# Transcriptomic Signatures of Mechanotransduction in Osteogenesis and a Machine Learning-Based Predictive Model

A[1†], B[1,2,3,4*]
1., China
† These authors contributed equally to this work.
Corresponding author:

**Disclosures**

●Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

●Funding statement

●Conflict of interest disclosure

The authors declare no conflict of interest.

## ABSTRACT

Traditional approaches in bone biology have relied on machine learning models that classify samples based on their overall osteogenic state, leading to limitations such as oversimplified sample labeling, data standardization biases, and batch effects. This study introduces a novel gene-centric classification framework that shifts the focus from predicting sample phenotypes to directly identifying genes involved in osteogenesis. By pre-labeling genes based on their known correlation with osteoporosis, we circumvent the need for large sample sizes and data normalization, effectively mitigating batch effects. Integrating bioinformatics and machine learning analyses of murine and human transcriptomic data, we developed a robust model that successfully predicted osteogenic differentiation-related genes. Validation using microgravity-responsive transcriptomic data led to the identification of six key hub genes (ETV4, ENPEP, ETS1, LRRFIP1, PLAUR, and PTX3) that are responsive to microgravity and promote osteogenic differentiation. This study provides valuable insights into the molecular mechanisms of osteogenesis and offers potential therapeutic targets for combating bone loss.

**Keywords:** machine learning, osteogenesis, mechanotransduction

## Introduction

Osteoporosis, a systemic skeletal disease characterized by reduced bone mineral density (BMD) and deterioration of bone microarchitecture, poses a significant global health challenge by increasing fracture risk and associated morbidity [(1)]. As the global population ages, its prevalence is projected to rise, imposing a substantial economic and societal burden, particularly in nations like China with rapidly growing elderly demographics [(2)]. Consequently, there is an urgent need for strategies focused on the early diagnosis and prevention of bone loss to preserve the quality of life in aging populations.

A powerful biological analogue for age-related bone loss is the osteopenia induced by mechanical unloading, such as that experienced under microgravity conditions [(3)]. Osteogenic cells reside in a dynamic mechanical environment, constantly sensing and responding to a complex milieu of biophysical cues from the extracellular matrix (ECM) and neighboring cells [(4)]. These cues, including substrate stiffness, fluid shear stress, and tensile strain, are fundamental regulators of osteogenic cell behavior, dictating critical processes like proliferation, migration, and differentiation [(5-9)]. Therefore, elucidating the underlying mechanisms of mechanotransduction—the process by which cells convert mechanical stimuli into biochemical responses—holds immense promise for identifying novel therapeutic targets to combat bone loss.

The molecular cascade of mechanotransduction is well-documented and can be conceptualized as a tripartite process: mechanosensing, intracellular signaling, and transcriptional regulation [(10)]. At the cell periphery, mechanosensors such as integrins, Piezo channels, primary cilia, and gap junctions detect physical forces and initiate downstream signaling [(12-15)]. For instance, calcium influx through Piezo1 channels can activate the ERK1/2 pathway and influence cytoskeletal organization in osteoblasts [(16)]. These signals converge on key transcriptional regulators like YAP/TAZ and MKL1, which translocate to the nucleus to modulate gene expression [(17)]. This culminates in the upregulation of osteogenic master genes, including *RUNX2*, *SP7* (Osterix), and *COL1A1*, thereby driving bone formation [(18)]. Furthermore, the LINC complex, which physically bridges the cytoskeleton and nucleoskeleton, is essential for transmitting mechanical forces directly to the nucleus, influencing chromatin organization and gene expression [(19-21)]. Despite this detailed understanding, a significant challenge remains: disentangling true nuclear mechanotransduction events from downstream consequences of cytoplasmic signaling pathways.

The rapid accumulation of high-throughput transcriptomic data offers an unprecedented opportunity to systematically explore the genetic networks governing mechanotransduction in osteogenesis. However, leveraging this wealth of information is fraught with challenges. Data heterogeneity, stemming from variations in experimental conditions, sequencing platforms, and batch effects, often precludes the direct integration and comparison of datasets [(22-24)]. Traditional differential analysis alone is frequently insufficient to reliably identify true biological signals amidst this technical noise. While machine learning (ML) has emerged as a powerful tool for biomarker discovery and disease prediction [(25-27)], its application to this problem is non-trivial.

