PLOS ONE

Analysis of potential genetic biomarkers and molecular mechanism of smoking-related postmenopausal osteoporosis using weighted gene co-expression network analysis and machine learning

Shaoshuo Li[1], Baixing Chen[2], Hao Chen[3], Zhen Hua[3], Yang Shao[3], Heng Yin[3], Jianwei Wang[3]*

1 Nanjing University of Chinese Medicine, Nanjing, P.R. China, **2** Department of Development and Regeneration, KU Leuven, University of Leuven, Leuven, Belgium, **3** Department of Traumatology & Orthopedics, Wuxi Affiliated Hospital of Nanjing University of Chinese Medicine, Wuxi, P.R. China

* wxwangjianwei1963@126.com

🔓 OPEN ACCESS

Abstract

Objectives

Smoking is a significant independent risk factor for postmenopausal osteoporosis, leading to genome variations in postmenopausal smokers. This study investigates potential bio-markers and molecular mechanisms of smoking-related postmenopausal osteoporosis (SRPO).

Materials and methods

The GSE13850 microarray dataset was downloaded from Gene Expression Omnibus (GEO). Gene modules associated with SRPO were identified using weighted gene co-expression network analysis (WGCNA), protein-protein interaction (PPI) analysis, and path-way and functional enrichment analyses. Feature genes were selected using two machine learning methods: support vector machine-recursive feature elimination (SVM-RFE) and random forest (RF). The diagnostic efficiency of the selected genes was assessed by gene expression analysis and receiver operating characteristic curve.

Results

Eight highly conserved modules were detected in the WGCNA network, and the genes in the module that was strongly correlated with SRPO were used for constructing the PPI net-work. A total of 113 hub genes were identified in the core network using topological network analysis. Enrichment analysis results showed that hub genes were closely associated with the regulation of RNA transcription and translation, ATPase activity, and immune-related signaling. Six genes (*HNRNPC*, *PFDN2*, *PSMC5*, *RPS16*, *TCEB2*, and *UBE2V2*) were selected as genetic biomarkers for SRPO by integrating the feature selection of SVM-RFE and RF.

Conclusion

The present study identified potential genetic biomarkers and provided a novel insight into the underlying molecular mechanism of SRPO.

1. Introduction

Osteoporosis is a systemic skeletal disorder. This disease is highly prevalent worldwide and is characterized by bone microstructure degeneration, reduction in bone mineral density (BMD), leading to increased bone fragility and decreased bone strength [1, 2]. It is reported that almost 50% of postmenopausal women develop osteoporosis [3]. Furthermore, a third of postmenopausal women have bone fractures due to osteoporosis [4]. The estimated cost of managing postmenopausal osteoporosis (PMOP) and related fractures in the United States in 2015 was over USD 15 billion [5], and PMOP has become a major public health problem worldwide [6].

Multiple factors are involved in PMOP by affecting the function of osteoblasts and osteo-clasts and regulating bone mineral homeostasis [7]. Estrogen secretion is decreased during menopause, resulting in the decline of ovarian function, increasing the risk of bone metabolic diseases [8, 9]. Estrogens modulate immune activity and the response of immune cells (T cells, B cells, and monocytes) to estrogen and its receptors [10]. Circulating B lymphocytes are strongly implicated in the pathogenesis of PMOP by producing cytokines that regulate the activity of osteoblasts and osteoclasts. In addition, the downregulation of MAPK3 and ESR1 in B cells decreases osteogenesis and increases osteoclastogenesis, demonstrating the importance of B cells in the etiology of PMOP [11].

Poor lifestyle habits are significant contributors to rapid bone loss in postmenopausal women [12]. In this context, smoking is a significant independent risk factor for osteoporosis ($P = 0.000$, OR = 1.911) [13]. Female smokers are almost twice as likely to have osteoporosis than non-smoking women [14]. Smoking may lead to changes in the microarchitecture of tra-becular bone and reduces the ability of the skeletal muscle to resist mechanical load and stress [15]. Moreover, smoking may induce harmful changes in the immune system and cause dis-eases via the dysregulation of impaired B cells. Smoking-related postmenopausal osteoporosis (SRPO) is an emerging area of research that assesses changes in gene expression levels in post-menopausal smokers.

With the rapid development of high-throughput microarray technologies, the identification of genomic variations and biological mechanisms has improved our understanding of disease pathogenesis and treatment [16, 17]. Weighted gene co-expression network analysis (WGCNA) is widely used to analyze gene expression microarray data, identify functional gene modules, and discover relationships between gene modules and disease traits [18–20]. WGCNA screens genes and divides them into modules, which in turn are correlated with spe-cific clinical phenotypes through Pearson correlation analysis. Machine learning algorithms have shown great promise in investigating the underlying relationship of high-dimensional data through supervised or unsupervised methods [21, 22]. Moreover, machine learning is use-ful to analyze high-dimension transcriptomic data and identify feature genes with biological significance [23–25]. However, no studies have analyzed genome variations in SRPO.

In this study, we performed a comprehensive analysis of gene expression patterns of circu-lating B cells from 20 postmenopausal female smokers with low or high BMD using bioinfor-matics and machine learning algorithms, including WGCNA, support vector machine-

recursive feature elimination (SVM-RFE), random forest (RF), protein-protein interaction (PPI) and functional analyses, and receiver operating characteristic (ROC) curve analysis. Six potential diagnostic biomarkers of SRPO were identified.

2. Materials and methods

2.1. Microarray data collecting and data preprocessing

The study flowchart is shown in Fig 1. The gene microarray dataset GSE13850 based on the Affymetrix Human Genome U133A (GPL96) platform, probe annotation files, and CEL files were downloaded from the Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo/). Quantile normalization, background correction, and probe summarization of raw data were performed using the robust multiarray average (RMA) algorithm [26]. If one gene matched more than one probe, the maximum value of the probe was selected and calculated. The GSE13850 dataset provided data on gene expression in circulating B cells of 20 postmenopausal female smokers (10 with high BMD and 10 with low BMD).

2.2. Construction of the WGCNA network

Phenotype-correlated gene modules associated with SRPO were identified by WGCNA. The top 5,000 genes with the highest expression levels were used to construct the WGCNA network using the WGCNA package in R [20]. First, Pearson's correlation matrices for all pairs of genes were calculated. The pairwise correlation coefficient between the pair of gene m and gene n with significance (S_{mn}) was defined as $S_{mn} = |cor(m,n)|$. These correlation matrices were transformed into a weighted adjacency matrix using the power function $a_{mn} = power(S_{mn}, \beta) = |S_{mn}|^{\beta}$ [26]. According to the average connectivity degree and standard of approximate scale-free topology network, an appropriate soft-thresholding power β was selected, and the adjacency matrix was transformed into a topological overlap matrix (TOM). TOM-based hierarchical clustering of gene modules was performed using the dynamic tree cut algorithm [27]. Gene modules with similar expression profiles were represented by different branches with appropriate colors, and the minimum module size was set as 40.

2.3. Correlation between gene modules and SRPO

The WGCNA algorithm uses module eigengene (ME) to evaluate relationships between gene modules and clinical traits. ME was defined as the major component computed by a principal component analysis that recapitulates the manifestation of genes from a specific module into a characteristic expression profile [28]. The Pearson correlation between ME and clinical traits was calculated to identify the module that was highly correlated with SRPO. The significance of Pearson correlation was assessed using a *t*-test, and the module with a *P*-value of less than 0.05 was considered to be significantly correlated with SRPO. Furthermore, gene significance (GS) and module membership (MM) were calculated for intramodular analysis. MM was the correlation between ME and the gene expression profile. GS was defined as the log10 transformation of the *P*-value (lg*P*) between gene expression and the clinical trait (GS = lg*P*). Module significance (MS) was defined as the average GS of all genes in a module. The module with the highest absolute MS was considered to be significantly correlated with SRPO. The module with the highest correlation with a clinical trait (osteoporosis) was selected as a research object.

