173 PMOP hub genes (Figure 2d). Therefore, the brown module was selected for subsequent analyses.
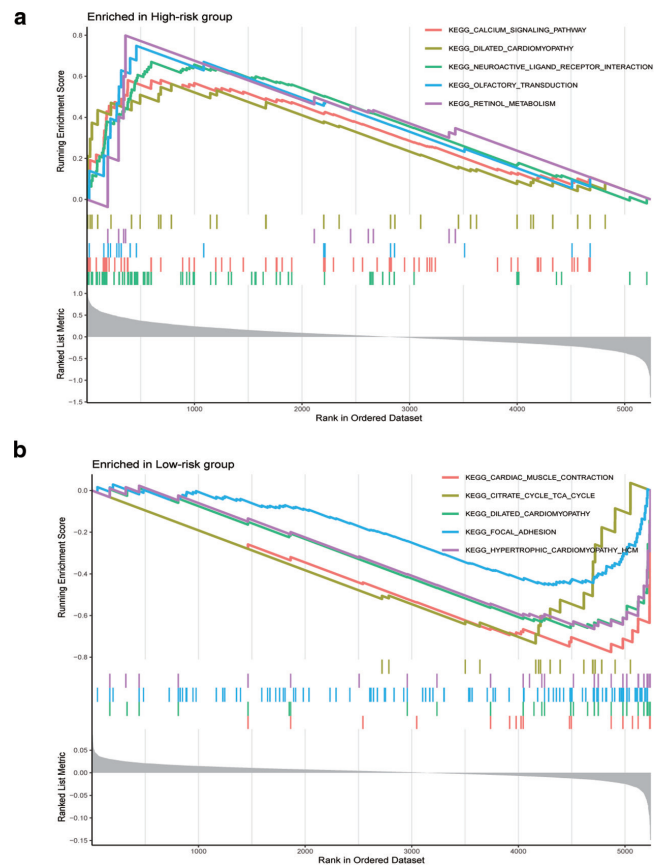
Functional enrichment analysis of PMOP hub genes. To further understand the potential functional mechanism of the 137 PMOP hub genes, we used the ClusterProfiler package to perform gene enrichment analysis. The results of the enrichment analysis showed that muscle system process, muscle contraction, and actin filament-based movement were more active in the biological gene process. In the cell component, these PMOP hub genes were mainly involved in myofibrils, contractile fibres, and sarcomeres. The molecular functions of 137 hub genes were mainly enriched in pathways such as ECM structural constituent and actin-binding (Figures 3a to 3c).

a

b

Fig. 7
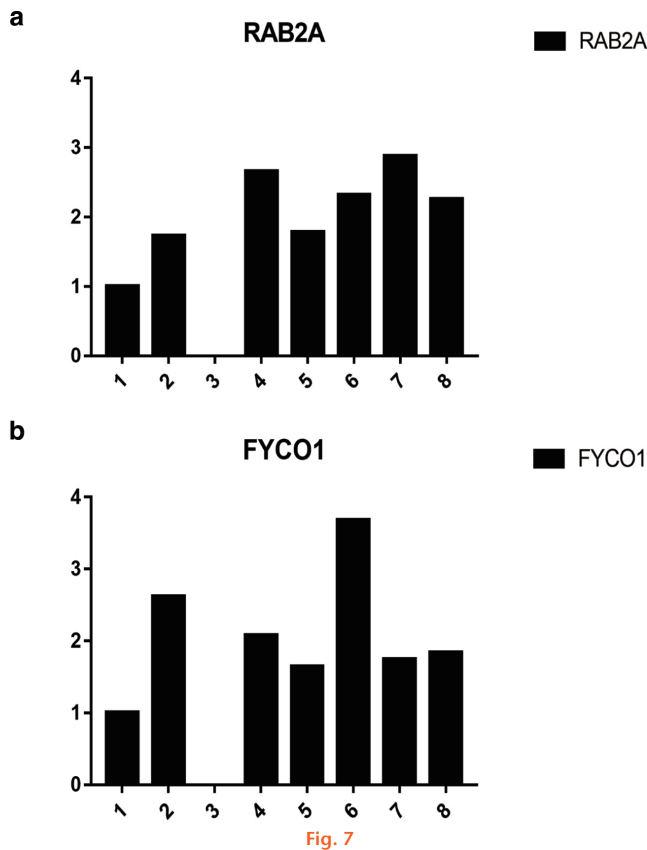
The gene expression of the gene signature in clinical samples. a) Member RAS oncogene family (*RAB2A*) was amplified in seven samples, except for sample 3; b) FYVE and coiled-coil domain autophagy adaptor 1 (*FYCO1*) was amplified in seven samples, except for sample 3.