Conventional ML approaches in genomics typically employ binary classification models trained on sample-level labels (e.g., diseased vs. healthy). This strategy faces several critical limitations. First, assigning simple binary labels to complex biological states can be an oversimplification, reducing model accuracy. Second, achieving consistency in sample features to mitigate batch effects often requires aggressive data filtering, which can drastically reduce sample size and compromise model robustness. Finally, even with preprocessing, residual batch variations can confound model interpretation, making it difficult to distinguish true biological signals from technical artifacts

[30]. To overcome these fundamental limitations, this study proposes a paradigm shift. We move away from the traditional sample-labeling framework and introduce a novel, gene-centric classification strategy. By labeling genes themselves based on their known biological roles, we test the feasibility of building a robust model that is inherently resilient to data heterogeneity and bypasses the confounding influence of batch effects. This approach aims to provide a more reliable and interpretable method for identifying key regulators of osteogenesis from large-scale, heterogeneous transcriptomic data.

## Materials and Methods

## Polished and Expanded Version

To develop a robust machine learning model for predicting osteogenesis-related genes, we assembled a comprehensive and heterogeneous transcriptomic compendium. This dataset comprised a total of 223 samples aggregated from 34 distinct public datasets (detailed in Supplemental Table 1), encompassing data from both microarray and RNA-sequencing (RNA-seq) platforms. The multi-source nature of this data necessitated a meticulous, platform-specific processing and harmonization workflow.

### Data Acquisition and Pre-processing

Our data integration strategy was tailored to the original data format. For the microarray data, we downloaded raw CEL files from 152 samples spanning 25 datasets. These files were processed and normalized using the oligo package in the R environment. For the RNA-seq data, we implemented a two-pronged approach. For 4 samples from the ERP003789 project, we began with raw .fastq files. Quality control was performed using FastQC, followed by alignment to the reference genome using STAR to generate sorted BAM files. Gene-level quantification was then carried out with HTSeq2 to produce an expression matrix. The remaining 67 RNA-seq samples were already processed and were directly downloaded as pre-formatted expression matrices from the GEO database.

### Data Integration and Normalization

To create a unified feature space for model training, we integrated all processed data into a single gene expression matrix. This matrix was filtered to include only the 9,661 genes that were consistently detected across all 223 samples, ensuring no missing values. To mitigate platform-specific biases and make expression values comparable, we applied a uniform normalization strategy. For the RNA-seq data, raw counts were first converted to Transcripts Per Million (TPM), with gene lengths retrieved via the biomaRt R package. Subsequently, all expression values across the entire integrated matrix (both microarray and RNA-seq derived) were transformed using a log2(TPM + 0.001) scaling. This final step resulted in a normalized, homogenized expression matrix ready for downstream machine learning applications.

## Machine Learning Model Development and Implementation

To construct predictive models for identifying genes that promote osteoblast differentiation, we employed three classical and interpretable machine learning algorithms: k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), and Logistic Regression (LR). All models were implemented within the R programming environment (v4.3.1) to ensure reproducibility. Specifically, the k-NN model was instantiated using the knn() function from the class package. The SVM model was constructed with the ksvm() function from the kernlab package, which provides a robust framework for kernel-based methods. For the logistic regression model, we utilized the glm() function from the base stats package, a standard approach for generalized linear models.

### Model Validation Using Microgravity-Responsive Transcriptomic Data

To validate our models and identify key osteogenic regulators under a physiologically relevant stress condition, we analyzed transcriptomic data related to microgravity. We selected two public datasets, GSE1367 and GSE4658, which profile the gene expression of 2T3 pre-osteoblasts following a 3-day exposure to simulated microgravity. These datasets are particularly valuable as they employ two distinct simulation devices: the Random Positioning Machine (RPM) and the Rotating Wall Vessel (RWV). The RWV simulates a low-shear, weightless environment through horizontal rotation, nullifying the net gravity vector over time. In contrast, the RPM achieves microgravity simulation by continuously and randomly reorienting the gravity vector, preventing gravitational settling. Differential expression analysis for both datasets was performed using GEO2R, an online bioinformatics tool from the NCBI. This analysis identified genes significantly altered under microgravity compared to normal gravity controls. The resulting outputs, including gene identifiers, log2 Fold Change (log2FC), p-values, and adjusted p-values, were downloaded for downstream visualization. Subsequently, volcano plots were generated using the ggplot2 package in R to visually summarize the significantly upregulated and downregulated genes in each dataset.