2.4. Construction of PPI networks

PPI networks were constructed to evaluate the relationship among genes in the selected modules using the Search Tool for the Retrieval of Interacting Genes version 11 (STRING V11,
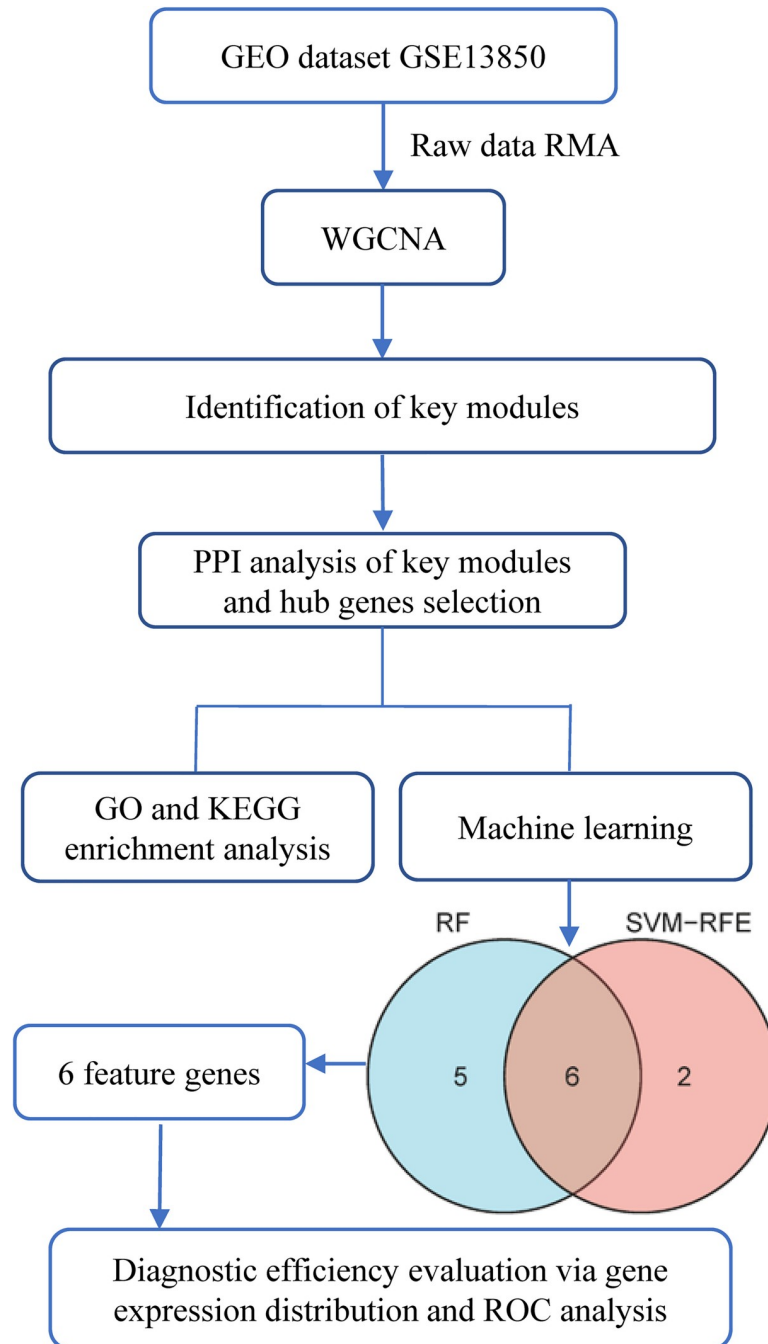
Fig 1. Workflow of the present study.

https://doi.org/10.1371/journal.pone.0257343.g001

https://string-preview.org/). The confidence level was set as >0.4, and the network was visualized using Cytoscape version 3.8.2 [29]. Hub genes are highly interconnected nodes and may play important roles in the PPI network. A topological network analysis, including betweenness centrality (BC), closeness centrality (CC), and degree centrality (DC), for screening hub genes was performed using the CytoNCA plugin for Cytoscape [30].

2.5. Function and pathway enrichment analyses

Gene Ontology (GO) enrichment analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the clusterProfiler [31] package in R to describe the possible biological functions of hub genes. Three categories of biological process (BP), cellular component (CC), and molecular function (MF) were included in the GO terms. A Benjamini–Hochberg adjusted P-value of less than 0.05 was considered to indicate significantly enriched GO terms and KEGG pathways.

2.6. Machine learning for feature selection

Feature genes associated with SRPO were selected using SVM-RFE and RF. SVM-RFE was an efficient feature selection algorithm and had shown promising power in the analysis of the genomics [32], metabolomics [33], proteomics [34], etc. During the performance, SVM-RFE iteratively removed the features with the smallest weight from a rank until all features were excluded. In each iteration, the current SVM-RFE model was evaluated by k-fold cross-validation. After that, the classifier model with the highest accuracy was constructed, and the best variables were found [35]. The RF algorithm used the variables to construct numerous decision trees and generated the most accurate classes of variables to individual trees. RF has also been widely used for detecting disease biomarkers [36, 37]. The SVM-RFE model was built using the R package caret version 6.0–88. RF was applied using the randomForest package version 4.6–14. Ultimately, the common genes obtained using both SVM-RFE and RF were combined for further analysis.

## 2.7. Evaluation of the diagnostic efficiency

The ability of feature genes to differentiate between SRPO patients and non-osteoporosis postmenopausal smokers was evaluated by gene expression and ROC curve analyses. The predictive efficiency was measured in the control group (ten samples from postmenopausal smokers with high BMD) and the SRPO group (ten samples from postmenopausal smokers with low BMD). A Benjamini–Hochberg adjusted P-value of less than 0.05 were considered to indicate significant differences in gene expression. The ROC curve was created using the pROC package version 1.17.0.1 in R. The genes with an area under the ROC curve (AUC)>0.7 were considered to have good diagnostic performance.

3. Results

3.1. Data collection and WGCNA analysis

Gene expression data and clinical data from the GSE13850 dataset were downloaded from the GEO database. Following data processing, the top 5,000 genes in circulating B cells were collected, and the WGCNA network was constructed. Subsequently, an appropriate soft-thresholding power β = 9 was adopted due to the signed R^2 of the scale-free topology network was 0.85 (Fig 2).

Eight gene modules were obtained using the dynamic tree cut algorithm (Fig 3A and 3B). The correlation between each module and osteoporosis was assessed by calculating the module–trait relationship and MS. First, the Pearson correlation between the ME of each module and osteoporosis was calculated and shown in the module–trait relationship heatmap (Fig 3C and Table 1). The blue module (module–trait relationships = 0.88, P-value = 7e-07) had the highest association with osteoporosis. After that, the MS of each module was calculated. We found that the blue module had the highest MS among all selected modules (Fig 3D). Hence, the 1078 genes in the blue module were significantly associated with SRPO, and these genes
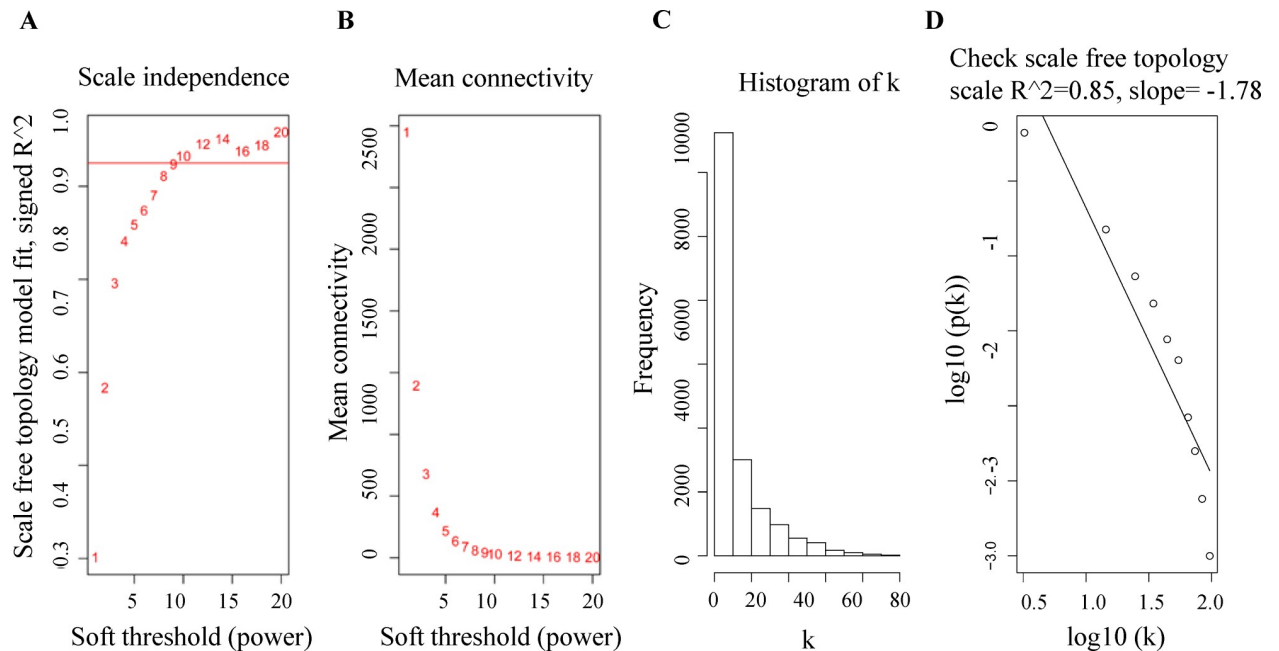
Fig 2. Construction of the weighted gene co-expression network of gene modules. (A) Analysis of the scale independence for the appropriate soft-thresholding power β. (B) Analysis of the mean connectivity for the appropriate soft-thresholding power β. (C) Histogram of connectivity distribution with an appropriate β = 9. (D) Checking the scale-free topology with an appropriate β = 9.

were selected for subsequent analysis in the PPI network. The clustering heatmap of the ME of the blue module and the scatterplots of GS vs. MM are presented in Fig 3E and 3F.