a

b

Fig. 8

Gene set enrichment analysis (GSEA) in high- and low-risk group samples. a) Samples of the high-risk group were enriched in the calcium signalling pathway, differentiated cardiology pathway, neuroactive live receiver interaction, olfactory production, retinol metabolism, and other pathways. b) Samples in the low-risk group were enriched in myocardial contraction, citrate cycle, dilated cardiomyopathy, adhesive spot, hypertrophic cardiomyopathy, and other pathways.

Screening of PMOP-related characteristic genes. We included 137 hub genes in the logistic regression univariate analysis and identified 39 variables with statistical significance. We then used two algorithms (LASSO and SVM-RFE) to screen PMOP-related characteristic genes based on 39 PMOP-related genes. First, according to PMOP-related genes, we used the LASSO algorithm to determine seven characteristic genes. We then screened a set of characteristic genes based on 39 genes by executing the SVM-RFE algorithm. Finally, we took the intersection of the PMOP feature genes determined by LASSO and the feature genes screened by the SVM-RFE algorithm and identified five feature genes, namely member RAS oncogene family (*RAB2A*), Syntrophin Alpha 1 (*SNTA1*), tumour protein p63 (*TP63*), receptor accessory protein 1 (*REEP1*), and FYVE and coiled-coil domain autophagy adaptor 1 (*FYCO1*) (Figures 4a to 4c).

Construction of a PMOP-related gene signature. Based on the identified PMOP-related signature genes, we used the LASSO regression model to find the variables used to construct the gene signatures, and employed ten-fold cross-validation to determine the best adjustment parameters related to the minimum generalization error. The results identified five variables: *RAB2A*, *SNTA1*, *TP63*, *REEP1*, and *FYCO1* (Figures 5a and 5b). Finally, we used a logistic

regression model to construct a PMOP-related gene signature to calculate the risk score, thereby assessing the risk of patients with PMOP and developing a nomogram to present it (Figure 5c). The equation used to calculate the risk score was the following: *risk score = -0.7899+(-1.5380) × (expression of RAB2A) + (1.2474) × (expression of SNTA1) + (0.9617) × (expression of TP63) + (0.7727) × (expression of REEP1) + (1.1511) × (expression of FYCO1)*.

Verification of the PMOP-related gene signature. The results of the ROC curve and C-index showed that the PMOP-related gene signature was highly accurate, showing that the risk score can accurately assess the risk of PMOP in patients. The AUC value was 0.781, and the C-index value was 0.7807 (Figure 5d). The calibration curve of this gene signature shows that the risk of PMOP predicted by the risk score is in good agreement with the actual risk of PMOP (Figure 5e). The DCA analysis results showed that, in our study, this gene signature was beneficial to patients within the threshold range of 9% to 92% (Figure 5f). In addition, we verified the accuracy of the gene signature in the GSE56815 cohort.

The ROC curve in Figure 5g shows that this signature had good predictive power (AUC = 0.797).

