

RESEARCH

Open Access



# Identification of shared gene signatures and pathways for diagnosing osteoporosis with sarcopenia through integrated bioinformatics analysis and machine learning

Xiaoli Zhou<sup>1,2†</sup>, Lina Zhao<sup>4,5†</sup>, Zepei Zhang<sup>1†</sup>, Yang Chen<sup>1</sup>, Guangdong Chen<sup>3</sup>, Jun Miao<sup>1\*</sup> and Xiaohui Li<sup>1,6\*</sup>

## Abstract

**Background** Prior studies have suggested a potential relationship between osteoporosis and sarcopenia, both of which can present symptoms of compromised mobility. Additionally, fractures among the elderly are often considered a common outcome of both conditions. There is a strong correlation between fractures in the elderly population, decreased muscle mass, weakened muscle strength, heightened risk of falls, and diminished bone density. This study aimed to pinpoint crucial diagnostic candidate genes for osteoporosis patients with concomitant sarcopenia.

**Methods** Two osteoporosis datasets and one sarcopenia dataset were obtained from the Gene Expression Omnibus (GEO). Differential expression genes (DEGs) and module genes were identified using Limma and Weighted Gene Co-expression Network Analysis (WGCNA), followed by functional enrichment analysis, construction of protein–protein interaction (PPI) networks, and application of a machine learning algorithm (least absolute shrinkage and selection operator (LASSO) regression) to determine candidate hub genes for diagnosing osteoporosis combined with sarcopenia. Receiver operating characteristic (ROC) curves and column line plots were generated.

**Results** The merged osteoporosis dataset comprised 2067 DEGs, with 424 module genes filtered in sarcopenia. The intersection of DEGs between osteoporosis and sarcopenia module genes consisted of 60 genes, primarily enriched in viral infection. Through construction of the PPI network, 30 node genes were filtered, and after machine learning, 7 candidate hub genes were selected for column line plot construction and diagnostic value assessment. Both the column line plots and all 7 candidate hub genes exhibited high diagnostic value (area under the curve ranging from 1.00 to 0.93).

**Conclusion** We identified 7 candidate hub genes (PDP1, ALS2CL, VLDLR, PLEKHA6, PPP1CB, MOSPD2, METTL9) and constructed column line plots for osteoporosis combined with sarcopenia. This study provides reference for potential peripheral blood diagnostic candidate genes for sarcopenia in osteoporosis patients.

**Keywords** Sarcopenia, Osteoporosis, Hub genes, Machine learning, Bioinformatics

<sup>†</sup> Xiaoli Zhou, Lina Zhao and Zepei Zhang contributed equally.

\*Correspondence:

Jun Miao  
mj6688@tju.edu.cn  
Xiaohui Li  
lxhhd@sina.com

Full list of author information is available at the end of the article



## Introduction

Sarcopenia, a progressive skeletal muscle disorder, is linked to increased risks of falls, fractures, physical disability, and mortality [1]. While commonly observed in older adults, sarcopenia can also manifest earlier in life. Rooted in age-related muscle changes, it significantly reduces muscle strength and mass, contributing to heightened fall risks and impaired daily activities. This often leads to disability, loss of independence, and even death. The substantial impact of sarcopenia on morbidity, mortality, and healthcare expenditure has spurred significant research and policy discussions, underscoring its critical importance. Decreased muscle strength, crucial for mobility, significantly increases fall prevalence among the elderly. This condition correlates closely with self-reported physical disability, transcending factors such as ethnicity, age, morbidity, obesity, income, or health behaviors [2]. Age-related decline in muscle strength not only diminishes functional capacity but also exacerbates disability, mortality, and other adverse health outcomes [3]. With the aging population, sarcopenia-related morbidity is expected to pose a significant healthcare challenge. Management strategies for sarcopenia include non-pharmacological interventions like resistance exercise and proper nutrition, notably protein intake and vitamin D supplementation. Resistance exercise, in particular, is a standard non-pharmacological treatment, supported by substantial evidence. While various pharmacological agents have shown efficacy, future research should focus on elucidating biological pathways, refining diagnostics, and developing superior treatment methods [4].

Osteoporosis and related fractures are prevalent among older adults, posing substantial morbidity and mortality risks. Bisphosphonates are the primary therapy, with additional options like denosumab, teriparatide, and selective estrogen receptor modulators available. Early identification and intervention for osteoporosis are crucial for mitigating its effects [5]. The coexistence of osteoporosis and sarcopenia, termed 'osteosarcopenia' [5], poses a dual challenge. Interactions between muscles and bones at various levels may contribute to osteosarcopenia's pathophysiology [6]. Understanding shared genes between these systems could offer novel treatment insights.

In this study, instrumental variables at the genome-wide significance level were utilized to assess the bi-directional causality between sarcopenia and osteoporosis. The results suggest a potential mutual influence between the two conditions. The study also highlights the frequency of osteosarcopenia and its association with increased fracture risks. Standardized classification of sarcopenia is crucial for accurately assessing its

relationship and consequences [7]. Muscles and bones, originating from mesodermal and ectodermal mesenchymal stem cells, share close anatomical proximity, facilitating mechanical and chemical signal exchange. Identifying shared crosstalk genes could offer novel prevention and treatment avenues [8]. After identifying shared genetic markers, validate the functionality of these genes using cellular or animal models, and determine their potential mechanisms in disease progression. Develop molecular diagnostic tests based on blood or tissue samples for screening the risk of muscle atrophy and osteoporosis. Develop gene therapy approaches or novel drugs targeting the expression or functionality of these genes.

