



OPEN ACCESS

EDITED BY

Nima Gharahdaghi,
University of Oxford, United Kingdom

REVIEWED BY

Jin Li,
Hainan Medical University, China
WanChen Cao,
Beijing University of Chinese Medicine, China

*CORRESPONDENCE

Renjie Tan,
✉ renjie.tan@hrbmu.edu.cn
Yina Zhang,
✉ zhangyina0209@126.com

RECEIVED 18 May 2025

REVISED 03 August 2025

ACCEPTED 24 October 2025

PUBLISHED 06 November 2025

CITATION

Li W, Xing Y, Jiang L, Meng J, Wang Y, Tan R and
Zhang Y (2025) Systems biology and
experimental validation indicate DDIT4, FOXO1,
and STAT3 as shared key genes linking
osteoporosis and sarcopenia.
Front. Genet. 16:1630705.
doi: 10.3389/fgene.2025.1630705

COPYRIGHT

© 2025 Li, Xing, Jiang, Meng, Wang, Tan and
Zhang. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Systems biology and experimental validation indicate DDIT4, FOXO1, and STAT3 as shared key genes linking osteoporosis and sarcopenia

Wenjing Li^{1,2}, Yuechen Xing³, Lihong Jiang², Jia Meng²,
Yue Wang³, Renjie Tan^{4*} and Yina Zhang^{1*}

¹Department of Geriatrics, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China, ²Department of General Practice, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China, ³Department of Occupational Health, School of Public Health, Harbin Medical University, Harbin, Heilongjiang, China, ⁴School of Interdisciplinary Medicine and Engineering, Harbin Medical University, Harbin, Heilongjiang, China

Background: With the aging population, osteoporosis and sarcopenia have emerged as two prevalent age-related degenerative diseases that pose significant public health challenges. Although clinical studies increasingly report the co-occurrence of these conditions, the underlying molecular mechanisms linking them remain poorly understood.

Methods: We adopted a systems biology approach to identify key biomarkers and explore their molecular roles in the interplay between osteoporosis and sarcopenia. Transcriptomic datasets were systematically analyzed to identify candidate genes. The expression patterns of core biomarkers were validated using independent datasets and *in vitro* cellular models of both diseases. Furthermore, a machine learning-based diagnostic framework was constructed using the identified biomarkers, and model interpretability was enhanced using Shapley Additive Explanations (SHAP).

Results: We identified DDIT4, FOXO1, and STAT3 as three central biomarkers that play pivotal roles in the pathogenesis of both osteoporosis and sarcopenia. Their expression patterns were consistently validated across multiple independent transcriptomic datasets, and their differential expression was further confirmed using quantitative reverse transcription polymerase chain reaction (RT-PCR) in disease-relevant cellular models. A diagnostic model constructed based on biomarker genes achieved high classification accuracy across diverse validation cohorts. Moreover, SHAP analysis quantified the individual contribution of each biomarker to the model's predictive performance.

Conclusion: This study uncovers key molecular links between osteoporosis and sarcopenia, highlighting DDIT4, FOXO1, and STAT3 as shared biomarkers. The findings provide novel insights into their common pathophysiology and lay the groundwork for developing more accurate diagnostic tools and targeted therapeutic strategies.

KEYWORDS

osteoporosis, sarcopenia, systems biology, disease biomarkers, machine learning, experimental validation

1 Introduction

With continuous social progress and improvements in healthcare, human life expectancy has been steadily increasing, making population aging a global trend. The prevalence of age-related diseases increases exponentially with age (Belikov, 2019). Over 50% of diseases have been identified as contributing to the global burden in adults due to aging (Chang et al., 2019). Age-related diseases have become a significant burden on human health; among them, osteoporosis and sarcopenia are increasingly common and concerning conditions (Guo et al., 2022; Yang et al., 2023). Osteoporosis is a common systemic skeletal disorder characterized by low bone mass and an increased propensity to fracture (Compston et al., 2019). Osteoporosis is influenced by both environmental and genetic factors, with genetic factors accounting for 50%–85% of the variability (Compston et al., 2019). Genome-wide association studies have identified approximately 100 genomic loci associated with bone density and other related phenotypes (Trajanoska and Rivadeneira, 2019). For example, EN1 has been reported as a determinant of bone density and fracture risk (Zheng et al., 2015). However, the causal mechanisms for many of these associations remain unclear (Boudin and Van Hul, 2017). Sarcopenia is defined as a progressive and generalized skeletal muscle disorder involving accelerated loss of muscle mass and function (Cruz-Jentoft and Sayer, 2019). The causes of sarcopenia can vary and include changes in hormones and growth factors, imbalances in protein metabolism, and inflammation. However, the mechanisms and pathways involved are still not fully understood. Additionally, there is no consensus on the diagnostic criteria for sarcopenia (Bruyère et al., 2016; Cruz-Jentoft and Sayer, 2019; Evans et al., 2024).

Notably, the coexistence of osteoporosis and sarcopenia in the elderly population is becoming increasingly common (Kirk et al., 2022; Laskou et al., 2022; Yang et al., 2023). Muscles and bones are not merely connected physically; they are closely related at both the physiological and pathological levels (Gielen et al., 2023; Kun et al., 2023). Some even suggest that osteoporosis and sarcopenia represent two manifestations of a single disease in different physiological systems (Gielen et al., 2023; Chen et al., 2024b). Recently, Liu et al. reported the involvement of STAT3 in both postmenopausal osteoporosis and sarcopenia (Liu et al., 2024). However, current research on the relationship between these two conditions remains limited (Yang et al., 2023), and there is insufficient evidence to conclusively establish molecular links between them. Furthermore, our understanding of this relationship is still incomplete. The precise diagnosis of osteoporosis and sarcopenia is particularly challenging given the lack of standardized diagnostic criteria for sarcopenia. Additionally, since identifying disease-specific biomarkers is essential for understanding diseases and the corresponding treatments (Byron et al., 2016; Wang et al., 2022; Mahmoodi Chahbatani et al., 2023), there is a need for an accurate approach to detect molecular biomarkers of osteoporosis and sarcopenia. To address this issue, our study incorporated gene expression microarray datasets from both osteoporosis and sarcopenia patients to conduct a systems biology analysis, aiming to identify the biomarkers and uncover the molecular mechanisms connecting the two diseases. Furthermore, inspired by the success of previous related studies (Liu et al., 2014; Lipman et al., 2022; Shu et al., 2022),

we developed a machine learning framework combined with computational and experimental validation to enhance the accurate identification of osteoporosis and sarcopenia.