The correlation between genes and immune cells to construct gene signature. In this study, we also used Spearman's rank correlation coefficient to examine the correlation between genes that construct gene signatures and immune cells. As shown in Figure 6, *TP63* showed a significantly negative correlation with B cells (p = 0.011), CD4 T cells (p = 0.009), CD8 T cells (p = 0.009), and myeloid-derived suppressor cells (MDSCs, p = 0.005), and a significantly positive correlation with dendritic cells (p = 0.001). *SNTA1* significantly negatively correlated with dendritic cells (p = 0.001), eosinophils (p = 0.003), gamma T cells (p = 0.001), and NKT cells (p = 0.002). *REEP1* showed a significantly negative correlation with CD8 T cells (p = 0.007), dendritic cells (p = 0.001), macrophages (p = 0.007), MDSCs (p = 0.001), and monocytes (p = 0.001), and a significantly positive correlation with immature dendritic cells (p < 0.001) and natural killer cells (p < 0.001). *RAB2A* showed a significantly negative correlation with monocytes (p < 0.001) and follicle-assisted cells (p < 0.001), and significantly positively correlated with immature dendritic cells (p < 0.001), mast cells (p = 0.019), natural killer cells (p = 0.014), and plasma-like dendritic cells (p < 0.001). *FYCO1* significantly inversely correlated with B cells (p = 0.002), macrophages (p = 0.002), and follicle-assisted cells (p < 0.001), and significantly positively correlated with immature dendritic cells (p < 0.001).

Verification of the gene expression of the gene signature in clinical samples. The level of expression of the genes in the gene signature of the clinical samples was analyzed. The results showed that *RAB2A* and *FYCO1* could be amplified except for sample 3 (Figures 7a and 7b; Supplementary Table ii; Supplementary Figure a). No amplification was detected in *SNTA1*, *TP63*, and *REEP1*, which may be related to the small sample size of the cohort.

Biological processes associated with risk scores. GSEA analysis further revealed the enrichment pathway between the high- and low-risk groups. The samples of the high-risk group were mainly enriched in the calcium signalling pathway, neuroactive ligand-receptor interaction, and other pathways (Figure 8a). The samples in the low-risk group were mainly enriched in myocardial contraction, dilated cardiomyopathy, focal adhesion, and other pathways (Figure 8b).

Discussion

OP, the most common bone disease globally, negatively impacts more than 200 million people.[35,36] Postmenopausal women are at a high risk of this disease. OP occurrence increases the risk of fracture. Annually, approximately 8.9 million fractures are caused by OP worldwide, posing a substantial burden on patients and society.[37] A combination of genetic and environmental factors leads to a complex disease known as PMOP.[38,39] In a recent study, the genetic polymorphism

of calcium-sensing receptor (CaSR) was reportedly associated with the prevalence of OP in elderly men.[40] Some genetic polymorphisms of the collagen type I alpha1 (COL1A1) gene are related to PMOP.[41,42] Zhu et al[43] found 50 new OP-related genes through a genome-wide association and transcriptome prediction model. Therefore, mining databases related to OP and identifying more characteristic genes can hint at new ideas for preventing and treating OP. In this study, we aimed to identify PMOP-related characteristic genes using unsupervised consistent cluster analysis, WGCNA analysis, and the LASSO and SVM-RFE algorithms, and then constructed a PMOP-related gene signature using a logistic regression model based on the characteristic genes to predict the risk of PMOP. This study aims to provide clinicians with a tool to predict the risk of PMOP in patients. The internal verification of the research cohort suggests that the PMOP gene signature has good predictive ability and applicability, ultimately providing a new way to improve the monitoring and treatment of PMOP.

In this study, we identified three different PMOP subtypes. PMOP cluster A was significantly related to basic TFs and DNA replication. The chemokine signalling pathways, MAPK signalling pathway, and VEGF signalling pathway were found to be significantly related to the PMOP cluster B. Pathways such as fatty acid metabolism, TGF-β signalling pathway, and ECM receptor interaction were more active in PMOP cluster C. In addition, we identified 1,146 PMOP-related and specifically upregulated genes from different PMOP subtypes, and obtained 137 PMOP hub genes through WGCNA analysis based on these genes. The Gene Ontology (GO) enrichment analysis results showed that 137 PMOP hub genes were significantly related to biological pathways such as muscle system process, muscle contraction, and actin filament-based movement. The unbalanced activity of osteoblasts and osteoclasts is the main reason for the development of PMOP. Myopenia is a decrease in muscle mass and function, and is considered one of the signs of the ageing process. The current view is that sarcopenia results from a variety of medical, behavioural, and environmental factors in the elderly. Similarly, bone fragility is known to depend on several pathogenic mechanisms, leading to reduced bone mass and bone strength. Muscle weakness, fear of falls, actual falls, and subsequent fractures are associated with concurrent myopenia and OP, resulting in limited activity, loss of autonomy, and reduced life expectancy. The bone and muscle organ systems are closely intertwined. The strongest mechanical force applied to the bone is the force generated by muscle contraction. These forces regulate bone density, strength, and microstructure. Therefore, decreased muscle strength leads to decreased bone strength, leading to OP.[44]