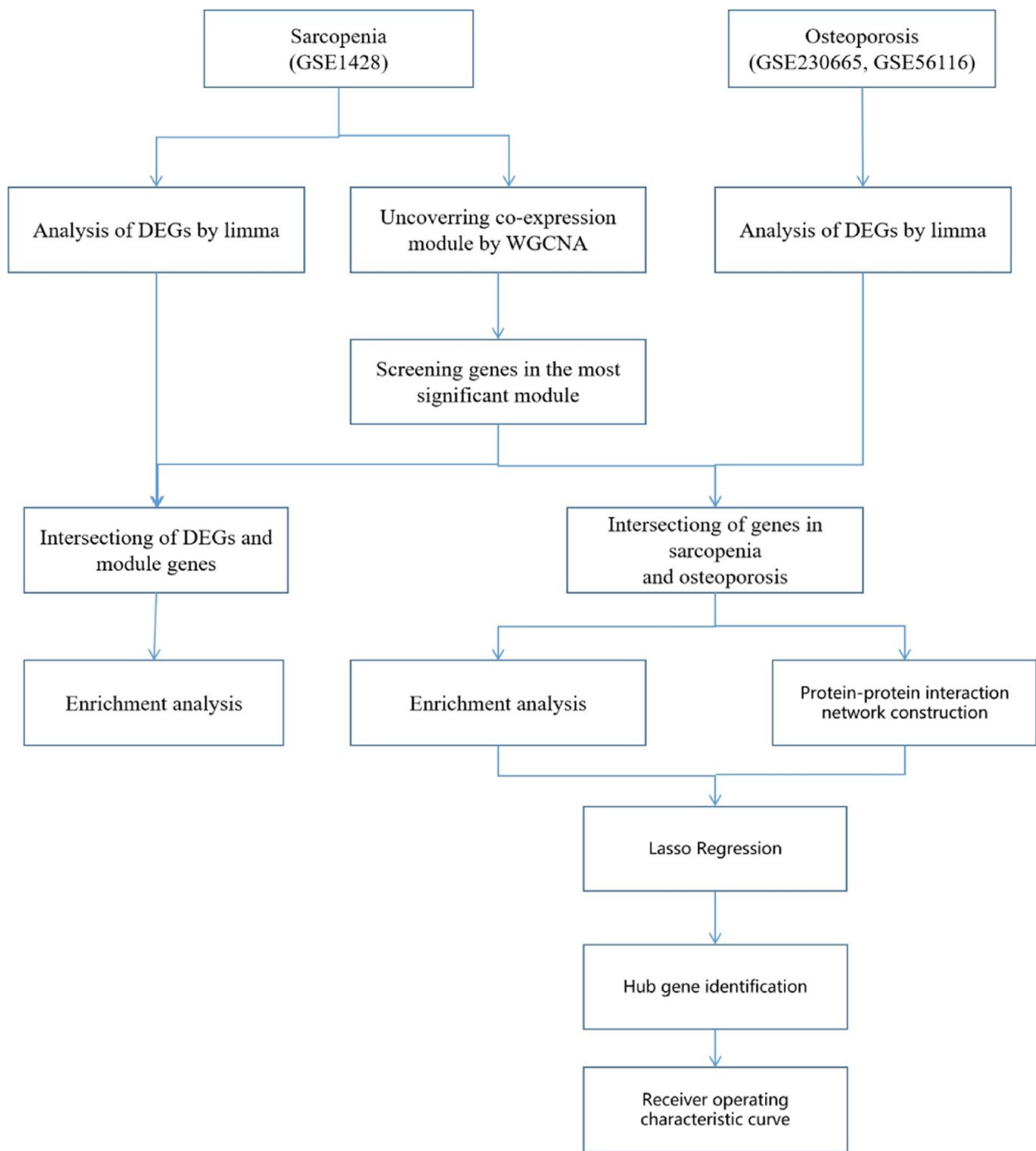
## Material and methods

### Data collection

The datasets GSE1428 and GSE230665, as well as GSE56116, were curated from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) [9]. For the microarray analysis of the GSE1428 [10] dataset, the GPL96 platform (Affymetrix Human Genome U133A Array) was utilized. GSE230665 [11] employed the GPL10332 platform (Agilent-026652 Whole Human Genome Microarray 4×44 K v2, Feature Number version). GSE56116 [11] utilized the GPL4133 platform (Agilent-014850 Whole Human Genome Microarray 4×44 K G4112F, Feature Number version). The GSE1428 dataset presents transcriptional responses related to sarcopenia, as provided by Giresi et al. On the other hand, the datasets GSE230665 and GSE56116, focusing on osteoporosis, were contributed by Ge, Li, and their respective collaborators. The GSE1428 dataset showcases the transcriptional responses associated with sarcopenia, while GSE230665 and GSE56116 pertain to datasets on osteoporosis (Fig. 1).

### Identification of DEGs between sarcopenia and osteoporosis

For the two original osteoporosis datasets, empirical Bayes methods [12] were applied to eliminate batch effects. The merged osteoporosis dataset and the sarcopenia dataset underwent the extraction of expression matrices, with the exclusion of genes and samples featuring missing values exceeding 50%. Subsequently, missing values were imputed using the "impute.knn" function from the R package "impute," setting the Number of neighbors to 10 for data completion. Furthermore, a log<sub>2</sub> transformation was applied to the data. In cases where multiple probes identified the same gene, the average expression was calculated. Finally, utilizing the Limma package, criteria of |log<sub>2</sub> Fold change (FC)| > 1.5 and  $P < 0.05$  were set as the standards for identifying DEGs. The definition of this threshold range refers to the study by Liu et al. [13].



**Fig. 1** Workflow of the whole study

**Weighted gene co-expression network analysis and module gene selection**

Exploring gene–gene correlations using systems biology strategy WGCNA [14]. Based on gene expression profiles, we computed the median absolute deviation for each gene and removed the bottom 50% of genes with

the smallest median absolute deviation. We utilized the R software package WGCNA’s goodSamplesGenes method to eliminate outlier genes and samples. Subsequently, we employed WGCNA to construct a scale-free co-expression network. Initially, Pearson’s correlation matrices and the average linkage method were applied to all pairwise

genes. Then, a weighted adjacency matrix was created using a power function  $A_{mn} = |C_{mn}|^\beta$  (where  $C_{mn}$  represents the Pearson's correlation between gene  $m$  and gene  $n$ , and  $A_{mn}$  denotes the adjacency between gene  $m$  and gene  $n$ ). The parameter  $\beta$  was chosen to be 12 for soft-thresholding, emphasizing strong correlations and penalizing weak ones. The adjacency matrix was transformed into a topological overlap matrix, measuring the network connectivity of a gene, defined as the sum of its adjacencies with all other genes for network gene ration, and the corresponding dissimilarity (1-TOM) was computed. To classify genes with similar expression profiles into gene modules, average linkage hierarchical clustering was performed based on the TOM-based dissimilarity measure, with a minimum module size of 100 for the gene dendrogram. We set the sensitivity to 3. To further analyze the modules, we calculated the dissimilarity of module eigen genes, selected a cut line for the module dendrogram, and merged some modules. Additionally, we merged modules with a distance less than 0.25, resulting in the identification of 2 co-expression modules.

#### Functional enrichment analysis

The Gene Ontology (GO) [15] system provides structured and computable information about the functions of genes and gene products. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [16] is a widely used database for gene system research. Using gene annotations from the R package *org.Hs.eg.db* (version 3.1.0), and obtaining the latest KEGG Pathway gene annotations from the KEGG rest API (<https://www.kegg.jp/kegg/rest/keggapi.html>) as background, genes were mapped to the background set. Functional enrichment analysis was performed using the R package *clusterProfiler* (version 3.14.3) to obtain results of gene set enrichment. The minimum gene set was set to 5, and the maximum gene set was set to 5000. A P value of  $<0.05$  and a false discovery rate of  $<0.1$  were considered statistically significant. Two rounds of GO and KEGG analyses were conducted based on the intersection of DEGs in sarcopenia and the most significant module genes, as well as the intersection of DEGs in osteoporosis and the most significant module genes in sarcopenia.

#### Construction of protein–protein interaction network

To explore the interactions between protein-coding genes, we utilized the String database [17] (version 11.5; [www.string-db.org](http://www.string-db.org)), with a minimum interaction score set to 0.400. The obtained network from String was further modified using Cytoscape software. All interacting genes within the protein–protein interaction (PPI) network were selected for subsequent analysis.

#### Machine learning

To further screen candidate genes for diagnosing sarcopenia and osteoporosis, a machine learning algorithm was employed. LASSO [18] (Least Absolute Shrinkage and Selection Operator) is a regression method used for variable selection to improve prediction accuracy. It is also a regularization technique that enhances the predictive accuracy and interpretability of statistical models. Utilizing the R package *glmnet*, gene expression data was integrated, and regression analysis was performed using the lasso-cox method. A threefold cross-validation was set up to obtain the optimal model. By dividing the dataset into 3 parts and rotating 2 of them for model training while keeping 1 for testing, we iterated through this process. In each iteration, we evaluated the performance of the model under different  $\lambda$  values, primarily by observing the prediction errors of the model. Finally, we selected the  $\lambda$  value that minimized the cross-validation error as the optimal  $\lambda$ . The Lambda value was set to 0.0639847346226388. The genes obtained from this analysis were identified as candidate hub genes for diagnosing sarcopenia and osteoporosis.

#### Construction of column line plots and ROC curve evaluation

Construction of column line plots holds certain value in diagnosing clinical sarcopenia and osteoporosis. Using the candidate genes, column line plots were constructed using the R package *pROC* (version 1.17.0.1). The "Score" represents the score of the candidate genes, while "Total Score" represents the sum of scores for all the aforementioned genes. ROC curves were established to evaluate the diagnostic value of the candidate genes and column line plots for sarcopenia and osteoporosis.