### Statistical Analysis

Statistical rigor was maintained throughout the study. In the machine learning framework, model performance was quantified using the Area Under the Curve (AUC), which was calculated with the pROC package (v1.18.5) in R. For the analysis of experimental data, comparisons between two groups were conducted using an unpaired two-tailed Student's t-test. A p-value of less than 0.05 was considered statistically significant.

## Results

Construction of a Machine Learning-Based Model for Predicting Osteogenic Differentiation

In a paradigm shift from conventional machine learning approaches in bone biology, which typically classify samples based on their overall osteogenic state, our study introduces a novel gene-centric classification framework. Instead of asking "Is this sample differentiating into bone?", our model asks, "Is this gene a promoter of bone formation?". This fundamental reorientation of the analytical question is enabled by a meticulously constructed gene-labeling strategy.

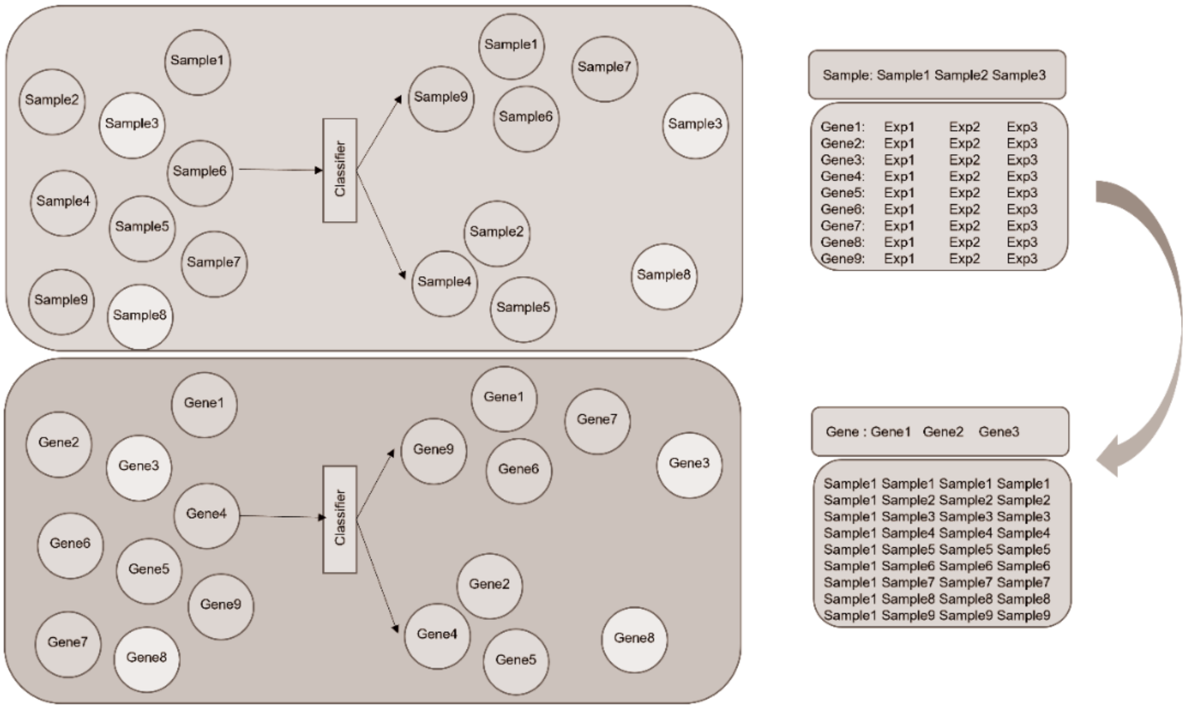
To establish robust training labels, we curated two distinct gene sets from the Gene Ontology (GO) database, restricting our search to *Homo sapiens* to align with our human cell line transcriptomic data. Genes annotated under the GO:0001649 (osteoblast differentiation) term were designated as the positive class, representing promoters of bone formation. Conversely, recognizing the well-established antagonistic relationship between osteogenic and adipogenic lineages in bone marrow mesenchymal stem cells, we leveraged this biological principle to define a biologically meaningful negative class. Genes annotated under GO:0045444 (adipocyte differentiation) were selected as the negative class.