## 3.2. Construction of the PPI network and enrichment analysis of hub genes

After removing the disconnected nodes, there were 998 nodes and 10940 edges in the constructed PPI network for genes in the blue module (Fig 4A). According to topological network analysis, PPI nodes are considered significant targets if the DC is greater than two-fold the median DC [38]. Thus, DC > 28 was set as the threshold, and significant nodes were identified to generate a subnetwork. Then, nodes where BC and CC values were greater than the median in the subnetwork (BC>158.81, CC>0.48) were considered a new core network containing hub genes. The core network containing 113 hub genes (nodes) and 1831 edges is shown in Fig 4B.

Functional enrichment analysis was performed to improve biological understanding of the hub genes identified in the PPI network. Regarding biological processes, GO analysis showed that hub genes were mainly involved in the regulation of mRNA transcription, regulation of cell cycle, protein targeting, and cellular response to hypoxia (Fig 5A). In the cellular component analysis, hub genes were mainly associated with ribosomal subunits, methylosome, and proteasome complexes (Fig 5B). Significantly enriched molecular functions were translation regulation, ATPase activity, hormone receptor binding, and protein binding (Fig 5C). KEGG pathway enrichment analysis showed that ribosome, apoptosis, mitophagy, HIF-1 signaling pathway, NF-kappa B signaling pathway, Th17 cell differentiation, and B cell receptor signaling pathway were the most significant processes in SRPO (Fig 5D).

3.3. Identification of feature genes using machine learning algorithms

Machine learning classification algorithms are being increasingly used to predict feature genes associated with diseases from the noise background. SVM-RFE and RF were used to predict
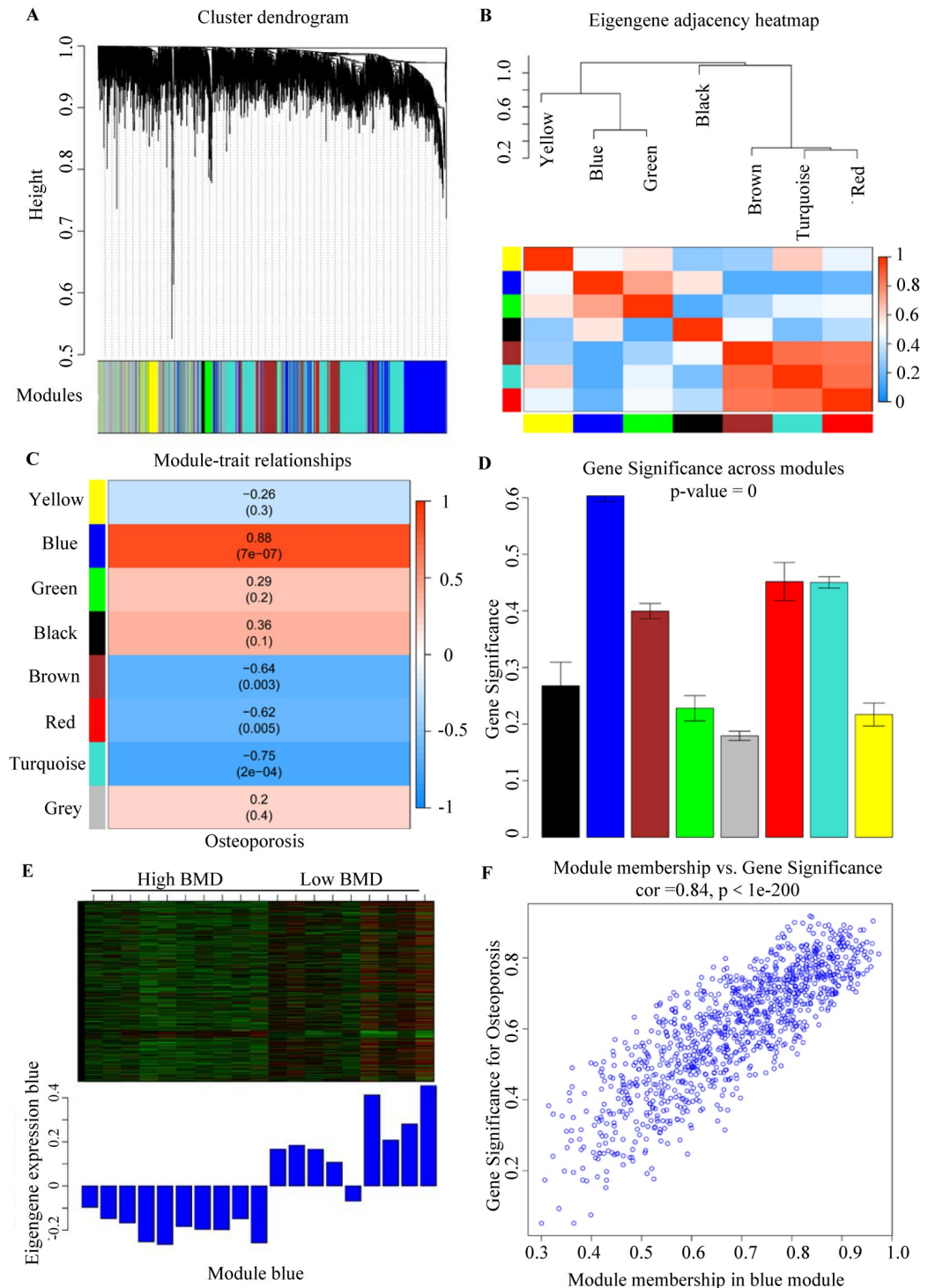
Fig 3. Identification of significant gene modules correlated with osteoporosis. (A) Cluster dendrogram of representative gene modules. (B) Clustering heatmap of module eigengenes. (C) Relationships of module eigengenes and osteoporosis. The number in the square at the top of each row is the correlation coefficient, and *P*-values are shown below. (D) Gene significance across modules. (E) Heatmap and bar graph of the eigengenes in module blue. (F) Scatterplot of gene significance vs. module membership in the blue module.

https://doi.org/10.1371/journal.pone.0257343.g003

Table 1. Correlation between modules and smoking-related postmenopausal osteoporosis.

| Modules	Gene count	Correlation	P-value
Blue	1078	0.88	7e-07
Turquoise	1728	-0.75	2e-04
Brown	771	-0.64	0.003
Red	78	-0.62	0.005
Black	60	0.36	0.1
Green	164	0.29	0.2
Yellow	225	-0.26	0.3
Grey	896	0.2	0.4

https://doi.org/10.1371/journal.pone.0257343.t001

feature genes associated with SRPO. First, an SVM-RFE classifier (Core: svmliner; Cross: 10-fold cross-validation; soft-margin; tuning parameter C = 1) was established based on 113 hub genes. Data from the control and SRPO groups were randomly divided into ten equal portions (training set: 9; test set: 1). During each of the ten iterations, SVM-RFE was applied to the training set to train the classifier with the selected features, and the trained classifier was applied to the test set to assess prediction accuracy. Then, the predictions from the ten iterations were combined to evaluate the accuracy of the classifier. Eight feature genes were validated using SVM-RFE (Fig 6A). Similarly, feature genes were screened by 10-fold cross-validation using RF algorithm. The RF classifier showed a least out-of-bag (OOB) error with the top 11 feature genes (Fig 6B). After integrating feature genes from SVM-RFE and RF, six feature genes closely associated with SRPO were obtained: *HNRNPC, PFDN2, PSMC5, RPS16, TCEB2*, and *UBE2V2* (Fig 6C).

3.4. Diagnostic efficiency of feature genes

The difference in expression pattern of the six feature genes between the SRPO and control groups was assessed. Gene expression was downregulated in the SRPO group, except for *UBE2V2* (Fig 7A). To identify if the feature genes influence SRPO diagnosis independently, ROC analysis was performed. The results showed that the ability of these genes to diagnose SRPO was high, with an AUC>0.9 (Fig 7B).