2 Materials and methods

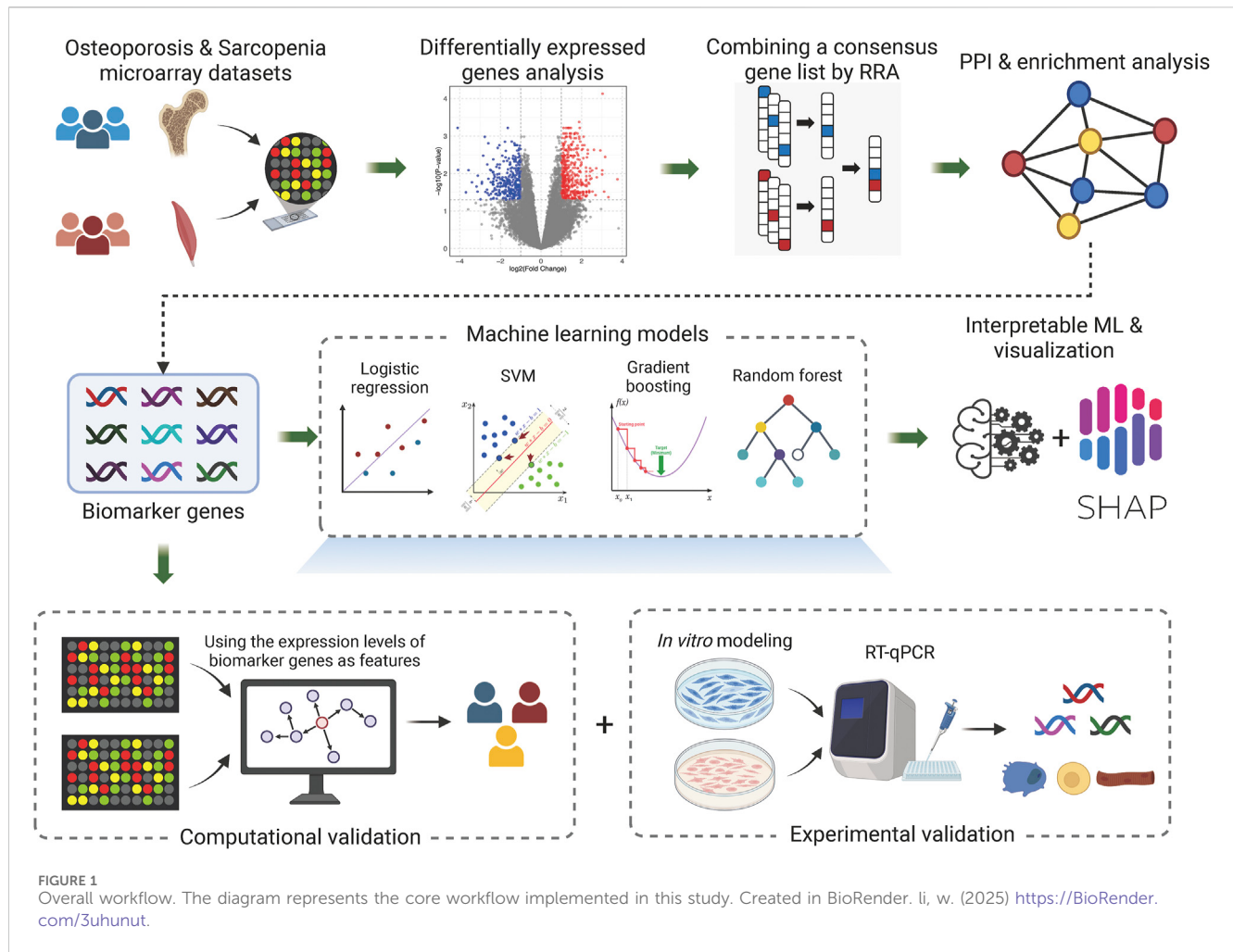
2.1 Overview

Figure 1 illustrates the overall workflow of this study. First, we analyzed multiple microarray datasets from osteoporosis and sarcopenia studies and conducted differentially expressed genes analysis for each individual dataset. Next, we employed a robust integration approach to rank the differentially expressed genes (DEGs) across different datasets within each disease. By combining data from multiple studies, we identified common DEGs shared between osteoporosis and sarcopenia. Subsequent network and enrichment analyses identified key biomarker genes that may play pivotal roles in both diseases, shedding light on their shared molecular mechanisms. Furthermore, leveraging these biomarker genes, we developed a machine learning framework to predict the presence of osteoporosis and sarcopenia. Finally, through computational analysis, *in vitro* modeling, and RT-qPCR validation, we confirmed the robustness of our findings, underscoring the potential of the identified genes as diagnostic biomarkers and therapeutic targets for these interrelated conditions.

2.2 Microarray datasets and data preprocessing

We downloaded osteoporosis and sarcopenia microarray datasets from the National Institutes of Health Gene Expression Omnibus (GEO) data repository (Barrett et al., 2013; Xu et al., 2023). ‘Osteoporosis’ and ‘Sarcopenia’ were used as keywords to query relevant datasets. We limited the entry type to ‘Series’, study type to ‘Expression profiling by array’, and top organisms to ‘*Homo sapiens*’ to accurately locate the datasets required for this study. We subsequently manually reviewed the selected datasets and excluded those involving non-target traits, non-transcriptomic platforms, or redundancy across datasets. This step ensured that the final dataset collection met the specific data requirements of our study. After identifying the appropriate microarray datasets, we downloaded the expression profile information, phenotypic data, and metadata using the ‘GEOquery’ R package (Sean and Meltzer, 2007). We then used the ‘AnnoProbe’ package to annotate the probe information in the expression profiles with gene names.

To minimize batch effects and ensure the reliability of the results, we first calculated the quantiles of the gene expression set at specified levels and assessed certain statistical properties to determine whether log transformation was necessary. This transformation helps stabilize the variance and makes the data more normally distributed. We subsequently used the ‘normalizeBetweenArrays’ function from the Limma package (Ritchie et al., 2015) to normalize the expression data. For genes with multiple corresponding probes, we calculated the average expression value of these probes as the overall gene expression level. Based on the phenotypic data, we classified the samples in each



dataset into case and control groups. Samples that were diagnosed with neither osteoporosis nor sarcopenia, as well as those with ambiguous phenotypic data, were classified as ‘unknown’ and excluded from downstream analyses. To further assess data consistency, we performed principal component analysis (PCA) both within each individual dataset and on the merged dataset used for machine learning model training. Any significant outliers indicating heterogeneity would be removed from downstream analysis.

2.3 Differentially expressed genes identification

We used the Limma package in R to conduct differential expression analysis for each microarray dataset. Limma fits linear models to the expression data based on case–control phenotypes and applies empirical Bayes moderation to estimate gene-wise variances. For each gene, it computes the log fold change (logFC), raw p-value, and adjusted p-value to account for multiple testing. Although DEGs in each dataset could be typically selected using default thresholds (e.g., adjusted p-value <0.05 and $|\log_{2}FC| \geq 1$), our study required integration across multiple datasets, which introduces challenges such as batch effects and incomplete or

inconsistent gene rankings. Therefore, instead of relying solely on conventional cutoffs, we employed the robust rank aggregation (RRA) method (Kolde et al., 2012), which evaluates the ranking of genes by their logFC or p-value across different datasets to identify genes that consistently rank highly in all datasets. We used this method to combine multiple ranked lists into a single consensus gene list and set the RRA p-value <0.05 as the significance cutoff when defining the DEGs for downstream analysis.

2.4 Protein–protein interaction analysis and hub gene selection

To systematically elucidate the pathogenesis of osteoporosis and sarcopenia from a systems biology perspective and further explore the potential connections between the two diseases, we conducted a protein–protein interaction (PPI) analysis on the significantly ranked genes obtained from the RRA analysis results. We used the STRING (Szklarczyk et al., 2023) online database to analyze the protein interactions of these significantly ranked genes. We selected a minimum required interaction score of 0.4 (medium confidence) as the threshold to select interacting proteins. Disconnected proteins in the network were ignored, resulting in the final PPI network. We then utilized the cytoHubba tool in Cytoscape to analyze the

resulting PPI network (Chin et al., 2014). To identify high-confidence key hub genes, which may serve as potential biomarker genes for molecular diagnosis, we selected the recommended MCC method along with two local-based methods (DMNC and Degree) and two global-based methods (Closeness and Betweenness). Finally, following the approach used in similar studies (Lv et al., 2022; Zhou et al., 2024), we employed Venn diagram analysis to identify high-confidence hub genes from the detected hub nodes.

2.5 Functional enrichment analysis

Functional enrichment analysis is a powerful tool for identifying and annotating the biological processes, molecular functions, cellular components, and signaling pathways associated with a given set of genes. We performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and gene ontology (GO) enrichment using the 'clusterProfiler' (Yu et al., 2012) R package on our selected gene sets. Additionally, we conducted comprehensive functional enrichment analysis using Metascape (Zhou et al., 2019) to further interpret the selected gene sets.

2.6 Machine learning model and its explainable visualization

We developed a machine learning-based framework to identify and validate molecular biomarkers for osteoporosis and sarcopenia. Using the selected hub genes as feature genes, we implemented machine learning models using Scikit-learn (version 1.6.1), a widely used Python library. We experimented with four machine learning models, including logistic regression, support vector machine (SVM), gradient boosting, and random forest. All models were trained using default parameter settings provided by Scikit-learn. Specifically, logistic regression was implemented with L2 regularization ($penalty='l2'$) using the 'lbfgs' optimizer. The SVM model utilized a radial basis function (RBF) kernel with a regularization parameter of $C = 1.0$. The GradientBoostingClassifier applied a learning rate of 0.1 and 100 estimators to achieve a balance between model complexity and generalization. The RandomForestClassifier was configured with 100 trees ($n_estimators = 100$) and used $max_features='sqrt'$, a commonly adopted strategy for classification tasks.