In this study, we used the LASSO and SVM-RFE algorithms to determine five PMOP-related feature genes based on PMOP hub genes. Later, to identify the five

characteristic genes (*RAB2A*, *SNTA1*, *TP63*, *REEP1*, and *FYCO1*) for developing a gene signature, the LASSO regression model was applied. The results of the ROC curve and C-index showed that our gene signature had a good predictive ability. The calibration curve showed promising calibration of genetic characteristics. The DCA results demonstrated the clinical applicability of the PMOP-related gene signature. The verification results of the GSE56815 cohort were consistent with those of our findings in this study. More importantly, qRT-PCR results showed that *RAB2A* and *FYCO1* were amplified in clinical samples. At present, there is still no relevant research to explore the mechanism between *RAB2A* and *PMOP*. However, Wang et al[45] found that *RAB2A* was associated with osteoarthritis. Gut-bone axis is a recently proposed concept that represents the impact of gut microbiota and their metabolites on bone health. Previous studies have shown that there is an association between gut microbiota and BMD. Gut microbiota play an important role in the development of OP.[46–48] Xiang et al[49] found that the nucleoside binding oligomerization domain 1 (Nod1) in the intestinal microbiota is located in dense core granules, and *RAB2A* is recruited on dense core granules, which affects the retention of chromogranin A in dense core granules, thus affecting the secretion of epinephrine under optimized stress. Therefore, we believe that *RAB2A* may also be a pathway regulated by similar intestinal microbiota, affecting bone loss and thus affecting PMOP. The mechanism of this interaction needs more research in the future. Transcription factor *p63* is a member of the transcription factor p53 family and is necessary for proper bone formation during development.[50] *P63* can produce two groups of *p63* isoforms: the tap63 isoform and the truncated ΔNp63 isoform.[51] Previous cumulative studies have shown that *p63* participates in endochondral bone formation by regulating cartilage formation, affecting endochondral ossification and bone formation.[50,52–54] Mesenchymal stem cells (MSCs) are immature adult stem cells that reside in a special bone niche. They produce osteoblasts, chondrocytes, and adipocytes during development and throughout mammalian adulthood. Immature human MSCs (hMSCs) can differentiate into the matrix-producing bone to form osteoblasts. Curtis et al[55] proved that the *p63* gene product is important for the osteoblast differentiation of hMSCs. Furthermore, the mechanism of action of the characteristic genes *SNTA1*, *REEP1*, and *FYCO1* in PMOP has not been reported. Therefore, we need to conduct additional experiments to study their potential mechanisms in PMOP.

In this study, we also analyzed the correlation between genes that construct gene signature and immune cells. The role of the immune system in various bone diseases, including OP, has been fully confirmed.[56] Bone health gets impacted by the immune cells affecting key factors or functional components of the bone mass regulators.[57] Our research results showed that the genes that build gene signature are significantly related to immune cells, such as CD4 T cells, macrophages, monocytes, natural killer cells, dendritic cells, and T follicular helper lymph cells. NKT cells are mainly involved in eliminating transformed cells and virus-infected cells.[58] NKT regulates the initiation and development of immune responses mediated by T and B cells by producing various growth factors and cytokines.[59] The differentiation and development of osteoclasts have been found to be regulated by the invariant NKT cells, as revealed from the earlier studies. In addition, NKT cells can also produce receptor activator of nuclear factor-kappa-B (RANK) ligand and macrophage colony-stimulating factor to induce osteoclast formation, thereby affecting OP development.[60,61] Macrophages have been shown to be important sources of local and systemic pro-inflammatory cytokines.[62,63] Increased pro-inflammatory cytokines can increase the differentiation and activity of osteoclasts by upregulating the RANK ligand, thereby posing a threat to bone health.[64] Our current understanding of the relationship between the immune system and OP is still very superficial, and more research is required. We believe that targeting the immune system may be an effective strategy for treating PMOP.