This data curation process yielded a high-quality training set. From the 135 genes identified for the positive osteogenic class, 84 were present in our expression matrix. Similarly, from the 116 genes identified for the negative adipogenic class, 77 were represented. This resulted in a final labeled dataset of 161 unique genes (84 positive, 77 negative) for model training.

This strategic labeling fundamentally transforms the model's architecture and objective. The primary unit of analysis shifts from the sample to the gene. Consequently, the feature variable is no longer a comprehensive gene expression profile of a single sample, but rather a cross-sample expression vector for an individual gene. As illustrated in Figure 1, this redefines the model's output: instead of predicting a sample's phenotype, the model now predicts a gene's functional role. It assesses whether a given gene's expression pattern across multiple samples is more indicative of an association with the positive label class (osteogenic differentiation) or the negative label class (adipogenic differentiation). This approach allows us to move beyond phenotype prediction to the direct, in-silico identification of candidate genes governing lineage commitment.

## A Machine Learning Framework for Prioritizing Candidate Genes in Osteogenesis

Following data collection and label assignment, we proceeded to develop our classification models. To ensure interpretability and transferability, we deliberately selected three classic and theoretically robust machine learning algorithms: k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), and Logistic Regression (LR). The core innovation of our methodology lies in the fundamental



Sample: Sample1 Sample2 Sample3

Gene1: Exp1 Exp2 Exp3
Gene2: Exp1 Exp2 Exp3
Gene3: Exp1 Exp2 Exp3
Gene4: Exp1 Exp2 Exp3
Gene5: Exp1 Exp2 Exp3
Gene6: Exp1 Exp2 Exp3
Gene7: Exp1 Exp2 Exp3
Gene8: Exp1 Exp2 Exp3
Gene9: Exp1 Exp2 Exp3

Gene : Gene1 Gene2 Gene3

Sample1 Sample1 Sample1 Sample1
Sample1 Sample2 Sample2 Sample2
Sample1 Sample3 Sample3 Sample3
Sample1 Sample4 Sample4 Sample4
Sample1 Sample5 Sample5 Sample5
Sample1 Sample6 Sample6 Sample6
Sample1 Sample7 Sample7 Sample7
Sample1 Sample8 Sample8 Sample8
Sample1 Sample9 Sample9 Sample9

Figure 1 A Comparative Analysis of Sample-Level and Gene-Level Classification Models. The green panel represents the conventional sample classification model, and the blue panel presents our novel approach. Various classification methods are listed on the right.

shift of the unit of analysis. Unlike traditional approaches where samples are the primary subject, our labels correspond directly to genes, and the feature vector describes the cross-sample expression profile of an individual gene (as illustrated in Figure 1). This unique data structure allowed us to bypass the conventional data normalization step. By utilizing raw expression values, we maximally preserved the inherent biological signals across various sources and platforms while effectively circumventing potential artifacts or over-correction of batch effects that normalization might introduce, thus ensuring the authenticity and distinctiveness of our feature vectors. To systematically evaluate model performance, we employed the confusion matrix as our primary evaluation tool, calculating several key metrics: Accuracy, F1-Score, and Cohen's Kappa coefficient. These metrics provided a comprehensive assessment of the models' predictive precision, class-wise balance, and consistency. The classification results are shown in Figure 2A-C. In terms of accuracy, k-NN, SVM, and LR achieved 77.6%, 81.0%, and 75.1%, respectively. The F1-scores were 0.775, 0.768, and 0788. Cohen's Kappa coefficient also indicated that SVM possessed superior consistency (k-NN: 0.921, SVM: 0.842, LR: 0.867). Furthermore, we assessed the discriminative power of the models by plotting the Receiver Operating Characteristic (ROC) curve and calculating the Area Under the Curve (AUC). All AUC values were significantly above 0.75, providing strong evidence that all models possess excellent classification performance in distinguishing between osteogenic and adipogenic genes. Based on a comprehensive consideration of all evaluation metrics, the Support Vector Machine (SVM) consistently outperformed the other models, achieving the highest scores across accuracy, F1-score, Kappa, and AUC. Therefore, SVM was selected as the optimal model for this study.