We trained the models using microarray expression profile datasets from samples with confirmed osteoporosis and sarcopenia diagnoses. We applied normalization before integrating multiple datasets for both diseases to ensure consistency across datasets. We randomly shuffled and split the data into training (80%) and validation (20%) sets to train and validate the models, respectively.

To further analyze the internal decision-making mechanisms of the machine learning model and to provide insights into the roles these biomarker genes play in precision diagnosis, we employed Shapley additive explanations (SHAP) for model explainability. SHAP is an interpretability framework based on game theory, which calculates SHAP values by fairly distributing the model's output among features using all possible feature combinations

(Lundberg and Lee, 2017). We used the beeswarm summary plot to offer a global overview of feature importance across all predictions in the dataset.

2.7 Model performance and computational validation

To objectively evaluate the model's performance, we obtained additional independent test sets. The refined model was evaluated on both the testing set and the independent datasets. Specifically, we used the refined model to predict the probabilities that each sample belonged to disease or control group. The state corresponding to the maximum probability value was selected as the predicted label. We classified a sample in the testing set as a true positive if its predicted label matched the corresponding true label. Cases without a matching predicted label were counted as false negatives, whereas predicted cases without a matching true label were counted as false positives. The performance of our model was evaluated using the receiver operating characteristic (ROC) curves and the areas under the curve (AUC).

In addition to validating the key genes using ROC curves, we analyzed the expression levels of the hub genes on the independent validation datasets. We compared the expression levels of these hub genes between the case and control samples and performed a *t*-test to statistically assess their differential expression.

2.8 In vitro modeling

To establish *in vitro* models for osteoporosis and sarcopenia, we used the mouse pre-osteoblastic cell line MC3T3-E1 and the mouse myoblast cell line C2C12 (both from ServiceBio, China). MC3T3-E1 cells were cultured in α -MEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37 °C with 5% CO₂. Upon reaching 70% confluence, cells were passaged using 0.25% trypsin, and passages 2 to 4 were used for experiments. To simulate osteoporotic conditions, cells were treated with dexamethasone at concentrations of 0, 0.1, 1, 10, 20, 40, 80, 160, and 320 μ g/mL for 24 h. Cell viability was evaluated using the CCK-8 assay, and absorbance was measured at 450 nm. The optimal treatment condition was determined to be 10 μ g/mL dexamethasone for 24 h. To verify model induction, quantitative RT-PCR was performed to assess the expression levels of osteogenic markers RUNX2 and osteoprotegerin (OPG), as well as the osteoclastogenesis-related factor RANKL (Liu et al., 2021).

Similarly, C2C12 myoblast cells were cultured in DMEM supplemented with 10% heat-inactivated FBS and 1% antibiotics under standard conditions. Upon reaching 70%–80% confluence, differentiation was induced by medium replacement, and cells were subsequently treated with dexamethasone at concentrations of 0, 3.125, 6.25, 12.5, 25, 50, 100, 200, and 400 μ g/mL for 24 h. Cell viability was assessed by adding 10 μ L of CCK-8 reagent per well, followed by incubation at 37 °C in the dark for 2.5 h, and absorbance was measured at 450 nm. Based on the dose-response curve, 12.5 μ g/mL dexamethasone for 24 h was selected as the optimal treatment. Quantitative RT-PCR was then conducted to evaluate the expression of muscle atrophy markers Atrogin-1 and MuRF1, and the

myogenic differentiation marker myogenin (MYOG), to confirm successful model establishment (Li et al., 2021; Kang et al., 2024).

2.9 Experimental validation

Following model establishment, we performed reverse transcription quantitative polymerase chain reaction (RT-qPCR) to validate the expression levels of identified biomarker genes in the induced osteoporosis and sarcopenia models, as well as their respective matched controls. The RT-qPCR primers targeting the coding regions of these genes were designed and synthesized by GenePharma (Shanghai, China). Total RNA was extracted using RNA extraction reagent (FastPure® Cell/Tissue Total RNA Isolation Kit V2, RC112). In the RT-qPCR reaction, the total volume was 20 μ L, including 10 μ L of 2 $\times$ Universal SYBR Green Fast qPCR Mix (ABclonal), 0.8 μ L of 10 nM primers, and 2 μ L of diluted cDNA. All reactions were performed in triplicate under the following conditions: initial denaturation at 95 °C for 3 min, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30s. The relative gene expression levels were normalized and analyzed using the $2^{-\Delta\Delta C_t}$ method.

3 Results

3.1 Data processing and differentially expressed genes selection

To identify biologically relevant differences in gene expression between disease and control groups, we aimed to identify DEGs and calculate corresponding p-value and logFC values. For this purpose, we sought to obtain a sufficient number of high-quality datasets to enable an in-depth investigation of the relationship between osteoporosis and sarcopenia. After searching public data repositories, we found that microarray datasets were more abundant, and these datasets had been validated by recent studies (Chen et al., 2022; Lv et al., 2022; Zhou et al., 2022; Chen et al., 2024a; Chen et al., 2024b; Liu et al., 2024). Consequently, we collected ten microarray gene expression datasets comprising a total of 356 samples (including 191 cases and 165 controls) from osteoporosis and sarcopenia cohorts (Supplementary Table S1). To the best of our knowledge, this study represents the largest integrated dataset to date for the co-disease analysis of osteoporosis and sarcopenia.

We next conducted PCA to assess data homogeneity (Supplementary Figure S1) and applied quality control procedures. During this process, we excluded samples associated with other diseases (e.g., from the GSE136344 dataset) to minimize potential bias and confounding effects. Subsequently, we independently normalized each dataset and performed differential expression analysis using the Limma package with default parameters to calculate p-values and logFC values.

To integrate the DEGs information across multiple datasets for each disease, we applied the RRA method. Notably, the RRA method aggregates DEGs rather than raw gene expression profiles. This approach effectively reduces batch effects and mitigates confounding issues arising from dataset integration, ensuring an

unbiased analysis. As a result, we identified a total of 1,540 significantly ranked genes across five datasets for osteoporosis and 801 significantly ranked genes across two datasets for sarcopenia. Finally, we found that 99 significantly ranked genes were common between the osteoporosis and sarcopenia datasets. These 99 intersecting genes were then selected for downstream analysis.