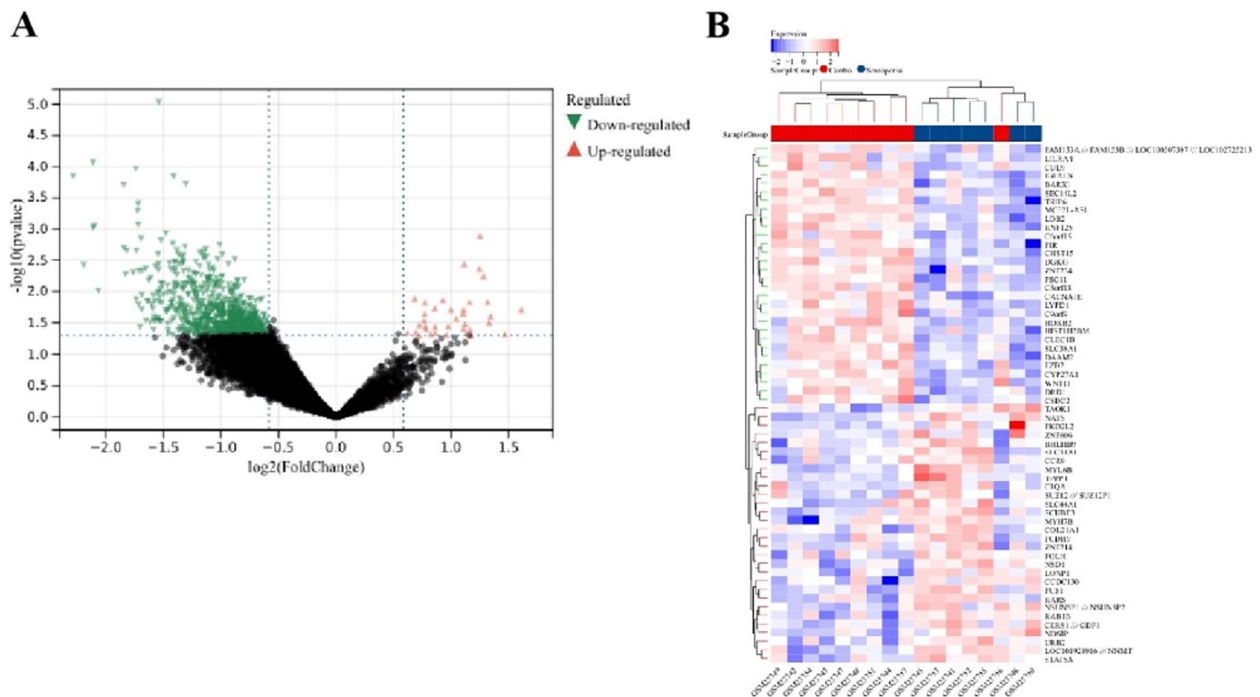
## Results

#### Identification of DEGs

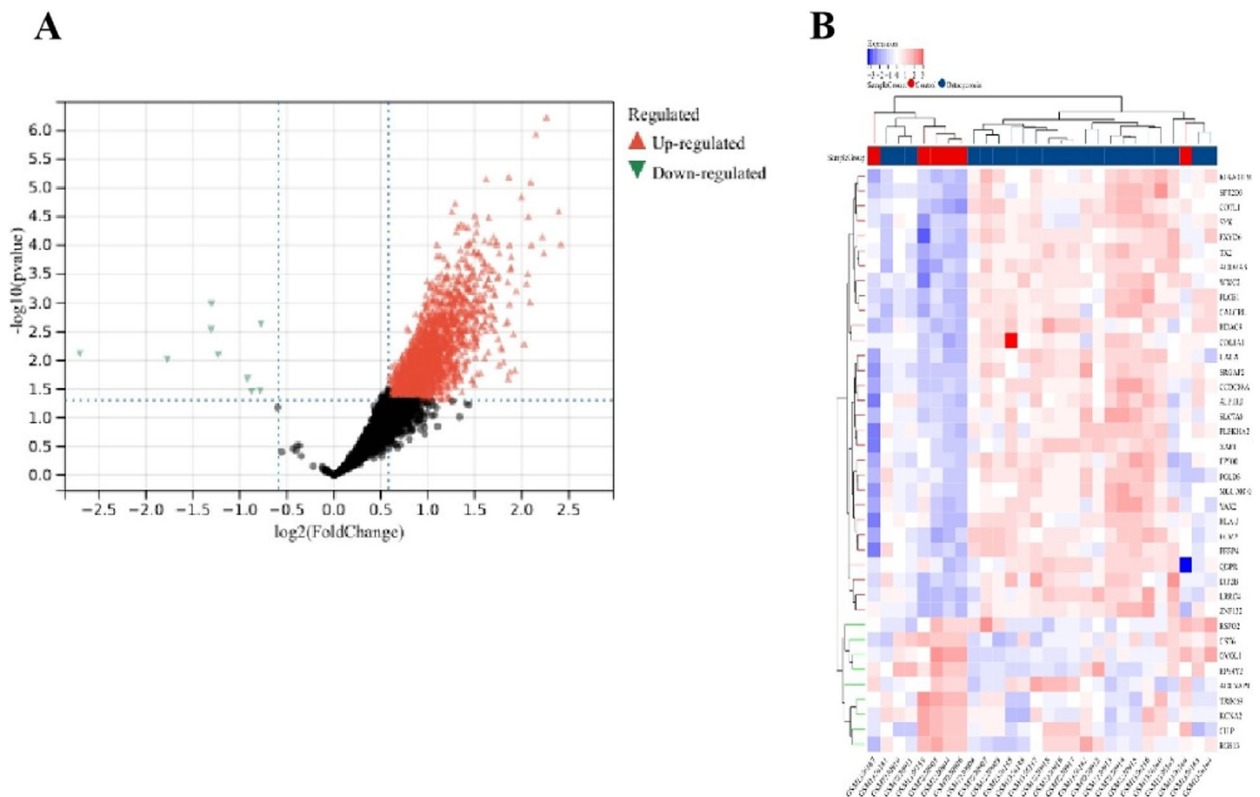
Identification of DEGs using Limma method revealed a total of 821 DEGs (337 upregulated, 484 downregulated) in the sarcopenia dataset. The heatmap and volcano plot of sarcopenia DEGs are shown in Fig. 2A-B. In the combined osteoporosis dataset, a total of 2067 DEGs were identified, with 2059 upregulated and 9 downregulated genes. The heatmap and volcano plot of osteoporosis DEGs are illustrated in Fig. 3A-B.

#### WGCNA analysis and identification of key modules

We selected  $\beta = 12$  (scale-free  $R^2 = 0.86$ ) as the "soft" threshold based on scale independence and average connectivity (Fig. 4A-B). The dendrogram (Fig. 4C) depicts the clustering of sarcopenia and control samples. Based on this, two gene co-expression modules were generated,