As an RNA-binding protein, heterogeneous nuclear ribonucleoprotein C (*HNRNPC*) is well known for regulating mRNA metabolism and RNA expression, splicing, and translation [39, 40]. In addition, *HNRNPC* regulates N6-methyladenosine (m6A) RNA methylation, which is crucial to neurogenesis, embryonic development, stress responses, and tumorigenesis [41, 42]. *TCEB2* (also known as *ELOB*) encodes the protein elongin B, a subunit of the transcription factor B complex and an adapter protein in the proteasomal degradation of target proteins through E3 ubiquitin ligases [43]. Proteasome 26S subunit ATPase 5 (*PSMC5*) interacts with several transcription factors, including nuclear hormone receptors, p53, c-Fos, and the basal transcription complex [44]. Moreover, *PSMC5* plays a proteasome-independent role in DNA repair, chromatin remodeling, and transcription activation and elongation [45, 46]. *PFDN2* is a component of β subunits of the URI prefoldin-like complex, which plays a critical role in maintaining cellular homeostasis [47]. Ubiquitin-conjugating enzyme E2 variant 2 (*UBE2V2*) mediates the transcriptional activation of target genes and controls cell differentiation, cell cycle, and DNA damage response [48]. Ribosomal protein S16 (*RPS16*), the basic component of the 40S ribosome, was reported to be associated with the defective mitochondrial translation [49]. These feature genes were closely associated with RNA transcription and translation, and important cellular activity in SRPO.

998 nodes, 10940 edges   283 nodes, 5229 edges   113 nodes, 1831 edges
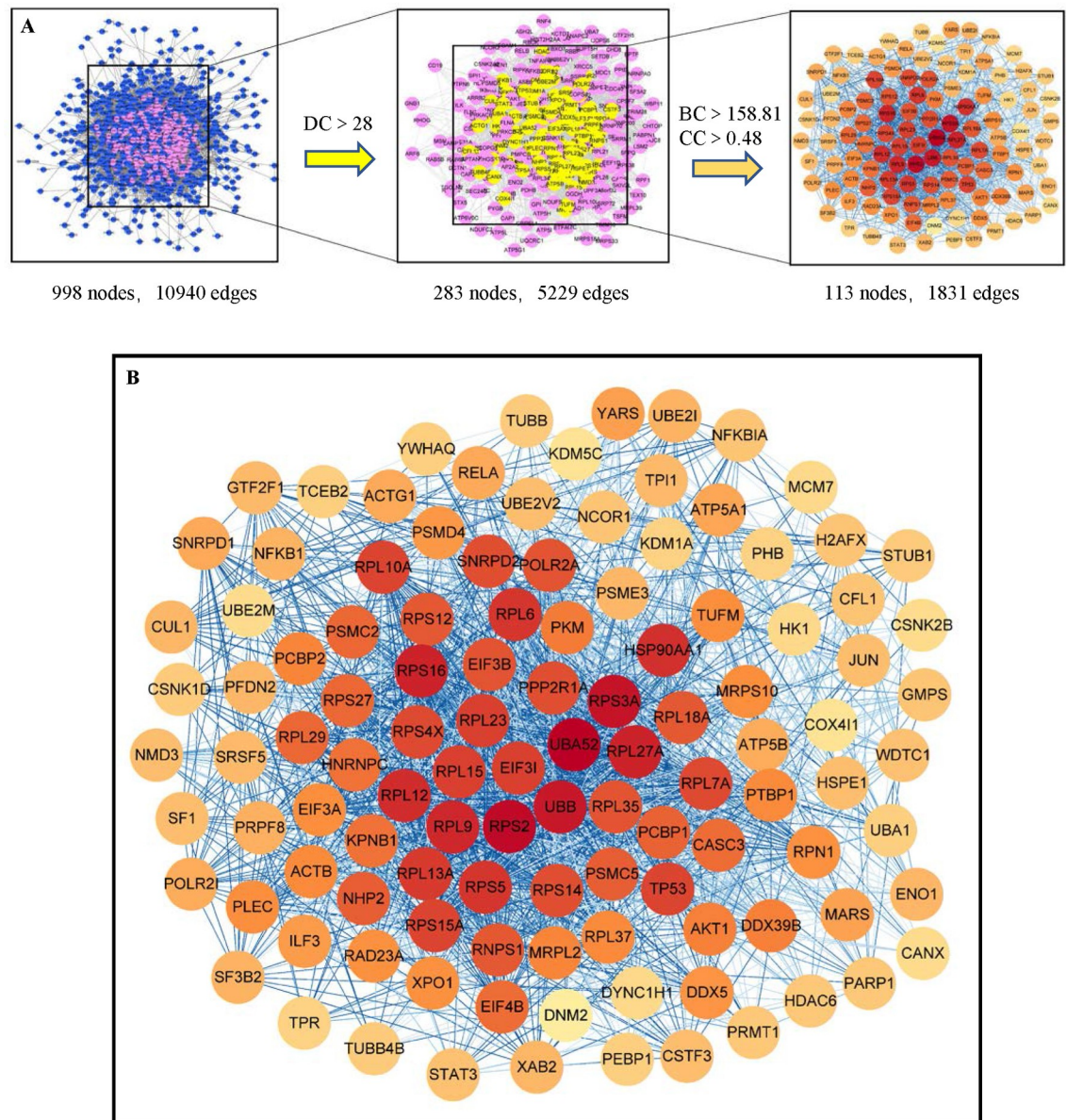
Fig 4. Protein-Protein Interaction (PPI) network of genes from the blue module. (A) Screening of hub genes. The screening criteria were degree centrality>28, betweenness centrality>158.81, and closeness centrality>0.48. (B) Core PPI network with 113 hub genes and 1831edges. The color of the nodes represented the value of degree. The darker (red) the color, the higher the degree.

https://doi.org/10.1371/journal.pone.0257343.g004

4. Discussion

There is increased public awareness of the harmful effects of exposure to cigarette smoking. However, although substantial progress has been made in tobacco control, cigarette smoking remains one of the most challenging global health issues to date [50, 51]. Postmenopausal smokers are at an increased risk of developing osteoporosis and osteoporotic fractures than non-smoking females [52]. Moreover, smoking-induced genetic alterations influence hormone secretion and bone metabolism in women [53, 54]. The molecular mechanism of occurrence and development of SRPO is incompletely understood, and identifying new biomarkers for SRPO diagnosis and treatment is crucial.
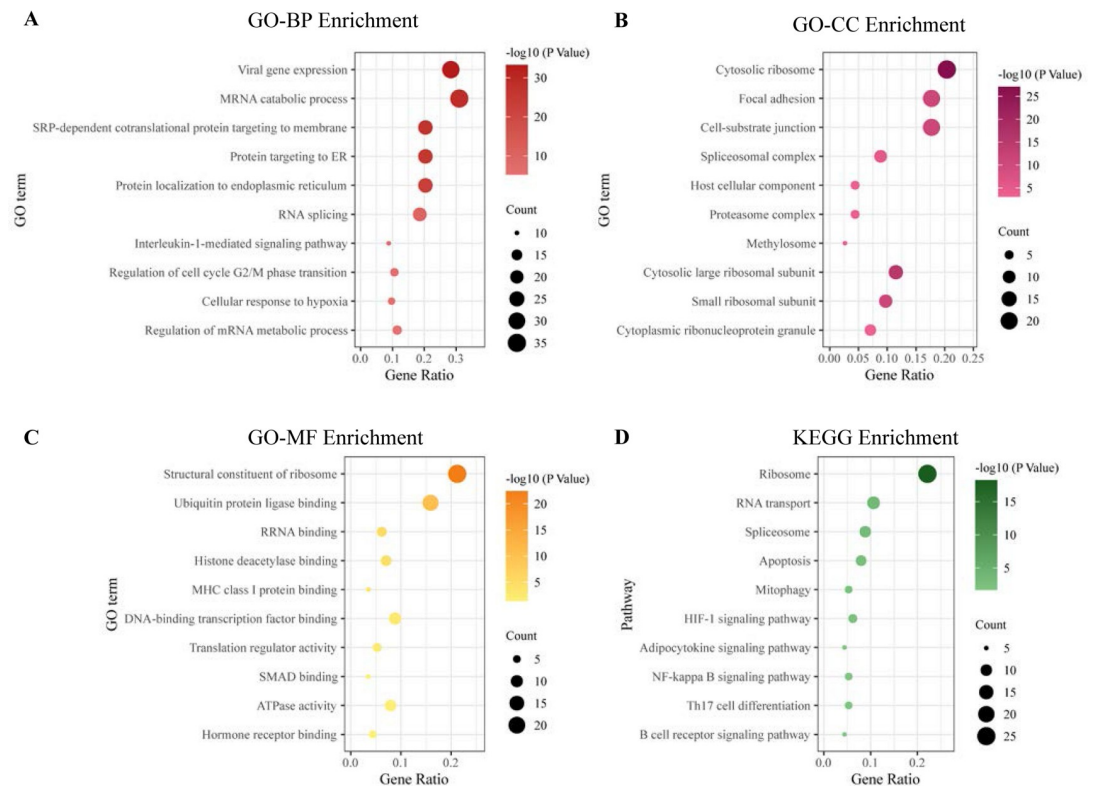
Fig 5. Functional enrichment analysis of hub genes. (A-C) Gene ontology enrichment analysis. (D) Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. BP, biological process; CC, cellular component, MF, molecular function.