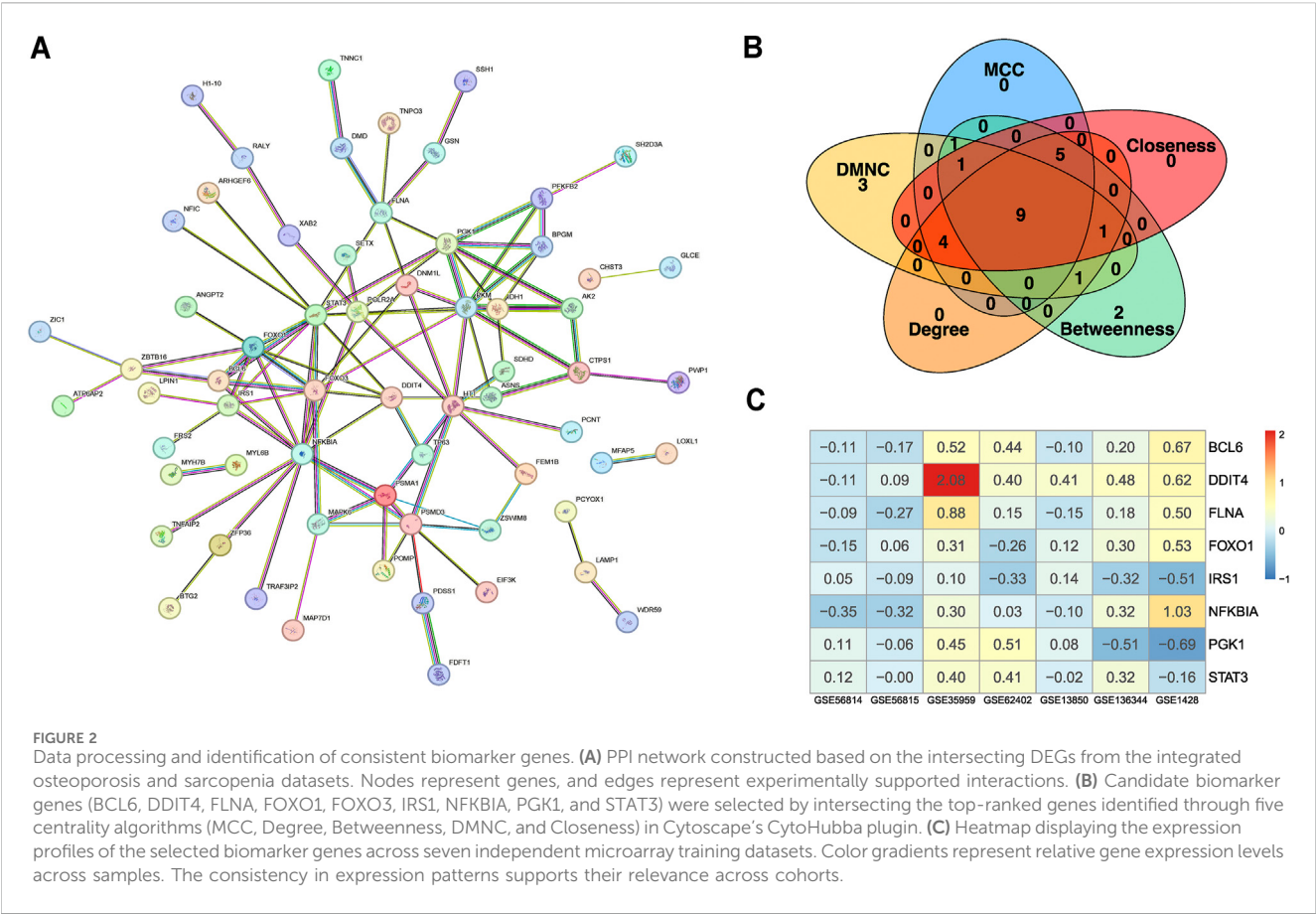
To uncover the functional roles and providing insights into biological processes, we conducted PPI network analysis of the 99 common significantly ranked genes using STRING (Figure 2A) and identified hub genes using the cytoHubba plugin in Cytoscape (Shannon et al., 2003; Chin et al., 2014). We next selected the top 20 genes using the MCC, Degree, Betweenness, DMNC, and Closeness methods (see 'Materials and Methods' section). Using a Venn diagram, we found that BCL6, DDIT4, FLNA, FOXO1, FOXO3, IRS1, NFKBIA, PGK1, and STAT3 were consistently identified by all five algorithms (Figure 2B). Analysis of the expression levels of these nine genes across five osteoporosis datasets and two sarcopenia datasets revealed that, with few exceptions, these genes generally exhibited moderate differential expression between disease and healthy samples (Figure 2C). This suggests that further pathway analysis is warranted.

3.2 Network and enrichment analysis

To explore the potential biological mechanisms underlying these observations, we performed KEGG pathway analysis. In addition to providing curated biological pathways that map molecular interactions, reactions, and relationships, the analysis revealed significant enrichment of these nine genes in eight pathways, including the FoxO signaling pathway, insulin resistance pathway, and longevity regulating pathway (Figure 3A). Gene ontology analysis revealed significant enrichment in biological processes such as myeloid cell differentiation, regulation of carbohydrate metabolic processes, and regulation of small molecule metabolic processes. Additionally, these genes were significantly enriched in molecular functions like chromatin DNA binding and DNA-binding transcription factor binding (Figure 3B). To enhance the reliability of our results, we also performed enrichment analysis using Metascape. This analysis revealed significant enrichment in pathways such as the gastrin signaling pathway, FoxO signaling pathway, cellular response to hormone stimulus, androgen receptor signaling pathway, DNA damage response only ATM dependent, DNA-templated transcription, and negative regulation of catabolic processes (Figures 3C,D). Notably, the FoxO signaling pathway was significantly enriched in both the KEGG and Metascape analyses.

3.3 Machine learning models and SHAP analysis

For precise molecular diagnosis of osteoporosis and sarcopenia, as well as validation of these identified hub genes as common biomarkers for both diseases, we trained the machine learning model using GSE35959, GSE62402, GSE13850, GSE56814, GSE56815, GSE1428, and GSE136344 datasets.



However, due to the absence of gene expression data for FOXO3 in certain datasets (missing in GSE62402 and GSE56814), we used the expression levels of the remaining eight consensus genes—BCL6, DDIT4, FLNA, FOXO1, IRS1, NFKBIA, PGK1, and STAT3—as features for the machine learning model framework. Before training the model, we standardized the data within each dataset to ensure comparability across different datasets. We used the ‘StandardScaler()’ function from Scikit-learn to normalize the samples. We used 80% of the data for training and reserved 20% for validation. We evaluated the model’s performance using 5-fold cross-validation.

To objectively assess model performance, we utilized the independent test datasets GSE7429 and GSE362, which contain data for osteoporosis and sarcopenia, respectively. The discriminative power of the identified biomarkers was first examined by evaluating each gene individually. Using their expression levels as the sole feature, we assessed their ability to distinguish between disease and healthy samples in the test datasets (Supplementary Figures S2, S3). The results indicated that none of the eight biomarker genes alone could provide robust classification performance for both diseases simultaneously.

We further expanded our analysis by employing logistic regression—a simple linear model—as a baseline approach to evaluate the classification performance of the biomarker genes without relying on complex machine learning techniques. As shown in Figures 4A,B, the independent test datasets showed that this baseline model achieved an AUC of 0.58 for osteoporosis

(GSE7429) and 0.91 for sarcopenia (GSE362), suggesting that, while the biomarker genes collectively exhibit moderate yet imbalanced predictive power, more advanced models may improve classification performance. Building on this, we applied random forest, gradient boosting, and SVM models to fully exploit the predictive potential of these biomarkers. The results demonstrated that integrating these genes into advanced machine learning frameworks effectively distinguished diseases from normal control groups (Figures 4A,B). The gradient boosting model achieved AUCs of 0.88 and 1.00 for osteoporosis and sarcopenia, respectively, while the SVM model attained AUCs of 0.83 and 0.96. Notably, the random forest model provided the most balanced and robust performance, with AUCs of 0.94 and 0.95 across the two independent datasets. Given these results, we selected random forest as the default model for our framework.