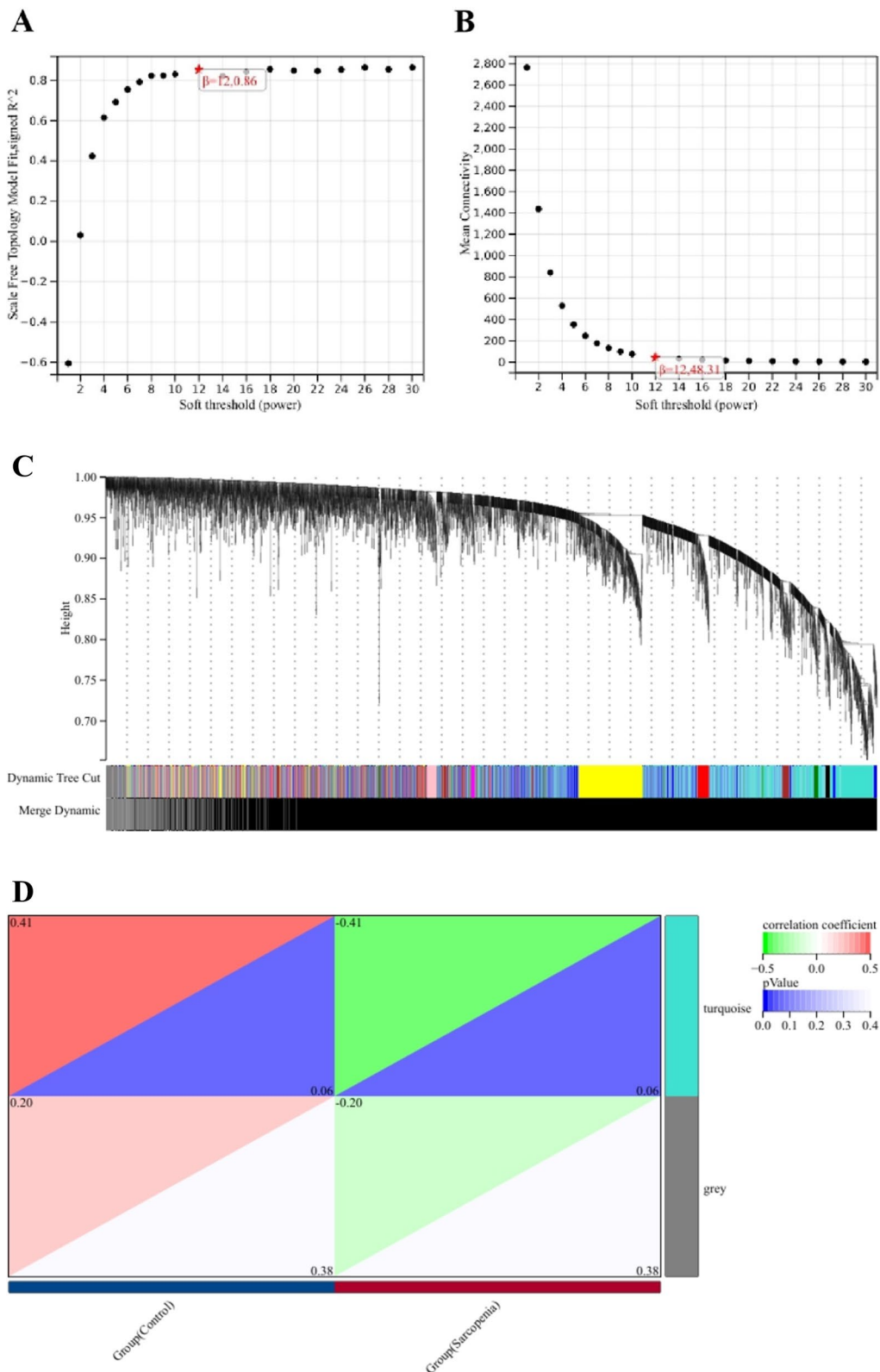


**Fig. 2** Heatmap and volcano plot for the DEGs identified from the sarcopenia dataset



**Fig. 3** Heatmap and volcano plot for the DEGs identified from the osteoporosis dataset





**Fig. 4** WGCNA of DEGs. **A-B** Estimation of the soft thresholding value for a scale-free co-expression network. **C** Cluster dendrogram of all DEGs. **D** Heatmap showing the correlation between modules and sarcopenia. The turquoise module is found to be significantly correlated with sarcopenia. The numbers in the top and bottom brackets represent the correlation coefficient and  $p$ -value, respectively.

as indicated by different colors in Fig. 4D. Among them, the turquoise module (424 genes) exhibited the highest correlation with sarcopenia (correlation coefficient = -0.41,  $P=0.06$ ) and was considered the key module for subsequent analysis.

### Functional enrichment analysis of sarcopenia

To assess whether the dataset GSE1428 reliably reflects the pathogenesis of sarcopenia, we further conducted functional enrichment analysis based on the intersection of Limma and WGCNA module genes. The intersection of 424 DEGs from the turquoise module with 821 genes yielded 16 common genes (Fig. 5A). KEGG analysis revealed that common genes were primarily enriched in "Metabolic pathways" and "Carbon metabolism" (Fig. 5B). GO analysis indicated that common genes were predominantly enriched in biological process (BP) terms, including "coenzyme metabolic process" and "purine ribonucleotide metabolic process" (Fig. 5C). Regarding cellular component (CC) ontology, CGs were mainly located in "mitochondrion," "mitochondrial part," and "mitochondrial matrix" (Fig. 5D). Molecular function (MF) analysis showed that "Ras guanyl-nucleotide exchange factor activity," "isocitrate dehydrogenase (NAD<sup>+</sup>) activity," and "L-aspartate transmembrane transporter activity" were the most significant terms within common genes (Fig. 5E).

### Enrichment analysis and node gene identification for osteoporosis and sarcopenia based on PPI networks

To further explore whether key genes associated with sarcopenia are also related to the pathogenesis of osteoporosis, we visualized the intersection of DEGs in osteoporosis and module genes in sarcopenia through a Venn diagram, identifying 60 genes (Fig. 6A). KEGG enrichment analysis revealed that these 60 genes were mainly enriched in "Human papillomavirus infection," "mTOR signaling pathway," and "Kaposi sarcoma-associated herpesvirus infection" (Fig. 6B). GO analysis showed that these genes were enriched in "Kaposi sarcoma-associated herpesvirus infection," "organonitrogen compound biosynthetic process," and "cellular amide metabolic process" in BP; "cytosol," "cytosol," and "nuclear chromosome" in CC; and "transcription coregulator activity," "transcription coregulator activity," and "ubiquitin-like protein ligase activity" in MF (Fig. 6C-E).

After confirming the filtered genes, we constructed a PPI network to identify interacting node genes for subsequent machine learning filtering. Figure 6F displays the PPI network, where 30 genes can interact with each other. These genes are sorted by node degree in Fig. 6G

### Identifying candidate hub genes through machine learning

The LASSO regression machine learning algorithm was applied to select candidate genes for column line plot construction and diagnostic value assessment. From Fig. 7A-B, it can be observed that the LASSO regression algorithm identified 7 potential candidate biomarkers for final validation.