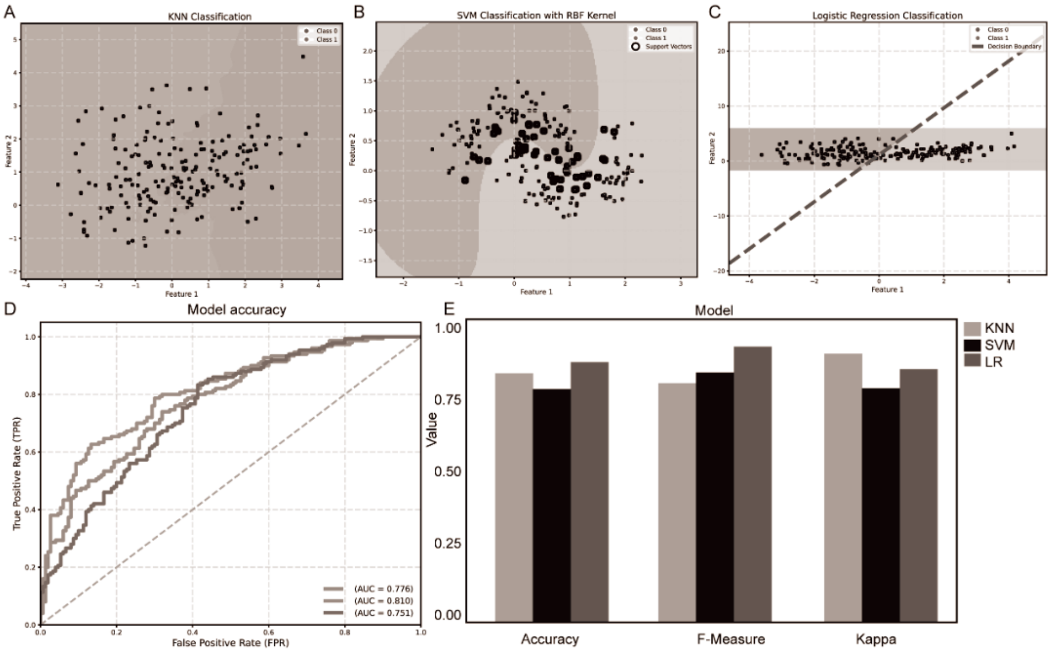


Figure 2 Evaluation of the four machine learning models. (A-C) Classification diagrams of the three models; (D) Model AUC curve; (E) The accuracy, F-measure, and Kappa values of four models.

### To screen genes involved in mechanotransduction during osteoblastic differentiation

To investigate the regulatory mechanisms of osteogenic differentiation under microgravity and to validate the practical utility of our mode The accuracy, F-measure, and Kappa values of four models. l, we conducted a systematic screening using transcriptomic data. We recruited two key datasets from the Gene Expression Omnibus (GEO): GSE1367 and GSE4658. These datasets contain gene expression profiles of osteoprogenitor cells exposed to simulated microgravity conditions using a Rotating Wall Vessel (RWV) and a Random Positioning Machine (RPM), respectively.Initially, we performed differential expression analysis to identify microgravity-responsive genes. In the GSE1367 dataset, 127 genes were significantly dysregulated under microgravity compared to the normal gravity (1g) control (Figure 3A). Similarly, 124 differentially expressed genes were identified in the GSE4658 dataset (Figure 3B). To isolate a more robust core set of microgravity-responsive genes, we took the intersection of the results from both datasets, identifying 93 genes that were significantly altered under both distinct simulation conditions.Subsequently, we applied our previously constructed machine learning model to further screen these differentially expressed genes. The model successfully identified 49 candidate genes from the set of 93 that were highly associated with osteogenic differentiation.

Finally, by integrating the results from the differential expression analysis (93 genes) with the predictions from our machine learning model (49 genes), we pinpointed their intersection. This integrative approach led to the final identification of six key hub genes (Figure 3C): ETV4, ENPEP, ETS1, LRRFIP1, PLAUR, and PTX3. These six genes are not only responsive to microgravity but were also predicted by our model to be potential promoters of osteogenic differentiation.To further validate the biological functions of these key genes, we performed Gene Ontology (GO) and Gene Set Enrichment Analysis (GSEA). The functional enrichment results (Figure 3D-E) demonstrated that these six genes were significantly enriched in biological pathways related to osteogenic differentiation, skeletal system development, and bone mineralization. This provides strong evidence supporting their potential critical role in promoting osteogenesis.