To quantify the contributions of biomarker genes in our random forest model, we employed SHAP analysis to explain the model’s performance. SHAP beeswarm summary plot presented a dense summary of how the biomarker genes in the dataset influenced the output of the random forest model. Specifically, the SHAP values (on the x-axis) illustrate the impact of the biomarker genes (on the y-axis) on the random forest model’s predictions, with each dot representing the contribution of the gene’s expression level in a specific sample to the model’s classification decision. This provided a measure of each feature’s contribution to the model’s output, with higher means absolute SHAP values indicating greater influence. As shown in Figure 4C, SHAP analysis revealed that DDIT4, STAT3,

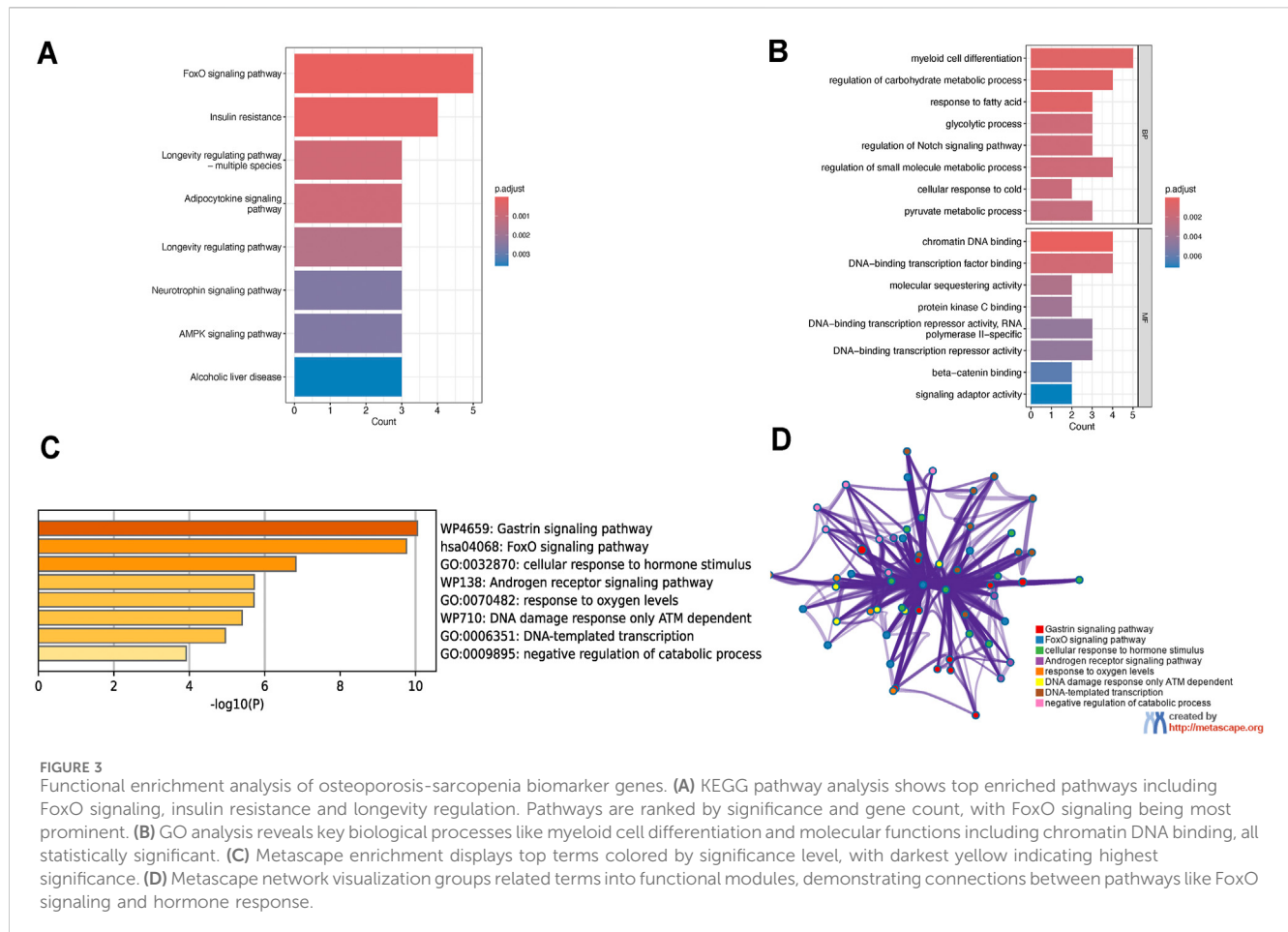


FIGURE 3

Functional enrichment analysis of osteoporosis-sarcopenia biomarker genes. **(A)** KEGG pathway analysis shows top enriched pathways including FoxO signaling, insulin resistance and longevity regulation. Pathways are ranked by significance and gene count, with FoxO signaling being most prominent. **(B)** GO analysis reveals key biological processes like myeloid cell differentiation and molecular functions including chromatin DNA binding, all statistically significant. **(C)** Metascape enrichment displays top terms colored by significance level, with darkest yellow indicating highest significance. **(D)** Metascape network visualization groups related terms into functional modules, demonstrating connections between pathways like FoxO signaling and hormone response.

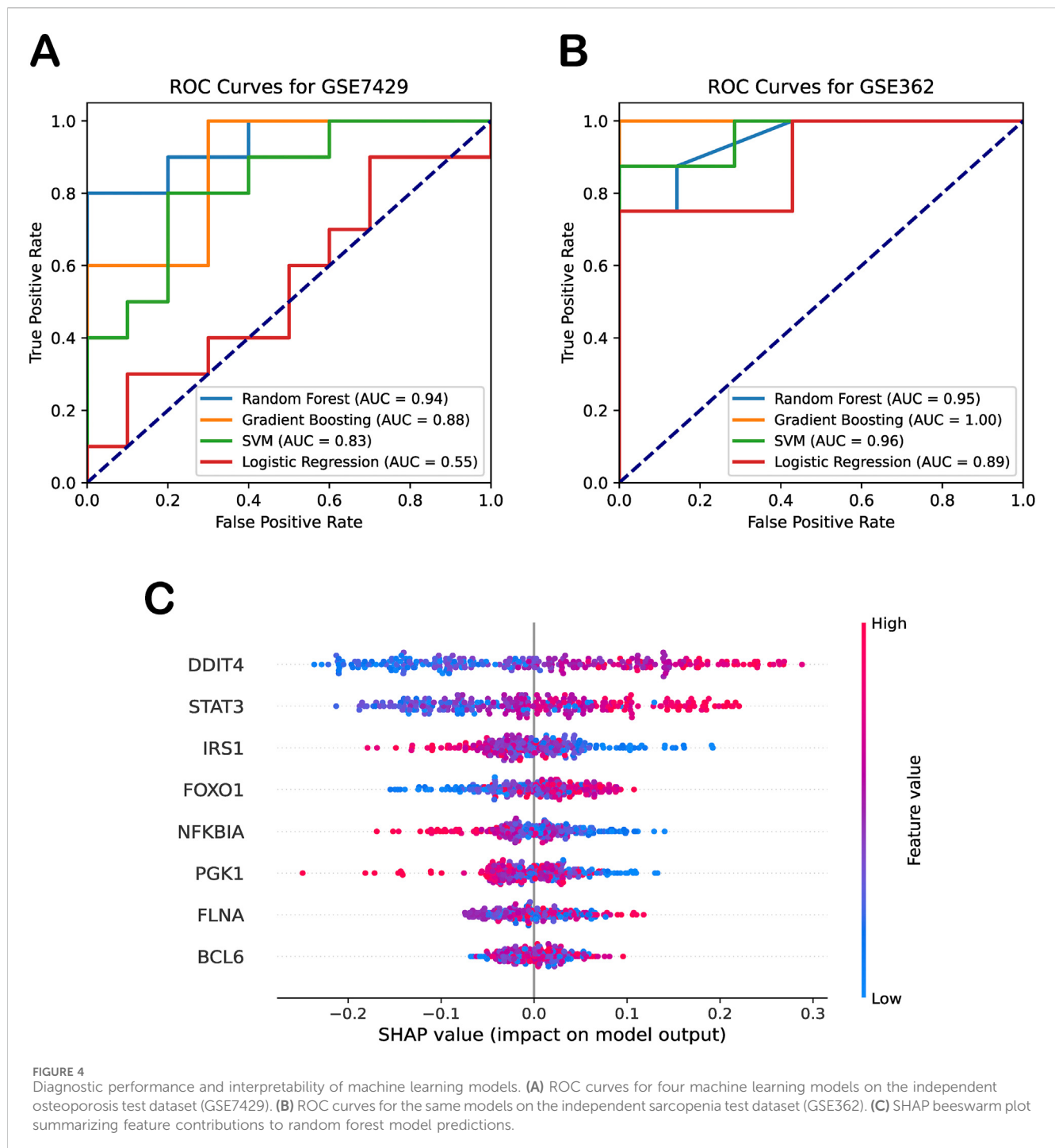
and FOXO1 had significant impact on the model. Their effects were positive, meaning that higher expression levels contributed to the classification of disease, whereas lower expression levels favored the classification of healthy samples. In contrast, IRS1 and PGK1 exhibited an opposite trend to some extent. Moreover, we observed that the SHAP analysis using the training dataset yielded results consistent with our validation findings on the independent test set. These results further support the robustness and reliability of our machine learning model in identifying key biomarkers for osteoporosis and sarcopenia.