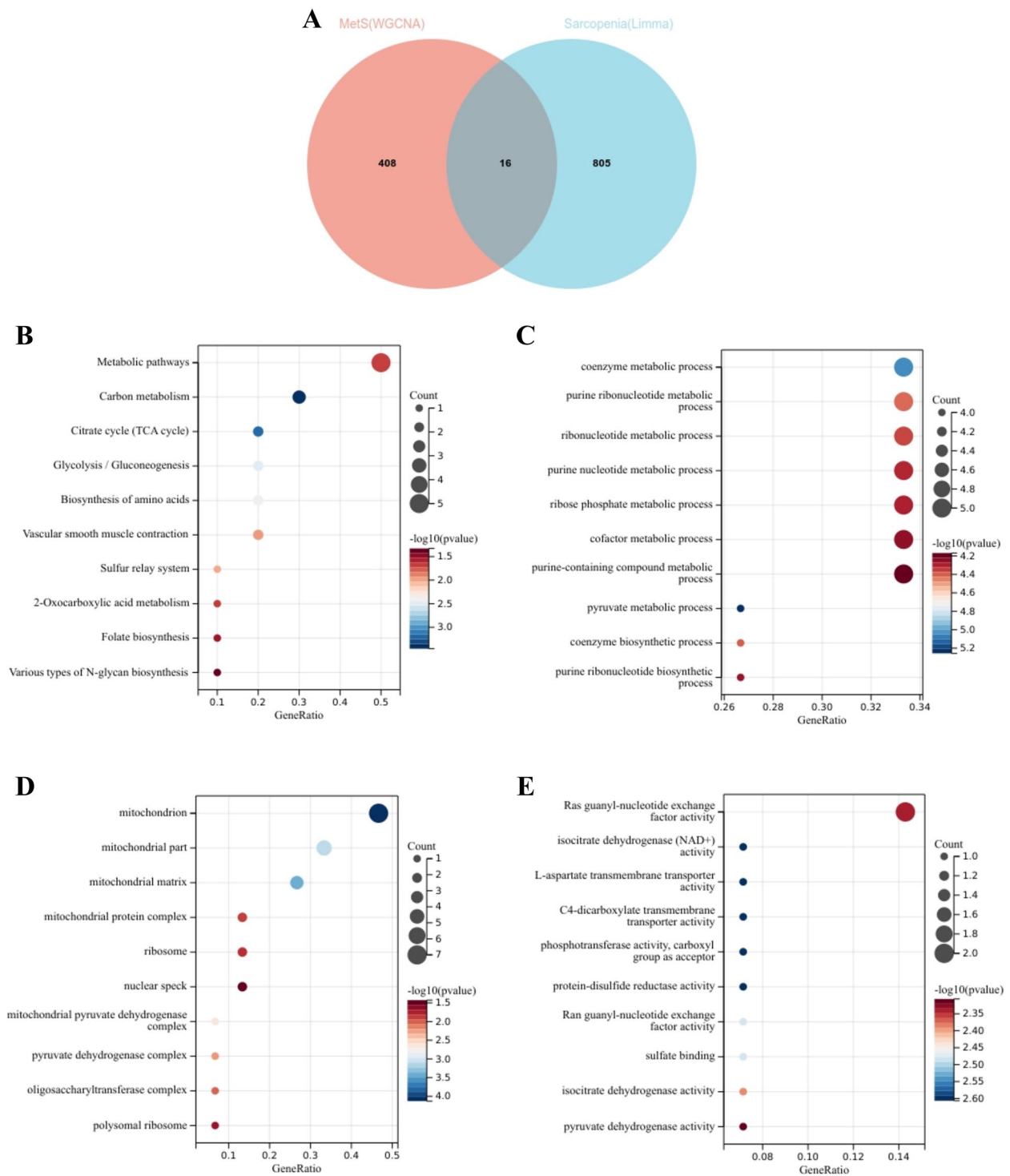
### Diagnostic Value Evaluation

Based on the 7 candidate hub genes, a column line plot was constructed (Fig. 8A), and ROC curves were established to evaluate the diagnostic specificity and sensitivity of each gene and the column line plot. The area under the curve (AUC) and its 95% confidence interval (CI) were calculated for each item. The results are as follows: PDP1 (AUC 0.96, CI 1.00~0.88), ALS2CL (AUC 0.80, CI 1.00~0.58), VLDLR (AUC 0.83, CI 1.00~0.5), PLEKHA6 (AUC 0.93, CI 1.00~0.83), PPP1CB (AUC 0.82, CI 1.00~0.63), MOSPD2 (AUC 0.73, CI 1.00~0.44), METTL9 (AUC 0.82, CI 1.00~0.58), and the column line plot (AUC 0.98, CI 1.00~0.93) (Fig. 8B-I). All candidate genes exhibited high diagnostic value for sarcopenia combined with osteoporosis, with the column line plot demonstrating the highest diagnostic value.

### Discussion

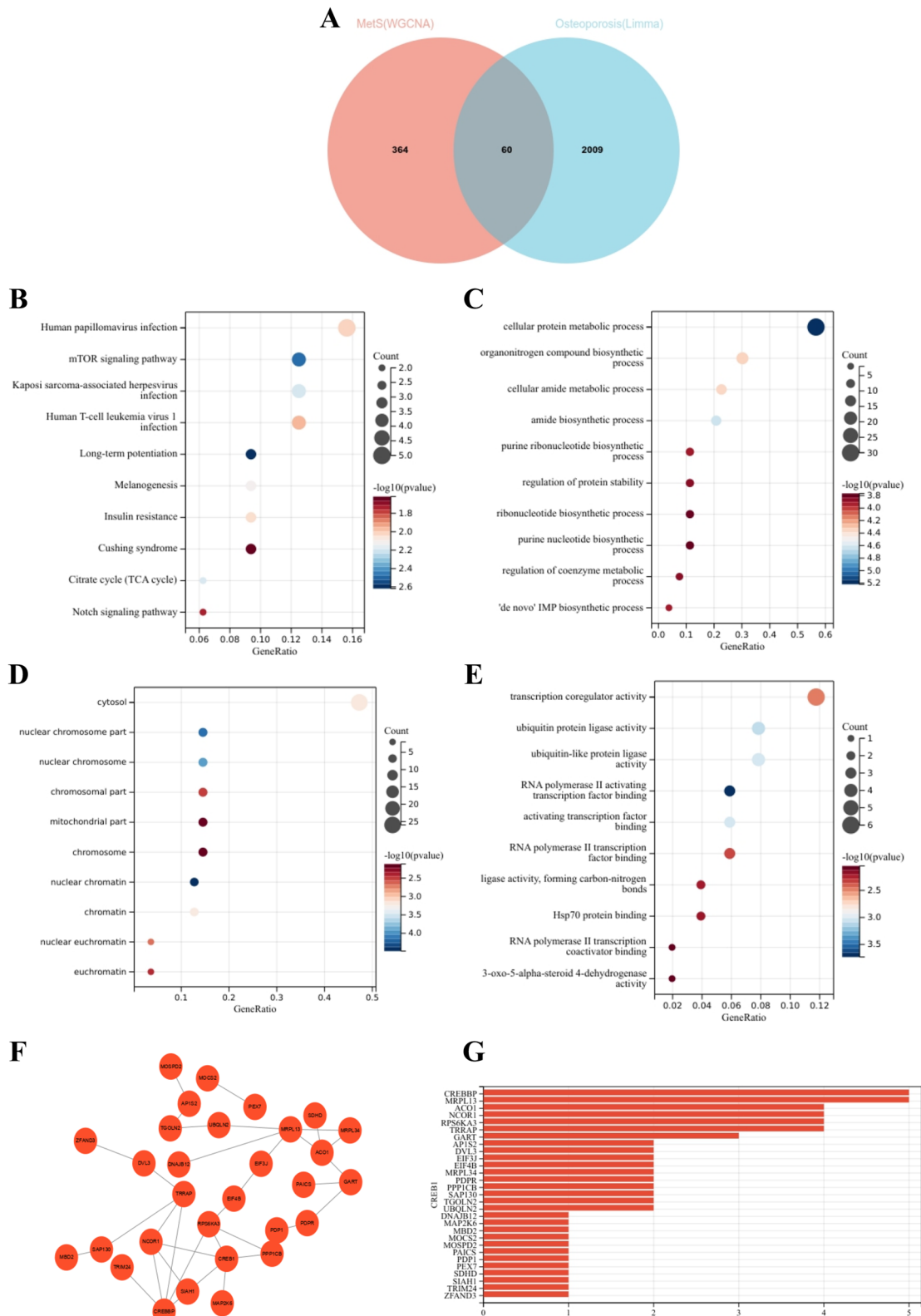
Osteosarcopenia, characterized by the coexistence of osteopenia/osteoporosis and sarcopenia, has emerged as a significant health concern, imposing a substantial global health burden. According to the World Health Organization, osteopenia and osteoporosis are defined by T scores equal to or less than -1 and -2.5 standard deviations, respectively, below the peak bone mass of a young, healthy cohort or in the presence of a minimal-trauma fracture. This skeletal condition results in the deterioration of bone microarchitecture and compromises bone strength [19]. Conversely, sarcopenia is identified by cut-off values indicating low muscle mass, strength, and/or functional capacity [20]. Both osteosarcopenia and sarcopenia share common risk factors [21] and exhibit strong associations with frailty, falls, fractures, hospitalizations, and mortality [21–23], contributing to a significant increase in healthcare expenditure. The coexistence of these conditions underscores the intricate interplay between skeletal and muscular health and emphasizes the need for comprehensive approaches to address their shared impact on overall well-being.

In this study, we utilized a series of integrated bioinformatics analyses and machine learning methods to construct a nomogram and evaluate the diagnostic value of osteoporosis in sarcopenia patients.

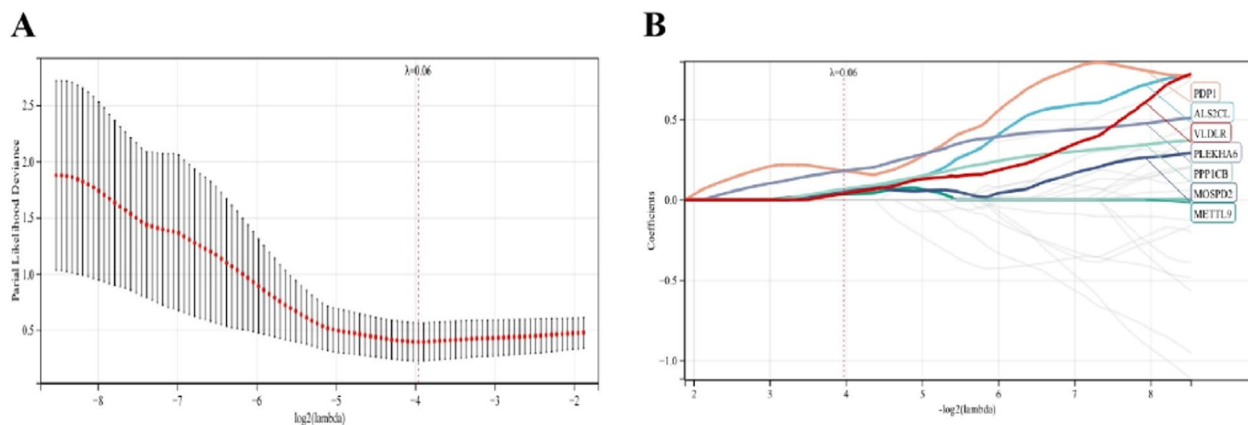


**Fig. 5** Enrichment analysis of the intersection of genes in sarcopenia. **A** Venn diagram shows that 16 genes are identified from the intersection of DEGs via Limma and green module genes via WGCNA. **B** KEGG pathway analysis of the intersection of genes. Different colors represent various significant pathways and related enriched genes. **C-E** GO analysis of the intersection of genes, including biological process, cellular component, and molecular function, respectively. The y-axis represents different GO terms, the x-axis represents gene ratio enriched in relative GO terms, the circle size refers to gene numbers, and the color represents  $p$  value.