The GSEA analysis results also suggest that the samples in the high-risk group are mainly enriched in the calcium signalling pathway, neuroactive ligand-receptor interaction, and other pathways. In contrast, samples in the low-risk group were mainly enriched in myocardial contraction, dilated cardiomyopathy, focal adhesion, and other pathways. Andreev et al[65] found that macrophage inducible C-type lectin (Mincle) was activated by the damage-related molecular pattern released by dying bone cells, and enhanced osteoclastogenesis through immunoreceptor tyrosine-based activation motif (ITAM)-based calcium signalling and oxidative phosphorylation induction. A study by Cao et al[66] showed that focal adhesion protein Kindlin-2 plays a pivotal role in the early development of cartilage adhesion plaques. Kindlin-2 deletion promotes osteocyte apoptosis, impairing osteocyte diffusion.

Nonetheless, this study had some limitations. First, although we performed global standardization on the integrated data, there may still be heterogeneity between different datasets. Second, the number of clinical samples included in this study was limited. Finally, only *RA2AB* and *FYCO1* were amplified in qRT-PCR, while *SNTA1*, *REEP1*, and *TP63* were not amplified. This may be the problem of sample quality or the problem of experimental design. Therefore, in future research, we shall use a larger sample size and perform additional experiments to further explore the function of PMOP-related gene signature.

In conclusion, in this study we identified five characteristic genes related to PMOP and constructed a PMOP-related gene signature based on these genes. These analyses can provide new insights into the molecular

mechanisms of PMOP. More importantly, our research results show *FYCO1* and *RAB2A* may facilitate the formation of new and improved treatment strategies for PMOP and help monitor the disease better.

Supplementary material

ë Tables showing primer sequences in quantitative real-time polymerase chain reaction (qRT-PCR) and expression levels of RAB2A and FYCO1 in clinical samples. Figure showing qRT-PCR analysis.

References

1. **Kanis JA**, **McCloskey EV**, **Johansson H**, **Oden A**, **Melton LJ**, **Khaltaev N**. A reference standard for the description of osteoporosis. *Bone*. 2008;42(3):467–475.

2. **Zhang Y**, **Liu H**, **Zhang C**, **et al**. Endochondral ossification pathway genes and postmenopausal osteoporosis: Association and specific allele related serum bone sialoprotein levels in Han Chinese. *Sci Rep*. 2015;5:16783.

3. **Ma M**, **Chen X**, **Lu L**, **et al**. Identification of crucial genes related to postmenopausal osteoporosis using gene expression profiling. *Aging Clin Exp Res*. 2016;28(6):1067–1074.

4. **Li S**, **Mao Y**, **Zhou F**, **Yang H**, **Shi Q**, **Meng B**. Gut microbiome and osteoporosis: a review. *Bone Joint Res*. 2020;9(8):524–530.

5. **Guebeli A**, **Platz EA**, **Paller CJ**, **McGlynn KA**, **Rohrmann S**. Relationship of sex steroid hormones with bone mineral density of the lumbar spine in adult men. *Bone Joint Res*. 2020;9(3):139–145.

6. **Meng J**, **Zhang D**, **Pan N**, **et al**. Identification of miR-194-5p as a potential biomarker for postmenopausal osteoporosis. *PeerJ*. 2015;3:e971.

7. **Chow S-H**, **Chim Y-N**, **Wang J-Y**, **Wong R-Y**, **Choy V-H**, **Cheung W-H**. Inflammatory response in postmenopausal osteoporotic fracture healing. *Bone Joint Res*. 2020;9(7):368–385.

8. **McClung MR**, **O'Donoghue ML**, **Papapoulos SE**, **et al**. Odanacatib for the treatment of postmenopausal osteoporosis: results of the LOFT multicentre, randomised, double-blind, placebo-controlled trial and LOFT Extension study. *Lancet Diabetes Endocrinol*. 2019;7(12):899–911.