3.4 Computational and experimental validation

To verify the accuracy of the biomarker genes identified by our machine learning model, we conducted both computational and experimental validations. For computational validation, we assessed the significance of the eight consensus biomarker genes across independent datasets. Differential expression analysis in osteoporosis and sarcopenia datasets revealed that seven biomarkers—BCL6, DDIT4, FOXO1, IRS1, NFKBIA, PGK1, and STAT3—were differentially expressed in the independent osteoporosis dataset GSE84500. Meanwhile, DDIT4, FOXO1, and STAT3 exhibited significant differential expressions in the sarcopenia dataset GSE362 (Figure 5A).

To experimentally validate the expression patterns of eight consensus biomarker genes, we established *in vitro* models of osteoporosis and sarcopenia using MC3T3-E1 and C2C12 cell lines, respectively. The osteoporosis model exhibited characteristic alterations in key regulatory factors of osteogenic differentiation (RUNX2), OPG, and RANKL, while the sarcopenia model demonstrated specific expression changes in MYOG and ubiquitin ligases Atrogin-1 and MuRF1. Significant differential expressions of these molecules confirmed the validity of model establishment (Supplementary Figure S4). Using the established models, we performed RT-qPCR to quantify gene expression levels in cells affected by osteoporosis and sarcopenia. Results revealed significant differential expressions of DDIT4, FOXO1, NFKBIA, PGK1, and STAT3 in the osteoporosis model. Correspondingly, the sarcopenia model showed marked differential expressions of BCL6, DDIT4, FLNA, FOXO1, NFKBIA, PGK1, and STAT3 (Figure 5B).

Overall, our findings demonstrate the consistency between computational validation and RT-qPCR experimental results. Our proposed systems biology framework and machine learning model can effectively identify reliable biomarkers. By combining computational and experimental validations, we confirmed that at least DDIT4, FOXO1, and STAT3 were significantly upregulated in both osteoporosis and sarcopenia, providing strong support for precise disease diagnosis, further exploration of comorbidity mechanisms, and potential therapeutic



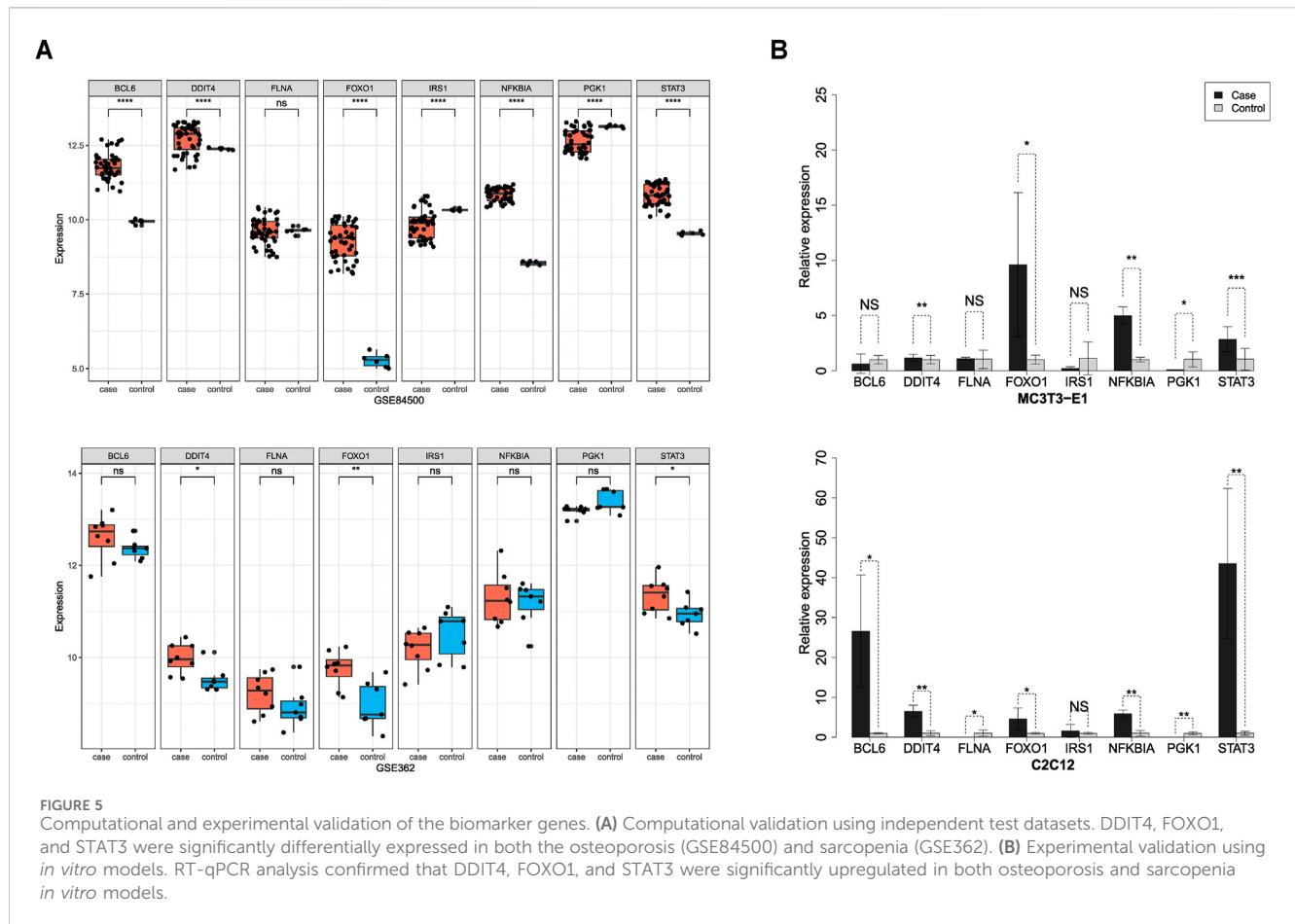
interventions. Moreover, these findings highlight the complexity of osteoporosis and sarcopenia as age-related diseases influenced by multiple genes, with pathway interactions and protein networks playing critical roles in disease progression.

4 Discussion

Emerging evidence suggests a close relationship between osteoporosis and sarcopenia; however, the molecular mechanisms linking them remain poorly defined. To address this gap, we applied

a systems biology framework to analyze transcriptomic data from both diseases. Our integrative analysis identified key genes—BCL6, DDIT4, FLNA, FOXO1, IRS1, NFKBIA, PGK1, and STAT3—potentially involved in their pathogenesis.

Functional enrichment analysis revealed that the FoxO signaling pathway was significantly and consistently enriched across both KEGG and Metascape platforms, highlighting its potential as a central regulatory axis linking osteoporosis and sarcopenia. FoxO transcription factors, particularly FOXO1, are key regulators of cellular homeostasis and longevity. In skeletal muscle, FOXO1 upregulates atrophy-related genes such as Atrogin1 and



MuRF1, thereby promoting muscle protein degradation and contributing to sarcopenia (Sandri et al., 2004; O'Neill et al., 2018). In bone tissue, FOXO1 also plays essential roles in osteoblast differentiation, redox balance, and bone remodeling, with its dysregulation associated with reduced bone mass and impaired repair (Siqueira et al., 2011). Beyond FOXO1 itself, DDIT4 emerged as another biomarker tightly linked to the FoxO pathway. DDIT4 is a stress-responsive gene that negatively regulates the mTOR pathway, thereby indirectly promoting FoxO transcriptional activity under conditions of oxidative or metabolic stress (Gharibi et al., 2016; Altav et al., 2025). This link may explain how systemic stress contributes to degeneration in both muscle and bone tissues. Additionally, STAT3, while not a canonical component of the FoxO pathway, interacts with it via shared downstream targets and stress response circuits. STAT3 is involved in IL-6-mediated inflammatory signaling, which has been implicated in both muscle wasting and bone resorption. Notably, several studies suggest crosstalk between STAT3 and FoxO proteins in regulating oxidative stress and cell survival, pointing toward their convergence in age-related musculoskeletal disorders (Levy and Loomis, 2007; Milner et al., 2015; Guadagnin et al., 2018). Collectively, these findings underscore that FoxO signaling does not act in isolation but rather serves as a molecular hub, integrating multiple upstream regulators (e.g., DDIT4) and intersecting with inflammatory pathways (e.g., STAT3), thereby contributing to the shared pathogenesis of osteoporosis and sarcopenia. This not only

provides a mechanistic explanation for the co-occurrence of these conditions but also suggests the FoxO pathway as a promising target for therapeutic intervention.