**Fig. 6** Enrichment analysis of common genes from osteoporosis with sarcopenia and the identification of node genes from PPI network. **A** Venn diagram shows that 60 common genes are identified from the intersection of genes in osteoporosis using Limma and sarcopenia using WGCNA. **B** KEGG analysis of 60 common genes. **C-E** GO analysis (biological process, cellular component, and molecular function) of 60 common genes. **F** PPI network reveals that 30 genes interact with each other. **G** The column shows the gene nodes of 26 genes in PPI network.



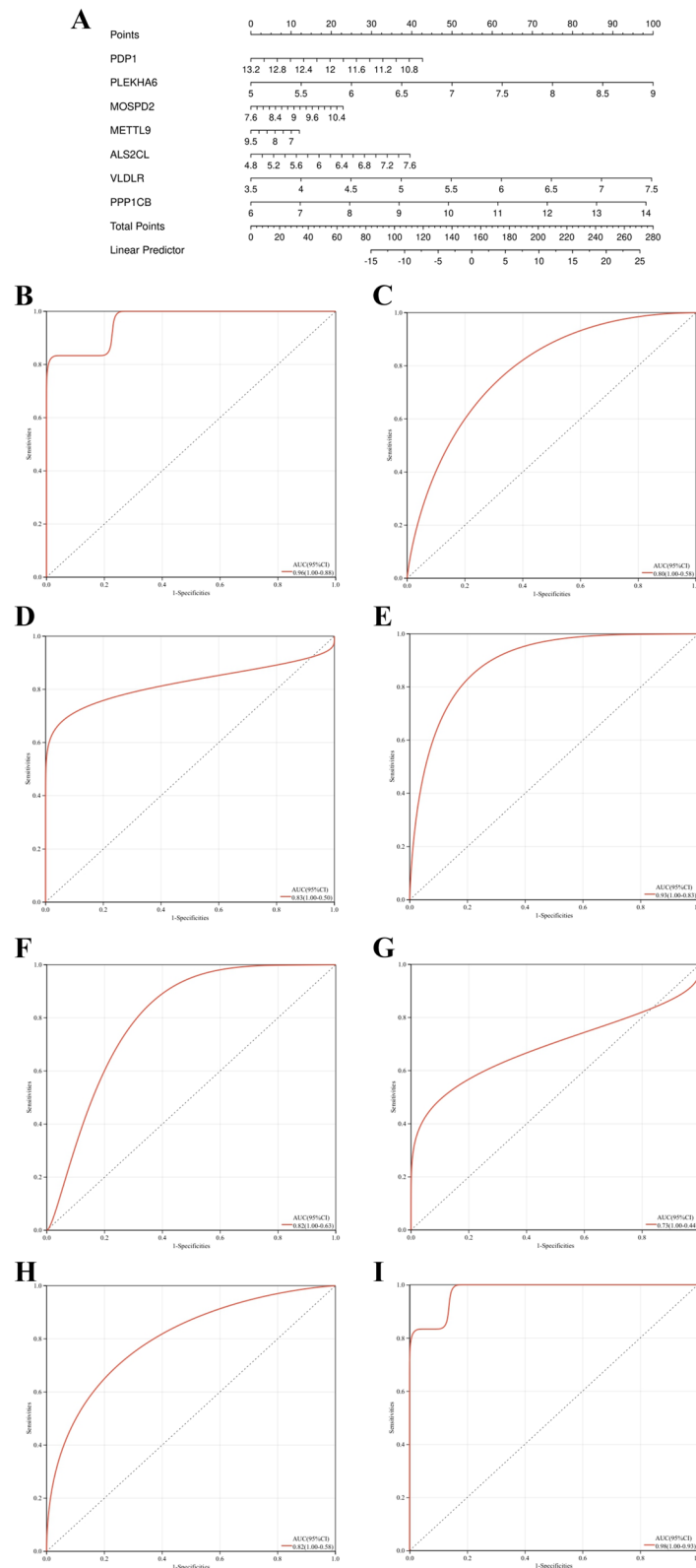
**Fig. 7** Machine learning in screening candidate diagnostic biomarkers for osteoporosis with sarcopenia. **A-B** Biomarkers screening in the Lasso model. The number of genes ( $n=7$ ) corresponding to the lowest point of the curve is the most suitable for osteoporosis with sarcopenia diagnosis.

A notable finding is the identification of 7 key candidate genes (PDP1, ALS2CL, VLDLR, PLEKHA6, PPP1CB, MOSPD2, and METTL9), and the development of a nomogram for diagnosing osteoporosis in sarcopenia patients.

The sarcopenia patient dataset used in this study all comes from peripheral blood samples. Therefore, we only need to collect peripheral blood samples from sarcopenia patients and evaluate the expression of the 7 identified immune-related genes to infer the probability of sarcopenia patients developing osteoporosis. This is an efficient and practical clinical approach. The use of peripheral blood testing in diagnosing various diseases is also widely accepted. Furthermore, although we confirmed that gene expression levels can serve as independent diagnostic markers, we plan to develop a more comprehensive diagnostic model by transforming them into scores and considering all 7 markers. The expression of each gene is quantified and converted into a score, with an increase in score indicating a higher linear prediction factor. When the linear prediction factor is high, we can conduct early monitoring and intervention in sarcopenia patients, which is more valuable for implementing osteoporosis diagnosis in sarcopenia. Using machine learning to identify pseudo-gene features for bone sarcoma prognosis, these four pseudo-gene features not only serve as promising indicators for predicting prognosis and survival rates but also represent potential markers for monitoring treatment regimens [24].

PDP1 encodes a protein that is one of the three components (E1, E2, and E3) of the large pyruvate dehydrogenase complex. PDP1 plays a crucial role in protein phosphorylation and has been implicated in various diseases [25, 26]. Research has shown that miR-18a-3p improves cartilage matrix remodeling and suppresses

inflammation in osteoarthritis by targeting PDP1 [27]. In pancreatic cancer, PDP1 promotes cancer proliferation and invasion by regulating the MAPK/mTOR signaling pathway [28]. Additionally, PDP1 is associated with osteosarcoma progression, patient prognosis, and chemosensitivity, making it a potential biomarker for osteosarcoma [29]. Given its role in multiple diseases, PDP1 is considered a potential diagnostic target for osteoporosis in sarcopenic patients. ALS2CL encodes a 108-kD protein with specific but relatively weak Rab5-GEF activity and strong Rab5-binding properties. Co-expression of ALS2CL and Rab5A in HeLa cells results in a unique tubulation phenotype of endosome compartments, indicating ALS2CL's involvement in modulating Rab5-mediated endosome dynamics [30]. VLDLR, or Very Low Density Lipoprotein Receptor, belongs to the low-density lipoprotein receptor family, with high expression levels in the brain, heart, skeletal muscle, and adipose tissue, while its expression in the liver is very low under physiological conditions. It plays a crucial role in controlling serum triglycerides and the development of non-alcoholic fatty liver disease. Previous studies have demonstrated the involvement of VLDLR in regulating the onset of various diseases. Research has found that homozygous loss-of-function mutations in VLDLR lead to dysequilibrium syndrome, a non-progressive cerebellar ataxia syndrome associated with intellectual disability [31]. PLEKHA6, along with other members of the WW-*PLEKHA* family, plays a role in the trafficking and retention of transmembrane proteins, including nectins, Tspan33, and the copper pump ATP7A, at cell-cell junctions and lateral membranes. Its C-terminal region and coiled-coil region promote its localization at adherens junctions of epithelial cells. This suggests that PLEKHA6 is involved in maintaining cell-cell adhesion and potentially regulates