9. **Paschalis EP**, **Gamsjaeger S**, **Hassler N**, **et al**. Vitamin D and calcium supplementation for three years in postmenopausal osteoporosis significantly alters bone mineral and organic matrix quality. *Bone*. 2017;95:41–46.

10. **Bandeira L**, **Bilezikian JP**. Novel therapies for postmenopausal osteoporosis. *Endocrinol Metab Clin North Am*. 2017;46(1):207–219.

11. **Li Q**, **Tang T**, **Zhang P**, **et al**. Correlation of IL-31 gene polymorphisms with susceptibility and clinical recurrence of bladder cancer. *Fam Cancer*. 2018;17(4):577–585.

12. **Mullin BH**, **Prince RL**, **Dick IM**, **et al**. Identification of a role for the ARHGEF3 gene in postmenopausal osteoporosis. *Am J Hum Genet*. 2008;82(6):1262–1269.

13. **Jemtland R**, **Holden M**, **Reppe S**, **et al**. Molecular disease map of bone characterizing the postmenopausal osteoporosis phenotype. *J Bone Miner Res*. 2011;26(8):1793–1801.

14. **Akalin PK**. Introduction to bioinformatics. *Mol Nutr Food Res*. 2006;50(7):610–619.

15. **Wilkerson MD**, **Hayes DN**. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–1573.

16. **Langfelder P**, **Horvath S**. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

17. **Zhang B**, **Horvath S**. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4(1):Article17.

18. **Guyon I**, **Weston J**, **Barnhill S**, **Vapnik V**, **Chowdhury NA**. Gene Selection for Cancer Classification using Support Vector Machines. Seminar on Data mining and data integration in Bioinformatics. Barnhill Bioinformatics, Savannah, Georgia, USA: AT&T Labs, Atlanta, Georgia, USA; 2022. https://classes.cs.uoregon.edu/10S/cis607dmb/Nafisa.pdf (date last accessed 21 June 2022).

19. **Lee SC**, **Tan HT**, **Chung MCM**. Prognostic biomarkers for prediction of recurrence of hepatocellular carcinoma: current status and future prospects. *World J Gastroenterol*. 2014;20(12):3112–3124.

20. **Brzeszczyńska J**, **Brzeszczyński F**, **Hamilton DF**, **McGregor R**, **Simpson AHRW**. Role of microRNA in muscle regeneration and diseases related to muscle dysfunction in atrophy, cachexia, osteoporosis, and osteoarthritis. *Bone Joint Res*. 2020;9(11):798–807.

21. **Wong RMY**, **Choy VMH**, **Li J**, **et al**. Fibrinolysis as a target to enhance osteoporotic fracture healing by vibration therapy in a metaphyseal fracture model. *Bone Joint Res*. 2021;10(1):41–50.

22. **Xin W**, **Yuan S**, **Wang B**, **Qian Q**, **Chen Y**. Hsa_circ_0066523 promotes the proliferation and osteogenic differentiation of bone mesenchymal stem cells by repressing PTEN. *Bone Joint Res*. 2021;10(8):526–535.

23. **Chan AWH**, **Zhong J**, **Berhane S**, **et al**. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J Hepatol*. 2018;69(6):1284–1293.

24. **He W**, **Peng B**, **Tang Y**, **et al**. Nomogram to predict survival of patients with recurrence of hepatocellular carcinoma after surgery. *Clin Gastroenterol Hepatol*. 2018;16(5):756–764.

25. **Zhang D**, **Ge Z**, **Ma X**, **et al**. Genetic association study identified a 20 kb regulatory element in WLS associated with osteoporosis and bone mineral density in Han Chinese. *Sci Rep*. 2017;7(1):13668.

26. **No authors listed**. Gene Expression Omnibus (GEO). National Center for Biotechnology Information (NCBI). 2022. https://www.ncbi.nlm.nih.gov/geo/ (date last accessed 17 June 2022).

27. **No authors listed**. ArrayExpress. European Bioinformatics Institute (EMBL-EBI). 2022. https://www.ebi.ac.uk/arrayexpress/ (date last accessed 17 June 2022).