To validate these findings, we examined gene expression across independent datasets and observed consistent differential expression of DDIT4, FOXO1, and STAT3 in both osteoporosis and sarcopenia cohorts, confirming their relevance as shared biomarkers. This systems-level insight not only uncovers potential mechanistic links between the two diseases but also informs future therapeutic targeting strategies. Moreover, we developed a machine learning-based classification framework leveraging these biomarkers, which achieved high accuracy in distinguishing between the two conditions. Finally, our *in vitro* experiments further validated the consistency of these findings, demonstrating that the integration of computational biology and experimental validation provides a robust approach for early and precise diagnosis of age-related musculoskeletal diseases.

Although our study contributes to establishing the molecular connection between osteoporosis and sarcopenia, it also has several limitations. First, due to current data constraints, we were unable to include single-cell or bulk RNA-seq transcriptomic analyses to explore the cellular composition and origins of the observed gene expression changes. A more fine-grained, cell-type-specific approach could offer deeper mechanistic insights into the pathogenesis of osteoporosis and sarcopenia. Collecting patient-derived samples or obtaining relevant single-cell data for analysis

will be a key priority in our future work. Second, although we identified several key genes potentially involved in both diseases, the exact molecular roles of these genes in the development and progression of osteoporosis and sarcopenia remain to be fully clarified. Future research should include more in-depth molecular biology experiments—such as gene editing or pathway perturbation studies—to better elucidate their mechanistic functions in this comorbidity. Third, our validation strategy combined computational analysis with *in vitro* modeling and RT-qPCR experiments using mouse cell lines. While MC3T3-E1 and C2C12 cells have been widely used in bone and muscle research and share high genetic conservation with human tissues, they cannot fully replicate the complexity of human disease pathology. Although *in vitro* systems offer the advantages of experimental control and reproducibility, the absence of animal models or clinical samples limits the translational applicability of our findings. Future studies incorporating patient-derived tissues or *in vivo* animal models will be essential to validate the clinical relevance of the identified biomarkers. Despite these limitations, we believe our study contributes to a better understanding of the shared molecular basis of osteoporosis and sarcopenia and may inform future research aimed at biomarker-driven diagnostics and the development of targeted therapeutic strategies.

Data availability statement

The datasets analyzed for this study can be found in the Gene Expression Omnibus database repository (See [Supplementary Table S1](#) for the detailed accession numbers). Our proposed framework is implemented in R and Python and is freely available from the GitHub repository at https://github.com/tanlaboratory/osteoporosis_sarcopenia_analysis.

Author contributions

WL: Writing – review and editing, Investigation, Formal Analysis, Writing – original draft, Data curation. YX: Writing – original draft, Validation, Formal Analysis. LJ: Supervision, Conceptualization, Writing – review and editing. JM: Data curation, Writing – original draft, Conceptualization. YW: Resources, Validation, Writing – review and editing, Supervision. RT: Supervision, Software, Investigation, Validation, Resources, Funding acquisition, Writing – review and editing, Conceptualization, Formal Analysis, Writing – original draft, Project administration, Data curation, Methodology, Visualization. YZ: Supervision, Conceptualization, Writing – review and editing, Project administration.

References

- Altob, G., Merry, B. J., Beckett, C. W., Raina, P., Lopes, I., Goljanek-Whysall, K., et al. (2025). Unravelling the transcriptomic symphony of muscle ageing: key pathways and hub genes altered by ageing and caloric restriction in rat muscle revealed by RNA sequencing. *BMC Genomics* 26 (1), 29. doi:10.1186/s12864-024-11051-1
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41 (Database issue), D991–D995. doi:10.1093/nar/gks1193
- Belikov, A. V. (2019). Age-related diseases as vicious cycles. *Ageing Res. Rev.* 49, 11–26. doi:10.1016/j.arr.2018.11.002
- Boudin, E., and Van Hul, W. (2017). Mechanisms in endocrinology: genetics of human bone formation. *Eur. J. Endocrinol.* 177 (2), R69–R83. doi:10.1530/EJE-16-0990
- Bruyère, O., Beaudart, C., Reginster, J. Y., Buckinx, F., Schoene, D., Hirani, V., et al. (2016). Assessment of muscle mass, muscle strength and physical performance in

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Heilongjiang Province Postdoctoral Research Startup Fund, Program for Young Talents of Basic Research in Universities of Heilongjiang Province (No. YQJH2023024), and the National Natural Science Foundation of China (No. 32470691).

Acknowledgements

We would like to thank Yongshuai Jiang, Haoxiu Sun, and Qinghua Jiang for their thoughtful discussions and kind support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1630705/full#supplementary-material>

clinical practice: an international survey. *Eur. Geriatr. Med.* 7 (3), 243–246. doi:10.1016/j.eurger.2015.12.009

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., and Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17 (5), 257–271. doi:10.1038/nrg.2016.10

Chang, A. Y., Skirbekk, V. F., Tyrovolas, S., Kassebaum, N. J., and Dieleman, J. L. (2019). Measuring population ageing: an analysis of the global burden of disease study 2017. *Lancet Public Health* 4 (3), e159–e167. doi:10.1016/S2468-2667(19)30019-2

Chen, Y., Zou, L., Lu, J., Hu, M., Yang, Z., and Sun, C. (2022). Identification and validation of novel gene markers of osteoporosis by weighted co expression analysis. *Ann. Transl. Med.* 10 (4), 210. doi:10.21037/ATM-22-229

Chen, Y., Zhang, Y., Zhang, S., and Ren, H. (2024a). Molecular insights into sarcopenia: ferroptosis-related genes as diagnostic and therapeutic targets. *J. Biomol. Struct. Dyn.*, 1–19. doi:10.1080/07391102.2023.2298390

Chen, Y., Zhang, Z., Hu, X., and Zhang, Y. (2024b). Epigenetic characterization of sarcopenia-associated genes based on machine learning and network screening. *Eur. J. Med. Res.* 29 (1), 54. doi:10.1186/S40001-023-01603-8

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8 (Suppl. 4), S11. doi:10.1186/1752-0509-8-S4-S11

Compston, J. E., McClung, M. R., and Leslie, W. D. (2019). Osteoporosis. *Lancet* 393 (10169), 364–376. doi:10.1016/S0140-6736(18)32112-3

Cruz-Jentoft, A. J., and Sayer, A. A. (2019). Sarcopenia. *Lancet* 393 (10191), 2636–2646. doi:10.1016/S0140-6736(19)31138-9

Evans, W. J., Guralnik, J., Cawthon, P., Appleby, J., Landi, F., Clarke, L., et al. (2024). “Sarcopenia: no consensus, no diagnostic criteria, and no approved indication—How did we get here?” in *GeroScience. Springer science and business media Deutschland GmbH*.