https://doi.org/10.1371/journal.pone.0257343.g005

We determined the gene expression profiles in circulating B cells from 20 postmenopausal smokers with low or high BMD. First, WGCNA was performed to select the gene modules with the strongest correlation with SRPO. Then, 1078 genes in the selected module were used to construct a PPI network. Topological network analysis identified a core PPI network and 113 hub genes. Functional enrichment analysis showed that these hub genes were closely associated with the development of SRPO via the control of several biological processes, including the regulation of RNA transcription and translation, hormone receptor binding, and NF-kappa B signaling pathway. Previous studies have shown that these biological processes and signaling pathways are implicated in bone metabolism and osteoporosis [55, 56]. The risk of missing important features was minimized by incorporating genes using two machine learning algorithms. SVM-RFE and RF were performed to screen six characteristic variables from these hub genes. Diagnostic efficiency analysis showed that the genes *HNRNPC*, *PFDN2*, *PSMC5*, *RPS16*, *TCEB2*, and *UBE2V2* were potential biomarkers for SRPO.

In a cigarette smoke-induced chronic obstructive pulmonary disease (COPD) animal model, *HNRNPC* was overexpressed in the lungs of cigarette smoke-exposed mice [57]. The dysregulation of *HNRNPC* is associated with telomere shortening in lung cells and circulating lymphocytes, impairing lung function and increasing COPD severity and mortality [58, 59]. In addition, the dysregulation of *HNRNPC* may increase the expression of the urokinase plasminogen activator receptor, resulting in inflammation and immune activation [60]. *TCEB2* plays an essential role in the development of acquired resistance to anti-angiogenic therapy in ovarian cancer cells via suppressing VEGF-A expression and promoting HIF-1α degradation [61]. The vascularization of bone tissue is tightly linked with bone formation in a spatial and
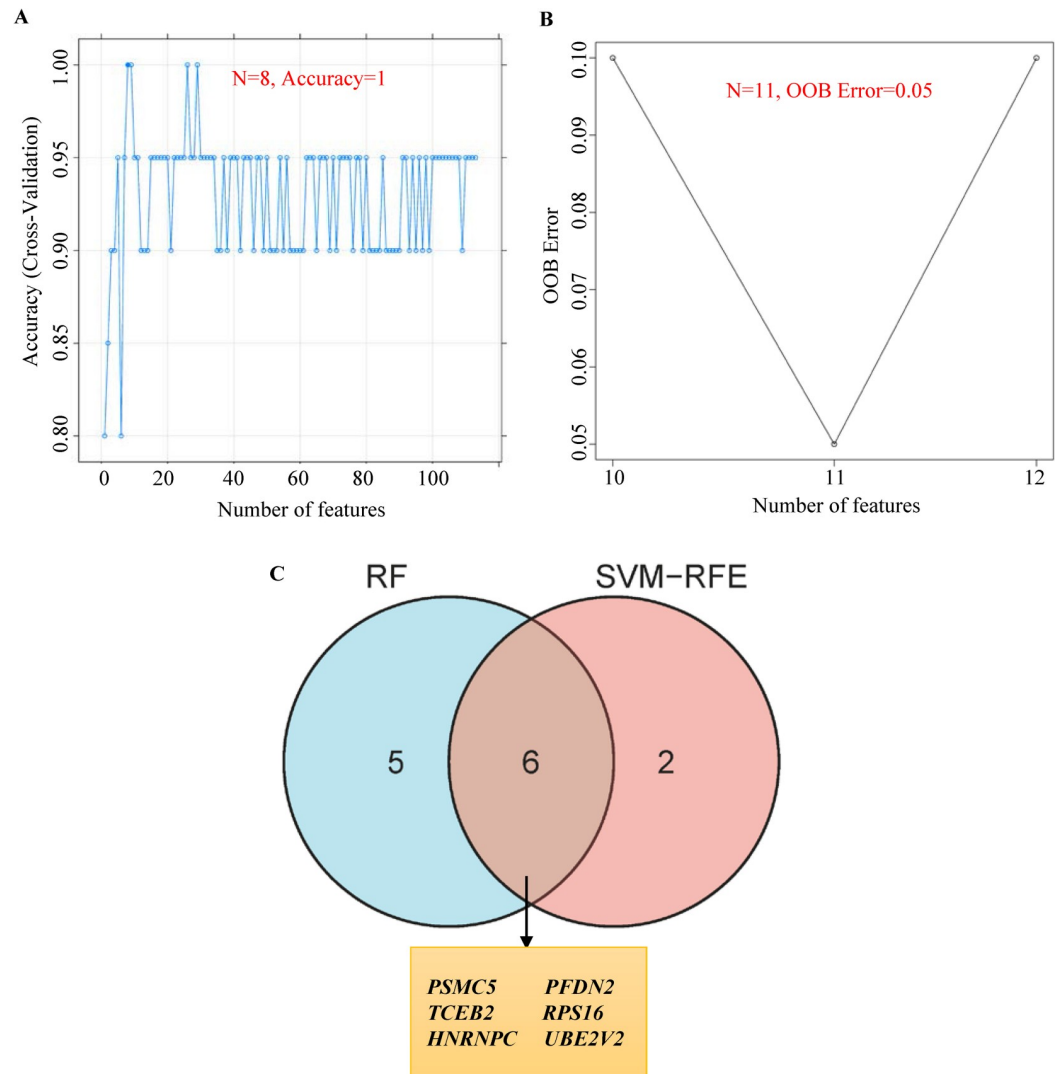
Fig 6. Feature genes selection. Using support vector machine-recursive feature elimination (SVM-RFE) (A) and random forest (RF) (B). (C) Venn plot of feature genes selected by RF and SVM-RFE.