28. **No authors listed**. Molecular Signatures Database v7.5.1. Gene Set Enrichment Analysis (GSEA). 2022. https://www.gsea-msigdb.org/gsea/msigdb/ (date last accessed 21 June 2022).

29. **Yu G**, **Wang L-G**, **Han Y**, **He Q-Y**. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.

30. **Tibshirani R**. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267–288.

31. **Huang M-L**, **Hung Y-H**, **Lee WM**, **Li RK**, **Jiang B-R**. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *ScientificWorldJournal*. 2014;2014:795624.

32. **Guo P**, **Luo Y**, **Mai G**, **et al**. Gene expression profile based classification models of psoriasis. *Genomics*. 2014;103(1):48–55.

33. **Kidd AC**, **McGettrick M**, **Tsim S**, **Halligan DL**, **Bylesjo M**, **Blyth KG**. Survival prediction in mesothelioma using a scalable Lasso regression model: instructions for use and initial performance using clinical predictors. *BMJ Open Respir Res*. 2018;5(1):e000240.

34. **Vickers AJ**, **Cronin AM**, **Elkin EB**, **Gonen M**. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.

35. **Srivastava M**, **Deal C**. Osteoporosis in elderly: prevention and treatment. *Clin Geriatr Med*. 2002;18(3):529–555.

36. **Langdahl BL**. Overview of treatment approaches to osteoporosis. *Br J Pharmacol*. 2021;178(9):1891–1906.

37. **Burge R**, **Dawson-Hughes B**, **Solomon DH**, **Wong JB**, **King A**, **Tosteson A**. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res*. 2007;22(3):465–475.

38. **Karasik D**, **Cohen-Zinder M**. Osteoporosis genetics: year 2011 in review. *Bonekey Rep*. 2012;1:114.

39. **Farber CR**, **Lusis AJ**. Future of osteoporosis genetics: enhancing genome-wide association studies. *J Bone Miner Res*. 2009;24(12):1937–1942.

40. **Di Nisio A**, **Rocca MS**, **Ghezzi M**, **et al**. Calcium-sensing receptor polymorphisms increase the risk of osteoporosis in ageing males. *Endocrine*. 2018;61(2):349–352.

41. **Aitkulova A**, **Akilzhanova A**, **Abilova Z**, **Zhumatova Z**, **Akilzhanova G**, **Zholdybayeva E**. Collagen type I alpha1 (COL1A1) gene polymorphism and bone mineral density in postmenopausal Kazakh women. *Cent Asian J Glob Health*. 2014;3(Suppl):144.

42. **Rojano-Mejía D**, **Coral-Vázquez RM**, **Espinosa LC**, **et al**. JAG1 and COL1A1 polymorphisms and haplotypes in relation to bone mineral density variations in postmenopausal Mexican-Mestizo Women. *AGE (Dordr)*. 2013;35(2):471–478.

43. **Zhu M**, **Yin P**, **Hu F**, **et al**. Integrating genome-wide association and transcriptome prediction model identifies novel target genes for osteoporosis. *Osteoporos Int*. 2021;32(12):2493–2503.

44. **Cederholm T**, **Cruz-Jentoft AJ**, **Maggi S**. Sarcopenia and fragility fractures. *Eur J Phys Rehabil Med*. 2013;49(1):111–117.

45. **Wang X**, **Yu Y**, **Huang Y**, **et al**. Identification of potential diagnostic gene biomarkers in patients with osteoarthritis. *Sci Rep*. 2020;10(1):13591.

46. **Cheng B**, **Wen Y**, **Yang X**, **et al**. Gut microbiota is associated with bone mineral density: an observational and genome-wide environmental interaction analysis in the UK Biobank cohort. *Bone Joint Res*. 2021;10(11):734–741.

47. **Behera J**, **Ison J**, **Tyagi SC**, **Tyagi N**. The role of gut microbiota in bone homeostasis. *Bone*. 2020;135:115317.

48. **Al-Hourani K**, **Tsang SJ**, **Simpson AHRW**. Osteoporosis: current screening methods, novel techniques, and preoperative assessment of bone mineral density. *Bone Joint Res*. 2021;10(12):840–843.