**Fig. 8** 8 Nomogram construction and the diagnostic value evaluation. **A** The visible nomogram for diagnosing osteoporosis with sarcopenia. **B-I** The ROC curve of each candidate gene (PDP1, ALS2CL, VLDLR, PLEKHA6, PPP1CB, MOSPD2, METTL9) and nomogram show the significant osteoporosis with sarcopenia diagnostic value.

signaling pathways associated with adherens junctions [32]. PPP1CB, located on chromosome 2p23.2, encodes a subunit of PPP1 involved in various cellular functions, including glycogen metabolism, cell division, and muscle contraction [33–36]. Recent studies have identified PPP1CB as the myosin light chain phosphatase responsible for Ca<sup>2+</sup>-transient rise and enhanced cell shortening in cardiomyocytes [37]. MOSPD2, a member of the VAP family, facilitates contact between the endoplasmic reticulum and various cellular organelles [38]. Unlike other VAP family members, MOSPD2 contains an additional cytoplasmic domain called CRAL-TRIO, which may be involved in lipid transport [39]. Research suggests that MOSPD2 is a key regulator of inflammation-driven monocyte migration and a potential therapeutic target for CNS inflammatory diseases [40]. METTL9, a methyltransferase, plays a crucial role in histone methylation and is implicated as an oncogene in various cancers [41, 42]. Targeting METTL9 significantly inhibits the growth of hepatocellular carcinoma patient-derived xenografts [43] and correlates with increased metastatic activity in human gastric cancer [44].

### Limitation

In summary, this study identified candidate hub genes for diagnosing osteoporosis combined with sarcopenia using integrated bioinformatics and machine learning approaches. However, limitations include reliance on publicly available datasets, potential selection bias in gene identification, limited generalizability to diverse populations, cross-sectional data analysis, and the need for further experimental validation to elucidate the functional mechanisms of the identified genes. These findings provide a foundation for potential peripheral blood diagnostic markers but require additional validation and clinical translation for practical application in healthcare settings. Examine the expression levels of these candidate genes in clinical samples using techniques such as real-time quantitative PCR or immunohistochemistry. Compare the expression differences of these genes between osteoporosis patients and healthy controls, as well as between muscle atrophy patients and healthy controls. Use statistical methods to determine the presence of significant correlations and evaluate the feasibility of these genes as potential biomarkers. Investigate the functions of these genes through cellular or animal models, especially their effects on bone and muscle tissue. For example, the impact of these genes on bone and muscle development and maintenance can be studied through gene knockout or overexpression. Conduct clinical cohort studies to track the disease progression and treatment response of patients with osteoporosis accompanied by

muscle atrophy, and assess the potential of these candidate genes as predictive or prognostic markers. Carry out drug intervention trials to evaluate the efficacy of drug treatments targeting these genes for osteoporosis with muscle atrophy.

### Conclusion

Our study systematically identified seven candidate hub genes (PDP1, ALS2CL, VLDLR, PLEKHA6, PPP1CB, MOSPD2, and METTL9) through a combination of various bioinformatics analyses and machine learning algorithms, and provided a nomogram for diagnosing sarcopenia associated with osteoporosis. The research offers reference for potential peripheral blood diagnostic candidate genes for sarcopenia related to osteoporosis.

### Abbreviations

AUC	Area under the curve
BP	Biological processes
CC	Cellular components
CI	Confidence interval
DEGs	Differentially expressed genes
GEO	Gene Expression Omnibus
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
MF	Molecular functions
PDP1	Pyruvate dehydrogenase phosphatase catalytic subunit 1
PPI	Protein-protein interaction
ROC	Receiver operating characteristic
TOM	Topological overlap matrix
WGCNA	Weighted Gene Co-expression Network Analysis

### Acknowledgements

None.

### Authors' contributions

Study Design: Xiaoli Zhou, Guangdong Chen, Zepei Zhang, Jun Miao. Data Collection: Xiaoli Zhou, Yang Chen, Zhiyong Qian, Jun Miao. Statistical Analysis: Xiaoli Zhou, Guangdong Chen. Data Interpretation: Yang Chen, Lina Zhao, Xiaohui Li. Manuscript Preparation: Xiaoli Zhou, Guangdong Chen, Zepei Zhang, Jun Miao. Literature Search: Xiaoli Zhou, Guangdong Chen

### Funding

This work was supported by the Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-026A), the Hebei Province High-Level Talent Funding Project (C20221122), and the Key Project of Tianjin Natural Science Foundation (23JCZDJC00830).

### Availability of data and materials

The datasets analysed during the current study are available in Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>): GSE1428, GSE230665, GSE56116.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

**Author details**

<sup>1</sup>Department of Spine Surgery, Tianjin Hospital, Tianjin University, Tianjin 300211, China. <sup>2</sup>Department of Toxicology, Tianjin Centers for Disease Control and Prevention, Tianjin 300011, China. <sup>3</sup>Department of Orthopaedics, Cangzhou Central Hospital, Hebei 061001, China. <sup>4</sup>The Third Central, Tianjin Key Laboratory of Extracorporeal Life Support for Critical Diseases, Clinical College of Tianjin Medical University, Nankai University Affinity the Third Central Hospital, Artificial Cell Engineering Technology Research Center, Tianjin Institute of Hepatobiliary Disease, Tianjin 300170, China. <sup>5</sup>Department of Anaesthesiology, Tianjin Hospital, Tianjin 300211, China. <sup>6</sup>Department of Joint Surgery, Tianjin Hospital, Tianjin University, Tianjin 300211, China.