### Discussion

Machine learning has firmly established itself as a cornerstone of modern genomic analysis, providing unparalleled insights into the complex circuitry of biological systems and the molecular etiology of human diseases. Its impact has been particularly profound in oncology, where the analysis of large-scale, high-dimensional datasets has enabled the development of sophisticated models for cancer subtyping, prognostic prediction, and therapeutic response assessment. These advancements have not only deepened our understanding of cancer biology but have also paved the way for more personalized and effective treatment strategies.

In stark contrast, the application of ML in the field of bone metabolism has lagged, primarily constrained by the limited size and dimensionality of publicly available genomic datasets. To address this data scarcity, researchers have proposed integrating multi-

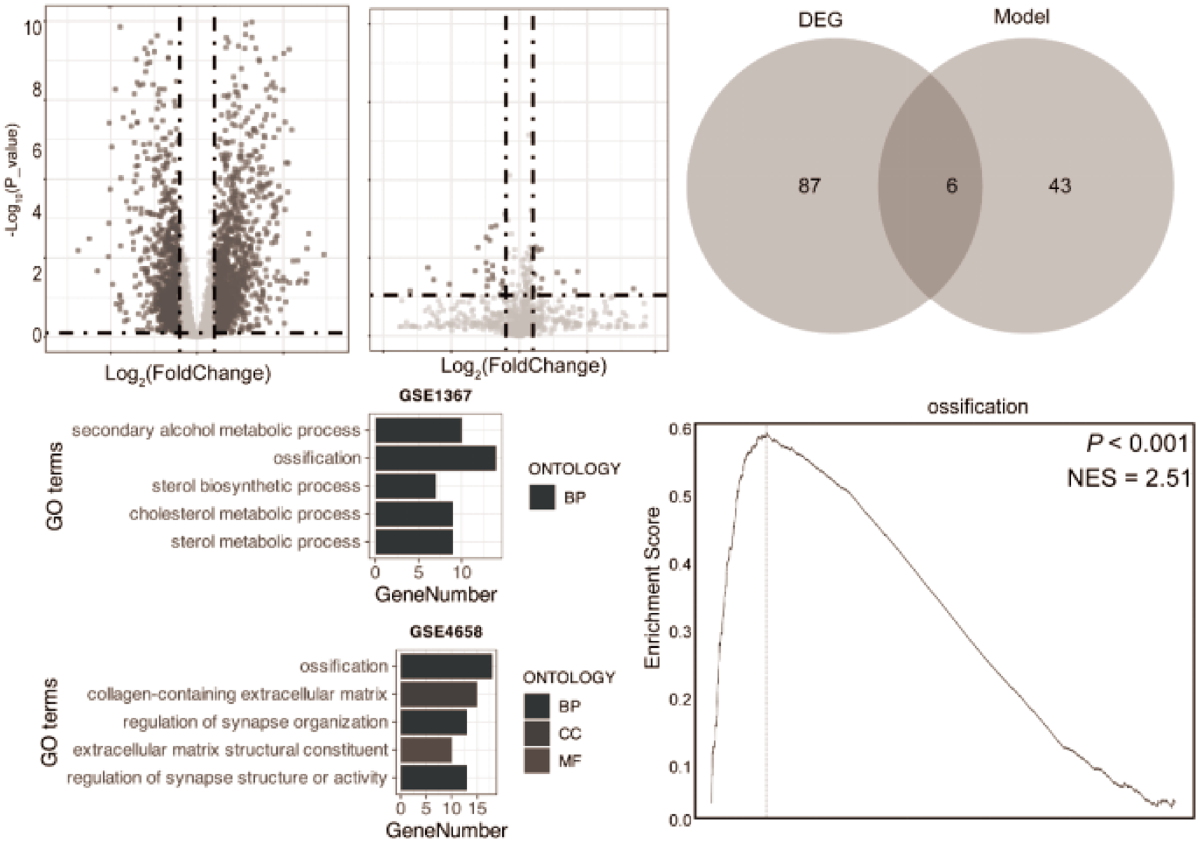


Figure 3 Among genes down-regulated in microgravity, ENPEP/Enpep is identified by 4 models as the gene promoting osteogenic differentiation. A-C Volcano plots and intersections of GSE4658 and GSE1367 differential expression genes. D-F Volcano plots and intersections of SM and OM group differential expression genes of GSE100930 in microgravity. G-H The predictions of human or mouse gene candidates by all four ML models, respectively. I Intersections of human and mouse genes predicted to promote osteogenic differentiation by four ML models.