temporal relationship known as angiogenesis-osteogenesis coupling [62]. Many factors, including HIF-1 and VEGF, regulate bone vascularization and angiogenic-osteogenic coupling in the bone microenvironment [63]. In this respect, the dysregulation of *TCEB2* may contribute to SRPO by impairing this coupling. *PSMC5* regulates ERK1/2 signaling transmission by remodeling the Shoc2 scaffold complex [64]. The activation of the ERK1/2 signaling cascade regulates the function of osteoblasts and osteoclasts, promoting inflammation and osteogenesis [65, 66]. *PFDN2* is closely associated with several diseases, such as Alzheimer's disease, colon cancer, and myelodysplastic syndromes, via different mechanisms [67–69]. The presence of antibodies against *PFDN2* is associated with an increased risk of type 2 diabetes through autoimmune activation and/or pro-inflammatory signals, which are involved in the regulation of bone homeostasis [70]. *UBE2V2* contributes to the development and progression of many cancers, including prostate, oropharyngeal, and breast cancers, via promoting cell proliferation, suppressing cell apoptosis, and regulating immune signaling [71–73]. Moreover, *UBE2V2* is an independent prognostic indicator for lung adenocarcinoma, which is closely
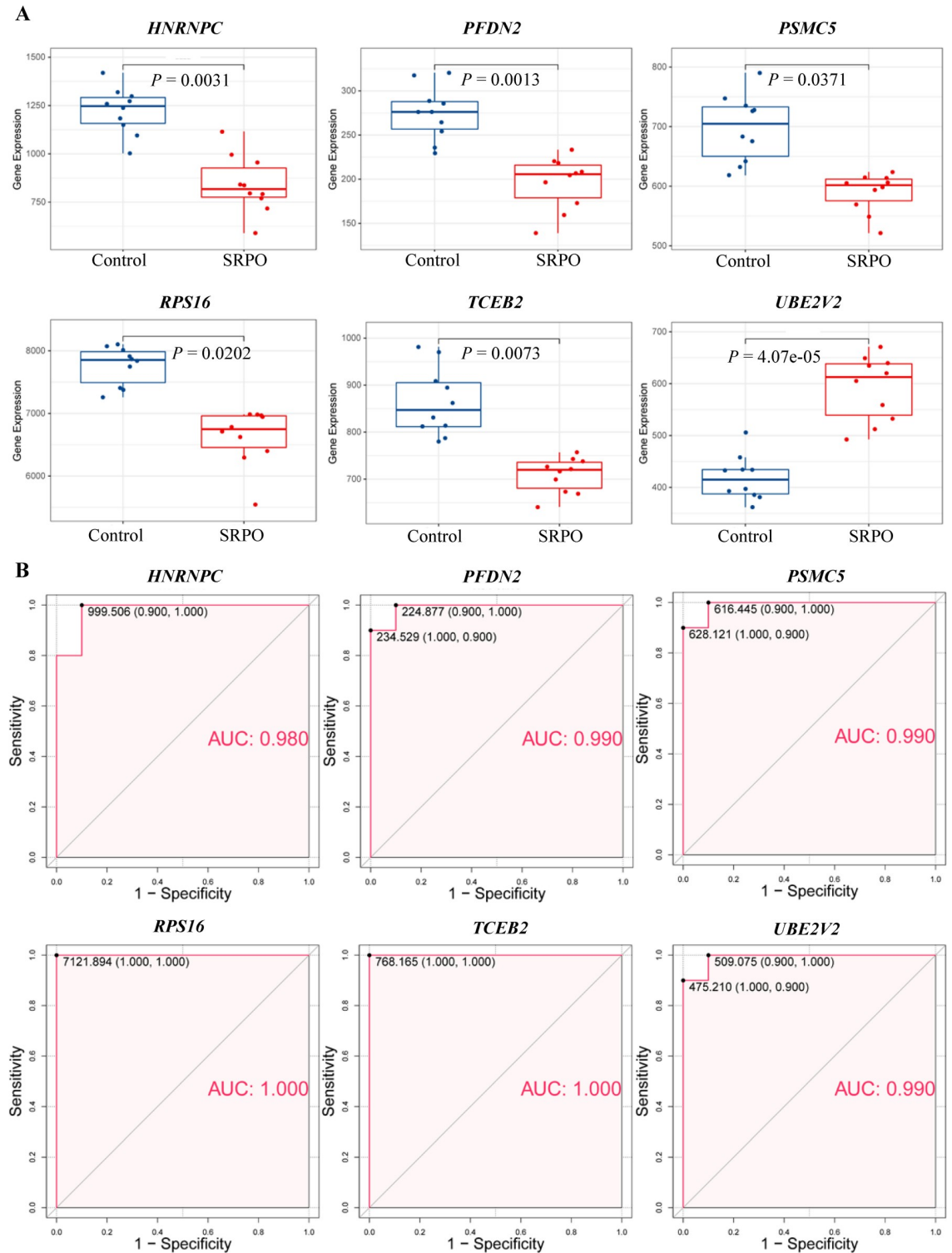
Fig 7. Diagnostic efficiency evaluation of feature genes. (A) Gene expression of six feature genes (*HNRNPC*, *PFDN2*, *PSMC5*, *RPS16*, *TCEB2*, and *UBE2V2*) in women with smoking-related postmenopausal osteoporosis and controls. (B) Receiver operating characteristic curve analysis.

related to the mutational processes of cigarette smoking [74, 75]. *RPS16* contributes to facilitate tumor progression of glioma via the PI3K/AKT signaling [76]. Previous studies have indicated that the PI3K/AKT signaling pathway is an important factor in the occurrence of osteoporosis by regulating the activity of osteoblasts and osteoclasts [77, 78].

WGCNA can identify genes with clinical significance and cluster genes associate with pathological processes based on medical and biological background. Machine learning algorithms have shown objective assessment and optimal accuracy in feature selection. The present study is the first to perform a comprehensive strategy of machine learning algorithms and WGCNA to identify potential biomarkers of SRPO. Although our results are consistent with the literature, the reliability of this study needs to be verified by further experiments. This study has limitations. First, the smoking history, frequency, and status of individuals in the study were not well known, which might cause uncontrolled factors in data analysis. Second, the identified biomarkers were not functionally and externally validated. Third, the small sample size may have limited the power of the study. Additional studies on the association of these biomarkers with SRPO are warranted.

5. Conclusion

The present study identified six genes (*HNRNPC*, *PFDN2*, *PSMC5*, *RPS16*, *TCEB2*, and *UBE2V2*) as potential biomarkers for SRPO using WGCNA and machine learning algorithms, providing a novel insight into the diagnosis and treatment of SRPO. However, these biomarkers need to be validated by clinical trials.

Supporting information

S1 Data. Gene expression of samples.
(XLSX)

S2 Data. Genes in each module.
(XLSX)

S3 Data. PPI network.
(XLSX)

S4 Data. Results of CytoNCA analysis.
(XLSX)

S5 Data. GO enrichment of hub genes.
(XLSX)

S6 Data. KEGG pathway enrichment.
(XLSX)

S7 Data. Feature selection of machine learning.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Shaoshuo Li.

Data curation: Shaoshuo Li, Baixing Chen.

Formal analysis: Shaoshuo Li, Baixing Chen, Hao Chen.

Investigation: Zhen Hua, Yang Shao.

Methodology: Shaoshuo Li.

Project administration: Heng Yin, Jianwei Wang.

Software: Shaoshuo Li.

Supervision: Shaoshuo Li.

Validation: Shaoshuo Li.

Visualization: Shaoshuo Li.

Writing – original draft: Shaoshuo Li, Baixing Chen.

Writing – review & editing: Shaoshuo Li, Heng Yin, Jianwei Wang.

References

1. Kanis JA, McCloskey EV, Johansson H, Oden A, Melton LJ 3rd, Khaltaev N. A reference standard for the description of osteoporosis. Bone. 2008; 42(3):467–75. Epub 2008/01/09. https://doi.org/10.1016/j.bone.2007.11.001 PMID: 18180210.

2. Kanis JA, Melton LJ 3rd, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research. 1994; 9(8):1137–41. https://doi.org/10.1002/jbmr.5650090802 PMID: 7976495.

3. Taguchi A, Ohtsuka M, Nakamoto T, Naito K, Tsuda M, Kudo Y, et al. Identification of post-menopausal women at risk of osteoporosis by trained general dental practitioners using panoramic radiographs. Dento maxillo facial radiology. 2007; 36(3):149–54. https://doi.org/10.1259/dmfr/31116116 PMID: 17463099.

4. Brown C. Osteoporosis: Staying strong. Nature. 2017; 550(7674):S15–s7. https://doi.org/10.1038/550S15a PMID: 28976955.

5. 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. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research. 2007; 22(3):465–75. https://doi.org/10.1359/jbmr.061113 PMID: 17144789.

6. Qian GF, Yuan LS, Chen M, Ye D, Chen GP, Zhang Z, et al. PPWD1 is associated with the occurrence of postmenopausal osteoporosis as determined by weighted gene co-expression network analysis. Molecular medicine reports. 2019; 20(4):3202–14. https://doi.org/10.3892/mmr.2019.10570 PMID: 31432133

7. Yuan FL, Xu RS, Jiang DL, He XL, Su Q, Jin C, et al. Leonurine hydrochloride inhibits osteoclastogenesis and prevents osteoporosis associated with estrogen deficiency by inhibiting the NF-κB and PI3K/Akt signaling pathways. Bone. 2015; 75:128–37. https://doi.org/10.1016/j.bone.2015.02.017 PMID: 25708053.

8. Faienza MF, Ventura A, Marzano F, Cavallo L. Postmenopausal osteoporosis: the role of immune system cells. Clinical & developmental immunology. 2013; 2013:575936. https://doi.org/10.1155/2013/575936 PMID: 23762093

9. Gohlke-Bärwolf C. Coronary artery disease—is menopause a risk factor? Basic research in cardiology. 2000; 95 Suppl 1:I77–83. https://doi.org/10.1007/s003950070014 PMID: 11192358.

10. Phiel KL, Henderson RA, Adelman SJ, Elloso MM. Differential estrogen receptor gene expression in human peripheral blood mononuclear cell populations. Immunology letters. 2005; 97(1):107–13. https://doi.org/10.1016/j.imlet.2004.10.007 PMID: 15626482.

11. Xiao P, Chen Y, Jiang H, Liu YZ, Pan F, Yang TL, et al. In vivo genome-wide expression study on human circulating B cells suggests a novel ESR1 and MAPK3 network for postmenopausal osteoporosis. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research. 2008; 23(5):644–54. https://doi.org/10.1359/jbmr.080105 PMID: 18433299

12. Zhu K, Prince RL. Lifestyle and osteoporosis. Current osteoporosis reports. 2015; 13(1):52–9. https://doi.org/10.1007/s11914-014-0248-6 PMID: 25416958.