49. **Xiang C**, **Chen P**, **Zhang Q**, **et al**. Intestinal microbiota modulates adrenomedullary response through Nod1 sensing in chromaffin cells. *iScience*. 2021;24(8):102849.

50. **Yang A**, **Schweitzer R**, **Sun D**, **et al**. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature*. 1999;398(6729):714–718.

51. **Yang A**, **Kaghad M**, **Caput D**, **McKeon F**. On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet*. 2002;18(2):90–95.

52. **Mills AA**, **Zheng B**, **Wang XJ**, **Vogel H**, **Roop DR**, **Bradley A**. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature*. 1999;398(6729):708–713.

53. **Li F**, **Lu Y**, **Ding M**, **et al**. Putative function of TAP63α during endochondral bone formation. *Gene*. 2012;495(2):95–103.

54. **Lu Y**, **Abbassi S**, **Li F**, **et al**. Distinct function of P63 isoforms during embryonic skeletal development. *Gene*. 2013;519(2):251–259.

55. **Curtis KM**, **Aenlle KK**, **Frisch RN**, **Howard GA**, **Van WA**. TAp63γ and ΔNp63β promote osteoblastic differentiation of human mesenchymal stem cells: regulation by vitamin D3 Metabolites. *PLoS One*. 2015;10(4):e0123642.

56. **Nakashima T**, **Takayanagi H**. Osteoimmunology: crosstalk between the immune and bone systems. *J Clin Immunol*. 2009;29(5):555–567.

57. **Srivastava RK**, **Dar HY**, **Mishra PK**. Immunoporosis: immunology of osteoporosis-role of T cells. *Front Immunol*. 2018;9:657.

58. **Lanier LL**. NK cell recognition. *Annu Rev Immunol*. 2005;23(1):225–274.

59. **Martín-Fontecha A**, **Thomsen LL**, **Brett S**, **et al**. Induced recruitment of NK cells to lymph nodes provides IFN-gamma for T(H)1 priming. *Nat Immunol*. 2004;5(12):1260–1265.

60. **Hu M**, **Bassett JHD**, **Danks L**, **et al**. Activated invariant NKT cells regulate osteoclast development and function. *J Immunol*. 2011;186(5):2910–2917.

61. **Hayday AC**. Gammadelta T cells and the lymphoid stress-surveillance response. *Immunity*. 2009;31(2):184–196.

62. **Kinne RW**, **Bräuer R**, **Stuhlmüller B**, **Palombo-Kinne E**, **Burmester GR**. Macrophages in rheumatoid arthritis. *Arthritis Res*. 2000;2(3):189–202.

63. **Lioté F**, **Boval-Boizard B**, **Weill D**, **Kuntz D**, **Wautier JL**. Blood monocyte activation in rheumatoid arthritis: increased monocyte adhesiveness, integrin expression, and cytokine release. *Clin Exp Immunol*. 1996;106(1):13–19.

64. **Michalski MN**, **McCauley LK**. Macrophages and skeletal health. *Pharmacol Ther*. 2017;174:43–54.

65. **Andreev D**, **Liu M**, **Weidner D**, **et al**. Osteocyte necrosis triggers osteoclast-mediated bone loss through macrophage-inducible C-type lectin. *J Clin Invest*. 2020;130(9):4811–4830.

66. **Cao H**, **Yan Q**, **Wang D**, **et al**. Focal adhesion protein Kindlin-2 regulates bone homeostasis in mice. *Bone Res*. 2020;8:2.

Author information:
- W. Yuan, MD, Associate Professor
- M. Yang, MD, Professor
- Y. Zhu, MD, Professor
  Department of Orthopedics, The First Hospital of China Medical University, Shenyang, China.

Author contributions:
- W. Yuan: Conceptualization, Data curation, Funding acquisition, Methodology, Writing – original draft.
- M. Yang: Formal analysis, Funding acquisition, Investigation, Validation, Writing – review & editing.
- Y. Zhu: Conceptualization, Investigation, Supervision, Validation, Writing – review & editing.