omic data or incorporating clinical and imaging information to enrich the feature space. Alternatively, transfer learning techniques have been suggested to leverage knowledge from data-rich domains to improve model performance in smaller, target datasets. Despite these potential solutions, a fundamental bottleneck remains: the reliance on large, consistently labeled datasets for supervised learning.

To navigate these challenges in the context of osteogenesis research, our study introduces a novel machine learning-based pipeline specifically designed to alleviate the dependence on sample consistency and mitigate the confounding effects of batch effects. While this gene-centric approach represents a significant conceptual advance, we acknowledge that it is not without its own limitations, which present clear avenues for future refinement.

First, the current gene-labeling strategy is exclusively dependent on predefined Gene Ontology (GO) terms. This reliance on existing, albeit curated, knowledge confines our labeled dataset to a relatively small cohort of 161 genes. This scarcity of high-quality labeled instances inherently restricts the model's ability to learn complex patterns and may limit its ultimate predictive accuracy. Future iterations could expand the labeled gene set by incorporating information from pathway databases (e.g., KEGG, Reactome) or literature mining, thereby creating a more comprehensive and robust training foundation.

Second, the models employed in this study are classical machine learning algorithms that operate within a supervised learning paradigm. While interpretable and effective, their performance is intrinsically capped by the quantity and quality of the available labels. A promising future direction is the integration of deep learning architectures. Models such as Graph Neural Networks (GNNs), which can learn from the topology of gene interaction networks, or self-supervised approaches that learn representations from unlabeled data, could circumvent the need for manual labeling altogether. We envision the development of a hierarchical, multi-tiered framework that synergistically combines the interpretability of traditional ML with the feature-learning power of deep learning. In such a system, an initial deep learning model could generate high-quality gene embeddings or predict pseudo-labels for a larger set of genes, which could then be used to train a more accurate and transparent classical classifier. This hybrid approach could unlock the full potential of limited genomic data in bone biology and beyond.

## References

1. Zeng Q, Li N, Wang Q, Feng J, Sun D, Zhang Q, et al. The Prevalence of Osteoporosis in China, a Nationwide, Multicenter DXA Survey. J Bone Miner Res. Oct 2019;34(10):1789-97.

2. Si L, Winzenberg TM, Jiang Q, Chen M, Palmer AJ. Projection of osteoporosis-related fractures and costs in China: 2010-2050. Osteoporos Int. Jul 2015;26(7):1929-37.

3. Goldsmith M, Crooks SD, Condon SF, Willie BM, Komarova SV. Bone strength and composition in spacefaring rodents: systematic review and meta-analysis. NPJ Microgravity. Apr 13 2022;8(1):10.

4. Martino F, Perestrelo AR, Vinarsky V, Pagliari S, Forte G. Cellular Mechanotransduction: From Tension to Function. Front Physiol. 2018;9:824.

5. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. Cell. Aug 25 2006;126(4):677-89.

6. Weyts FAA, Bosmans B, Niesing R, van Leeuwen JPTM, Weinans H. Mechanical control of human osteoblast apoptosis and proliferation in relation to differentiation. Calcified Tissue Int. Apr 2003;72(4):505-12.

7. Ponik SM, Triplett JW, Pavalko FM. Osteoblasts and osteocytes respond differently to oscillatory and unidirectional fluid flow profiles. J Cell Biochem. Feb 15 2007;100(3):794-807.

8. Mullen CA, Haugh MG, Schaffler MB, Majeska RJ, McNamara LM. Osteocyte differentiation is regulated by extracellular matrix stiffness and intercellular separation. J Mech Behav Biomed. Dec 2013;28:183-94.