13. Bijelic R, Milicevic S, Balaban J. Risk Factors for Osteoporosis in Postmenopausal Women. Medical archives (Sarajevo, Bosnia and Herzegovina). 2017; 71(1):25–8. https://doi.org/10.5455/medarh.2017.71.25-28 PMID: 28428669

14. Korkor AB, Eastwood D, Bretzmann C. Effects of gender, alcohol, smoking, and dairy consumption on bone mass in Wisconsin adolescents. WMJ: official publication of the State Medical Society of Wisconsin. 2009; 108(4):181–8. PMID: 19753823.

15. Brook JS, Balka EB, Zhang C. The smoking patterns of women in their forties: their relationship to later osteoporosis. Psychological reports. 2012; 110(2):351–62. https://doi.org/10.2466/13.18.PR0.110.2.351-362 PMID: 22662390

16. Yan S. Integrative analysis of promising molecular biomarkers and pathways for coronary artery disease using WGCNA and MetaDE methods. Molecular medicine reports. 2018; 18(3):2789–97. https://doi.org/10.3892/mmr.2018.9277 PMID: 30015926

17. Gruebner O, Sykora M, Lowe SR, Shankardass K, Galea S, Subramanian SV. Big data opportunities for social behavioral and mental health research. Social science & medicine (1982). 2017; 189:167–9. https://doi.org/10.1016/j.socscimed.2017.07.018 PMID: 28755794.

18. Giulietti M, Occhipinti G, Righetti A, Bracci M, Conti A, Ruzzo A, et al. Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data. Frontiers in oncology. 2018; 8:450. https://doi.org/10.3389/fonc.2018.00450 PMID: 30370253

19. Abu-Jamous B, Kelly S. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. Genome biology. 2018; 19(1):172. https://doi.org/10.1186/s13059-018-1536-8 PMID: 30359297

20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559 PMID: 19114008

21. Ahuja K, Rather GM, Lin Z, Sui J, Xie P, Le T, et al. Toward point-of-care assessment of patient response: a portable tool for rapidly assessing cancer drug efficacy using multifrequency impedance cytometry and supervised machine learning. Microsystems & nanoengineering. 2019; 5:34. https://doi.org/10.1038/s41378-019-0073-2 PMID: 31645995

22. Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 2019; 571(7763):95–8. https://doi.org/10.1038/s41586-019-1335-8 PMID: 31270483.

23. Bogard N, Linder J, Rosenberg AB, Seelig G. A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. Cell. 2019; 178(1):91–106.e23. https://doi.org/10.1016/j.cell.2019.04.046 PMID: 31178116

24. Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, et al. Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into Streptococcus pyogenes pathogenesis. Nature genetics. 2019; 51(3):548–59. https://doi.org/10.1038/s41588-018-0343-1 PMID: 30778225.

25. Huang Y, Liu H, Zuo L, Tao A. Key genes and co-expression modules involved in asthma pathogenesis. PeerJ. 2020; 8:e8456. https://doi.org/10.7717/peerj.8456 PMID: 32117613

26. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005; 4:Article17. https://doi.org/10.2202/1544-6115.1128 PMID: 16646834.

27. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics (Oxford, England). 2008; 24(5):719–20. https://doi.org/10.1093/bioinformatics/btm563 PMID: 18024473.

28. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. Methods in molecular biology (Clifton, NJ). 2014; 1084:193–226. https://doi.org/10.1007/978-1-62703-658-0_11 PMID: 24061923

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003; 13 (11):2498–504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

30. Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. Bio Systems. 2015; 127:67–72. https://doi.org/10.1016/j.biosystems.2014.11.005 PMID: 25451770.

31. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012; 16(5):284–7. https://doi.org/10.1089/omi.2011.0118 PMID: 22455463

32. Wei S, Lu J, Lou J, Shi C, Mo S, Shao Y, et al. Gastric Cancer Tumor Microenvironment Characterization Reveals Stromal-Related Gene Signatures Associated With Macrophage Infiltration. Frontiers in genetics. 2020; 11:663. https://doi.org/10.3389/fgene.2020.00663 PMID: 32695142

33. Grissa D, Pétéra M, Brandolini M, Napoli A, Comte B, Pujos-Guillot E. Feature Selection Methods for Early Predictive Biomarker Discovery Using Untargeted Metabolomic Data. Frontiers in molecular biosciences. 2016; 3:30. https://doi.org/10.3389/fmolb.2016.00030 PMID: 27458587

34. Zhang F, Petersen M, Johnson L, Hall J, O'Bryant SE. Recursive Support Vector Machine Biomarker Selection for Alzheimer's Disease. Journal of Alzheimer's disease: JAD. 2021; 79(4):1691–700. https://doi.org/10.3233/JAD-201254 PMID: 33492292.

35. Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, et al. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2012; 910:149–55. https://doi.org/10.1016/j.jchromb.2012.05.020 PMID: 22682888.

36. Toth R, Schiffmann H, Hube-Magg C, Büscheck F, Höflmayer D, Weidemann S, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. Clinical epigenetics. 2019; 11 (1):148. https://doi.org/10.1186/s13148-019-0736-8 PMID: 31640781

37. Guo L, Wang Z, Du Y, Mao J, Zhang J, Yu Z, et al. Random-forest algorithm based biomarkers in predicting prognosis in the patients with hepatocellular carcinoma. Cancer cell international. 2020; 20:251. https://doi.org/10.1186/s12935-020-01274-z PMID: 32565735

38. Xu T, Ma C, Fan S, Deng N, Lian Y, Tan L, et al. Systematic Understanding of the Mechanism of Baicalin against Ischemic Stroke through a Network Pharmacology Approach. Evidence-based complementary and alternative medicine: eCAM. 2018; 2018:2582843. https://doi.org/10.1155/2018/2582843 PMID: 30647760

39. Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell. 2013; 152(3):453–66. https://doi.org/10.1016/j.cell.2012.12.023 PMID: 23374342

40. Park YM, Hwang SJ, Masuda K, Choi KM, Jeong MR, Nam DH, et al. Heterogeneous nuclear ribonucleoprotein C1/C2 controls the metastatic potential of glioblastoma by regulating PDCD4. Molecular and cellular biology. 2012; 32(20):4237–44. https://doi.org/10.1128/MCB.00443-12 PMID: 22907752

41. Niu Y, Zhao X, Wu YS, Li MM, Wang XJ, Yang YG. N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. Genomics, proteomics & bioinformatics. 2013; 11(1):8–17. https://doi.org/10.1016/j.gpb.2012.12.002 PMID: 23453015

42. Wang LC, Chen SH, Shen XL, Li DC, Liu HY, Ji YL, et al. M6A RNA Methylation Regulator HNRNPC Contributes to Tumorigenesis and Predicts Prognosis in Glioblastoma Multiforme. Frontiers in oncology. 2020; 10:536875. https://doi.org/10.3389/fonc.2020.536875 PMID: 33134160

43. Stebbins CE, Kaelin WG Jr., Pavletich NP. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science (New York, NY). 1999; 284(5413):455–61. https://doi.org/10.1126/science.284.5413.455 PMID: 10205047.

44. St-Arnaud R. Dual functions for transcriptional regulators: myth or reality? Journal of cellular biochemistry. 1999;Suppl 32– 33:32–40. PMID: 10629101.

45. Rousseau E, Kojima R, Hoffner G, Djian P, Bertolotti A. Misfolding of proteins with a polyglutamine expansion is facilitated by proteasomal chaperones. The Journal of biological chemistry. 2009; 284 (3):1917–29. https://doi.org/10.1074/jbc.M806256200 PMID: 18986984

46. Sulahian R, Sikder D, Johnston SA, Kodadek T. The proteasomal ATPase complex is required for stress-induced transcription in yeast. Nucleic acids research. 2006; 34(5):1351–7. https://doi.org/10.1093/nar/gkl012 PMID: 16517940

47. Chaves-Pérez A, Thompson S, Djouder N. Roles and Functions of the Unconventional Prefoldin URI. Advances in experimental medicine and biology. 2018; 1106:95–108. https://doi.org/10.1007/978-3-030-00737-9_7 PMID: 30484155.

48. Zhao Y, Long MJC, Wang Y, Zhang S, Aye Y. Ube2V2 Is a Rosetta Stone Bridging Redox and Ubiquitin Codes, Coordinating DNA Damage Responses. ACS central science. 2018; 4(2):246–59. https://doi.org/10.1021/acscentsci.7b00556 PMID: 29532025

49. Miller C, Saada A, Shaul N, Shabtai N, Ben-Shalom E, Shaag A, et al. Defective mitochondrial translation caused by a ribosomal protein (MRPS16) mutation. Annals of neurology. 2004; 56(5):734–8. https://doi.org/10.1002/ana.20282 PMID: 15505824.