Received: 8 February 2024 Accepted: 28 May 2024  
Published online: 03 June 2024

**References**

- Dhillon RJ, Hasni S. Pathogenesis and Management of Sarcopenia. *Clin Geriatr Med.* 2017;33(1):17–26.
- Baumgartner RN, Koehler KM, Gallagher D, Romero L, Heymsfield SB, Ross RR, Garry PJ, Lindeman RD. Epidemiology of sarcopenia among the elderly in New Mexico. *Am J Epidemiol.* 1998;147(8):755–63.
- Roubenoff R. Origins and clinical relevance of sarcopenia. *Canadian J Appl Physiol.* 2021;26(1):78–89.
- Cho MR, Lee S, Song SK. A Review of Sarcopenia Pathophysiology, Diagnosis, Treatment and Future Direction. *J Korean Med Sci.* 2022;37(18):e146.
- Peterson JA. Osteoporosis overview. *Geriatr Nurs.* 2021;22(1):17–23.
- Gielen E, Dupont J, Dejaeger M, Laurent MR. Sarcopenia, osteoporosis and frailty. *Metabolism.* 2023;145:155638.
- Nielsen BR, Abdulla J, Andersen HE, Schwarz P, Suetta C. Sarcopenia and osteoporosis in older people: a systematic review and meta-analysis. *European geriatric medicine.* 2018;9(4):419–34.
- Liu M, Wang Y, Shi W, Yang C, Wang Q, Chen J, Li J, Chen B, Sun G. PCDH7 as the key gene related to the co-occurrence of sarcopenia and osteoporosis. *Front Genet.* 2023;14:1163162.
- Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013;41:D991–5.
- Giresi PG, Stevenson EJ, Theilhaber J, Koncarevic A, Parkinson J, Fielding RA, Kandarian SC. Identification of a molecular signature of sarcopenia. *Physiol Genomics.* 2005;21(2):253–63.
- Xie L, Feng E, Li S, Chai H, Chen J, Li L, Ge J. Comparisons of gene expression between peripheral blood mononuclear cells and bone tissue in osteoporosis. *Medicine.* 2023;102(20):e33829.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England).* 2007;8(1):118–27.
- Liu F, Huang Y, Liu F, Wang H. Identification of immune-related genes in diagnosing atherosclerosis with rheumatoid arthritis through bioinformatics analysis and machine learning. *Front Immunol.* 2023;14:1126647.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
- The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019;47(D1):D330–8.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605–12.
- Yang C, Delcher C, Shenkman E, Ranka S. Machine learning approaches for predicting high cost high need patient expenditures in health care. *Biomed Eng Online.* 2018;17(Suppl 1):131.
- Kirk B, Al Saedi A, Duque G. Osteosarcopenia: A case of geroscience. *Aging Med (Milton (NSW)).* 2019;2(3):147–56.
- Cruz-Jentoft A, J, Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A. A., Schneider, S. M., Sieber, C. C., Topinkova, E., Vandewoude, M., Visser, M., Zamboni, M., and Writing Group for the European Working Group on Sarcopenia in Older People 2 (EWG-SOP2), and the Extended Group for EWG-SOP2. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing.* 2019;48(4):601.
- Sepúlveda-Loyola W, Phu S, Bani Hassan E, Brennan-Olsen SL, Zanker J, Vogrin S, Conzade R, Kirk B, Al Saedi A, Probst V, Duque G. The Joint Occurrence of Osteoporosis and Sarcopenia (Osteosarcopenia): Definitions and Characteristics. *J Am Med Dir Assoc.* 2020;21(2):220–5.
- Greco EA, Pietschmann P, Migliaccio S. Osteoporosis and Sarcopenia Increase Frailty Syndrome in the Elderly. *Front Endocrinol.* 2019;10:255.
- Yoo JI, Kim H, Ha YC, Kwon HB, Koo KH. Osteosarcopenia in Patients with Hip Fracture Is Related with High Mortality. *J Korean Med Sci.* 2018;33(4):e27.
- Liu F, Xing L, Zhang X, Zhang X. A Four-Pseudogene Classifier Identified by Machine Learning Serves as a Novel Prognostic Marker for Survival of Osteosarcoma. *Genes.* 2019;10(6):414.
- Fan J, Shan C, Kang HB, Elf S, Xie J, Tucker M, Gu TL, Aguiar M, Lonning S, Chen H, Mohammadi M, Britton LM, Garcia BA, Alečković M, Kang Y, Kaluz S, Devi N, Van Meir EG, Hitosugi T, Seo JH, Chen J. Tyr phosphorylation of PDP1 toggles recruitment between ACAT1 and SIRT3 to regulate the pyruvate dehydrogenase complex. *Mol Cell.* 2014;53(4):534–48.
- Miriayala S, Subramanian T, Panchatcharam M, Ren H, McDermott MI, Sunkara M, Drennan T, Smyth SS, Spielmann HP, Morris AJ. Functional characterization of the atypical integral membrane lipid phosphatase PDP1/PPAPDC2 identifies a pathway for interconversion of isoprenols and isoprenoid phosphates in mammalian cells. *J Biol Chem.* 2010;285(18):13918–29.
- Feng X, Lu J, Wu Y, Xu H. MiR-18a-3p improves cartilage matrix remodeling and inhibits inflammation in osteoarthritis by suppressing PDP1. *The journal of physiological sciences : JPS.* 2022;72(1):3.
- Li Y, Shen J, Cheng CS, Gao H, Zhao J, Chen L. Overexpression of pyruvate dehydrogenase phosphatase 1 promotes the progression of pancreatic adenocarcinoma by regulating energy-related AMPK/mTOR signaling. *Cell Biosci.* 2020;10:95.
- Song Y, Zhang J, Zhang L, Zhang S, Shen C. PDP1 Promotes Cell Malignant Behavior and Is Associated with Worse Clinical Features in Ovarian Cancer Patients: Evidence from Bioinformatics and In Vitro Level. *Comput Math Methods Med.* 2022;2022:7397250.
- Hadano S, Otomo A, Suzuki-Utsunomiya K, Kunita R, Yanagisawa Y, Showguchi-Miyata J, Mizumura H, Ikeda JE. ALS2CL, the novel protein highly homologous to the carboxy-terminal half of ALS2, binds to Rab5 and modulates endosome dynamics. *FEBS Lett.* 2004;575(1–3):64–70.
- Boycott KM, Flavelle S, Bureau A, et al. Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification [J]. *Am J Human Genet.* 2005;77(3):477–83.
- Sluysmans S, Méan I, Jond L, Citi S. WW, PH and C-Terminal Domains Cooperate to Direct the Subcellular Localizations of PLEKHA5, PLEKHA6 and PLEKHA7. *Frontiers in cell and developmental biology.* 2021;9:729444.
- Cho YL, Min JK, Roh KM, Kim WK, Han BS, Bae KH, Lee SC, Chung SJ, Kang HJ. Phosphoprotein phosphatase 1CB (PPP1CB), a novel adipogenic activator, promotes 3T3-L1 adipogenesis. *Biochem Biophys Res Commun.* 2015;467(2):211–7.
- Korrodí-Gregório L, Esteves SL, Fardilha M. Protein phosphatase 1 catalytic isoforms: specificity toward interacting proteins. *Translational research : the journal of laboratory and clinical medicine.* 2014;164(5):366–91.
- Printen JA, Brady MJ, Saltiel AR. PTG a protein phosphatase 1-binding protein with a role in glycogen metabolism. *Science.* 1997;275(5305):1475–8.
- Alessi D, MacDougall LK, Sola MM, Ikebe M, Cohen P. The control of protein phosphatase-1 by targeting subunits. The major myosin phosphatase in avian smooth muscle is a novel form of protein phosphatase-1. *Eur J Biochem.* 1992;210(3):1023–35.
- Aoyama H, Ikeda Y, Miyazaki Y, Yoshimura K, Nishino S, Yamamoto T, Yano M, Inui M, Aoki H, Matsuzaki M. Isoform-specific roles of protein phosphatase 1 catalytic subunits in sarcoplasmic reticulum-mediated Ca(2+) cycling. *Cardiovasc Res.* 2011;89(1):79–88.
- Di Mattia T, Wilhelm LP, Ikhlef S, Wendling C, Spehner D, Nominé Y, Giordano F, Mathelin C, Drin G, Tomasetto C, Alpy F. Identification of MOSPD2, a novel scaffold for endoplasmic reticulum membrane contact sites. *EMBO Rep.* 2018;19(7):e45453.



39. Chiapparino A, Maeda K, Turei D, Saez-Rodriguez J, Gavin AC. The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *Prog Lipid Res.* 2016;61:30–9.
40. Yacov N, Kafri P, Salem Y, Propheta-Meirani O, Feldman B, Breitbart E, Mendel I. MOSPD2 is a therapeutic target for the treatment of CNS inflammation. *Clin Exp Immunol.* 2020;201(2):105–20.
41. Davydova E, Shimazu T, Schuhmacher MK, Jakobsson ME, Willemsen HLD, Liu T, Moen A, Ho AYY, Malecki J, Schroer L, Pinto R, Suzuki T, Grønsberg IA, Sohtome Y, Akakabe M, Weirich S, Kikuchi M, Olsen JV, Dohmae N, Umehara T, Falnes PØ. The methyltransferase METTL9 mediates pervasive 1-methylhistidine modification in mammalian proteomes. *Nature communications.* 2021;12(1):891.
42. Lv M, Cao D, Zhang L, Hu C, Li S, Zhang P, Zhu L, Yi X, Li C, Yang A, Yang Z, Zhu Y, Zhang K, Pan W. METTL9 mediated N1-histidine methylation of zinc transporters is required for tumor growth. *Protein Cell.* 2021;12(12):965–70.
43. Bi F, Qiu Y, Wu Z, Liu S, Zuo D, Huang Z, Li B, Yuan Y, Niu Y, Qiu J. METTL9-SLC7A11 axis promotes hepatocellular carcinoma progression through ferroptosis inhibition. *Cell death discovery.* 2023;9(1):428.
44. Hara T, Tominaga Y, Ueda K, Mihara K, Yanagihara K, Takei Y. Elevated METTL9 is associated with peritoneal dissemination in human scirrhous gastric cancers. *Biochemistry and biophysics reports.* 2022;30:101255.

### **Publisher's Note**

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