9. Wang JH, Thampatty BP. An introductory review of cell mechanobiology. Biomech Model Mechanobiol. Mar 2006;5(1):1-16.

10. Liu ZS, Wang QL, Zhang JY, Qi SH, Duan YY, Li CY. The Mechanotransduction Signaling Pathways in the Regulation of Osteogenesis. Int J Mol Sci. Sep 2023;24(18):14326.

11. Lim XR, Harraz OF. Mechanosensing by Vascular Endothelium. Annu Rev Physiol. Feb 12 2024;86:71-97.

12. Oria R, Wiegand T, Escribano J, Elosegui-Artola A, Uriarte JJ, Moreno-Pulido C, et al. Force loading explains spatial sensing of ligands by cells. Nature. Dec 14 2017;552(7684):219-24.

13. Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ, et al. Piezo1 and Piezo2 Are Essential Components of Distinct Mechanically Activated Cation Channels. Science. Oct 1 2010;330(6000):55-60.

14. Chinipardaz Z, Liu M, Graves DT, Yang S. Role of Primary Cilia in Bone and Cartilage. J Dent Res. Mar 2022;101(3):253-60.

15. Jiang JX, Siller-Jackson AJ, Burra S. Roles of

gap junctions and hemichannels in bone cell functions and in signal transmission of mechanical stress. Front Biosci. Jan 1 2007;12:1450-62.

16. Zhang G, Li X, Wu L, Qin YX. Piezo1 channel activation in response to mechanobiological acoustic radiation force in osteoblastic cells. Bone Res. Mar 10 2021;9(1):16.

17. Panciera T, Azzolin L, Cordenonsi M, Piccolo S. Mechanobiology of YAP and TAZ in physiology and disease. Nat Rev Mol Cell Biol. Dec 2017;18(12):758-70.

18. Zhao X, Tang L, Le TP, Nguyen BH, Chen W, Zheng M, et al. Yap and Taz promote osteogenesis and prevent chondrogenesis in neural crest cells in vitro and in vivo. Sci Signal. Oct 25 2022;15(757):eabn9009.

19. Stroud MJ, Banerjee I, Veevers J, Chen J. Linker of nucleoskeleton and cytoskeleton complex proteins in cardiac structure, function, and disease. Circ Res. Jan 31 2014;114(3):538-48.

20. Kirby TJ, Lammerding J. Emerging views of the nucleus as a cellular mechanosensor. Nat Cell Biol. Apr 2018;20(4):373-81.

21. Tajik A, Zhang Y, Wei F, Sun J, Jia Q, Zhou W, et al. Transcription upregulation via force-induced direct stretching of chromatin. Nat Mater. Dec 2016;15(12):1287-96.

22. Foox J, Tighe SW, Nicolet CM, Zook JM, Byrska-Bishop M, Clarke WE, et al. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. Nat Biotechnol. Sep 2021;39(9):1129-40.

23. Jia HX, Tan SJ, Zhang YE. Chasing Sequencing Perfection: Marching Toward Higher Accuracy and Lower Costs. Genom Proteom Bioinf. Mar 11 2024;22(2):qzae024.

24. Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. Bmc Genomics. May 28 2024;25(1):528.

25. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv. Jul-Aug 2021;49:107739.

26. Shi W, Chen Z, Liu H, Miao C, Feng R, Wang G, et al. COL11A1 as an novel biomarker for breast cancer with machine learning and immunohistochemistry validation. Front Immunol. 2022;13:937125.

27. Wang H, Zhang Z, Cheng X, Hou Z, Wang Y, Liu Z, et al. Machine learning algorithm-based biomarker exploration and validation of mitochondria-related diagnostic genes in osteoarthritis. PeerJ. 2024;12:e17963.

28. Zhang YY, Wang Z, Jiang JJ, You HM, Chen JJ. Toward Improving the Robustness of Deep Learning Models via Model Transformation. Ieee Int Conf Autom. 2022:1-13.

29. Pfeifer SP. From next-generation resequencing reads to a high-quality variant data set. Heredity. Feb 2017;118(2):111-24.

30. Yu Y, Mai YB, Zheng YT, Shi LM. Assessing and mitigating batch effects in large-scale omics studies. Genome Biol. Oct 3 2024;25(1):254.