50. Alberg AJ, Shopland DR, Cummings KM. The 2014 Surgeon General's report: commemorating the 50th Anniversary of the 1964 Report of the Advisory Committee to the US Surgeon General and updating the evidence on the health consequences of cigarette smoking. American journal of epidemiology. 2014; 179(4):403–12. https://doi.org/10.1093/aje/kwt335 PMID: 24436362

51. Petzold AM, Balciunas D, Sivasubbu S, Clark KJ, Bedell VM, Westcot SE, et al. Nicotine response genetics in the zebrafish. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(44):18662–7. https://doi.org/10.1073/pnas.0908247106 PMID: 19858493

52. Oncken C, Allen S, Litt M, Kenny A, Lando H, Allen A, et al. Exercise for Smoking Cessation in Post-menopausal Women: A Randomized, Controlled Trial. Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco. 2020; 22(9):1587–95. https://doi.org/10.1093/ntr/ntz176 PMID: 31536112

53. Marom-Haham L, Shulman A. Cigarette smoking and hormones. Current opinion in obstetrics & gynecology. 2016; 28(4):230–5. https://doi.org/10.1097/GCO.0000000000000283 PMID: 27285958.

54. Yoon V, Maalouf NM, Sakhaee K. The effects of smoking on bone metabolism. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2012; 23(8):2081–92. https://doi.org/10.1007/s00198-012-1940-y PMID: 22349964.

55. Saad FA. Novel insights into the complex architecture of osteoporosis molecular genetics. Annals of the New York Academy of Sciences. 2020; 1462(1):37–52. https://doi.org/10.1111/nyas.14231 PMID: 31556133.

56. Tang N, Zhao H, Zhang H, Dong Y. Effect of autophagy gene DRAM on proliferation, cell cycle, apoptosis, and autophagy of osteoblast in osteoporosis rats. Journal of cellular physiology. 2019; 234 (4):5023–32. https://doi.org/10.1002/jcp.27304 PMID: 30203495.

57. Skerrett-Byrne DA, Bromfield EG, Murray HC, Jamaluddin MFB, Jarnicki AG, Fricker M, et al. Time-resolved proteomic profiling of cigarette smoke-induced experimental chronic obstructive pulmonary disease. Respirology (Carlton, Vic). 2021. https://doi.org/10.1111/resp.14111 PMID: 34224176.

58. Lee J, Sandford AJ, Connett JE, Yan J, Mui T, Li Y, et al. The relationship between telomere length and mortality in chronic obstructive pulmonary disease (COPD). PloS one. 2012; 7(4):e35567. https://doi.org/10.1371/journal.pone.0035567 PMID: 22558169

59. Albrecht E, Sillanpää E, Karrasch S, Alves AC, Codd V, Hovatta I, et al. Telomere length in circulating leukocytes is associated with lung function and disease. The European respiratory journal. 2014; 43 (4):983–92. https://doi.org/10.1183/09031936.00046213 PMID: 24311771.

60. Shetty S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. Molecular and cellular biochemistry. 2005; 272(1–2):107–18. https://doi.org/10.1007/s11010-005-7644-2 PMID: 16010978.

61. Deng Z, Zhou J, Han X, Li X. TCEB2 confers resistance to VEGF-targeted therapy in ovarian cancer. Oncology reports. 2016; 35(1):359–65. https://doi.org/10.3892/or.2015.4388 PMID: 26531153.

62. Schipani E, Maes C, Carmeliet G, Semenza GL. Regulation of osteogenesis-angiogenesis coupling by HIFs and VEGF. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research. 2009; 24(8):1347–53. https://doi.org/10.1359/jbmr.090602 PMID: 19558314

63. Zhu S, Yao F, Qiu H, Zhang G, Xu H, Xu J. Coupling factors and exosomal packaging microRNAs involved in the regulation of bone remodelling. Biological reviews of the Cambridge Philosophical Society. 2018; 93(1):469–80. https://doi.org/10.1111/brv.12353 PMID: 28795526.

64. Jang ER, Jang H, Shi P, Popa G, Jeoung M, Galperin E. Spatial control of Shoc2-scaffold-mediated ERK1/2 signaling requires remodeling activity of the ATPase PSMC5. Journal of cell science. 2015; 128 (23):4428–41. https://doi.org/10.1242/jcs.177543 PMID: 26519477

65. Sinha KM, Zhou X. Genetic and molecular control of osterix in skeletal formation. Journal of cellular biochemistry. 2013; 114(5):975–84. https://doi.org/10.1002/jcb.24439 PMID: 23225263

66. Prasadam I, Friis T, Shi W, van Gennip S, Crawford R, Xiao Y. Osteoarthritic cartilage chondrocytes alter subchondral bone osteoblast differentiation via MAPK signalling pathway involving ERK1/2. Bone. 2010; 46(1):226–35. https://doi.org/10.1016/j.bone.2009.10.014 PMID: 19853676.

67. Broer L, Ikram MA, Schuur M, DeStefano AL, Bis JC, Liu F, et al. Association of HSP70 and its co-chaperones with Alzheimer's disease. Journal of Alzheimer's disease: JAD. 2011; 25(1):93–102. https://doi.org/10.3233/JAD-2011-101560 PMID: 21403392

68. Xing S, Wang Y, Hu K, Wang F, Sun T, Li Q. WGCNA reveals key gene modules regulated by the combined treatment of colon cancer with PHY906 and CPT11. Bioscience reports. 2020; 40(9). https://doi.org/10.1042/BSR20200935 PMID: 32812032

69. Kim K, Park S, Choi H, Kim HJ, Kwon YR, Ryu D, et al. Gene expression signatures associated with sensitivity to azacitidine in myelodysplastic syndromes. Scientific reports. 2020; 10(1):19555. https://doi.org/10.1038/s41598-020-76510-7 PMID: 33177628

70. Chang DC, Piaggi P, Hanson RL, Knowler WC, Bogardus C, Krakoff J. Autoantibodies against PFDN2 are associated with an increased risk of type 2 diabetes: A case-control study. Diabetes/metabolism research and reviews. 2017; 33(8). https://doi.org/10.1002/dmrr.2922 PMID: 28731290

71. Hamzeh O, Alkhateeb A, Zheng JZ, Kandalam S, Leung C, Atikukke G, et al. A Hierarchical Machine Learning Model to Discover Gleason Grade-Specific Biomarkers in Prostate Cancer. Diagnostics (Basel, Switzerland). 2019; 9(4). https://doi.org/10.3390/diagnostics9040219 PMID: 31835700

72. Walline HM, Komarck CM, McHugh JB, Bellile EL, Brenner JC, Prince ME, et al. Genomic Integration of High-Risk HPV Alters Gene Expression in Oropharyngeal Squamous Cell Carcinoma. Molecular cancer research: MCR. 2016; 14(10):941–52. https://doi.org/10.1158/1541-7786.MCR-16-0105 PMID: 27422711

73. Santarpia L, Iwamoto T, Di Leo A, Hayashi N, Bottai G, Stampfer M, et al. DNA repair gene patterns as prognostic and predictive factors in molecular breast cancer subtypes. The oncologist. 2013; 18 (10):1063–73. https://doi.org/10.1634/theoncologist.2013-0163 PMID: 24072219

74. Hua ZD, Liu XB, Sheng JH, Li C, Li P, Cai XQ, et al. UBE2V2 Positively Correlates With PD-L1 Expression and Confers Poor Patient Survival in Lung Adenocarcinoma. Applied immunohistochemistry & molecular morphology: AIMM. 2021. https://doi.org/10.1097/PAI.0000000000000928 PMID: 33734107.

75. Lee JJ, Park S, Park H, Kim S, Lee J, Lee J, et al. Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. Cell. 2019; 177(7):1842–57.e21. https://doi.org/10.1016/j.cell.2019.05.013 PMID: 31155235.

76. Wang Z, Li J, Long X, Jiao L, Zhou M, Wu K. MRPS16 facilitates tumor progression via the PI3K/AKT/Snail signaling axis. Journal of Cancer. 2020; 11(8):2032–43. https://doi.org/10.7150/jca.39671 PMID: 32127931

77. Wang H, Zhao W, Tian QJ, Xin L, Cui M, Li YK. Effect of lncRNA AK023948 on rats with postmenopausal osteoporosis via PI3K/AKT signaling pathway. European review for medical and pharmacological sciences. 2020; 24(5):2181–8. https://doi.org/10.26355/eurrev_202003_20483 PMID: 32196569.

78. Zhang Y, Cao X, Li P, Fan Y, Zhang L, Li W, et al. PSMC6 promotes osteoblast apoptosis through inhibiting PI3K/AKT signaling pathway activation in ovariectomy-induced osteoporosis mouse model. Journal of cellular physiology. 2020; 235(7–8):5511–24. https://doi.org/10.1002/jcp.29261 PMID: 32017075.