Gharibi, B., Ghuman, M., and Hughes, F. J. (2016). DDIT4 regulates mesenchymal stem cell fate by mediating between HIF1α and mTOR signalling. *Sci. Rep.* 6, 36889. doi:10.1038/srep36889

Gielen, E., Dupont, J., Dejaeger, M., and Laurent, M. R. (2023). Sarcopenia, osteoporosis and frailty. *Metabolism* 145, 155638. doi:10.1016/j.metabol.2023.155638

Guadagnin, E., Mazala, D., and Chen, Y. W. (2018). STAT3 in skeletal muscle function and disorders. *Int. J. Mol. Sci.* 19 (8), 2265. doi:10.3390/ijms19082265

Guo, J., Huang, X., Dou, L., Yan, M., Shen, T., Tang, W., et al. (2022). Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduct. Target. Ther.* 7 (1), 391. doi:10.1038/s41392-022-01251-0

Kang, M., Kang, M., Yoo, J., Lee, J., Lee, S., Yun, B., et al. (2024). Dietary supplementation with *Lactisacibacillus rhamnosus* IDCC3201 alleviates sarcopenia by modulating the gut microbiota and metabolites in dexamethasone-induced models. *Food Funct.* 15 (9), 4936–4953. doi:10.1039/d3fo05420a

Kirk, B., Kuo, C.-L., Xiang, M., and Duque, G. (2022). Associations between leukocyte telomere length and osteosarcopenia in 20,400 adults aged 60 years and over: data from the UK biobank. *Bone* 161, 116425. doi:10.1016/j.bone.2022.116425

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 28 (4), 573–580. doi:10.1093/bioinformatics/btr709

Kun, E., Javan, E. M., Smith, O., Gulamali, F., de la Fuente, J., Flynn, B. I., et al. (2023). The genetic architecture and evolution of the human skeletal form. *Science* 381 (6655), eadf8009. doi:10.1126/science.adf8009

Laskou, F., Fuggle, N. R., Patel, H. P., Jameson, K., Cooper, C., and Dennison, E. (2022). Associations of osteoporosis and sarcopenia with frailty and multimorbidity among participants of the hertfordshire cohort study. *J. Cachexia, Sarcopenia Muscle* 13 (1), 220–229. doi:10.1002/jcsm.12870

Levy, D. E., and Loomis, C. A. (2007). STAT3 signaling and the Hyper-IgE syndrome. *N. Engl. J. Med.* 357 (16), 1655–1658. doi:10.1056/NEJME078197

Li, Z., Liu, C., Li, S., Li, T., Li, Y., Wang, N., et al. (2021). BMSC-derived exosomes inhibit dexamethasone-induced muscle atrophy via the miR-486-5p/FoxO1 axis. *Front. Endocrinol. (Lausanne)* 12, 681267. doi:10.3389/fendo.2021.681267

Lipman, A. R., Fan, X., Shen, Y., and Chung, W. K. (2022). Clinical and genetic characterization of CACNA1A-related disease. *Clin. Genet.* 102 (4), 288–295. doi:10.1111/CGE.14180

Liu, Y., Li, B., Tan, R., Zhu, X., and Wang, Y. (2014). A gradient-boosting approach for filtering *de novo* mutations in parent–offspring trios. *Bioinformatics* 30 (13), 1830–1836. doi:10.1093/bioinformatics/btu141

Liu, Y., Wang, L., Yang, Y., and Xiong, J. (2021). Silencing *Hoxa2* reverses dexamethasone-induced dysfunction of MC3T3-E1 osteoblasts and osteoporosis in rats. *Adv. Clin. Exp. Med.* 30 (5), 525–534. doi:10.17219/acem/133495

Liu, D., Wang, K., Wang, J., Cao, F., and Tao, L. (2024). Identification of the molecular link: STAT3 is a shared key gene linking postmenopausal osteoporosis and sarcopenia. *Bone Jt. Res.* 13 (8), 411–426. doi:10.1302/2046-3758.138.BJR-2023-0351.R2

Lundberg, S. M., and Lee, S.-I. (2017). “A unified approach to interpreting model predictions,” in *Proceedings of the 31st international conference on neural information processing systems* (Long Beach, California, USA: Curran Associates Inc.).

Lv, N., Zhou, Z., He, S., Shao, X., Zhou, X., Feng, X., et al. (2022). Identification of osteoporosis based on gene biomarkers using support vector machine. *Open Med. (Wars)* 17 (1), 1216–1227. doi:10.1515/med-2022-0507

Mahmoodi Chalbatani, G., Gharagouzloo, E., Malekraei, M. A., Azizi, P., Ebrahimi, A., Hamblin, M. R., et al. (2023). The integrative multi-omics approach identifies the novel competing endogenous RNA (ceRNA) network in colorectal cancer. *Sci. Rep.* 13 (1), 19454. doi:10.1038/s41598-023-46620-z

Milner, J. D., Vogel, T. P., Forbes, L., Ma, C. A., Stray-Pedersen, A., Niemela, J. E., et al. (2015). Immunobiology: early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. *Blood* 125 (4), 591. doi:10.1182/BLOOD-2014-09-602763

O'Neill, B. T., Bhardwaj, G., Penniman, C. M., Krumpoch, M. T., Suarez Beltran, P. A., Klaus, K., et al. (2018). FoxO transcription factors are critical regulators of diabetes-related muscle atrophy. *Diabetes* 68 (3), 556–570. doi:10.2337/db18-0416

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/NAR/GKV007

Sandri, M., Sandri, C., Gilbert, A., Skurk, C., Calabria, E., Picard, A., et al. (2004). Foxo transcription factors induce the atrophy-related ubiquitin ligase Atrogin-1 and cause skeletal muscle atrophy. *Cell* 117 (3), 399–412. doi:10.1016/S0092-8674(04)00400-3

Sean, D., and Meltzer, P. S. (2007). GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14), 1846–1847. doi:10.1093/BIOINFORMATICS/BTM254

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/GR.1239303

Shu, C., Green Snyder, L., Shen, Y., Chung, W. K., and Consortium, T. S. (2022). Imputing cognitive impairment in SPARK, a large autism cohort. *Autism Res.* 15 (1), 156–170. doi:10.1002/aur.2622

Siqueira, M. F., Flowers, S., Bhattacharya, R., Faibish, D., Behl, Y., Kotton, D. N., et al. (2011). FOXO1 modulates osteoblast differentiation. *Bone* 48 (5), 1043–1051. doi:10.1016/j.bone.2011.01.019

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids Res.* 51 (D1), D638–D646. doi:10.1093/NAR/GKAC1000

Trajanoska, K., and Rivadeneira, F. (2019). The genetic architecture of osteoporosis and fracture risk. *Bone* 126, 2–10. doi:10.1016/j.BONE.2019.04.005

Wang, D. C., Liu, X., Powell, C. A., and Wang, X. (2022). Novel paradigms of macrophage biology and function: identification of disease biomarkers and therapeutic targets. *Cell Biol. Toxicol.* 38 (4), 553–556. doi:10.1007/s10565-022-09749-7

Xu, R., Ma, L. L., Cui, S., Chen, L., and Xu, H. (2023). Bioinformatics and systems biology approach to identify the pathogenetic link between heart failure and sarcopenia. *Arq. Bras. Cardiol.* 120 (10), e20220874. doi:10.36660/ABC.20220874

Yang, J., Jiang, T., Xu, G., and Liu, W. (2023). Bibliometrics analysis and visualization of sarcopenia associated with osteoporosis from 2000 to 2022. *J. Pain Res.* 16, 821–837. doi:10.2147/JPR.S403648

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/OMI.2011.0118

Zheng, H. F., Forgetting, V., Hsu, Y. H., Estrada, K., Rosello-Diez, A., Leo, P. J., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* 526, 112–117. doi:10.1038/nature14878

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523–10. doi:10.1038/s41467-019-09234-6

Zhou, H., Jiang, J., Chen, X., and Zhang, Z. (2022). Differentially expressed genes and miRNAs in female osteoporosis patients. *Med. Baltim.* 101 (28), e29856. doi:10.1097/MD.00000000000029856

Zhou, X., Chen, Y., Zhang, Z., Miao, J., Chen, G., and Qian, Z. (2024). Identification of differentially expressed genes, signaling pathways and immune infiltration in postmenopausal osteoporosis by integrated bioinformatics analysis. *Heliyon* 10 (1), e23794. doi:10.1016/j.heliyon.2023.e23794